**RESEARCH ARTICLE**

# A Differentiable Throughput Model for Load-Aware Cellular Network Optimization Through Gradient Descent

**LUKAS ELLER** [1,2], (Graduate Student Member, IEEE),
**PHILIPP SVOBODA** [1,2], (Senior Member, IEEE),
**AND MARKUS RUPP** [1], (Fellow, IEEE)

[1] Institute of Telecommunications, TU Wien, 1040 Vienna, Austria
[2] Christian Doppler Laboratory for Digital Twin Assisted AI for Sustainable Radio Access Networks, Institute of Telecommunications, TU Wien, 1040 Vienna, Austria

Corresponding author: Lukas Eller (lukas.eller@tuwien.ac.at)

**ABSTRACT** The efficient operation of cellular networks requires careful tuning of configuration parameters, such as the transmit power or antenna tilts, to adequately balance interference while providing the necessary capacity to the connected UEs. As manual tuning of these parameters is typically unfeasible, several automated coverage and capacity optimization methods have been proposed. However, most existing solutions are either based on poorly scalable black-box optimization methods or solely consider interference management, while omitting the potential of congested cells. In this work, we instead propose a differentiable framework for cellular network optimization, centered around the end-user throughput, that enables load-aware tuning of network parameters through gradient descent. Hereby, we approach the problem from a data-driven perspective, and include dedicated model subcomponents derived from monitoring data, which enable the calibration to site-specific traffic patterns and KPI measurements. We validate our approach for joint transmit power optimization in a real-world network layout with ≈ 150 cells in two frequency bands. In our evaluation, the gradient descent-based optimization reliably reduces the outage ratio for different levels of demand, while the black-box baseline struggles to explore the large search space. Our results further reveal substantial differences between the proposed load-aware and commonly used SINR-based objectives, for which we repeatedly obtain unbalanced network configurations with severely congested cells. In contrast, the proposed end-user throughput objective promotes a balanced network configuration, providing adequate resources to the connected UEs while also limiting interference.

**INDEX TERMS** Coverage and capacity optimization, cellular network optimization, gradient descent, radio planning, load balancing, monitoring data, minimization of drive tests, 4G, 5G, 6G, wireless networks.

## I. INTRODUCTION

The efficient operation of cellular networks requires careful selection and tuning of configuration parameters, in order to provide the necessary capacity to connected user equipments (UEs), while also adequately balancing interference. For large network deployments, the manual configuration of parameters like transmit power or antenna down-tilt is typically unfeasible, so that several automated coverage and capacity optimization (CCO) methods have been proposed [1], [2], [3]. Still, the optimization remains challenging, as the parameters have to be tuned jointly over all cells

The associate editor coordinating the review of this manuscript and approving it for publication was Peng-Yong Kong.

serving the respective area, given that they are inherently linked, e.g., changing the transmit power impacts the interference levels for neighboring cells, and potentially also the number of connected UEs and thus the cell load (CL) [4]. Hence, the optimization does not only require efficient methods, scaling well to large network deployments with the corresponding high number of tunable parameters, but also mandates a suitable objective function, representing the end-user performance, while taking all of these complex dependencies into account. Due to the challenge of this modeling task, most work in the literature treats the problem from a completely data-driven perspective, utilizing black-box optimization methods such as Bayesian optimization (BO) or reinforcement learning (RL), which iteratively probe potential configurations without accounting for the internal structure of the objective function [2], [3]. These approaches promise to omit the, potentially error-prone propagation modeling task, by operating on crowdsourced measurements of key performance indicators (KPIs), such as the reference signal received power (RSRP), the reference signal received quality (RSRQ) or the reference signal (RS)-signal to interference and noise ratio (SINR) [4]. These KPIs characterize the conditions from the UE perspective and can be obtained from monitoring sources such as minimization of drive-tests (MDT), hence providing direct feedback for online optimization [2], [3].

However, these model-free approaches require extensive probing of potential network configurations, which can lead to poor scalability for high dimensional search spaces, accompanied by a high risk of degrading end-user performance during the initial exploration phase [5]. In contrast, it has been shown in [6], that a suitable model-based formulation of the SINR-based objective enables the efficient and scalable optimization of antenna orientations through gradient descent (GD). Compared to model-free methods, the GD-based approach proposes suitable update directions in each step instead of exploring the configuration space through exhaustive search. While the differentiable formulation in [6] required certain simplified assumptions, recent work on differentiable propagation modeling, either through deep learning [7] or in the form of dedicated differentiable ray tracing (RT) [8] suggests, that end-to-end gradient information is becoming more accessible for state-of-the-art radio simulation methods, potentially forming a complete ecosystem for GD-based tuning of the network configuration.

Still, we identify certain aspects that require further analysis to enable a comprehensive framework for GD-based network optimization. Foremost, this concerns the optimization target itself, for which we find the commonly used SINR-based objective unsuitable, as it is unaware of the possibility of congested cells. In this work, we thus propose a differentiable framework centered around the shared end-user throughput, which ensures adequate interference management, but is also inherently load-aware. As sketched in Fig. 1, we further follow a hybrid approach using common model-based assumptions, while including monitoring data for
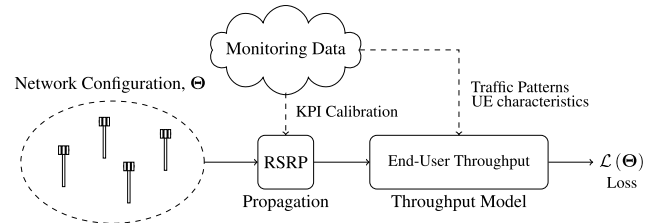


**FIGURE 1.** Our differentiable throughput model is based on standardized KPIs to allow for calibration to monitoring data. The end-to-end differentiable formulation enables the scalable and efficient optimization through GD.

calibration of KPIs and to capture site-specific traffic patterns and UE characteristics. Through the end-to-end differentiable formulation, we then obtain suitable update directions for the network configuration through backpropagation.

### A. CONTRIBUTION
We summarize our contribution as follows:

*i)* We derive a framework for load-aware cellular network optimization based on GD. In contrast to purely SINR-based optimization, our scheme centers around the end-user throughput, and is thus aware of limited resources shared among the connected UEs. Accordingly, the objective avoids congested cells and promotes configurations with a balanced distribution of UEs across the available cells. Through a straightforward model extension, we can also incorporate a GD-based scheme that facilitates active load balancing.

*ii)* Moreover, we follow a hybrid approach and derive data-driven mappings from real-world monitoring data, which form subcomponents of our differentiable framework. These components address the complex task of modeling user demand and act as a data-driven abstraction over site-specific traffic patterns. They further enable the calibration of our model to distinct UE characteristics and KPI measurements.

*iii)* We validate the proposed framework by studying large-scale GD-based transmit power optimization for a real-world network deployment with $\approx$ 150 sectors in two frequency bands. Hereby, we consider BO as a model-free baseline. Throughout the evaluation, we further characterize the relation between load balancing and network optimization, showing that both objectives are strongly connected.

For the interested reader, we also make a *tensorflow* [9] implementation of the differentiable throughput model publicly available.[1]

### B. RELATED WORK
Typically, CCO centers around antenna parameters that can be controlled remotely, hence offering the possibility to retune the network configuration after deployment [1]. Specifically the optimal setting of the transmit power and the antenna downtilt is of interest [2], [3], [5], [10],

---

[1]https://squid.nt.tuwien.ac.at/gitlab/leller/differentiable_throughput_model

[11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. Overall, we can distinguish two strategies, either approaching the topic from a model-based or a model-free perspective.

Here, the promise of model-free methods is to directly include monitoring data into the optimization process, most notably from MDT, which plays an increasingly important role in the operation of cellular networks [22], [23]. In [2], the authors successfully utilize such MDT data for automated tuning of the antenna downtilt based on a RL approach. Similarly, the use of RL variants has further been validated in [10], [11], and [12], with multi-agent systems gaining more interest to address the high dimensional action space required for joint optimization over large areas [13], [14], [15], [16]. Despite these advances, the required training phase of such model-free approaches still poses a severe practical limitation, most notably in terms of sample-efficiency and risk-aversion, which are ongoing research topics [5], [24]. For network optimization, the sample-efficiency of model-free approaches is commonly addressed through BO [25], which has shown similar performance compared to RL in [3], while substantially reducing the number of required probing points. For BO, the primary challenge is the scalability to higher dimension, for which several reformulated variants have recently been proposed [26], [27]. Also for CCO optimization, promising results have been reported for a large number of antenna parameters in [17], [18], and [19]. However, this required combining BO with evolutionary algorithms in [17] and [18], and an effective reduction of the search space in [19].

