Autonomous Self-Optimization of Coverage and Capacity in LTE Cellular Networks

Alexander Engels, *Student Member, IEEE*, Michael Reyer, *Member, IEEE*, Xiang Xu, *Student Member, IEEE*, Rudolf Mathar, *Member, IEEE*, Jietao Zhang, and Hongcheng Zhuang

***Abstract*—Self-organizing networks promise significant expenditure savings for operators when rolling out modern cellular network infrastructure, such as Long-Term Evolution (LTE) and LTE-Advanced systems. Savings in capital expenditures (CAPEX) and operational expenditures (OPEX) can be achieved in both the network deployment and network operation phase. Particularly, self-organized optimization of network coverage and network capacity is a key challenge to cope with the boost in mobile data traffic that is expected in the next years and to benefit from the growing market. We present a traffic-light-related approach to autonomous self-optimization of tradeoff performance indicators in LTE multitier networks. Introducing a low-complexity interference approximation model, the related optimization problem is formulated as a mixed-integer linear program and is embedded into a self-organized network operation and optimization framework. Tuning site activity, transmission power, and antenna downtilt are parameters of eNodeBs and Home eNodeBs. The optimization procedure is carried out considering time-variant optimization parameters that are automatically adapted with respect to changes in the network. Simulation-based evaluation of representative case studies demonstrates applicability and the benefit potential of our overall concept.**

***Index Terms*—Coverage and capacity optimization (CCO), multiobjective optimization, self-optimization, self-organizing networks (SONs), self-planning.**

I. Motivation and Background

According to recent market studies, cellular networks have to cope with the boost in data traffic in the next years [1]: The overall mobile data traffic is expected to grow from 0.6 EB in 2011 by one magnitude in 2015 and by a factor of 18 in 2016. By that time, the forecast expects smartphones and tablets to be accountable for more than 50% of the traffic and, particularly, high-data-rate services, such as mobile video, to dominate the requested mobile services. To support these future mobile demands and to maximally benefit from the exploding market, most operators introduce modern fourth-generation (4G) wireless communication systems based on the *Long-Term Evolution/System Architecture Evolution (LTE/SAE)* standard [2] and its extension *LTE-Advanced* [3]. The LTE rollout comes along with many opportunities to reduce cost and complexity for deploying and operating 4G wireless access networks. From the operator perspective, hardware deployment and installation cost (capital expenditures) and cost for network management, physical resource rental, and equipment maintenance [operational expenditures (OPEX)] highly determine network profitability and, therefore, are strong criteria for rollout decisions (see [4] and [5]).

Self-organizing networks (SONs) are a key concept to take full advantage of the mentioned opportunities, particularly in the context of 4G *multitier networks (MTNs)*. A high degree of process automation promises significant OPEX reduction, and the application of self-triggered advanced optimization techniques (self-optimization) shall improve system capacity, coverage, and service quality without manual intervention. The strong interest in this topic is reflected by many recent activities in the scope of SONs under the guidance of the *Next Generation Mobile Networks (NGMN) Alliance* and the *Third-Generation Partnership Project (3GPP)*. While the NGMN Alliance mainly provides economical and technical guidelines [6], [7], 3GPP is responsible for the standardization of related network components (see [8] and [9]). Concerning *self-organized LTE MTNs (SO-MTNs)*, relevant 3GPP specifications are covered by different releases: LTE Release 8 contains fundamental specifications for LTE systems that are currently deployed by most operators [2], [10]. In addition to LTE system fundamentals, it covers basic specifications for *Home eNodeB (HeNB)* components, self-establishment of network equipment, and automatic neighbor relation list management [11]. However, it has been decided that SON algorithms themselves are not standardized in Release 8 [12]. LTE Release 9 contains specifications for enhanced HeNB and studies on self-organization for HeNB, self-healing, and self-organized *coverage and capacity optimization (CCO)*. Further releases (10–12) continuously extend the specification of LTE system features, particularly with respect to future *LTE Advanced* systems [3], [13], [14]. In [15], 3GPP defines the CCO-related technical requirements and specifies scenarios for CCO assessment, e.g., coverage hole situations or coverage overflow constellations when neighboring cells have too much overlap and when system capacity is degraded.

Several initiatives have been established to investigate and to contribute to self-optimization and self-configuration in wireless communication networks [16]: As part of the *Celtic Initiative* [17], the *Celtic GANDALF* project contributed at a very early stage, for example, to automated troubleshooting

Manuscript received May 31, 2012; revised October 22, 2012 and January 23, 2013; accepted March 19, 2013. Date of publication April 2, 2013; date of current version June 12, 2013. This work was supported by the Huawei Innovation Research Program under Contract YJCB2010010RE. The review of this paper was coordinated by Dr. F. Gunnarsson.

A. Engels, M. Reyer, X. Xu, and R. Mathar are with the Institute for Theoretical Information Technology, RWTH Aachen University, 52074 Aachen, Germany (e-mail: engels@ti.rwth-aachen.de; reyer@ti.rwth-aachen.de; xu@ti.rwth-aachen.de; mathar@ti.rwth-aachen.de).

J. Zhang and H. Zhuang are with the Communications Technologies Lab of the Central Research Institute, Huawei Technologies Co., Ltd., Shenzhen, 518129, China (e-mail: jtzhang@huawei.com; zhc@huawei.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2013.2256441

and automatic tuning of network parameters [18]. The *End-to-End Efficiency* project, which is funded by the European Union (EU) within the *Seventh Framework Programme* (FP7), covers some SON-related use cases, such as handover optimization or intercell interference coordination [19]. Within the same EU program, the SOCRATES project was established considering SON aspects, such as integrated handover parameter optimization and load balancing, automatic generation of initial insertion parameters, and cell outage management [20]. According to the SON use case specifications provided by SOCRATES [21], NGMN Alliance [7], and in [16], our work is mainly related to *self-planning* with respect to automatic CCO. This task is basically located at the intersection of *self-configuration* and *self-optimization*. Since we consider cells that are permanently switched off for CCO evaluation, our work might also be related to the context of *self-healing*. Particularly, by tuning transmission power and antenna downtilt, we consider a subset of optimization parameters that is supposed to be the most effective not only for CCO [16] but for self-healing purposes as well [21].

Presenting an approach to *closed-loop autonomous self-optimization of coverage and capacity* [16] that is applicable centralized (whole network) or semicentralized (cluster-wise), our work fulfills the following 3GPP requirements: First, human intervention is not necessary once optimization parameters are initially specified. In fact, the optimization parameters that are sensitive to dynamic changes in the network, e.g., the traffic distribution and related interference parameters, are automatically updated within the autonomously running CCO loop. Second, static optimization components, e.g., the performance measure thresholds for CCO triggering and the available configuration space, are fully customizable by the operator. Third, different cells and network areas are independently customizable. In our approach, periodic or aperiodic detection of degraded system performance automatically triggers optimization procedures that autonomously improve the performance by (re-) configuration of basic control parameters, namely, transmission power, antenna tilt, and transmitter activity (switch on/off). Since maximization of coverage and maximization of capacity, generally, are tradeoff tasks [22], we consider a typical multi-objective optimization problem, as discussed in [23] and [24]. To cope with the contradicting objectives, we develop a traffic-light-related decision scheme that optimizes MTN coverage and capacity either jointly or in a hierarchical manner if significant performance degradation need to be resolved. Furthermore, we introduce a novel interference approximation model that allows for reducing the overall computational complexity and that leads to a linear formulation of the CCO problem.

Related work considering CCO is presented in [25] and [26], where reinforcement learning methods are applied to iteratively improve an expert system that defines rule-based decisions for self-organized configuration of cell parameters. In contrast to our CCO framework, such a model-free approach does not require any *a priori* information since it does not consider an explicitly formulated system model but is based only on the observations it gets as feedback from system measurement reports. On the other hand, the missing system model makes it hard to interpret the decisions taken by a model-free approach, and consequently, it is difficult to influence model-free approaches straightforwardly. Therefore, in our work, we consider a model-based optimization approach that explicitly reflects underlying system dependence relations but that adapts its parameters automatically with respect to the observation feedback from the dynamic system.

Further related work is the self-optimization of HeNBs with respect to coverage optimization and interference control, considered within the SOCRATES project [20]. Similar to the work presented in [27], the considered macrocell (MC) environment is static, and the eNB control parameters are not jointly optimized. On the other hand, *cell outage management* in SOCRATES [20] and self-organized eNB downtilt optimization in [25] do not consider joint HeNB configuration. Our approach, however, considers both tiers, i.e., MCs and femtocells (FCs), jointly for optimization. Most of the related work in the SON area considers a hexagonal grid network model [28]. Since we observed effects that were specifically caused by that artificial topology, e.g., a very high frequency reuse potential due to cell symmetry, we switched to a more realistic irregular network layout that implements propagation characteristics of an urban environment.

The rest of this paper is organized as follows: In Section II, we describe the considered LTE MTN model and the measures and metrics that are used for system performance assessment and system optimization. Particularly, we introduce our novel interference approximation model as part of the optimization-specific system model. We present the traffic-light-related approach to autonomous CCO and the corresponding optimization model in Section III. In Section IV, we carry out a simulation-based proof of concept by demonstrating the performance of our approach in representative case studies. Finally, we conclude this paper and discuss future work in Section V.

II. System Model

We consider downlink transmission in an LTE MTN, i.e., LTE eNodeB (eNB) related MCs and HeNBs that are deployed indoors and that are associated with FCs. All cells are operated within the same network, and all transmitters utilize the same access technology [orthogonal frequency-division multiple access (OFDMA)] in the same frequency band. Such a system is also often referred to as a *heterogeneous network (HetNet)* [3], but we will keep the phrase *MTN* since it emphasizes the considered MC–FC topology. OFDMA FCs have been proposed not only to overcome the indoor coverage problem but also to deal with the growth of traffic within MCs [29]. We assume FCs to provide coverage of approximately 10–50 m for stationary or low-mobility *user entities (UEs)* that are located at home or in small offices [30]. The corresponding HeNBs are connected to the backhaul via a broadband connection such as optical fiber or digital subscriber line, which basically enables for centralized coordination.

In addition to an *open-access* policy for FC usage, we consider *cochannel FC deployment*, which provides the largest amount of available transmission bandwidth to eNBs and HeNBs. However, full spectrum sharing can cause strong cochannel (cross-tier) interference, and hence, *interference*

control becomes even more important compared with a single-tier network topology. Concerning autonomous CCO, it is a challenge to incorporate intercell and cross-tier interference into the optimization model without introducing too much complexity. In Section II-B, we propose an approximation model that describes multitier interference effects in a computationally efficient way.

Each user in the system requests a certain data rate and is associated with a predefined *priority level*. While the data rate is directly related to the requested service, the UE priority can correspond to either the service class or the user type, e.g., Voice over Internet Protocol (VoIP) services might have a higher priority than data services, or business clients may be favored over private customers.

For system optimization, we consider the following control parameters: Each eNB can apply several *antenna downtilts* (electrical), i.e., it can increase its footprint by tilting up or increase received signal power (RSP) near its location by tilting down. Furthermore, each eNB may use different antennas and adjust its transmission power within a certain range. HeNBs are equipped with a static antenna and can control their transmission power. Since we do not consider power allocation for particular physical resource blocks (PRBs), all power modifications affect the whole transmission band.