These challenges pose the question of whether a complete black-box formulation is preferable, given that it omits the internal structure of well-understood mechanisms in CCO. Foremost, this concerns the availability of gradients, which enables the scalable and efficient tuning of antenna orientations and the transmit power for the model-based approaches in [6] and [20]. Here, the main challenge is to obtain an end-to-end differentiable formulation, also including the propagation model. However, recent work on deep learning-based propagation modeling [7], [28], and differentiable RT [8], [29], [30], suggests that a gradient-based framework for network optimization, using state of the art propagation modeling, is within reach. Here, we see the possibility of a hybrid approach, providing gradients for well-understood mechanisms while accounting for the possibility of direct feedback from sources like MDT. In this work, we thus follow the methodology from [2] and [4] by only relying on standardized KPIs and utilization metrics that are directly available to operators. Here, the data-driven perspective is particularly helpful to obtain throughput estimates, which require an understanding of UE characteristics and, most importantly, traffic patterns.

In fact, the load aspect is often omitted in CCO, with methods like [3], [6], [11], [13], [19], and [20] solely considering interference management. Similarly, this also applies for the theoretical empty cell throughput, as in
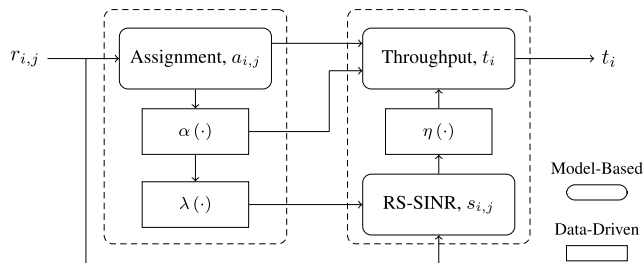


**FIGURE 2.** Schematic block diagram of the throughput model. As all components are differentiable with respect to their inputs, we can obtain the gradient of the end-user throughput with respect to $r_{i,j}$.

**TABLE 1.** Summary of notation for the throughput model.

| | | | | | |
|---|---|---|---|---|---|
| $r_{i,j}$ | ... | RSRP | $a_{i,j}$ | ... | cell assignment |
| $s_{i,j}$ | ... | RS-SINR | $\eta_{i,j}$ | ... | spectral efficiency |
| $u_j$ | ... | connected UEs | $\alpha_j$ | ... | parallely active UEs |
| $\lambda_j$ | ... | cell load | $t_i$ | ... | end-user throughput |
| $i \in \mathcal{U}$ | ... | index for UE | $j \in \mathcal{C}$ | ... | index for cell |

[10], [14], [17], and [18], which omits the possibility of congestion and is thus inherently load-unaware. However, the importance of an adequate distribution of UEs among the available cells is apparent from the design of load balancing and handover algorithms [31], [32], [33]. Further, it has been shown in [34] and [35] that the joint optimization of the user association strategy and the antenna tilt achieves superior results for energy management tasks. Similar results are also reported for the model-based approaches in [36], [37], [38], and [39], where the authors again consider the optimization of network parameters and UE association jointly. Overall, this connection is not surprising, as adaptations of the network configuration inherently change the UE assignment, potentially also resulting in congestion. While the approaches in [36], [37], [38], and [39] partially utilize derivatives, they are inherently model-based and thus not directly comparable to the hybrid approach considered in our work. For instance, [36], and also the gradient-free schemes in [34] and [35] model the network traffic on a flow level, while [37], [38], [39] similarly require information about the requested transmission rate per UE. As in [2] and [4], we instead estimate the throughput solely from metrics that are directly available to operators. From this real-world data, we derive the data-driven subcomponents of our model. Showing that such a hybrid approach can still be optimized through efficient scalable GD-based methods represents a major contribution of our work.

The remainder of the paper is organized as follows: We first introduce the differentiable throughput model in Sec. II before we discuss the data-driven mappings derived from monitoring data in Sec. III. Sec. IV then details the considered large-scale optimization scenario on a real-world network deployment. We summarize the evaluation results in Sec. V and further study the difference between load-aware and unaware optimization in Sec. VI, by considering a model extension that enables active load balancing. We finally end with a discussion and conclusion in Sec. VII and VIII.
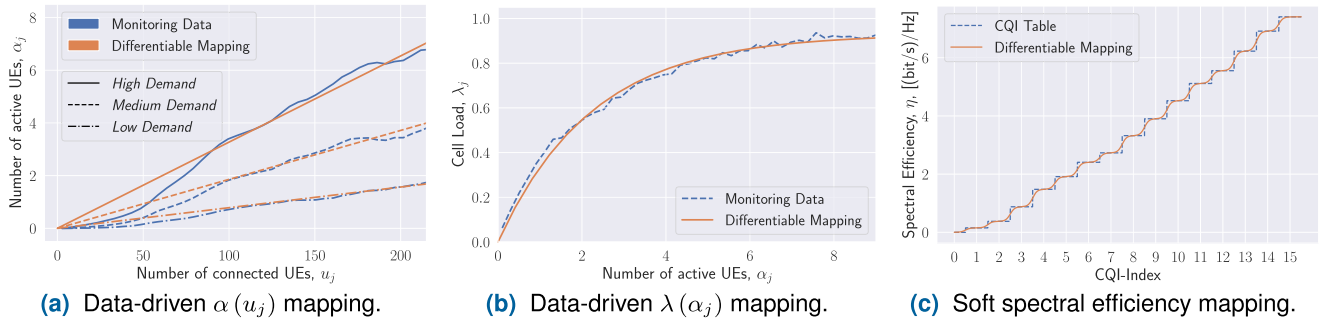
**FIGURE 3.** The subcomponents of the throughput model enable calibration to site-specific traffic patterns or distinct UE fleets.

## II. DIFFERENTIABLE THROUGHPUT MODEL

Our GD-based framework is centered around a differentiable end-user throughput model that is based on KPI measurements. As such, we consider $i \in \mathcal{U}, \mathcal{U} = \{1, \dots, U\}$ UEs and $j \in \mathcal{C}, \mathcal{C} = \{1, \dots, C\}$ cells, such that $r_{i,j}$ denotes the RSRP of cell $j$ observed by UE $i$ in dBm. We then target an expression for the end-user throughput $t_i$ for UE $i$, that is fully differentiable with respect to these RSRP values $r_{i,j}$. As sketched in Fig. 2, we build the model from differentiable components, such that the gradients can be obtained through backpropagation. Hereby, we use common model-based assumptions for the cell assignment $a_{i,j}$, the RS-SINR $s_{i,j}$ and the resulting end-user throughput $t_i$, but include data-driven mapping for the CL $\lambda_j$, the number of active UEs $\alpha_j$, and also the spectral efficiency $\eta_{i,j}$.

In the following, we discuss all of these model-based components, before the data-driven mappings are detailed in Sec. III. A summary of the used notation for the throughput model derivation is further provided in Tab. 1.

### A. DIFFERENTIABLE SOFT CELL ASSIGNMENT

We first derive a soft, and thus differentiable, formulation of the cell assignment, by introducing $a_{i,j} \in (0, 1)$, denoting the probability of assignment for UE $i$ to cell $j$. We compute:

$$\Delta r_{i,j} = \max_{j' \in \mathcal{C}} (r_{i,j'}) - r_{i,j}, \tag{1}$$

representing the RSRP offset to the strongest observed cell per UE, and subsequently apply a softmax function, mapping each $\Delta r_{i,j}$ value to an assignment probability:

$$a_{i,j} = \operatorname*{softmax}_{j' \in \mathcal{C}} \left(-\Delta r_{i,j'}\right)_j. \tag{2}$$

Accordingly, $a_{i,j}$ is largest for the strongest cell observed by UE $i$, but non-zero for cells with smaller $r_{i,j}$ values. This not only ensures a differentiable formulation, but also realistically accounts for the uncertainty introduced by handover effects. Further $a_{i,j}$ fulfills the properties $\sum_{j=1}^{C} a_{i,j} = 1$ and $\sum_{i,j=1}^{U,C} a_{i,j} = U$, such that a differentiable expression for $u_j$, the number of assigned UEs for cell $j$ can be obtained as $u_j = \sum_{i=1}^{U} a_{i,j}$. As sketched in Fig. 2, $u_j$ then acts as the basis to compute the number of active UEs $\alpha_j = \alpha\left(u_j\right)$ and the CL $\lambda_j = \lambda\left(\alpha_j\right)$, using the mappings described in Sec. III. Our model hence incorporates the distribution of the

load among the available cells, based on the RSRP values $r_{i,j}$. In Sec. VI, we will also discuss an extension where we alternatively interpret $a_{i,j}$ as an optimization variable for fixed $r_{i,j}$. This optimized cell assignment then characterizes further improvements when actively balancing the load in the considered deployment.