A. Link Quality Computation

According to the LTE system specification, 16 *channel quality indicators (CQIs)* are distinguished [2]. Each CQI corresponds to a supported modulation scheme and code rate for downlink transmission, i.e., we can compute spectral efficiency in terms of bits per second per hertz for each CQI. The smallest nonzero spectral efficiency in present LTE systems is 0.25 bit/s/Hz for quadrature phase-shift keying (QPSK) and code rate 1/8, and the largest spectral efficiency is 4.8 bit/s/Hz for 64-quadrature amplitude modulation (QAM) and code rate 4/5. The *system link budget specification* defines what RSP and what receiver reference sensitivity (RS) level are required to support a certain CQI, such that the receiver can decode the data with a transport block error probability below 10%. The RS typically considers thermal noise (-174 dBm/Hz) multiplied by the transmission bandwidth, the receiver noise figure (9 dB), an implementation margin (2.5 dB for QPSK, 3 dB for 16-QAM, and 4 dB for 64-QAM), and a diversity gain (-3 dB; see [10]). Different QoS requirements of a UE can be modeled by modifying the corresponding CQI specifications accordingly.

To compute the link-wise RSP information with respect to a certain transmitter configuration, we utilize the ray optical approach presented in [31] and [32]. Comparing the achievable RSP with CQI-specific RS requirements, we select the maximum supportable CQI, and the corresponding spectral efficiency describes the maximum supported link quality. If UE t has rate demand r_t and is served by *either eNB or HeNB ((H)eNB)* a, which supports spectral efficiency e_{at} on its link to t, the required bandwidth for successful downlink transmission is computed as

$$b_{at} = \frac{r_t}{e_{at}}. \tag{1}$$

B. Interference Approximation Model

Since interference is one of the main limiting factors for network coverage, capacity, and performance [21], it has to be suitably considered for network operation and for network optimization. Resource allocation in terms of power and spectrum allocation is the key component to apply *intercell interference coordination (ICIC)* [7]. Practical schemes for interference coordination are, for example, interference mitigation techniques or soft frequency reuse [29], [33], [34]. For network operation (simulation), we consider a system that applies *interference mitigation*, i.e., certain parts of the transmission band might be blocked to mitigate interference effects to other cells. For incorporation into network optimization (CCO), we introduce the following approximation model that describes the according resource consumption in a computationally efficient way.

Generally, resource allocation considering ICIC is performed on the basis of *signal-to-interference-plus-noise ratio (SINR)* information, i.e.,

$$\mathrm{SINR}_{atn} = \frac{p_{an} g_{atn}}{\sum_{a' \neq a} p_{a'n} g_{a'tn} + \sigma_{an}^2}$$

where p_{an} denotes the transmission power of (H)eNB a on *PRB* n, g_{atn} is the channel gain on that PRB experienced by UE t, and σ_{an}^2 is the (thermal) noise power [35]. As resource allocation of other (H)eNBs a' appears as an interference term in the denominator, optimal resource allocation is a computationally hard combinatorial problem for practical problem sizes [29], [36]. To evade this level of additional complexity on top of CCO, we propose the following approximation model that partitions the bandwidth, which is consumed at a transmitter entity into resource b^{SRV} that is allocated for serving user rates on the basis of the *signal-to-noise ratio (SNR)* and into resource b^{ITF} that is reserved (blocked) additionally for ICIC, i.e., the total bandwidth consumption of an (H)eNB a is modeled as linear superposition. Thus,

$$b_a = b_a^{\mathrm{SRV}} + b_a^{\mathrm{ITF}}. \tag{2}$$

Component b_a^{SRV} is the sum over the bandwidth that is allocated to users that are served by a, where the user-specific amount is computed according to (1). The blocked bandwidth b_a^{ITF} is computed as a linear combination over all resources that are allocated by interfered transmissions and where a link from a' to user t is interfered with impact factor $q_{aa't}^{\mathrm{ITF}} \in [0, 1]$. The impact factor basically depends on the SNR difference and on the frequency reuse potential of the related cells. Fig. 1 shows a realization of the proposed approximation, assuming an equal SNR for all transmission links and all interference links, a constant interference impact factor of 1/2, and that each eNB has to allocate two PRBs for serving an assigned UE. The depicted allocation is obtained by assuming resource-allocation algorithms that benefit from the frequency reuse potential, e.g., eNB 1 blocks the same PRBs to mitigate interference to UEs 2 and 3.

In a HetNet, there are typically many FCs, i.e., HeNBs, located near each other and, hence, interfered by the same eNB. Consequently, the transmit efficiency of those FCs can be significantly increased by blocking bandwidth for FC interference

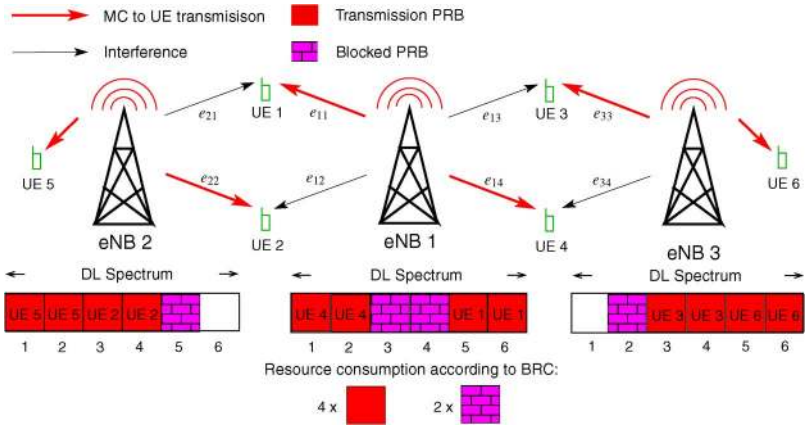

Fig. 1. Exemplary PRB allocation at eNBs and approximation of resource consumption for eNB 1.

mitigation in the related MC. Inter-femtocell interference is expected to be low since the HeNB coverage area is generally very small. This property allows for a reuse of the same PRBs at the considered HeNBs without significant SINR degradation. Consequently, it is beneficial for the system performance to align the bandwidth that is blocked for interference mitigation at the interfering eNB jointly over all affected FCs. We incorporate these considerations into our approximation model as follows: Each eNB additionally blocks bandwidth for FC interference mitigation according to the maximum consumed bandwidth at potentially interfered FCs. The bandwidth consumption at HeNBs is determined by applying the approximation model, considering FCs only. The amount of blocked bandwidth typically varies from eNB to eNB due to the spatial diversity of rate demand and FC distribution. Since our interference approximation model is an estimator for the *expected average amount of bandwidth* that has to be reserved in the spectrum to support the overall UE rates, we denote it as the *bandwidth reservation concept (BRC)*.

Please note that the BRC considers resources as a continuous variable and that it does not describe how and where the consumed resources are allocated in the spectrum. Skipping the combinatorial problem of PRB assignment, our approach allows for computationally efficient modeling of the HetNet interference situation, particularly when highly loaded cells operate near the limit of available bandwidth.

C. System Performance Measures

For system performance assessment, we consider the measures (performance counters) that are listed in Table I. The measures either refer to the assessment of the related *key performance index (KPI)* at one time step or they describe the performance for the whole observation period, i.e., typically for one operation cycle. The metrics for system optimization refer to the overall observation period only. This is reasonable since the considered optimization parameters are not varying over an operation cycle. On the basis of the aforementioned UE attributes, we define the measures from Table I as follows.

For optimization, we apply the *cumulated priorities of covered UEs* as a metric to describe system coverage performance. Generally, a UE is covered if it experiences a certain minimum SINR. We consider the *sum rate of served UEs* as a metric

TABLE I
SYSTEM PERFORMANCE MEASURES

Measure (Metric)	Related KPI	Reference period
System operation (simulation)		
Cumulated prior. of covered UEs	Coverage	Observation period
Cumulated prior. of served UEs	Capacity	Observation period
Sum rate of served UEs	Capacity	Observation period
Coverage indicator	Cov. & Cap.	One time step
Capacity indicator	Capacity	One time step
System optimization		
Cumulated prior. of covered UEs	Coverage	Observation period
Sum rate of served UEs	Capacity	Observation period

to maximize system capacity. Here, a UE is served if the link quality and the available resources at the assigned transmitter are large enough to provide at least a certain minimum data rate to the UE. Both metrics are used in the CCO objectives in Section III-C.

For system simulation, we define UE coverage identically as for the optimization. We consider the average of cumulated priorities of covered UEs over all time steps as the *cumulated priorities of covered UEs* for an observation period. This measure might be compared directly with the optimization metric. Analogously, we compute the *cumulated priorities of served UEs* for an observation period. During simulation, some covered UEs may not be served if there is not enough bandwidth for serving all users jointly. Hence, if the latter measure is smaller than the cumulated priorities of covered UEs, it indicates capacity problems. The *sum rate of served UEs* is averaged over the observation period to facilitate comparison with the optimization metric.

The measures defined so far are aggregated over time, i.e., one value assesses the related KPI for the whole observation period. To investigate the time-variant performance behavior during simulation, we introduce the following two *indicators* that describe the coverage and capacity status at a certain time step. We consider

$$\frac{\text{no. of served calls}}{\text{no. of served calls} + \text{no. of unattended calls}} \tag{3}$$

as a *coverage indicator*. In contrast to the coverage measure from before, here, a UE is defined as covered if its call, i.e., its service-related data rate demand, can be served by an (H)eNB. Hence, indicator (3) takes into account the actual user

distribution and the specific user rate demands and the present resource allocation at (H)eNBs. Correspondingly, the *unattended calls* comprise all users that experience an insufficient SINR (noncovered users) and the users that cannot be served due to a lack of available transmission bandwidth.

Since we consider a system that applies interference mitigation techniques, we distinguish a bandwidth that is utilizable (free) for serving users from a bandwidth that is blocked to mitigate interference to other cells. Thus, the *free bandwidth* is given by the difference between the *total available transmission bandwidth* and the *blocked bandwidth*. Defining the ratio of the bandwidth that is utilized for serving users and the free bandwidth as a *utilization ratio* at an (H)eNB, we might consider

$$1-\frac{\text{bandwidth utilized for serving users}}{\text{free bandwidth}}=1-\text{utilization ratio}$$

as an indicator for the cell capacity status. This indicator can become quite low although only a few users in the cell are served. That effect arises if the amount of bandwidth that is blocked to mitigate interference to UEs in other cells gets very large, and consequently, the utilization ratio increases. Since this measure might overemphasize the influence of surrounding cells in terms of the blocked bandwidth, we consider

$$1-\omega\ \text{utilization ratio}-(1-\omega)\frac{\text{free bandwidth}}{\text{total bandwidth}} \qquad (4)$$

as a *cell capacity indicator* instead. Keeping in mind that the free bandwidth is defined as the total bandwidth minus the blocked bandwidth, the last term adds a partial amount of blocked bandwidth—according to the chosen parameter ω—as contribution to the cell capacity. This partial amount of blocked bandwidth might be interpreted as *potential capacity*.

Since algorithms for *radio resource management (RRM)* typically aim at serving all requested user rates, we assume that the maximal utilization of available bandwidth is equivalent to the maximization of the served sum rate when cells are highly loaded.

In addition to the KPI measures previously introduced, spectral efficiency might be be taken into account to assess system capacity. However, we do not consider that measure since our implemented RRM algorithms are not intended to maximize spectral efficiency, and consequently, the simulation results would not show any improvement that is (potentially) achieved by applying our CCO approach.

III. Self-Optimization of Coverage and Capacity

Considering the MTN framework described in Section II, we apply the following approach to joint *CCO*. The optimization model comprises three key components.

1) The *objective function* represents the KPI metrics that shall be maximized by tuning according control parameters, i.e., the optimization variables.
2) The optimization *constraints* model system dependence relations and system restrictions, mostly in a technical sense.
3) Traffic-related *input parameters* describe spatial radio conditions for different system configurations and the distribution of user rate demand.