### B. RS-SINR & SPECTRAL EFFICIENCY

In accordance with $a_{i,j}$, we compute the RS-SINR for each possible cell assignment combination of UE $i$ and cell $j$ as

$$s_{i,j} = r_{i,j} - 10 \log_{10} \left( \sum_{j' \in \mathcal{C} \setminus j} \left(\lambda_{j'} r_{i,j'|mW}\right) + \zeta_{|mW} \right). \tag{3}$$

In (3), we also account for the CL $\lambda_j$, which scales the interference terms depending on the current load, while $r_{i,j|mW}$ and $\zeta_{|mW}$ denote the received power and the noise power in linear scale. Note, that the expression is again differentiable[2] with respect to $r_{i,j}$, and also $\lambda_j$. Hence, gradients are being passed to those components during backpropagation in Fig. 2. We then follow the approach from [2], and map each RS-SINR value $s_{i,j}$ to the channel quality indicator (CQI), and subsequently the spectral efficiency $\eta_{i,j}$, but use the differentiable mapping $\eta_{i,j} = \eta\left(s_{i,j}\right)$ discussed in Sec. III.

### C. END-USER THROUGHPUT & SHARED RESOURCES

In the final step, we then derive the end-user throughput. Under the assumption of an empty cell, the throughput for UE $i$ amounts to the expectation of the spectral efficiency $\eta_{i,j} = \eta\left(s_{i,j}\right)$ over the assignment probability $a_{i,j}$:

$$t_i^{(\text{empty})} = B_{\text{RB}} \cdot N_{\text{RB}} \cdot \sum_{j=1}^{C} a_{i,j} \cdot \eta\left(s_{i,j}\right), \tag{4}$$

scaled by the total number of available resource blocks (RBs) $N_{\text{RB}}$ and their respective bandwidth $B_{\text{RB}}$. However, the RBs in (4) have to be shared among the connected UEs in the respective cell. Accounting for these shared resources and assuming a round-robin scheme with $\alpha_j$, the number of

---

[2]Note, that (3) can be implemented using the numerically stable logsumexp function. The gradient for the SINR is also derived in [6].
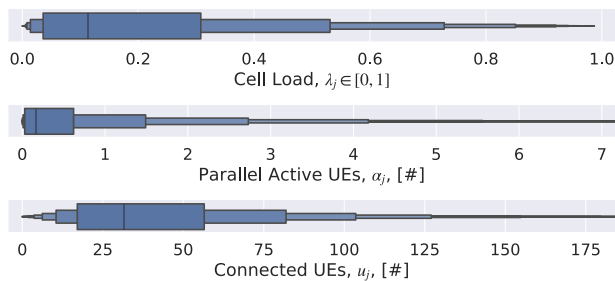
**FIGURE 4.** Statistics of raw data samples from the monitoring source, used to derive data-driven subcomponents.

**TABLE 2.** Coefficients $k^{(\alpha)}$ for different levels of demand.

|  | Low Demand | Medium Demand | High Demand |
|---|---|---|---|
| $k^{(\alpha)}$ | $0.78 \times 10^{-2}$ | $1.86 \times 10^{-2}$ | $3.27 \times 10^{-2}$ |

parallelly active UEs per cell, yields the following expression:

$$t_i = B_{\mathrm{RB}} \cdot N_{\mathrm{RB}} \cdot \sum_{j=1}^{C} \frac{a_{i,j} \cdot \eta\left(s_{i,j}\right)}{1 + \alpha\left(u_j\right)}. \quad (5)$$

Note, that (5) closely follows the model derived in [4, Cha. 6], with the plus one term accounting for the fully active UE under consideration. Further, neither $u_j \geq 0$ nor $\alpha_j \geq 0$ will typically be integers due to the soft formulation, so that we do not model the immediate scheduling of RBs, but rather the average behaviour over a longer time frame. Clearly, (5) differs from the *load-unaware* empty cell throughput in (4), as the formulation does not only promote interference management through $s_{i,j}$ and ultimately $\eta_{i,j}$, but also accounts for limited resources. We will show later in this work, that a *load-aware* objective based on (5) does indeed promote network configurations with an adequate distribution of UEs, hence avoiding overloaded cells.

## III. DATA-DRIVEN ABSTRACTIONS

To enable calibration to specific network deployments, the throughput expression derived in Sec. II incorporates the data-driven subcomponents $\alpha\left(u_j\right)$, $\lambda\left(\alpha_j\right)$ and $\eta\left(s_{i,j}\right)$.

In our case, these mappings, sketched in Fig. 3, are directly derived from network traces extracted for the operational LTE network that also acts as the basis for the optimization use case in Sec. IV. This allows us to reach the required level of abstraction without incorporating any possibly faulty model-based assumption for the critical components characterizing the site-specific traffic patterns. These components can also easily be recalibrated to other network configurations, updated UE generations, or different 3GPP releases. The statistics of the raw data used for characterizing the traffic patterns are summarized in Fig. 4 and stem from an internal base station (BS) tool[3] monitoring the cell utilization. In particular, this monitoring data includes

the CL $\lambda_j$, the number of parallelly active UEs $\alpha_j$, and the number of connected UEs $u_j$. In order to correctly capture the behavior of the considered real-world network for our throughput model, we extract the relations between these individual quantities. As such monitoring data is generally noisy, we fit our models to the median trend, which describes clear and consistent relations for the quantities in Fig. 4. From these relations, we derive the mappings $\alpha\left(u_j\right)$, $\lambda\left(\alpha_j\right)$ and $\eta\left(s_{i,j}\right)$:

*Active UEs Mapping:* We first consider $\alpha\left(u_j\right)$, the number of connected UEs that are active in parallel during one time interval — compare [4, Cha. 6]. Here, the monitoring data in Fig. 3a suggests, that a linear relation of the form

$$\alpha_j = \alpha\left(u_j\right) = k^{(\alpha)} \cdot u_j \quad (6)$$

can adequately describe $\alpha_j$ for a wide range of the number of connected UEs $u_j$. Hence, we obtain the coefficients $k^{(\alpha)}$ summarized in Tab. 2, each characterizing the ratio of active UEs for a distinct level of demand.[4] In the evaluation, we will consider all three levels, which result in distinct network configurations under our *load-aware* objective.

*Cell Load Mapping:* Similarly, we derive the relation in Fig. 3b, describing the influence of the number of active UEs on the CL $\lambda\left(\alpha_j\right)$, denoting the ratio of assigned RBs [4]. Here, a linear relation is inappropriate, as the CL is bound to the interval $\lambda_j \in [0, 1]$ and saturates with a higher number of active UEs $a_j$. In principle, several regression methods are suitable, however, a simple exponential fit closely follows the trend in Fig. 3b. As such, we obtain

$$\lambda_j = \lambda\left(\alpha_j\right) = w_0 \cdot \left(1 - e^{-\alpha_j/w_1}\right), \quad (7)$$

with the parameterization $w_0 = 0.93$ and $w_1 = 2.25$. Clearly, this mapping is smooth and differentiable so that the gradients required for backpropagation in Fig. 2 are available.

*Spectral Efficiency Mapping:* Using the same principle, we also derive the data-driven subcomponent characterizing the spectral efficiency $\eta\left(s_{i,j}\right)$. Following the approach in [2], we first derive a mapping from RS-SINR to CQI, before we obtain the spectral efficiency $\eta_{i,j}$ through the tabular mapping from the 3GPP standard [40, p. 220]. Hereby, we transform the table into the soft differentiable form shown in Fig. 3c, by approximating the steps through smooth sigmoid functions. For the mapping from RS-SINR we then retrieve an empirical relation from MDT data to capture average channel conditions and UEs characteristics as in [2]. The procedure is detailed in Appendix A, where we also discuss necessary preprossessing steps. Clearly, also this mapping is differentiable and can easily be adapted to different 3GPP releases or updated UE generations.

With the differentiable throughput model and the data-driven subcomponents specified, we discuss the application to GD-based network optimization in the following.

---

[3]The same data source has also been studied in [4, Cha. 6], where the relation of the parallelly active UEs to the assigned RBs used in (5) is detailed. Note, that the data source reports 15 min averages for all quantities.