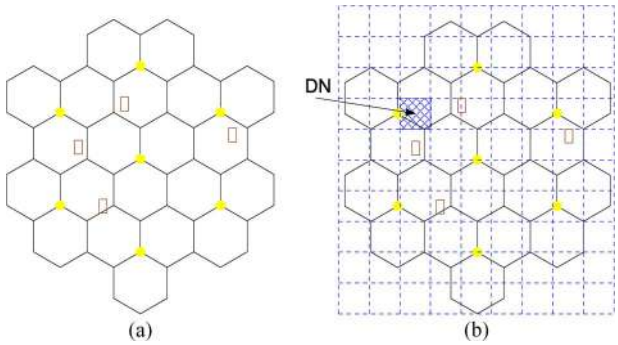

Fig. 2. DN generation principle. (a) Considered network area with (rectangles) buildings. (b) Area is divided into patches of equal size.

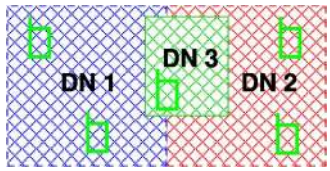

Fig. 3. DN separation for (relevant) buildings.

The spatial radio conditions are obtained by the RSP and link quality computation according to Section II-A. Alternatively, this information might be derived from system observations and UE measurements, e.g., according to the *X-map estimation* approach proposed in [20].

A. Demand Prediction Model

To take into account the (future) situation of UE locations and demands, we adapt the *demand node (DN) concept* in [37]. It provides an approach to abstract from single UEs and their mobility to reduce the number of UEs that have to be considered. *DNs* model the spatial distribution of aggregated UEs and their joint rate demand and priority with respect to a certain reference time period. This concept is very useful when computation time and memory are critical resources and UE abstraction is reasonable, e.g., for transmitter location planning or anticipative network configuration. DN distribution, priority, and rate demand are very important parameters that have to be accurately chosen to model the (future) de facto behavior. Generally, DNs can be extracted from information provided by operators or according to simulation statistics.

For proof-of-concept purposes, we generate DN parameters according to the derived simulation statistics. Figs. 2 and 3 show how DNs are created: First, the network area in Fig. 2(a) is divided into equal patches according to Fig. 2(b). Initially, each patch corresponds to a DN. Since indoor users are of special interest, buildings are represented by their own DNs and are consequently separated from surrounding DNs. Fig. 3 visualizes this principle for two patches that cover a building; separation here leads to three resulting DNs. DN priority p_t and DN rate demand r_t can be generated as follows: Considering value v (p_t or r_t), we assume knowledge of observation v^{past} from previous operation cycle(s) and information v^{future} for the next operation cycle, e.g., from historical data. Such data might include individual forecast information provided by the operator monitoring center and consider related prediction

TABLE II
INPUT PARAMETERS AND OPTIMIZATION VARIABLES

Symbol & domain	**Parameter** description
$\mathcal{S}^{\mathrm{MC}}$, $\mathcal{F}^{\mathrm{FC}}$, $\mathcal{T}$	Index set of eNBs (MCs), HeNBs (FCs), DNs.
$\mathcal{S}$, $\mathcal{F}$	Index set of different (H)eNB configurations.
$s \in \mathcal{S}, f \in \mathcal{F}$, $a \in \mathcal{S} \cup \mathcal{F}, t \in \mathcal{T}$	Representative indices for an eNB, HeNB, either eNB or HeNB, and DN.
r_t, $r_t^{\mathrm{MIN}} \in \mathbb{R}_{\geq 0}$	Requested data rate at the DN and minimum required data rate for serving the DN.
$p_t \in \mathbb{R}_{>0}$	DN priority level, related to either the service type or the customer type.
B^{MC}, $B^{\mathrm{FC}} \in \mathbb{R}_{\geq 0}$	Total transmission bandwidth available at (H)eNBs, constant for all possible configurations.
e_{st}, $e_{ft} \in \mathbb{R}_{\geq 0}$	Supported spectral efficiency from eNB to DN and from HeNB to DN, computed according to Section II-A.
e_{MIN}, $e_{\mathrm{MAX}} \in \mathbb{R}_{>0}$	Minimum required spectral efficiency for transmission (CQI 1) and maximally achievable spectral efficiency in the system.
$\delta_{\mathrm{ITF}} \in \mathbb{R}_{>0}$	Minimum (SINR) threshold for pairwise spectral efficiency ratios.
$p_a^{\mathrm{MAX}} \in \mathbb{R}_{>0}$	Maximum feasible sum of covered priorities at the (H)eNB a.
Symbol & domain	**Variable** description
y_s, $y_f \in \{0,1\}$	Decision variables for the selected configuration at (H)eNBs.
z_{st}, $z_{ft} \in \{0,1\}$	Decision variables for transmission related assignment of DNs to (H)eNBs.
$\tilde{z}_{st}$, $\tilde{z}_{ft} \in \{0,1\}$	Decision variables for coverage related assignment of DNs to (H)eNBs.
$\tilde{z}_t, z_t \in \{0,1\}$	Decision variables indicating DN coverage and DN assignment for data transmission.
b_{st}, $b_{ft} \in \mathbb{R}_{\geq 0}$	Amount of allocated bandwidth at (H)eNBs for serving assigned DNs.
b_s^{ITF}, $b_f^{\mathrm{ITF}} \in \mathbb{R}_{\geq 0}$	Amount of blocked bandwidth at (H)eNBs to cope with interference to other cells.
b_s, $b_f \in \mathbb{R}_{\geq 0}$	Amount of overall consumed bandwidth (service plus blocked) at (H)eNBs.
$r_t^{\mathrm{SRV}} \in \mathbb{R}_{\geq 0}$	Effectively served data rate at the DN, depending on the corresponding signal quality and the particular bandwidth allocation.

algorithms [38] or information from the marketing department regarding exceptional events such as the launch of new services [16]. Introducing *reliability indicator* $\mu \in [0,1]$ that describes the level of confidence we have in prediction accuracy, we compute

$$v = (1-\mu)v^{\mathrm{past}} + \mu v^{\mathrm{future}}.$$

Prediction accuracy is evaluated by monitoring procedures during the operation cycles. In case of significant differences between predictions and realizations, the DN generation process can be adapted accordingly.

B. Notation

We introduce notations according to Table II to describe CCO-related input parameters and optimization variables. For each selectable configuration at an eNB (MC) $i \in \mathcal{S}^{\mathrm{MC}}$ or HeNB (FC) $j \in \mathcal{F}^{\mathrm{FC}}$, we create element s or f in configuration index set $\mathcal{S}$ or $\mathcal{F}$, respectively. Setting related variable y_s or y_f to one corresponds to the selection of the associated configuration. We denote the set of configuration indexes that belong to the same eNB $i \in \mathcal{S}^{\mathrm{MC}}$ or HeNB $j \in \mathcal{F}^{\mathrm{FC}}$ by $\mathcal{C}_i$ and $\mathcal{C}_j$, respectively. Hence, it holds that $\mathcal{S} = \bigcup_{i \in \mathcal{S}^{\mathrm{MC}}} \mathcal{C}_i$ and $\mathcal{F} = \bigcup_{j \in \mathcal{F}^{\mathrm{FC}}} \mathcal{C}_j$. The configuration of eNBs provides *transmission power* and *antenna downtilt* as control parameters, whereas HeNB configuration considers the *transmission power* only. To ensure the selection of exactly one configuration at each transmitter, we generally apply

$$\sum_{s \in C_i} y_s = 1,\ i \in \mathcal{S}^{\mathrm{MC}}\ (\text{exactly one config. per eNB}) \quad (5)$$

$$\sum_{f \in C_j} y_f = 1, j \in \mathcal{F}^{\mathrm{FC}}\ (\text{exactly one config. per HeNB}). \quad (6)$$

For the sake of simplicity, we use notation *transmitter* $s \in \mathcal{S}$, $f \in \mathcal{F}$ although this actually means a certain configuration at the corresponding transmitter. Since many expressions in the following are related to transmitters $s \in \mathcal{S}$ and to transmitters $f \in \mathcal{F}$, we introduce wildcard symbol $a \in \mathcal{S} \cup \mathcal{F}$ for short notation. Utilizing the DN concept in Section III-A to abstract from single UEs, it is facilitating to interpret DNs as UEs: Each DN t requests a certain data rate r_t and is associated with priority p_t, as described in Section II. Serving a DN requires its assignment and a sufficiently large amount of available resources to fulfill at least its minimum data rate demand r_t^{MIN}. As resources, we consider bandwidth b_{at} that has to be allocated for serving t, and that is computed according to (1).

Since we want to distinguish coverage-related terms from capacity-related terms, we consider decision variables $\tilde{z}_{st}$ and $\tilde{z}_{ft}$ that indicate which DN is covered by what (H)eNB and decision variables s_{st} and z_{ft} that describe the (H)eNB-to-DN assignment for transmission. Variables b_{st} and b_{ft} represent the *amount* of bandwidth that is allocated at the (H)eNB to serve DN t. We define

$$\begin{aligned}
\mathcal{S} * \mathcal{T} &= \{(s,t) \in \mathcal{S} \times \mathcal{T} : e_{st} \geq e_{\mathrm{MIN}}\} \\
\mathcal{F} * \mathcal{T} &= \{(f,t) \in \mathcal{F} \times \mathcal{T} : e_{ft} \geq e_{\mathrm{MIN}}\} \\
\mathcal{S}_t &= \{s \in \mathcal{S} : (s,t) \in \mathcal{S} * \mathcal{T}\} \\
\mathcal{F}_t &= \{f \in \mathcal{F} : (f,t) \in \mathcal{F} * \mathcal{T}\} \\
\mathcal{T}_s &= \{t \in \mathcal{T} : (s,t) \in \mathcal{S} * \mathcal{T}\} \\
\mathcal{T}_f &= \{t \in \mathcal{T} : (f,t) \in \mathcal{F} * \mathcal{T}\}
\end{aligned}$$

to exclude variables for combinations that are irrelevant due to an insufficient link quality (CQI 0). Finally, $\mathcal{T}_i = \bigcup_{s \in C_i} \mathcal{T}_s$ describes all DNs that can be assigned to eNB $i \in \mathcal{S}^{\mathrm{MC}}$ for at least one configuration of i. Parameter e_{MIN} is chosen such that each transmission link supports CQI 1 or higher. Likewise, e_{MAX} is set to the maximally achievable spectral efficiency in the system. Both eNBs and HeNBs can utilize a maximum transmission bandwidth of $B^{\mathrm{MC}} = B^{\mathrm{FC}}$. To specify reasonable values for the maximum feasible sum of covered priorities at transmitters, we consider average user priority $\bar{p}_a$ that might be derived from (simulation) statistics. Furthermore, we define a *coverage-related minimum target rate* r^{COV}. Assuming an average spectral efficiency e_a for each entity of a transmitter category (eNB or HeNB) and an average available transmission bandwidth B_a, we suggest setting the maximum feasible sum of covered priorities to

$$p_a^{\mathrm{MAX}} = \frac{B_a e_a}{r^{\mathrm{COV}}} \bar{p}_a. \quad (7)$$

C. Joint Coverage and Capacity Maximization

We consider

$$\max\left\{\lambda_{\mathrm{COV}}\sum_{t\in\mathcal{T}}p_t\tilde{z}_t+\lambda_{\mathrm{CAP}}\sum_{t\in\mathcal{T}}r_t^{\mathrm{SRV}}\right\} \tag{8}$$

over variables y_a, $\tilde{z}_{at}$, z_{at}, b_{at}, b_a^{ITF}, b_a, $\tilde{z}_t$, z_t, and r_t^{SRV} as an objective function for *Joint Coverage* and *Capacity MAXimization (JoCoCaMAX)*, where $\tilde{z}_t$ is the binary indicator for coverage of DN t, and r_t^{SRV} describes the effectively served rate [see (10) and (14)]. According to the discussion in Section II-C, we assume the maximization of covered priorities to maximize coverage and the maximization of the effective sum rate to maximize capacity. By introducing weighting factors λ_{COV} and λ_{CAP}, we apply a *scalarization* approach to cope with the multiobjective optimization problem [24].