[4]In addition to the median, Fig. 3 also includes the 0.75 and 0.25 percentiles to represent the cases of high and low demand.
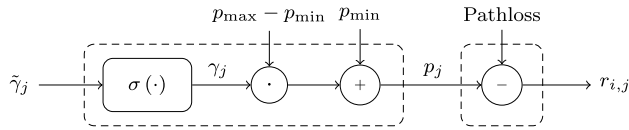
**FIGURE 5.** We tune the transmit power $p_j$ within a defined interval. Gradients of $r_{i,j}$ for the latent variable $\tilde{\gamma}$ are available.

## IV. LOAD-AWARE NETWORK OPTIMIZATION

For network optimization, the differentiable throughput model from Sec. II can, in principle, be applied to a wide range of antenna parameters, as long as the gradients for the RSRP $r_{i,j}$ are available, as in [6] and [8]. Throughout this work, we, however, limit the analysis to the transmit power, which does not require dedicated propagation modeling, and instead focus on a detailed validation of the proposed model. In particular, we want to characterize how the *load-aware* objective differs from the commonly used *load-unaware* optimization centered around the SINR. Hereby, we specifically focus on the adequate distribution of UEs among the available cells. In Sec. VI, we will further discuss an extension that enables *active load balancing* for a fixed network configuration, which allows us to actively reassign the connected UEs.

In the following, we first define the problem statement in detail and specify the used loss function. Then, we discuss the real-world network deployment used for validation.

### A. LOSS FUNCTION & EVALUATION METRICS

To differentiate the proposed *load-aware* optimization from the *load-unaware* case, we specify two distinct loss functions, which use the different throughput variants from Sec. II-C. In particular, the *load-aware* variant considers the optimization based on the shared end-user throughput from (5). We formulate the loss as a penalty for the ratio of UEs not fulfilling a lower throughput threshold $t^{\text{(thresh)}}$ yielding:

$$\mathcal{L}_{\text{outage}}\left(\cdot\right) = \frac{1}{U}\sum_{i=1}^{U} \sigma\left(t^{\text{(thresh)}} - t_i\right). \tag{8}$$

Note, that the soft outage formulation through the sigmoid function $\sigma$ ensures a differentiable expression. Further, (8) promotes basic fairness,[5] as it does not incentivize providing excessive rate to a single UE. To allow for a direct comparison of the two variants, we use the same formulation for the *load-unaware* objective, but replace the shared end-user throughput $t_i$ in (8), with $\frac{1}{1+\alpha(U/C)} \cdot t_i^{\text{(empty)}}$. Here, $t_i^{\text{(empty)}}$ is the empty cell rate from (4), and the scaling term in front ensures that both objectives are comparable for a given threshold $t^{\text{(thresh)}}$. Hence, the objective considers the throughput that can be achieved, when the UEs are evenly distributed among the available cells. As the constant scaling term is extrinsic to the optimization, this *load-*

---

[5]Instead of a fixed threshold, basic fairness can also be incorporated through a logarithmic utility function as discussed in detail in [32]. This results in diminishing returns for particularly high values of the throughput.
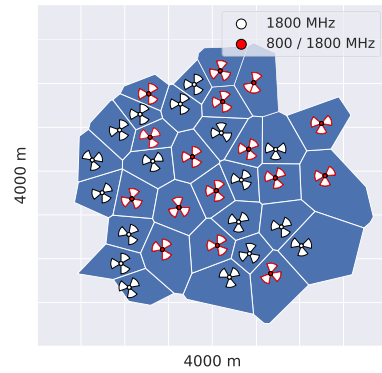


**FIGURE 6.** Optimization scenario based on real-world deployment, with $\approx 150$ cells in two frequency bands with distinct propagation conditions. The optimization region, indicated in blue, stretches over an area of $\approx 8\ \text{km}^2$.

*unaware* reference objective is still completely unaware of the possibility of congested cells.

For the evaluation, we further introduce dedicated metrics to study these two optimization targets in detail. In particular, we assess the interference management by keeping track of the evolution of the RS-SINR for a given UE $i$, which we compute as the expectation over the cell assignment $s_i = \sum_{j=1}^{C} a_{i,j} \cdot s_{i,j}$, similar to (4). Note, that we compute the expectation over both bands to enable a direct comparison with the respective throughput values. Moreover, we also consider a metric of the balanced distribution of UEs among the available cells, in the form of the Gini coefficient $g_{\text{coef}} \in [0, 1]$, which is a common fairness measure [41]. In our case, the coefficient can be computed as:

$$g_{\text{coef}}\left(u\right) = \frac{1}{2UC}\sum_{j=1}^{C}\sum_{j'=1}^{C}\left|u_j - u_{j'}\right|. \tag{9}$$

We interpret (9) as a congestion measure, as it describes the dispersion of the number of connected UEs over all available cells. In particular, $g_{\text{coef}} = 0$ represents the case where the UEs are perfectly distributed among the available cells with a constant $u_j = U/C$. In contrast, $g_{\text{coef}} = 1$ resembles the extreme case, where all $U$ UEs are assigned to a single cell, while all other cells are empty. For the *load-aware* objective function, we thus expect a reduction in the Gini coefficient during optimization, as a balanced UEs distribution avoids congested cells.

### B. GD-BASED TRANSMIT POWER TUNING

In order to optimize the transmit power through GD, we consider the scheme sketched in Fig. 5. As such, we introduce a latent optimization variable $\tilde{\gamma}_j$ for each cell $j$, and map it to an activation parameter $\gamma_j$ through a sigmoid function resulting in $\gamma_j = \sigma\left(\tilde{\gamma}_j\right) \in (0, 1)$. Subsequently, the $\gamma_j$ parameter controls the transmit power for cell $j$:

$$p_j = p_{\min} + \gamma_j \cdot \left(p_{\max} - p_{\min}\right), \tag{10}$$

by mapping it to the allowed interval $p_j \in (p_{\min}, p_{\max})$. Subtracting the pathloss then yields the RSRP values $r_{i,j}$,
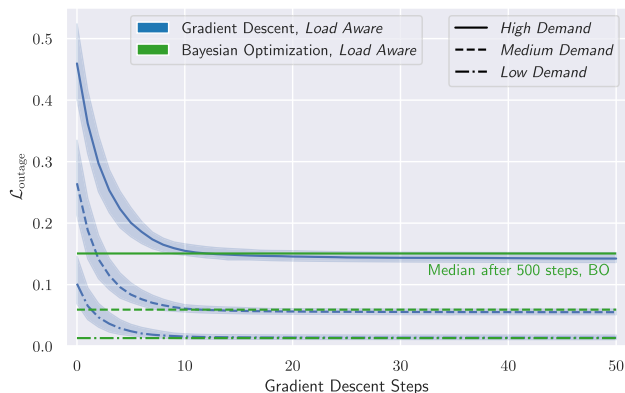
**FIGURE 7. Statistics of optimization history over 100 runs for *load-aware* GD. Resulting outage ratios $\mathcal{L}_{\text{outage}}$ are within narrow intervals and comparable to global-optimization BO.**

which acts as the entry point to the differentiable throughput model in Fig. 2. For joint transmit power optimization we thus have to specify the tunable vector $\tilde{\gamma}$, consisting of all $C$ trainable parameters. Hereby we start from a random initial transmit power configuration with $p_j$ drawn uniformly at random from the allowed interval. Subsequently we optimize the latent variable vector:

$$\tilde{\gamma}^{(*)} = \arg\min \, \mathcal{L}_{\text{outage}}\left(\tilde{\gamma}\right), \tag{11}$$

through the gradient $\nabla_{\tilde{\gamma}}$ of the *load-aware* loss function. Likewise, we will also consider the *load-unaware* objective, to characterize how the two optimization targets differ.