Generally, coverage of a DN and its priority value by an (H)eNB a requires the selection of the particular transmitter and configuration combination. Furthermore, covered DNs are counted only once in the objective function, i.e., constraints (5), (6), and

$$\tilde{z}_{at}\le y_a, \qquad \forall(a,t)\in(\mathcal{S}\cup\mathcal{F})*\mathcal{T} \tag{9}$$

$$\tilde{z}_t=\sum_{s\in\mathcal{S}_t}\tilde{z}_{st}+\sum_{f\in\mathcal{F}_t}\tilde{z}_{ft}\le 1, \qquad \forall t\in\mathcal{T} \tag{10}$$

must hold. We model the minimum SINR condition for coverage by transmitter a as a minimum link quality constraint, i.e.,

$$\frac{e_{at}}{e_{a't}}\ge(\tilde{z}_{at}+y_{a'}-1)\delta_{\mathrm{ITF}}, \quad \begin{matrix}\forall(a,t)\in(\mathcal{S}\cup\mathcal{F})*\mathcal{T}\\ \forall a'\in\mathcal{S}_t\cup\mathcal{F}_t\end{matrix} \tag{11}$$

with respect to all potentially interfering transmitters a' and where the constraint becomes a tautology for all decision variables $y_{a'}$ that equal zero. Note that for $\lambda_{\mathrm{ITF}}=1$, we allow best link coverage only.

Since we want to obtain feasible solutions also for the case that all covered DNs (UEs) become active and request some data rate, we limit the sum of maximal coverable priorities at transmitters by

$$\sum_{t\in\mathcal{T}_a}p_t\tilde{z}_{at}\le p_a^{\mathrm{MAX}}, \qquad \forall a\in\mathcal{S}\cup\mathcal{F} \tag{12}$$

where p_a^{MAX} is predefined according to (7) and might differ for various transmitters.

For the capacity (transmission)-related decision variables z_{st} and z_{ft}, we consider constraints analogously to (9) and (10). Generally, DN rate demand can be served by either an MC or an FC, allocating the required amount of bandwidth b_{at} that is computed according to (1). The maximally served rate is limited by the actual requested rate r_t, i.e.,

$$b_{at}\le\frac{r_t}{e_{at}}z_{at}, \qquad \forall(a,t)\in(\mathcal{S}\cup\mathcal{F})*\mathcal{T} \tag{13}$$

but the total DN rate demand does not necessarily have to be fulfilled. However, the *effectively served rate*

$$r_t^{\mathrm{SRV}}=\sum_{s\in\mathcal{S}_t}e_{st}b_{st}+\sum_{f\in\mathcal{F}_t}e_{ft}b_{ft}\ge r_t^{\mathrm{MIN}}z_t, \qquad \forall t\in\mathcal{T} \tag{14}$$

has to exceed at least the minimum rate requirement r_t^{MIN}; otherwise, the DN cannot be assigned to a serving station. This problem can arise if the available (remaining) bandwidth resources at potentially serving (H)eNBs are not sufficient due to the limitation constraint, i.e.,

$$b_a=\sum_{t\in\mathcal{T}_a}b_{at}+b_a^{\mathrm{ITF}}\le B+(1-y_a)\cdot\infty, \quad \forall a\in\mathcal{S}\cup\mathcal{F} \tag{15}$$

where we assume $B=B^{\mathrm{MC}}=B^{\mathrm{FC}}$, and the infinity term on the right-hand side is introduced to avoid feasibility problems for nonselected configurations. The interference-related bandwidth consumption b_a^{ITF} is computed according to the interference approximation model discussed in Section II-B, i.e., all FCs $f\in\mathcal{F}$ block an amount of bandwidth

$$b_f^{\mathrm{ITF}}=\sum_{\substack{(f',t)\in\mathcal{F}*\mathcal{T},\\ f'\neq f}}q_{ff't}^{\mathrm{ITF}}b_{f't} \tag{16}$$

to mitigate interference to users that are served by other FCs f'. Analogously, all MCs $s\in\mathcal{S}$ block

$$b_s^{\mathrm{ITF}}=\sum_{\substack{(s',t)\in\mathcal{S}*\mathcal{T},\\ s'\neq s}}q_{ss't}^{\mathrm{ITF}}b_{s't}+\max_{\substack{f\in\mathcal{F}_s,\\ y_f=1}}\left\{\frac{e_{sf}}{e_{\mathrm{MAX}}}b_f\right\} \tag{17}$$

where we assume that all DNs served by an FC and all inter-FC interfered FCs are located nearby the HeNB and that the interference impact factor in the cross-tier term, hence, scales with the eNB signal strength to the HeNB location.

Finally, the constraints

$$\sum_{t\in\mathcal{T}_i}p_t\tilde{z}_t\ge p_i^{\mathrm{MIN}}, \quad \sum_{t\in\mathcal{T}_i}r_t^{\mathrm{SRV}}\ge r_i^{\mathrm{MIN}}, \quad \forall i\in\mathcal{S}^{\mathrm{MC}} \tag{18}$$

ensure a minimum supported level of coverage and capacity at the considered eNBs. We apply these constraints during the optimization procedure to restrict the potential degradation of coverage and capacity. Particularly, preserving KPI values from previous optimization steps allows for a monotonous improvement of solutions in iteratively conducted CCO processing.

Since all considered terms are linearly formulated and the optimization variables are Boolean or from the positive continuous domain, the presented model is a *mixed-integer linear program (MILP)*. We utilize state-of-the-art MILP solvers such as *CPLEX* [39] or *Gurobi Optimizer* [40] to compute (optimal) CCO solutions.

D. Optimization Tradeoff

Generally, maximization of coverage and maximization of capacity are tradeoff tasks, and we apply a scalarization approach to solve the joint optimization problem.

Considering JoCoCaMAX for $\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}>0$ in (8), every computed solution is *Pareto optimal* [24]. Since the weight vector $(\lambda_{\mathrm{COV}},\lambda_{\mathrm{CAP}})$ gives the normal of the tangential hyperplane at the associated Pareto optimal point, the particular setting of the weighting factors determines which Pareto optimal points are found. If the coverage status and the capacity status are both at a sufficient level, we apply JoCoCaMAX with respect to the preservation of achieved KPIs according to (18) and

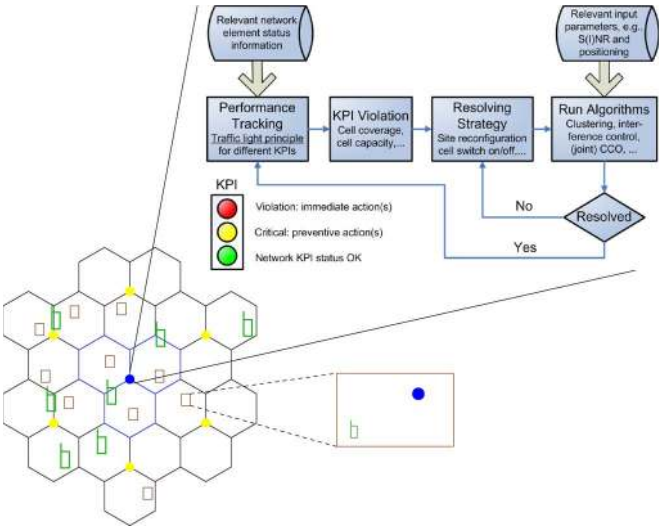

Fig. 4. Operation scheme for an SO-MTN, considering (hexagons) eNB MCs, (rectangles) HeNBs in buildings, and (mobiles) UEs.

considering customizable weightings λ_{COV}, $\lambda_{\mathrm{CAP}} > 0$ to further improve both KPIs.

Otherwise, we apply the following approach to cope with insufficient KPIs whenever one or both KPIs are significantly degraded: By choosing according weighting factors λ_{COV}, $\lambda_{\mathrm{CAP}} \in \{0, 1\}$ for *single target maximization* [24] of a primary (worst case) KPI, we counteract insufficient system performance in a hierarchical manner. Nevertheless, we restrict the potential degradation of each achieved KPI status in (18) by defining parameters r_i^{MIN} and p_i^{MIN} accordingly for all eNBs $i \in \mathcal{S}^{\mathrm{MC}}$. This leads to the variants *Restricted Coverage MAXimization (RCovMAX)* and *Restricted Capacity MAXimization (RCapMAX)*, which both optimize a single (worst case) eNB with respect to the preservation of achieved KPI performance for all other eNBs. This approach is an integral part of the optimization procedure described in the following section.

E. Autonomous Traffic-Light-Based System Control

Considering an MTN in its operational phase, we propose a scheme for autonomous self-organized system control and optimization guided by Fig. 4: All MTN cells monitor the following relevant network status information, which serve as input to track system performance in terms of the KPIs introduced in Section II-C:

- number of served calls and number of unattended calls;
- requested data rate demand and UE priorities;
- utilized and blocked bandwidth.

The observation of insufficient system performance might automatically trigger optimization procedures, i.e., aperiodic optimization. Such a mechanism requires complex routines for event detection and network status classification, particularly, it has to be defined for how long a detection phase has to be to provide reliable detection results. Hence, in this paper, we apply time-triggered optimization, i.e., periodic optimization.

The proposed MTN control and optimization scheme is applicable on single eNB level, on a cluster level, or on a full network level. We define a *cluster* as a group of eNBs such that the strongest interferers to the contained eNBs are

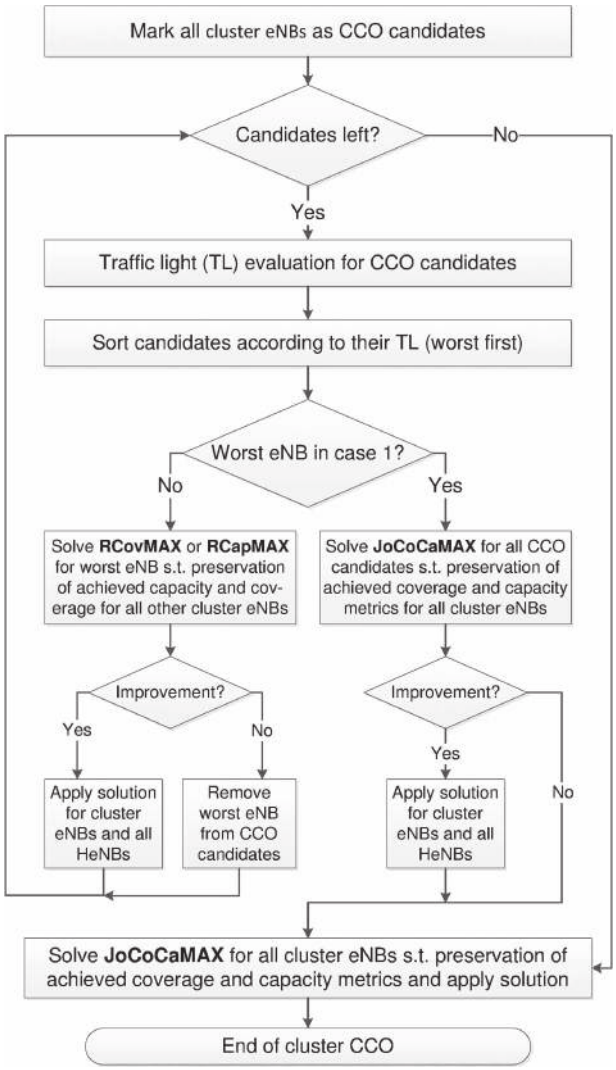

Fig. 5. Processing scheme for cluster-wise closed-loop autonomous optimization of coverage and capacity. The related traffic light cases and the according optimization preferences are shown in Fig. 6.

elements of the group. Since the interference situation depends on the particular user and traffic distribution, clustering is performed as an entry point of the CCO routines and utilizes the available DN information and the interference approximation model in Section II-B. For sufficiently long trigger intervals, it makes sense to apply CCO at least in a semicentralized way, i.e., *cluster-wise* established control instances (master units) collect all necessary information and *perform optimization centrally* for their cluster. We expect that this approach provides a beneficial tradeoff between computational complexity and optimization quality. The chance to find the global optimum for the network, however, increases with the internal cardinality of the considered clusters.