### C. RADIO SIMULATION ENVIRONMENT

To ensure a realistic scenario, we base our evaluation on the real-world macro-cell LTE deployment shown in Fig. 6, with $C = 147$ cells serving an urban area of $\approx$ 8 km$^2$ in the city of Vienna, Austria. For all of these cells, the respective operator provided us with detailed information about the network configuration parameters, i.e., we have exact knowledge of the BS position, antenna height, and the horizontal and vertical antenna orientation. Overall, the considered network deployment consists of two distinct carrier frequencies, each with 20 MHz bandwidth. As such, 48 cells transmit at 800 MHz, while the remaining 99 cells belong to the 1800 MHz band. As apparent in Fig. 6, the different propagation conditions are accounted for in the network deployment, with the low attenuation 800 MHz band ensuring basic coverage, while the higher pathloss 1800 MHz carrier offers the necessary capacity in a more densely deployed configuration. When computing the RS-SINR following (3), we treat both of these bands separately as they do not generate interference among each other. However, we consider them jointly for the cell assignment in (2), to account for offloading of UEs between carriers which is critical in real-world operation. In total, we consider $U = 10\,000$ UEs, uniformly distributed within the Voronoi[6]

---

[6]Note, that we consider interference generated by cells outside of the Voronoi area, while excluding them for the cell assignment. Thus, all considered UEs have to be served by the $C$ cells shown in Fig. 6.
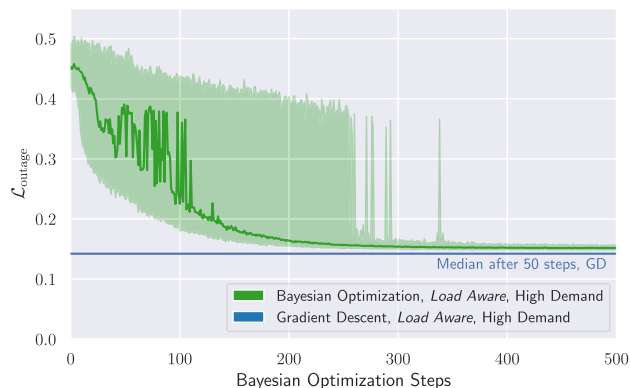


**FIGURE 8. Statistics of the outage ratio for BO over 100 runs in the *load-aware* case with high demand. Compared to GD, the optimization procedure is drastically prolonged.**

**TABLE 3. Median outage ratio over 100 runs for *load-aware* optimization after 50 steps for GD and 500 steps for BO.**

|  | Gradient Descent | Bayesian Optimization |
|---|---|---|
| *High Demand* | 0.14 | 0.15 |
| *Medium Demand* | 0.06 | 0.06 |
| *Low Demand* | 0.01 | 0.01 |

areas shown in Fig. 6. Overall, this amounts to $\approx$ 70 connected UEs per cell, which is within the range of the monitoring data in Fig. 4. For all UEs positions, we then conduct extensive radio simulations using the network planning tool from [7], directly trained on end-user RSRP measurements. Hence, our simulations are well validated for the considered region and make use of a detailed 3D building model of the environment to capture the relevant propagation mechanisms in urban areas. To adequately describe indoor locations, we further introduce an additional 10 dB penetration loss, which is in line with extensive MDT data analysis in [23]. We further use a constant noise term of $\zeta = -120$ dBm, and set $N_{\text{RB}} = 100$ with $B_{\text{RB}} = 180$ kHz according to the LTE standard definition for a 20 MHz carrier.

Overall, both in terms of scale and complexity, the considered deployment is well suited for a realistic evaluation of the proposed scheme.

### V. PERFORMANCE EVALUATION AND RESULTS

Based on the realistic scenario described in Sec. IV-C, we now evaluate our differentiable network optimization framework. Here, we first validate the GD-based transmit power optimization scheme by comparing the proposed approach to a global-optimization BO baseline. Subsequently, we focus on differentiating our *load-aware* objective from the *load-unaware* reference objective. Finally, we will characterize further achievable improvements by optimally distributing the UEs through *active load balancing*. If not stated otherwise, we will use a fixed constant throughput threshold of $t^{(\text{thresh})} = 10$ MBit/s during the evaluation. Further, we consider a feasible transmit power interval of $p_{\min} = -15$ and $p_{\max} = 15$ dBm for the network optimization scenario, denoted in power per LTE resource
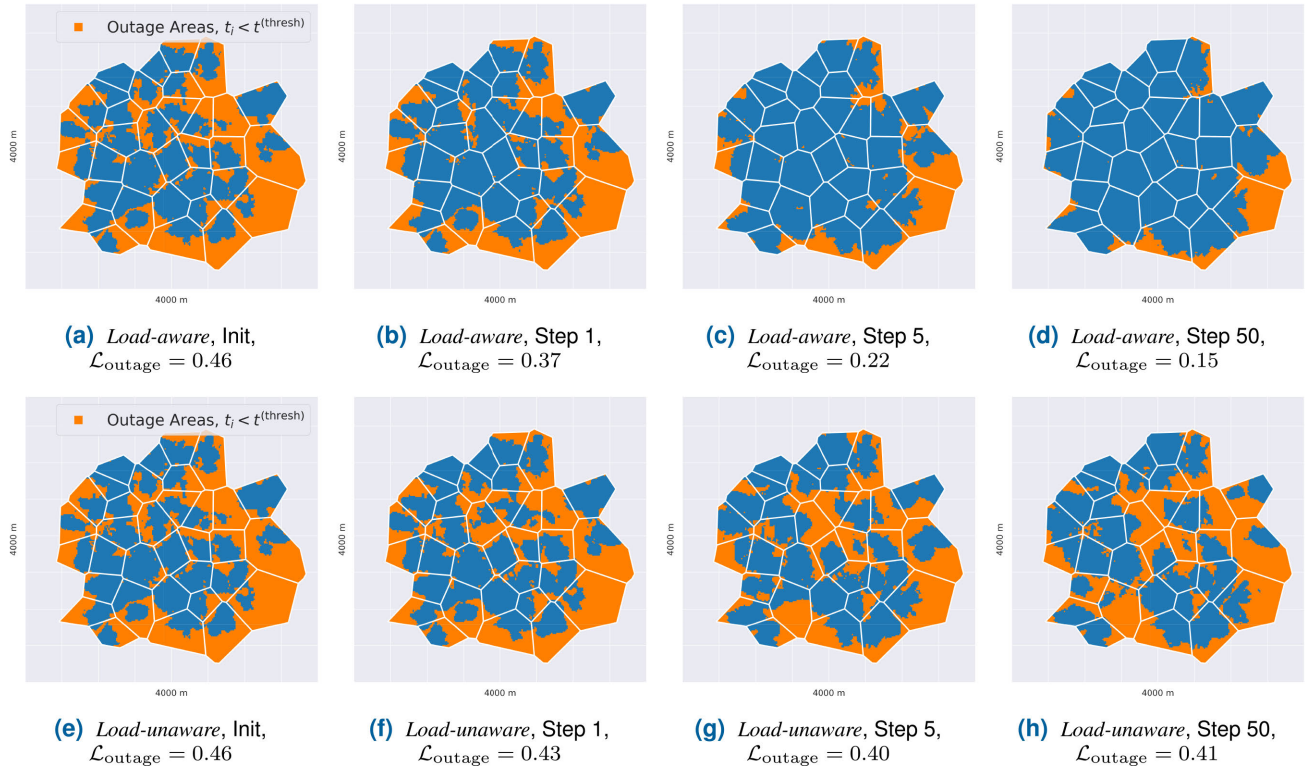
**FIGURE 9.** Exemplary outage areas for high demand during GD optimization based on the *load-aware* or *load-unaware* objective.

element. Throughout this work, we use the well-known Adam optimizer for the GD variant [42], with the BO implementation from [43] acting as the baseline.

### A. VALIDATION & COMPARISON WITH BO BASELINE

We first focus on the overall validation of the GD-based minimization of the differentiable *load-aware* objective from (8). As described in Sec. IV-B, we generally start from a random network configuration by drawing $p_j$ uniformly from the feasible transmit power interval. To assess the role of the initialization for the GD-based scheme, we conduct 100 repeated runs of the optimization procedure, each starting from a distinct randomly drawn configuration. This is necessary, as we can, in general, not expect to reach the global optimum using the GD-based scheme, but rather target a sufficiently well performing local one. This is in contrast to the considered BO baseline, which, in principle, targets a global optimum, but typically struggles with high dimensional search spaces.

The resulting optimization history[7] for GD is shown in Fig. 7, highlighting the median as well as the 5 and 95 percentile of the outage ratio $\mathcal{L}_{\text{outage}}$ for the cases of low, medium and high demand. For all levels of demand, the loss decreases consistently, reaching the optimized configuration after only a small number of around ten GD steps. Also the role of the

---

[7]Note, that we select a constant step-size of $1 \times 10^{-1}$ for the Adam optimizer in the GD variant, and use the default settings for the BO baseline.

**TABLE 4.** Comparison of metrics at initialization and after *load-aware* or *unaware* optimization for high demand case.

| | Med. Throughput, $t_i$ [MBit/s] | Med. RS-SINR, $s_i$ [dB] | Congestion Meas. $g_{\text{coef}} \in [0, 1]$ |
|---|---|---|---|
| Initialization | 10.1 | 3.30 | 0.71 |
| *Load-aware* | **21.0** | 6.68 | **0.43** |
| *Load-unaware* | 11.1 | **7.24** | 0.73 |

initialization seems manageable, as the confidence intervals of the optimized outage ratios are sufficiently narrow.