If the optimization loop is triggered, the following consecutive subroutines are executed.

1) Generate predictions for DN priorities and DN rate demands according to Section III-A.
2) Compute input parameters for CCO and for the interference approximation model with respect to the generated DN information.

Case	KPIs	Green	Yellow	Red	Primary Maximization Target
1	Coverage Capacity	● ●			Joint Coverage and Capacity
2	Coverage Capacity	●	●		Capacity
3	Coverage Capacity		● ●		Coverage
4	Coverage Capacity		● ●		Coverage
5	Coverage Capacity	●		●	Capacity
6	Coverage Capacity		●	●	Capacity
7	Coverage Capacity	●		●	Coverage
8	Coverage Capacity		●	●	Coverage
9	Coverage Capacity			● ●	Coverage

Fig. 6. Traffic-light-related performance cases (best first) and corresponding optimization preferences.

3) Apply the traffic-light-related CCO loop in Fig. 5, considering the optimization preferences in Fig. 6.
4) If the CCO results (potentially) improve the MTN performance, the corresponding optimal configurations are applied at all (H)eNBs.

Particularly, the first two steps autonomously adapt the optimization-related parameters every time optimization is triggered. Since all further steps depend on the DN information generated in step one, this task is very important. Fig. 5 shows the optimization loop that is applied to implement the traffic-light-related CCO: We first compute the KPI traffic lights according to the JoCoCaMAX evaluation for the currently selected (H)eNB configurations at CCO candidates and with respect to the generated DN information. The selected configurations are fixed for the evaluation process, which leads to a significantly reduced solution space and makes this step very fast. If the KPI-related optimization metric exceeds a predefined traffic light threshold, the resulting light indicates a green, yellow, or red KPI status. Note that this is the expected future traffic light status if the DN information includes the prediction component, as introduced in Section III-A. The coverage and capacity metrics obtained from the evaluation step define the feasible settings for the minimum required number of covered DN priorities p_i^{MIN} and the minimum required amount of served rate r_i^{MIN} for all evaluated eNBs.

Depending on the particular performance status, we choose the according optimization strategy and the related optimization parameters: We apply the hierarchical optimization strategy mentioned in Section III-D whenever one KPI level is significantly degraded, i.e., for status 2–9 in Fig. 6. We suggest to prefer coverage as the primary KPI whenever the performance of this KPI is not indicated by a green traffic light. Particularly, for status 7–9 (insufficient coverage), we allow the capacity in solutions to potentially degrade to a red traffic light in favor of having the maximal degree of freedom available for coverage maximization. This concept is implemented by choosing the minimum parameters for constraints (18) accordingly. The optimization loop shown in Fig. 5 tries to maximally improve the worst performing eNB according to a *climbing-up principle* and removes it from the CCO candidate list if an improvement is not possible. Finally, JoCoCaMAX is applied one time subject to the preservation of the achieved coverage and capacity metrics. After this step, the optimization loop terminates.

Generally, the KPI performance of all eNBs is preserved in each optimization step by considering constraints (18) with respect to the KPI metrics achieved in the evaluation step for current configurations. Only for single eNB optimization, the secondary KPI might degrade to a lower level, but the degradation is bounded by a status-related minimum value. For objective function (8), we do not consider eNBs that are not an element of the CCO candidate list or not the worst eNB. However, a minimum performance of those eNBs is guaranteed by applying the constraining approach previously described.

IV. Simulation Results

We demonstrate performance and behavior of the proposed concepts and models for autonomous self-organized optimization of coverage and capacity by simulative evaluation of representative case studies. Our intention is particularly to provide a proof of concept by demonstrating applicability and achievable performance gains. With respect to this purpose, we consider a reduced set of possible transmitter configurations to keep the complexity appropriate, and we apply our approach to one cluster, i.e., we consider the whole network as one cluster. Investigation of further improvements by considering a larger configuration state space and the analysis of potential performance loss due to more clusters are open for future work.

A. RRM and Scheduling

For our simulation framework, we implemented RRM algorithms that are sufficient for proof-of-concept purposes. Particularly, our RRM is not intended to provide all features that RRM typically comprises [41]. Our RRM implementation covers admission control, (H)eNB station assignment, PRB allocation, rate allocation, and interference mitigation.

We apply the following model to avoid the scheduling of resources over time (time scheduling): PRBs are the smallest elements of resource allocation assigned by the (H)eNB RRM. A PRB comprises 12 subcarriers (SCs) with a 15-kHz bandwidth each. PRBs are nonconsecutively assigned to UEs such that the UE rate demands are fulfilled. UEs with a low rate demand may not require a full PRB for a certain period of time, but they need one PRB from time to time. We model this time scheduling by frequency scheduling, i.e., we allow separate allocation of SCs over an extended time duration and, hence, consider finer granularity in the frequency domain. This approach is exemplarily motivated in Fig. 7, where we assume that serving UE1 requires the allocation of a 45-kHz transmission bandwidth. This is realized by an assignment of every fourth PRB, where T describes the smallest time interval before PRB reallocation is possible. We transform this scheduling over time into an assignment in the frequency domain such that UE1 gets the first, fifth, and ninth subcarrier instead of the first, fifth, and ninth PRB. This assignment is equivalently realizable in the time domain for choosing the extended time duration as

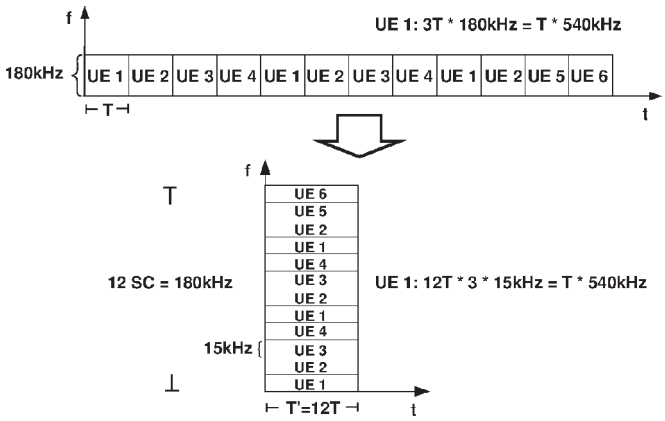

Fig. 7. Modeling scheduling over time by subcarrier allocation.

$T' = 12T$ and if we assume T' to be sufficiently small such that channel conditions do not change (significantly) over this period.

The following RRM processing loop is independently applied for each simulation step.

1) Greedy Initialization: First, an initial assignment of users to cells is performed. This is done by a simple greedy algorithm that assigns each user to the cell with the best pilot SINR, assuming a full interference situation.

2) Reassignment and Drop Users: In general, the user distribution is nonuniform, and therefore, cells experience different loads. This fact might even lead to overloaded cells, i.e., not all users assigned to such a cell can be served. In that case, we try to reassign the users that are not served to neighboring cells that are not overloaded. Users that do need a small amount of additional resources are preferred for reassignment. This procedure is repeated until there are no overloaded cells or if there are no further candidates for reassignment. If there are no candidates left but there still exists an overloaded cell, some users in that cell need to be dropped. The first execution of this block assumes no interference, whereas all subsequent calls take into account the current interference situation.

3) Resource Reservation: Interference mitigation is conducted by asking neighboring cells not to use particular resources (subcarriers). This procedure is applicable for real systems since users typically know the neighboring cells and their related signal strength. They share this information with their serving cell such that the cell might trigger a request for resource reservation if needed. For application to real systems, the communication overhead can be kept low by assessing the potential benefit of the reservation before asking the neighboring cells to block resources. However, for our simulation purposes, this can be neglected. Overall, this step allows for the elimination of interference to the reserved resources, i.e., subcarriers, from (some) neighboring cells.

4) Resource Allocation: Using the information from the preceding steps, the conditionally optimal resource allocation is performed for all cells in parallel. Parallel execution is possible since usage and reservation of resources are fixed at this stage. If it turns out that no feasible allocation exists, more users need to be reassigned or dropped, and all according steps are repeated.

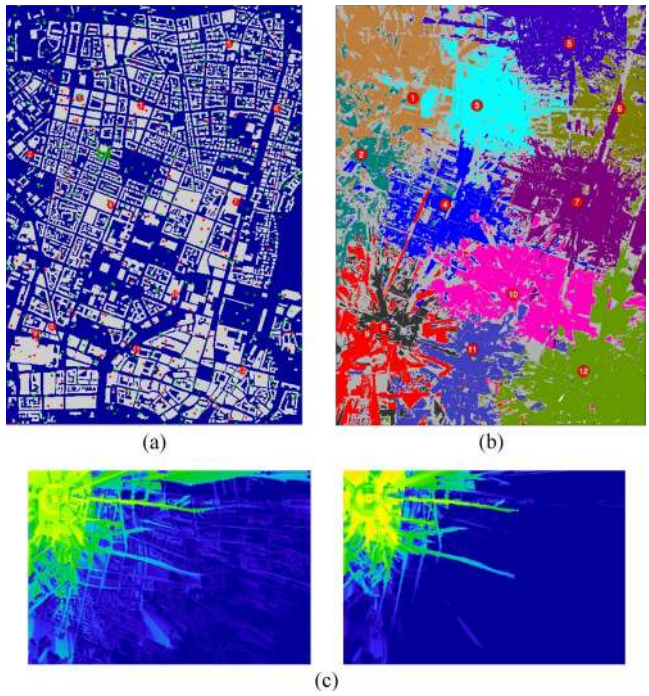

Fig. 8. Considered MTN in Munich, Germany, and visualization of signal strength (RSP) distribution for two different antenna configurations of an eNB. (a) Considered building map and MTN transmitters. (b) Initial MCs; light-gray areas indicate handover zones. (c) eNB applying (left) 0° and (right) 5° antenna downtilt.

B. Simulation Setup

We consider an LTE MTN according to Section II and apply an RRM and scheduling implementation, as described in the previous section. As target network area, we choose the urban environment of Munich, Germany, that is shown in Fig. 8(a): eNB antennas (large red circles) are mounted on top of buildings (gray polygons), and HeNBs (small red circles) are located indoors; users are depicted in small green circles and can move outdoors as pedestrians or vehicles or they stay inside of a building. Table III specifies the relevant system and simulation parameters. The 12 eNB site locations are derived from a previously conducted MTN deployment stage; the according (initial) MCs are shown in Fig. 8(b). The light-gray areas indicate potential handover zones, i.e., the strongest two signals do not differ by more than 2 dB. HeNBs are randomly distributed over buildings, following a uniform distribution while keeping a minimum distance of 35 m to the surrounding eNBs. The reference period associated with one simulation step is 1 s. This is particularly relevant for UE service requirements, UE mobility, and RRM procedures.