This is in stark contrast to the results for black-box BO. Here, the optimization history in Fig. 8, again showing the median, the 5 and 95 percentiles over 100 independent runs, reports a substantially prolonged optimization phase, accompanied by a significant risk of probing poorly performing configurations. As expected, applying the scheme directly to operational networks, as often suggested in the literature, would thus significantly deteriorate end-user performance in practice. Ultimately, the final outage ratios in Tab. 3 report similar performance for BO and GD, showing that the differentiable throughput formulation guides the GD scheme towards reasonably well performing configurations, which can not be significantly improved upon by further search-space exploration through BO. As standard BO is known to struggle for such high-dimensional spaces, we further evaluated two state-of-the-art scalable BO variants, which did, however, not yield significant improvements in the final outage ratio and the number of function evaluations.
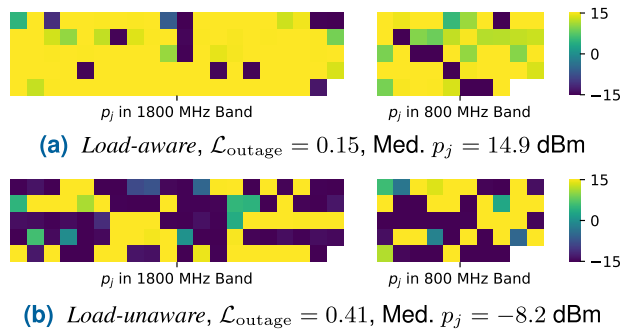
**FIGURE 10.** Overview of transmit power setting $p_j$ after optimization. Comparison of *load-aware* and *unaware*. Each pixel represents one cell, with a consistent ordering to enable comparison.



**FIGURE 11.** Overview of number of connected UEs per cell $u_j$ after optimization. Comparison of *load-aware* and *unaware*. Each pixel represents one cell, with a consistent ordering to enable comparison.
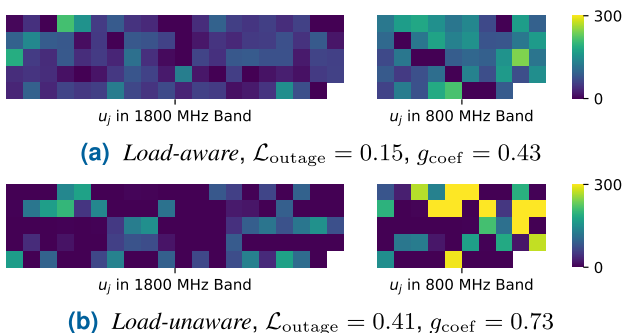
We discuss these results in detail in Appendix B, where we also characterize the substantial computational complexity of BO. In contrast, the GD-based scheme is particularly efficient, both in terms of computational complexity and the required number of function evaluations, showing that gradient information is highly valuable for network optimization.

### B. LOAD-AWARE VS. UNAWARE OPTIMIZATION

With the overall feasibility of GD-based optimization of the transmit power validated, we now focus on studying the parameter settings promoted by the objective in (8) in detail. In particular, we want to assess how the obtained configurations for *load-aware* and *load-unaware* optimizations differ.

To study this aspect, we consider an exemplary optimization procedure for a distinct initialization under high end-user demand. We visualize the outage areas[8] for such an initial configuration in Figs. 9a and 9e, amounting to an overall outage ratio of $\mathcal{L}_{\text{outage}} = 0.46$. Clearly, the throughput threshold is violated throughout the considered area, most notably at the cell edges indicated by the Voronoi tessellation. The poor configuration of the network is also apparent from the metrics in Tab. 4, showing a low median throughput and SINR. This is accompanied by a relatively large congestion measure $g_{\text{coef}}$, suggesting that the current configuration causes an uneven distribution of UEs among the cells. The *load-aware* optimization with GD, then yields

[8]Note, that we use nearest neighbor interpolation to construct dense outage maps from sparse UEs positions.
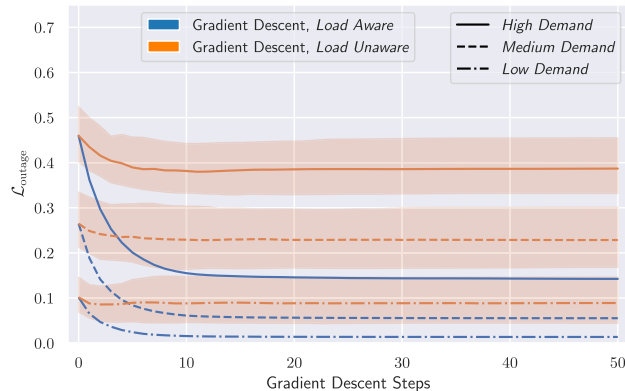
**FIGURE 12.** Comparison of optimization history over 100 runs for *load-aware* and *unaware*. The exclusive focus on interference reduction consistently leads to high outage ratios.

a consistent reduction of the throughput violations in Figs. 9b and 9c, resulting in the substantially lower outage ratio of $\mathcal{L}_{\text{outage}} = 0.15$ after 50 steps. The corresponding outage map is shown in Fig. 9d, with the outage areas effectively limited to the lower right region, where the cell deployment is less dense — compare Fig. 6. Note, that also the metrics in Tab. 4 are improved during the optimization. As such, we observe an increase in the median throughput and also the RS-SINR, accompanied by a substantial reduction of the congestion measure.

These trends are in stark contrast to the obtained configuration, when optimizing the transmit power based on the *load-unaware* objective from Sec. IV-A. Here, the optimization history is shown in Figs. 9e to 9h, again starting from the same initial configuration. Interestingly, minimizing the *load-unaware* objective only slightly reduces the outage ratio from (8). As such, we obtain $\mathcal{L}_{\text{outage}} = 0.41$ for the configuration shown in Fig. 9h, with considerable throughput violations throughout the optimization area. While this behavior seems counter-intuitive at first, it can be explained by the metrics in Tab. 4, with only a minor increase of the median throughput from the already low initial value. This is combined with a substantial increase for the RS-SINR, as is to be expected for the *load-unaware* objective, which solely targets interference management. However, the exclusive focus on the RS-SINR improvement seems to be partially conflicting with a balanced UE distribution described by the congestion measure $g_{\text{coef}}$. As such, Tab. 4 reports an increase from the already high congestion measure for the initialization, showing that only insufficient resources are available for large parts of the considered UEs.

To further validate this observation, we study the transmit power configurations obtained for the two objectives. Fig. 10 visualizes the obtained configuration through heatmaps, with each pixel representing the obtained $p_j$ settings for one of the 147 cells in a consistent order to enable comparison. Hereby, we split the two bands and report the 99 cells for the 1800 MHz band separately from the 48 transmitting at 800 MHz. This visualization later allows to directly compare
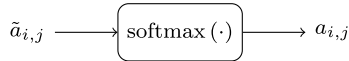
**FIGURE 13.** For a fixed configuration $r_{i,j}$, we can actively balance the cell assignment through the latent variable $\tilde{a}_{i,j}$.

the transmit power setting $p_j$ with the number of connected UEs $u_j$. For the *load-aware* case in Fig. 10a, the majority of cells are close to the upper limit of the feasible transmit power interval $p_{\max}$. Again, this is in stark contrast to the transmit power configuration for the *load-unaware* variant in Fig. 10b. Here, the interference reduction objective seems to promote drastically lower transmit power configurations, with a substantial number of cells set to $p_{\min}$. While the significant reduction of the transmit power in Fig. 10c seems to be beneficial for interference management, it also results in a configuration where large parts of the considered area are served by only a subset of the available cells. By definition, this poses a high risk of overloaded cells, as the available resources have to be shared among a higher number of connected UEs.

To highlight the close relation of the transmit power setting $p_j$ and cell assignment, Fig. 11 depicts the number of connected UEs $u_j$ for all cells, using the same visualization[9] as for the transmit power. The comparison of Figs. 11a and 11b clearly reveals a much more balanced cell assignment for the *load-aware* optimization, compared to the *load-unaware* variant, where we observe a substantial number of congested cells. Notably, most of these overloaded cells are transmitting in the 800 MHz band, which is not surprising when considering the preferable propagation conditions compared to the higher attenuated 1800 MHz carrier, leading to increased cell sizes at similar transmit power settings. This visual inspection is also in line with the results for $g_{\text{coef}}$ in Tab. 4, which reports drastically different values for the *load-aware* and *unaware* case, and hence acts as an adequate congestion measure.