UEs can be located indoors and outdoors and show dynamic behavior in terms of rate demand and mobility. UEs enter the scenario with a randomly chosen lifetime such that the expectation of active UEs stays constant during simulation. After a UE has exceeded its lifetime, it disappears, and new UEs might enter the scenario. The UE *traffic profile* describes the requested mobile service, i.e., the type of service, the data rate demand to meet the QoS requirements, and the priority level. Table IV shows the considered services and their proportion in the overall traffic. Basically taken from the recommendations in

TABLE III
SYSTEM PARAMETERS FOR SIMULATION

System parameter	Setting
Simulation area	2.5 km×3.5 km, 5 m resolution
Carrier frequency	2 GHz
Effective transmission bandwidth	9 MHz (50 LTE PRBs)
Shadowing standard deviation	8 dB
Wall penetration loss	10 dB
Number of eNBs (macrocells)	12
eNB Tx power profile	{$-\infty$ (off), 40 (low), 43 (medium), 46 (high)} dBm
eNB antenna downtilt profile	{0° (low), 5° (high)}
Propagation model eNB	ray optical, omnidirectional
eNB antenna gain and noise figure	14 dBi and 5 dB
Number of HeNBs (femtocells)	200
HeNB Tx power profile	{$-\infty$ (off), 3 (normal), 10 (high)} dBm
Propagation model HeNB	Free space + wall penetration, omnidirectional
Noise figure HeNB	8 dB
HeNB distribution	random, uniformly distributed over buildings
Minimum distance HeNB to eNB	35 m
Average number of UEs	215
Number of hotspot UEs	60
UE outdoor-to-indoor ratio	80%
Vehicular-to-pedestrian ratio	20%
UE initial distribution	random, uniformly distributed
UE antenna gain and noise figure	0 dBi and 9 dB

TABLE IV
USER TRAFFIC PROFILE

Service	Priority	Data rate [kbit/s]	Proportion [%]
VoIP	1	$64-128$	40
Web	1	$128-512$	50
Data	1	$128-2000$	10

[42], we modified the traffic profile, aiming at the generation of an MTN that operates at the limits of its capabilities most of the time, and that is temporarily overloaded. The effective service distribution is chosen according to the proportions in Table IV. Since we consider priority one for all services, the total sum of coverable priorities equals the number of active UEs in the system.

The applied UE *mobility model* determines how a UE moves across the simulation area over time. Indoor UEs move according to a *random walk model* and do not leave the building. Outdoor UEs are either pedestrians that move according to the random walk model or vehicles that follow a *random waypoint model*. In the considered urban scenario, the waypoints are modeled by an appropriate set of road points that describe the irregular course of the roads. At each crossing point, the UE randomly chooses the direction to go on; the way back is excluded from that decision. All related mobility parameters are listed in Table V.

To generate locally bounded overload situations, we introduce a *moving traffic hotspot model*: Specifying a certain number of hotspot UEs and a hotspot radius, those UEs are located circularly around a *hotspot center UE* and stay within the given radius. The hotspot center UE is a vehicle UE that moves along roads according to a configurable speed. All hotspot UEs become active when the predefined hotspot activity period starts and change their state to inactivity if the hotspot period is over. The considered hotspot moves along roads in the middle part of the MTN area.

TABLE V
USER MOBILITY PARAMETERS

Parameter	Setting
Time per simulation step	1 s
Pause time	Deterministic at 0 s
Random walk direction	Uniform $[0, 2\pi)$
Velocity pedestrian/indoor	Uniform $[0, 1.5]$ m/s
Velocity vehicle	Uniform $[10, 20]$ m/s
Velocity traffic hotspot	3 m/s

TABLE VI
OPTIMIZATION PARAMETERS

Parameter	Setting
DN prediction reliability μ	1
Yellow and red coverage TL threshold	0.99 and 0.95
Yellow and red capacity TL threshold	0.98 and 0.95
SINR coverage threshold δ_{ITF}	0.6
Coverage bound for MCs p_s^{MAX} and FCs p_f^{MAX}	125 and 15
Interference impact factors $q_{aa't}^{\mathrm{ITF}}$	1
Utilization weight ω	0.9

C. Simulation and Optimization Procedure

Dynamic UEs are simulated according to the previous section. With respect to the applied (H)eNB configurations, the implemented RRM algorithms are performed to serve UE rate demands. We compute all information that is required to evaluate the system performance measures according to Section II-C. Furthermore, we compute the average DN rate demands and the average DN priorities over the considered operation cycle for each serving station. This information serves as input data for the traffic-light-related system control from Section III-E. Overall, we consider round about 2400 DNs that are generated in a preprocessing step according to Section III-A using quadratic patches of 100 m. The DNs are subdivided into outdoor DNs and indoor DNs by roughly one third to two thirds.

For system optimization (CCO), we consider the following configuration state space: Each of the 12 eNBs can apply 0° (low) or 5° (high) *antenna downtilt* for an omnidirectional antenna pattern [see Fig. 8(c)]. Furthermore, each eNB can adjust its transmission power according to the available power profile that is specified in Table III. All 200 HeNBs provide just one antenna pattern (omnidirectional) and support the power profile that is specified in the according part of Table III. Hence, eNBs and HeNBs can select from a set of seven and three different configurations, respectively. We consider further optimization parameters that are listed in Table VI. Since we apply the CCO configuration results for identically repeated traffic simulations, we have a perfectly reliable traffic forecast, which corresponds to $\mu = 1$ in the DN model in Section III-A. This allows for investigation of the achievable performance gains, assuming best case input parameter selection. The traffic light thresholds for coverage and capacity refer to the optimization preferences

in Fig. 6. The coverage threshold parameter from Table II and the utilization weight in the capacity measure (4) are chosen as $\lambda_{\mathrm{ITF}} = 0.9$ and $\omega = 0.9$, respectively. The impact factors $q^{\mathrm{ITF}}_{aa't}$ introduced in Section II-B are set to 1, indicating that the frequency reuse potential for interference mitigation is assumed to be low in the dense urban scenario. Finally, we choose tradeoff weighting factors λ_{COV} and λ_{CAP}, such that they get the same weight if all traffic is covered and slightly adapt them with respect to the imbalanced preferences for coverage and capacity.

Note that optimization parameters are defined once in advance to the start of the closed-loop operation scheme and that they are fully customizable with respect to operator preferences. Further modification of parameters, however, is possible before each operation cycle, but it requires human intervention, which should take place only in emergency situations. With this parameter settings, the optimization loop is triggered, and an optimal configuration for that period is determined. After applying the obtained optimal configurations at (H)eNBs, we go back to the beginning of the considered operation cycle and repeat the UE simulation. Following this approach, we are able to provide a fair comparison between nonoptimized and optimized configuration results.

D. Proof-of-Concept Results

In the following, we present three exemplary case studies and briefly discuss the accuracy of our proposed interference approximation model. The first case study shows that in case of a short but strong traffic variation within an observation period, an optimization process with long-term perspective can fail to react on short-term events (degradation). By the second example, we demonstrate the benefit potential of our approach when the degradation periods are treated within suitably sized optimization intervals. However, this example also shows that even if a degradation period is not perfectly separated, there is still significant potential for performance improvement by appropriate reconfiguration. The third case study serves as an example to investigate the behavior of our approach if it has to cope with a coverage and capacity tradeoff situation. Finally, the last part of this section illustrates the accuracy of our interference approximation model.

1) Obliteration of Events (Degradation): In the first case study, we investigated CCO for an operation cycle of 1 h and a traffic hotspot activity of 5 min. According to Section IV-C, the full operation cycle was monitored before optimization was triggered. Afterward, the UE simulation was identically repeated but with optimized configurations. Comparing the behavior of the nonoptimized system with that of the optimized system shows the following. On average, we achieve almost no improvement by applying CCO. Contrariwise, the KPI deficiencies at eNBs in the sphere of influence of the hotspot slightly increase. The reason for this behavior is quite intuitive: By generating DN information that is averaged over 1 h, the influence of the 5 min of hotspot activity is obliterated, and hence, the system is mainly optimized for the period of hotspot inactivity. Consequently, this leads to performance degradation during the hotspot activity period. We conclude that averaging over periods with significant variation in user behavior has to be treated carefully. More precisely, we suggest that such periods should be separately considered if the duration of one period is considerably larger than the other period.

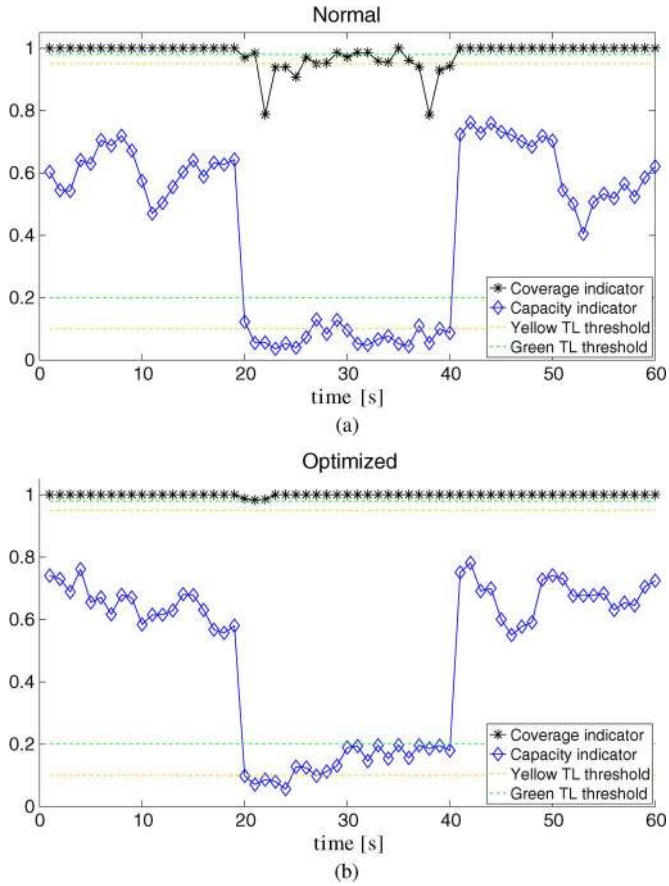

Fig. 9. Second-wise evaluation of the (upper) coverage indicator and the (lower) capacity indicator at the eNB (4) that is most affected by the traffic hotspot. The traffic hotspot is active from 20 to 40 s. (a) eNBs apply default configurations. (b) eNBs apply configurations according to CCO results.

TABLE VII
COVERAGE AND CAPACITY INDICATORS FOR THE eNB THAT IS MOST AFFECTED BY THE TRAFFIC HOTSPOT

Traffic light status	Default config.	CCO config.
Red coverage indicator	8	0
Yellow coverage indicator	8	0
Red capacity indicator	16	6
Yellow capacity indicator	5	15

2) Handling Temporary Degradation (Traffic Hotspot): For the second case study, we considered an operation cycle of 1 min and hotspot activity in the period from 20 to 40 s. The CCO procedure is triggered every 5 s such that periods of significantly different UE behavior are fully separated. Although realistic trigger periods and the hotspot duration are typically much longer than the considered values, we expect the results to scale with the ratio of hotspot duration and the length of trigger periods. However, scaling is limited by the case of full separation. Coverage and capacity indicators of the eNB that is most affected by the traffic hotspot (eNB 4) are illustrated for optimized and nonoptimized configurations in Fig. 9: The hotspot activity period is easy to identify since coverage and capacity

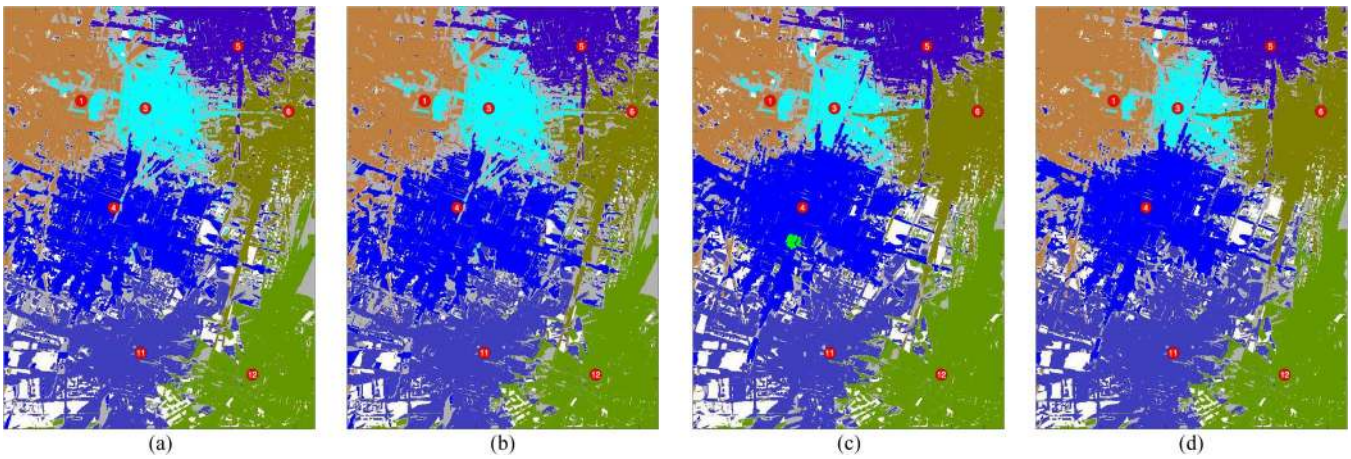

Fig. 10. CCO-related eNB cell footprints while handling coverage problems (switched off eNBs) and capacity problems (traffic hotspot). (a) Switched off eNBs cause coverage holes (white spaces). (b) Remaining eNBs predominantly compensate the degraded network coverage. (c) Reconfiguration with respect to a temporary traffic hotspot (bunch of spots below eNB 4). (d) Combination of coverage- and capacity-triggered reconfiguration.

indicators significantly decrease during that time. However, in the optimized case, the coverage indicator is uncritical, whereas the capacity indicator, although improving, remains in the critical zone. Considering the depicted KPI traffic light threshold lines, we count the number of time instances where the KPIs are above or below the corresponding lines and show the results in Table VII. The coverage problems are fully resolved by applying the optimized configurations, whereas the capacity deficiencies are lowered by lifting up the indicator from red to yellow for ten time steps. This improvement is mainly achieved by reducing the interference from eNBs that surround the eNB that serves the hotspot UEs. In the optimized system configuration, the related eNBs increase their antenna downtilt, and the serving eNB increases its transmission power to the maximum. Fig. 9 shows high volatility and, sometimes, clearly worse values for the capacity indicator in the optimized case compared with the nonoptimized case. Both effects are caused by our RRM implementation: First, volatility is explained by the fact that the RRM is restarted for each time instance without taking into account former assignments. Second, the RRM is not designed to maximize the capacity indicator but stops processing when a feasible assignment has been found. Thus, our RRM does not take full advantage of the CCO benefits as long as it has not to cope with problems in covering and serving UEs. In other words, the monitored KPI values of the nonoptimized system can outperform the values achieved by the optimized MTN but *only* when the nonoptimized system does not suffer from any KPI deficiencies.

3) Coverage and Capacity Tradeoff: Since in the second case study the unattended calls in (3) result mostly from a lack of available resources (bandwidth), the coverage indicator mainly accounts for capacity problems. Therefore, we present this case study that emphasizes coverage problems due to insufficient RSP: We switch off eNB 2 and eNBs 7–10 permanently to cause artificial coverage problems, particularly in the potential influence zone of eNBs 4, 11, and 12 [see Fig. 10(a)]. Otherwise, we keep the simulation setup from the second case study including hotspot activity in the period from 20 to 40 s. First, we evaluate the scenario for the initial configuration that is derived from MTN deployment (see Section IV-B). Second, CCO is triggered *one time* with respect to the UE traffic over the overall simulation period. Third, the investigated simulation period of 60 s is separated into three isochronous cycles for operation and optimization such that the traffic hotspot activity is exactly covered by the middle cycle. In this cycle, in addition to coverage problems, there arises a lack of capacity. Hence, tradeoff handling becomes a serious problem for CCO in the second cycle.

TABLE VIII
SYSTEM-WIDE COVERED USERS, SERVED USERS, AND SERVED RATE FOR CASE STUDY 3)

Performance indicators / Configuration	Number of users which are not covered	covered	served	Average served rate [Mbit/s]
1st cycle (coverage compensation)				
Initial	3	4152	3815	53.382
Optimized once	0	4155	3957	56.255
Optimized per cycle	0	4155	3944	55.932
2nd cycle (hotspot activity)				
Initial	5	5255	4106	53.288
Optimized once	0	5260	4265	55.502
Optimized per cycle	3	5257	4763	67.568
3rd cycle (regular operation)				
Initial	2	3958	3631	52.392
Optimized once	2	3958	3828	56.569
Optimized per cycle	2	3958	3840	57.252

Table VIII shows that the initial configuration can be significantly improved in terms of coverage and capacity performance. Furthermore, one-time CCO and cyclic CCO reveal similar behavior for the first and third cycles. Marginal divergences are expectable since our CCO approach utilizes a model of the system and does not reflect it perfectly. In the second cycle, however, we observe a significant difference: Applying the cyclic CCO results reduces the coverage slightly while it considerably improves the capacity, i.e., the number of served users and the served rate. Contrarily, the one-time optimization has to take into account the coverage problems from the first and third cycles jointly, which reduces the degrees of freedom for improving the capacity in the second cycle.

TABLE IX
ACTIONS PERFORMED IN THE MTN FOR CASE STUDY 3) DUE TO AUTONOMOUSLY RUNNING CCO

eNB	1	3	4	5	6	11	12
Initial configuration (tilt / power)	high / med.	high / med.	low / med.	low / med.	low / med.	low / med.	high / med.
1st cycle (coverage compensation)							
Main CCO problems	Most eNBs suffer from capacity problems and eNBs 11 and 12 have additional coverage problems.						
Resulting configuration (tilt / power)	high / high	high / high	low / high	low / high	low / high	low / high	high / high
Main effects	eNBs keep initial tilt but power up, 40% FCs operate at high power. eNBs 1, 4 and 11 reduce coverage holes.						
2nd cycle (hotspot activity)							
Main CCO problems	All eNBs have capacity problems but coverage is tolerable. eNB 4 is mostly affected by the traffic hotspot.						
Resulting configuration (tilt / power)	high / low	high / low	high / high	high / high	high / high	low / med.	high / high
Main effects	eNBs 4, 5, and 6 tilt down to increase near distance capacity while eNBs 1, 3, and 11 reduce their interference.						
3rd cycle (regular operation)							
Main CCO problems	Coverage problems dominate at eNBs 1, 4, and 11 but also degraded network capacity is still a problem.						
Resulting configuration (tilt / power)	high / med.	high / low	high / high	high / high	high / high	low / high	high / high
Main effects	eNBs 1 and 11 power up to compensate coverage holes. The number of high power FCs slightly increases.						

To illustrate how our CCO approach basically works, we discuss the details in Table IX with respect to the configuration-related cell footprints shown in Fig. 10. Note that white spaces indicate *coverage holes*, where the RSP is below −121 dBm. The initial configuration leads to cells where coverage holes exist, particularly at the cell border of eNBs 11 and 12 [see Fig. 10(a)]. Consequently, the coverage holes at eNBs 11 and 12 and severe capacity problems have to be considered in the first optimization cycle. Interestingly, the CCO solution for the first cycle keeps the same tilts but powers up all eNBs, supported by many FCs that operate at high power. This solution improves the coverage situation significantly—compare the white spaces in Fig. 10(b)—and allows for the reduction of the capacity problems. In the second cycle, the active hotspot causes capacity problems, particularly at eNB 4. Due to the reconfiguration results from the first cycle, the coverage performance is still tolerable. As expectable, eNB 4 applies the high downtilt and operates at full power to serve most of the hotspot users. Furthermore, all stations except for eNB 11 use a high downtilt to mitigate interference to their neighboring cells and, particularly, to the eNB 4 coverage zone. Overall, the CCO results for the second cycle lead to a system footprint that reveals more coverage problems as in the first cycle, which is indicated by increasing white spaces. For the last cycle, the coverage problems become more important again, whereas capacity is still an issue. The coverage problems, however, cannot be fully resolved without reducing the capacity significantly. Consequently, CCO leads, similarly to the first cycle, to power changes at eNBs and to a slightly increased capacity.

This case study clearly points out the coverage and capacity tradeoff strategy that is applied in our CCO approach. Coverage performance is the preferred optimization objective but only up to the point when the capacity performance becomes (more) crucial. In this case, we allow the coverage to decrease slightly in favor of gaining degrees of freedom for capacity improvement. The preferences, however, can be adapted, or even inverted, by choosing according optimization parameters. The operating network benefits from the CCO since the corresponding reconfiguration provides the opportunity to locally focus on the traffic hotspot, if necessary. This is, for instance, reflected by the resource consumption at eNB 4, which can increase the bandwidth for serving users from round about 66% for operation cycles one and three to more than 90% in cycle two. Moreover, the one-time CCO results demonstrate that even if the degradation period is not perfectly separated, there is still significant potential for improvements by applying our approach.

4) Interference Approximation Accuracy: To evaluate our interference approximation model from Section II-B, we compare the bandwidth that is allocated according to our model to the bandwidth that is actually allocated during simulation. We observe that the approximation gets more accurate the higher the load of the incorporated cells is. This is not a serious drawback, since on one hand, we are mostly interested in high load situations and, on the other hand, if a cell is only slightly loaded, neither interference nor reconfiguration is a crucial matter. As the proposed optimization model considers the SNR information to compute the required transmission bandwidth, the obtained value is, in general, lower than the value from actual simulation, where the SINR is considered. However, the approximation model estimates the *total* bandwidth consumption (transmission plus blocked) very accurately with a deviation that is less than 2% on average.

Moreover, we investigated the accuracy of our interference approximation model with respect to the previous case study 3). In this setup, all simulated MCs operate at full utilization for most of the time, which is predicted almost exactly by the approximation model. As this specific example represents a high-interference scenario, the bandwidth utilized at eNBs for actual data transmission is just around 2–3 MHz. The only exception is eNB 4, which mainly serves the hotspot users and utilizes about 8 MHz of bandwidth for data transmission during the hotspot activity period. Due to the applied interference approximation, the integrated CCO approach considers round about 75% of this amount for transmission and a correspondingly higher amount of blocked bandwidth.

However, particularly the exchange of blocked bandwidth to bandwidth that is utilized for serving users at the hotspot eNB

illustrates that the blocked bandwidth can serve as a kind of potential capacity if there is a capacity shortage situation. This observation is also reflected in the considered capacity indicator (4), where we reduce the capacity by means of the last addend if the free bandwidth is high, which means, by implication, that the blocked bandwidth is low.

V. Conclusion and Future Work

In this paper, we have presented concepts, models, and algorithms for self-organized autonomous optimization of coverage and capacity in LTE MTNs. A traffic-light-related control mechanism automatically triggers reconfiguration of (H)eNB transmission parameters to improve system performance, if necessary. Running autonomously, the proposed optimization loop takes into account the multitier topology of the network and reconfigures eNBs and HeNBs with respect to situation-aware adapted input parameters. Furthermore, we have introduced an interference approximation model that particularly allows for a linear formulation of the CCO problem. Our model-based optimization approach is fully customizable by choosing corresponding optimization parameters: The modification of coverage and capacity weighting factors in (8) enables a prioritization of the target KPIs. Furthermore, the hierarchical strategy to cope with the CCO tradeoff problem can be (re-)defined by setting the traffic-light-related optimization preferences [see Fig. 6]. All optimization parameters are customizable for each eNB, i.e., the CCO approach is independently configurable for different network areas. Although we can assume a correlation between downlink performance and uplink behavior of the system [43], all presented results hold in the first instance for the downlink perspective.