Note, that the observations for the exemplary initialization also hold for repeated runs as visualized in Fig. 12. For all levels of demand, the difference between the *load-aware* and *load-unaware* objective is substantial, also indicated by the wide confidence interval for the *load-unaware* objective. Overall, we would argue that this effect is relevant for most urban network deployments that are, in general, interference limited as a side-effect of the dense deployment required to serve the high number of UEs. In such a scenario, a SINR-based objective can trivially reduce the interference, but also risks deteriorating the end-user performance by omitting the possibility of congested cells. In contrast, this trade-off is natively incorporated in the proposed *load-aware* optimization based on the end-user throughput expression from (5).

In the following, we will show that the same objective can also be used for an extension that enables *active*

[9]Note, that we cap $u_j$ to a maximum of 300 for better visual inspection in Fig. 11. Still, the total number of UEs sums up to $U$ for both cases.
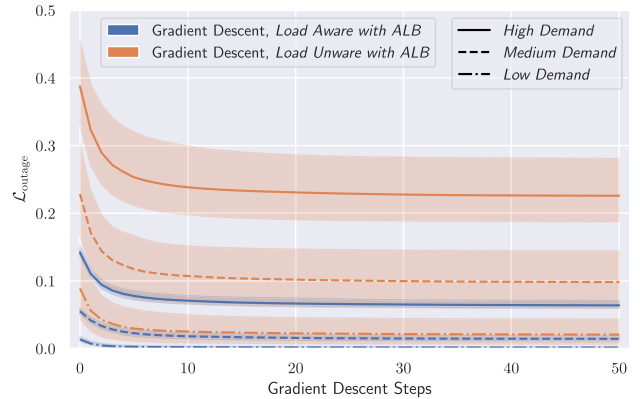


**FIGURE 14.** Outage ratio over GD-steps during *active load balancing (ALB)* starting from transmit power configurations optimized through the *load-aware* or *load-unaware* objectives.

**TABLE 5.** Median outage ratio over 100 runs after *active load balancing (ALB)* for the *load-aware* or *unaware* configuration.

|  | Load-aware with ALB | Load-unaware with ALB |
|---|---|---|
| *High Demand* | **0.06** | 0.23 |
| *Medium Demand* | **0.01** | 0.10 |
| *Low Demand* | **0.00** | 0.02 |

*load balancing* within our differentiable framework. This allows us to study the apparent relation between network optimization and load balancing more closely.

## VI. ACTIVE LOAD BALANCING EXTENSION

The results in Sec. V-B show a direct relation between transmit power settings and the UE distribution following the differentiable assignment model incorporated in the throughput expression from Sec. II-C. While the RSRP-based assignment is a common assumption that reflects the behavior of real networks sufficiently well, it can be suboptimal for edge-cases, where the active reassigning of UEs would allow for the discharging of highly loaded cells. Hence, we also consider an *active load balancing* extension of our differentiable framework, showing that the same objective function can also propose optimal assignments to distribute UEs among the available cells. Here, it will be of particular interest to study whether the observations from Sec. V-B hold, or whether the poor network configurations for the *load-unaware* objective can partially be compensated through an optimized reassignment of the considered UEs.

### A. GD-BASED OPTIMAL CELL ASSIGNMENT

When optimizing the network configuration, we determine the cell assignment directly from the obtained RSRP values $r_{i,j}$, following the model in Sec. II-A. However, for a fixed network configuration and thus fixed values of $r_{i,j}$, we can alternatively interpret the cell assignment as the optimization target. For this, we follow the scheme sketched in Fig. 13 and introduce a latent variable $\tilde{a}_{i,j}$, which directly determines the assignment probability $a_{i,j}$ through a softmax function:

$$a_{i,j} = \text{softmax}\left(\tilde{a}_{i,j}\right). \tag{12}$$

**(a)** *Load-aware with ALB*, Init, $\mathcal{L}_{\text{outage}} = 0.15$

**(b)** *Load-aware with ALB*, Step 1, $\mathcal{L}_{\text{outage}} = 0.12$

**(c)** *Load-aware with ALB*, Step 5, $\mathcal{L}_{\text{outage}} = 0.08$

**(d)** *Load-aware with ALB*, Step 50, $\mathcal{L}_{\text{outage}} = 0.07$

**(e)** *Load-unaware with ALB*, Init, $\mathcal{L}_{\text{outage}} = 0.41$

**(f)** *Load-unaware with ALB*, Step 1, $\mathcal{L}_{\text{outage}} = 0.35$

**(g)** *Load-unaware with ALB*, Step 5, $\mathcal{L}_{\text{outage}} = 0.25$

**(h)** *Load-unaware with ALB*, Step 50, $\mathcal{L}_{\text{outage}} = 0.21$
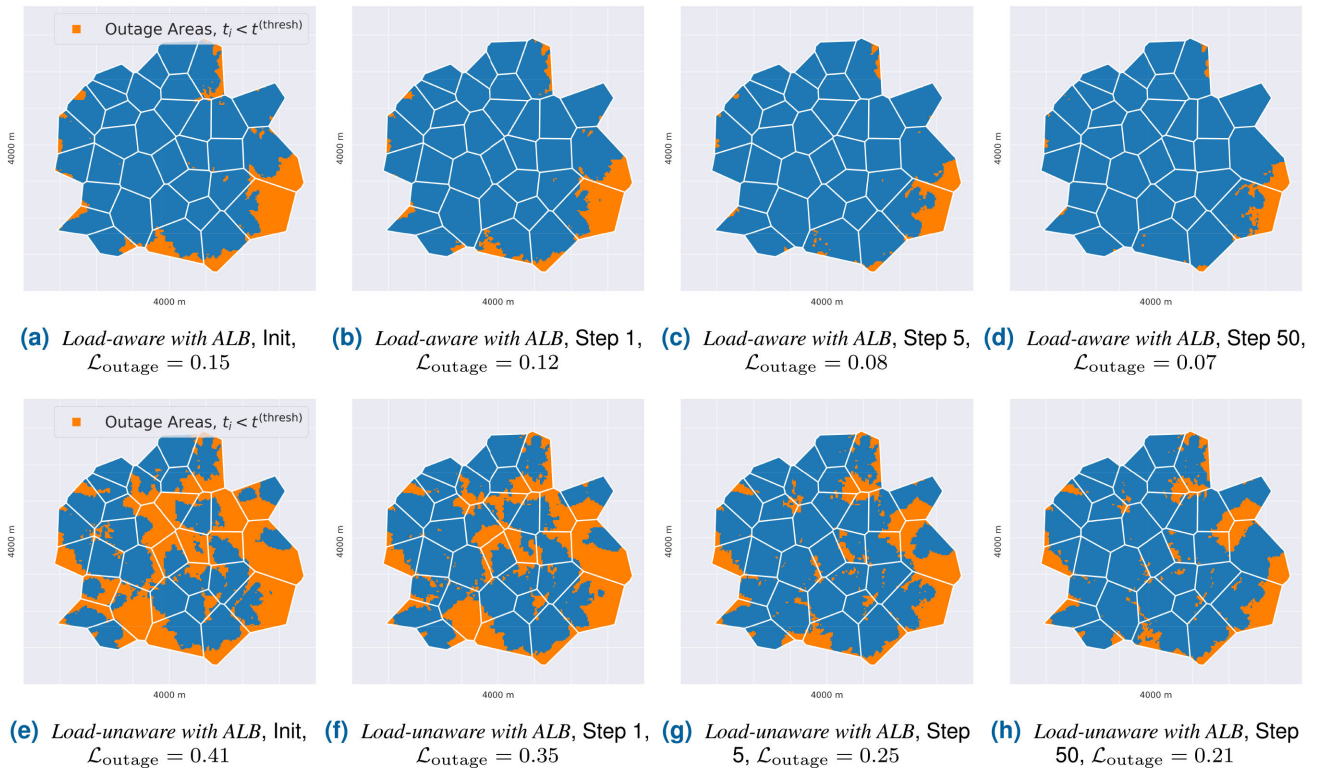
**FIGURE 15.** Exemplary evolution of outage areas for high demand case during *active load balancing (ALB)* via GD.

This adaptation can trivially be included in our GD-based framework, by using (12) as a replacement for the RSRP-based assignment in (2). Note, that for *active load balancing*, we have a substantially higher number of $U \times C$ trainable parameters. We collect all of these parameter in a matrix $\tilde{\mathbf{A}}$ and then follow a GD-based procedure to obtain:

$$\tilde{\mathbf{A}}^{(*)} = \arg\min \; \mathcal{L}_{\text{outage}}\left(\tilde{\mathbf{A}}; \tilde{\boldsymbol{\gamma}}^{(*)}\right). \tag{13}$$

As such, we consider *active load balancing* as a subsequent step for a fixed, already optimized, transmit power configuration $\tilde{\boldsymbol{\gamma}}^{(*)}$. This allows us to quantify further outage reductions possible by optimally reassigning UEs across available cells. While several initial values are possible for $\tilde{\mathbf{A}}$, we found that the RSRP-based assignment[10] for the considered network configuration $\tilde{\boldsymbol{\gamma}}^{(*)}$ worked sufficiently well in our scenario. By definition, the optimization in (13) requires the *load-aware* objective. However, we consider the extension for transmit power configurations $\tilde{\boldsymbol{\gamma}}^{(*)}$ optimized through both the *load-aware* and *unaware* variant. This way, we can characterize whether suboptimal configurations can partially be compensated by the subsequent *active load balancing* step.