By simulation-based evaluation of representative test cases, we have demonstrated applicability and the performance of our overall concept. We have achieved performance gains for situations that suffer from a lack of network capacity, from a lack of network coverage, and for mixed degradation constellations. For a higher degree of freedom in the configuration state space, e.g., by application of sectorized antennas, we expect further improvements. We have observed that an obliteration of short-term events, e.g., a temporary traffic hotspot, can happen due to averaging over (relatively) large monitoring periods and that such an effect can lead to further performance degradation. Although the application of very short optimization cycles is a solution to avoid such problems, it is generally not applicable in practice since the related (computational) complexity is too high. Hence, for future work, we suggest a combination of sufficiently short periodic optimization cycles and aperiodic trigger mechanisms that are sensitive to heavy degradation.

In addition to the aforementioned aspects, future research will particularly cover the application of adaptive clustering, an acceleration of optimization procedures by heuristics, and the analysis of spatial scalability, i.e., how does the network size affect the overall complexity and solution accuracy. Furthermore, we will implement an alternative method for DN generation and will analyze CCO robustness with respect to uncertainty in traffic predictions.

References

[1] Cisco Systems, Inc., Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016, Feb. 2012. [Online]. Available: www.cisco.com

[2] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2nd ed. New York, NY, USA: Academic, 2008.

[3] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*, 1st ed. New York, NY, USA: Academic, 2011.

[4] "ECOSYS Deliverable 19," Final Techno-Economic Results on Mobile Services and Technologies Beyond 3G, Sep. 2006. [Online]. Available: http://ecosys.optcomm.di.uoa.gr/deliverableslist.html

[5] A. Hoikkanen, "Economics of 3G long-term evolution: The business case for the mobile operator," in *Proc. IEEE WOCN*, Jul. 2007, pp. 1–5.

[6] NGMN Alliance, Next Generation Mobile Networks Beyond HSPA and EVDO, version 3.0, Dec. 2006. [Online]. Available: www.ngmn.org

[7] NGMN Alliance, NGMN Recommendation on SON and O&M Requirements, version 1.23, Dec. 2008. [Online]. Available: www.ngmn.org

[8] "3GPP Work Items on SON," Work Items on Self-Organising Networks, ver. 0.0.6, Oct. 2010. [Online]. Available: www.3gpp.org

[9] Self-configuring and self-optimizing network (SON) use cases and solutions, Third-Generation Partnership Project, Cedex, France, 3GPP TR 36.902, ver. 9.3.0. [Online]. Available: www.3gpp.org

[10] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution*. Hoboken, NJ, USA: Wiley, 2009.

[11] "3GPP Rel. 8," Overview of 3GPP Rel. 8, ver. 0.2.1, Sep. 2010. [Online]. Available: www.3gpp.org

[12] "3GPP TS 32.500," SON Concepts and Requirements, ver. 9.0.0, Dec. 2009. [Online]. Available: www.3gpp.org

[13] "3GPP Rel. 10," Overview of 3GPP Rel. 10, ver. 0.0.8, Sep. 2010. [Online]. Available: www.3gpp.org

[14] "3GPP Rel. 11," Overview of 3GPP Rel. 11, ver. 0.0.4, Sep. 2010. [Online]. Available: www.3gpp.org

[15] "3GPP TS 32.521," SON Policy Network Resource Model (NRM) Integration Reference Point (IRP) Requirements, ver. 10.1.0, May 2011. [Online]. Available: www.3gpp.org

[16] J. Ramiro and K. Hamied, *Self-Organizing Networks (SON): Self-Planning, Self-Optimization and Self-Healing for GSM, UMTS and LTE*, 1st ed. Hoboken, NJ, USA: Wiley, 2012.

[17] Celtic-Plus Initiative. [Online]. Available: http://www.celtic-initiative.org/

[18] GANDALF Project. [Online]. Available: http://www.celtic-initiative.org/Projects/Celtic-projects/Call2/GANDALF/gandalf-default.asp

[19] End-to-End Efficiency (E3) Project. [Online]. Available: http3://ict-e3.eu

[20] Final report on self-organisation and its implications in wireless access networks, SOCRATES Deliverable D5.9, Delft, The Netherlands, EU STREP SOCRATES (INFSO-ICT-216284). [Online]. Available: www.fp7-socrates.eu

[21] SOCRATES Deliverable D2.1, Use Cases for Self-Organising Networks, Mar. 2008. [Online]. Available: www.fp7-socrates.eu

[22] NGMN Alliance, NGMN Use Cases Related to Self Organising Network, Overall Description, ver. 2.02, Dec. 2008. [Online]. Available: www.ngmn.org

[23] R. Whitaker, L. Raisanen, and S. Hurley, "A model for conflict resolution between coverage and cost in cellular wireless networks," in *Proc. 37th HICSS*, Jan. 2004.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[25] R. Razavi, S. Klein, and H. Claussen, "Self-optimization of capacity and coverage in LTE networks using a fuzzy reinforcement learning approach," in *Proc. PIMRC*, Sep. 2010, pp. 1865–1870.

[26] M. Naseer ul Islam and A. Mitschele-Thiel, "Reinforcement learning strategies for self-organized coverage and capacity optimization," in *Proc. WCNC*, Apr. 2012, pp. 2818–2823.

[27] H. Claussen, L. Ho, and L. Samuel, "Self-optimization of coverage for femtocell deployments," in *Proc. WTS*, Apr. 2008, pp. 278–285.

[28] Simulation assumptions and parameters for FDD HeNB RF requirements, Third-Generation Partnership Project, Cedex, France, 3GPP R4-092042, Alcatel-Lucent, picoChip Designs, Vodafone. [Online]. Available: www.3gpp.org

[29] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 41–48, Sep. 2009.

[30] Y. Bai, J. Zhou, and L. Chen, "Hybrid spectrum usage for overlaying LTE macrocell and femtocell," in *Proc. GLOBECOM*, Honolulu, HI, USA, Dec. 2009, pp. 1–6.

[31] R. Mathar, M. Reyer, and M. Schmeink, "A cube oriented ray launching algorithm for 3D urban field strength prediction," in *Proc. ICC*, Glasgow, U.K., Jun. 2007, pp. 5034–5039.
[32] F. Schröder, M. Reyer, and R. Mathar, "Efficient implementation and evaluation of parallel radio wave propagation," in *Proc. 5th EuCAP*, Apr. 2011, pp. 2323–2327.
[33] "Soft Frequency Reuse Scheme for UTRAN LTE," Third-Generation Partnership Project, Sophia-Antipolis, France, 3GPP R1-050507, May 2005.
[34] J. Zhang, H. Tian, P. Tian, Y. Huang, and L. Gao, "Dynamic frequency reservation scheme for interference coordination in LTE—Advanced heterogeneous networks," in *Proc. VTC Spring*, May 2012, pp. 1–5.
[35] D. Tse and P. Visvanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[36] S. Görtzen and A. Schmeink, "Optimality of dual methods for discrete multiuser multicarrier resource allocation problems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3810–3817, Oct. 2012.
[37] K. Tutschku, "Demand-based radio network planning of cellular mobile communication systems," in *Proc. INFOCOM*, Mar./Apr. 1998, vol. 3, pp. 1054–1061.
[38] D. Tikunov and T. Nishimura, "Traffic prediction for mobile network using Holt–Winter's exponential smoothing," in *Proc. SoftCOM*, Sep. 2007, pp. 1–5.
[39] IBM ILOG, CPLEX Optimization 12.2. [Online]. Available: http://www.ilog.com/products/cplex
[40] Gurobi Optimization, Inc., Gurobi Optimizer 5.0. [Online]. Available: http://www.gurobi.com/products/gurobi-optimizer
[41] G. Lee, D. Park, and H. Seo, *Wireless Communications Resource Management*. Hoboken, NJ, USA: Wiley, 2009.
[42] "NGMN white paper," Next Generation Mobile Networks Radio Access Performance Evaluation Methodology, Jun. 2007.
[43] O. Grondalen, P. Gronsund, T. Breivik, and P. Engelstad, "Fixed WiMAX field trial measurements and analyses," in *Proc. 16th IST Mobile Wireless Commun. Summit*, Jul. 2007, pp. 1–5.

Alexander Engels (S'09) received the Diploma degree in mathematics from FH Aachen University of Applied Sciences, Aachen, Germany, in 2004 and the Diploma degree in computer science from RWTH Aachen University, in 2008, where he is currently working toward the Ph.D. degree.

Since 2004, he has been with the Institute for Theoretical Information Technology, RWTH Aachen University, first as a Nonscientific Staff Member and, from 2008 onward, as a Research Assistant. Being in charge of numerous research projects with industrial cooperation partners, his research interests particularly include radio network planning, optimization and heuristics, and self-organization in heterogeneous wireless networks.

Michael Reyer (S'07–M'09) received the Diploma degree in mathematics and the Ph.D. degree in electrical engineering from RWTH Aachen University, Aachen, Germany, in 2000 and 2008, respectively.

From 2000 to 2004, he was a Research Assistant with the Faculty of Mathematics, RWTH Aachen University. From 2004 to 2009, he was a Research Assistant and, since 2009, has been a Senior Engineer with the Institute for Theoretical Information Technology, RWTH Aachen University. He has regularly published in IEEE journals and conferences. His research interests include mobile communications, particularly resource allocation for orthogonal frequency-division multiplexing networks, wireless channel modeling, optimization, e.g., radio network planning, and radio wave propagation.

Xiang Xu (S'10) received the B.Sc. degree in electronic information science and technology from the University of Science and Technology of China, Hefei, China, in 2004 and the M.Sc. degree (with distinction) in digital communications from the University of Kiel, Kiel, Germany, in 2007. He is currently working toward the Ph.D. degree with the Institute for Theoretical Information Technology, RWTH Aachen University, Aachen, Germany.

He is currently a Research Assistant with the Institute for Theoretical Information Technology. His main research interests include digital transceiver design, channel modeling, and self-organizing heterogeneous networks.

Rudolf Mathar (M'98) received the Diploma and Ph.D. degrees from RWTH Aachen University, Aachen, Germany, in 1978 and 1981, respectively.

He was a Lecturer with Augsburg University, Augsburg, Germany, and with the European Business School, Brussels, Belgium. In 1989, he joined the Faculty of Natural Sciences, RWTH Aachen University. In 1999, he held the International IBM Chair in Computer Science with the Free University of Brussels. In 2004, he was the Head of the Institute for Theoretical Information Technology with the Faculty of Electrical Engineering and Information Technology, RWTH Aachen University. Since October 2011, he has been the Dean of the Faculty of Electrical Engineering and Information Technology, RWTH Aachen University. He is a Cofounder of two spin-off enterprises. He has extensively published in IEEE journals and conferences. His research interests include mobile communication systems, particularly, optimization, resource allocation, and access control.

Dr. Mathar received the prestigious Vodafone Innovation Award in 2002. In 2010, he was elected as a member of the North Rhine-Westphalia Academy of Sciences and Arts.

Jietao Zhang received the B.Sc. degree in electrical engineering from Shenzhen University, Shenzhen, China, and the M.Sc. and Ph.D. degrees in radio communications from the University of Liverpool, Liverpool, U.K.

In 2005, he joined Huawei Technologies Co., Ltd., Shenzhen, as a Research Staff with the Wireless Research Department, where his main research interests were in WiMAX Network Working Group standardization, cooperative communications, cognitive radios, and their applications in future wireless communication systems. He is currently with the Communications Technologies Lab, Huawei Technologies Co., Ltd. His current research interests include self-organizing networks and network intelligence techniques.

Hongcheng Zhuang received the Ph.D. degree in radio physics from Sun Yat-sen University, Guangzhou, China, in 2002.

In 2002, he joined Huawei Technologies Co., Ltd., Shenzhen, China. Since 2005, he has been leading standard and long-term research projects in wireless communications. His research interests include cellular systems and wireless local area networks, particularly system and protocol design, intelligent optimization, and radio resource management.