### B. ACTIVE LOAD BALANCING EVALUATION

For the evaluation, we start from the optimized transmit power configurations from Sec. V-A, and subsequently apply GD-based *active load balancing*. Here, we again use the

**TABLE 6.** Metrics after subsequent *active load balancing (ALB)* for the *load-aware* and *load-unaware* configuration.

| | Med. Throughput, $t_i$ [MBit/s] | Med. RS-SINR, $s_i$ [dB] | Congestion Meas. $g_{\text{coef}} \in [0, 1]$ |
|---|---|---|---|
| *Load-aware with ALB* | **20.0** | 5.96 | **0.39** |
| *Load-unaware with ALB* | 15.5 | **6.37** | 0.60 |

Adam optimizer,[11] but omit the BO baseline, as the high number of trainable parameters renders it unsuitable for determining the optimal cell assignment. Hence, we study the performance improvements compared to the case without the extension, instead of a global-optimization reference.

The optimization history for *active load balancing* is shown in Fig. 14, where we start from the configurations obtained either through the *load-aware* or the *unaware* transmit power configurations from Sec. V-B. In both cases, Fig. 14 shows, that the outage ratio can indeed be further reduced by adequately redistributing the UEs among the available cells. For both cases, we observe a consistent decrease of the outage ratio throughout the *active load balancing* procedure, indicating that the GD-based scheme can reliably identify sufficiently well performing assignment settings. However, Fig. 14 also shows substantial differences for the two cases, with a substantial gap between the *load-aware* and *unaware* configuration also remaining after *active load balancing*. The offset is also apparent from the median

---

[10]In particular, we set $\tilde{a}_{i,j} = -\Delta r_{i,j}$, such that $a_{i,j}$ is initialized with the RSRP-based assignment introduced in (2).

[11]Note, that we select a slightly higher learning rate of $5 \times 10^{-1}$ for *active load balancing*, which performed well in our evaluation.

**(a)** *Load-aware with ALB,* $\mathcal{L}_{\mathrm{outage}} = 0.07$, $g_{\mathrm{coef}} = 0.39$



**(b)** *Load-unaware with ALB,* $\mathcal{L}_{\mathrm{outage}} = 0.21$, $g_{\mathrm{coef}} = 0.60$
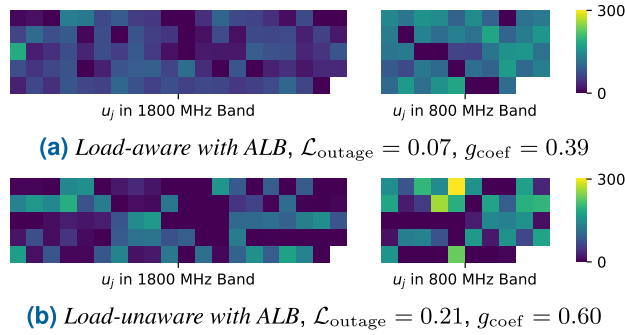
**FIGURE 16.** Connected UEs $u_j$ after *active load balancing (ALB)* for the *load-aware* and *load-unaware* configuration. Each pixel represents one cell, with a consistent ordering to enable comparison.



**FIGURE 17.** Overview of change $\Delta$ for selected metrics during optimization, with and without subsequent *active load balancing (ALB)*. Note, that metrics are reported relative to the median values for the random initializations.

performance summarized in Tab. 5, where the outage ratio for *load-unaware* remains large relative to the *load-aware* case.

We can again consider the exemplary initialization from Sec. V-B and study the *active load balancing* procedure in more detail. As such, Figs. 15a and 15e show the outage areas obtained after the *load-aware* and *load-unaware* optimization of the transmit power from Sec. V-B. Starting from these configurations, the GD-based *active load balancing* procedure is applied to the *load-aware* scenario in Figs. 15b and 15c, finally resulting in a reduced outage ratio of $\mathcal{L}_{\mathrm{outage}} = 0.07$ in Fig. 15d. Similarly, the *active load balancing* procedure also reduces the throughput violations for the configuration obtained through the *load-unaware* objective shown in Figs. 15e to 15h. After the reassignment, we then obtain the configuration in Fig. 15h, with an overall outage ratio of $\mathcal{L}_{\mathrm{outage}} = 0.21$. The reduction of the outage areas is also apparent from the metrics in Tab. 6. Here, the increase of the median throughput for the *load-unaware* case is mainly caused by a reduction of the congestion measure by more than 0.1 compared to Tab. 4. As expected, we also observe a minor decrease in the RS-SINR, as the *active load balancing* step partially reassigns UEs to weaker serving cells in order to mitigate congestion. Apparently, the offloading is successful, with the visualization of the cell utilization in Fig. 16 showing a substantially more balanced UEs assignment. In particular, this applies to the 800 MHz band where we observed many congested cells before *active load balancing* in Sec. V-B. In Fig 16, these cells are now discharged, with the UEs distributed among cells offering similar levels of RS-SINR. Still, there remains a large gap between the congestion measure for the *load-aware* and *load-unaware* case also after *active load balancing*.

We conclude, that the load balancing aspect has to be incorporated into the optimization, as poor transmit power configurations can only partially be compensated for by optimal handover decisions from the network side.

## VII. FINAL COMPARISON AND DISCUSSION
Throughout the evaluation, we observed substantial differences between the transmit power configurations promoted by the *load-aware* and *unaware* objective. For a final direct
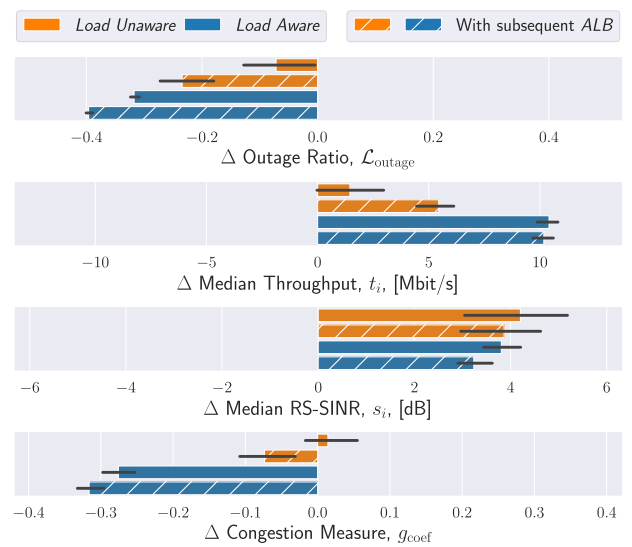
comparison, Fig. 17 summarizes the optimization history in the case of high demand for all four considered variants. In particular, we report the difference in the outage ratio, the median throughput, the median RS-SINR, and the congestion measure after optimization, compared to the median values for the randomly initialized transmit power configuration. For all these metrics, Fig. 17 shows the median change together with the 5 and 95 percentiles.

Again, we note that the outage ratio reduction is largest for the *load-aware* case and can further be increased through the subsequent *active load balancing* step. These trends are directly related to the improvement in median throughput, which is significantly lower for the *load-unaware* objective, even after the subsequent *active load balancing step*. All of this is accompanied by an increase of the RS-SINR, which is largest for the *load-unaware* case. In contrast, the *load-aware* optimization yields significantly smaller improvements, as it does not solely focus on interference management, but also accounts for the risk of congested cells. Accordingly, we also observe a decrease of the RS-SINR during the subsequent *active load balancing* step, as UEs are reassigned to discharge overloaded cells. Finally, the high outage ratios for the *load-unaware* optimization despite the superior RS-SINR can be explained by the increase of the median congestion measure, starting from the already high random initialization. While the congestion can partially be compensated by the subsequent *active load balancing* step, the large gap relative to the *load-aware* objective remains.

Based on the comparison from Fig. 17, we can thus conclude that the *load-unware* optimization of the network configuration is unsuitable as it consistently leads to congested cells and hence low end-user throughput. This congestion is not only caused by the strict RSRP-based assignment and can only partially be compensated by active
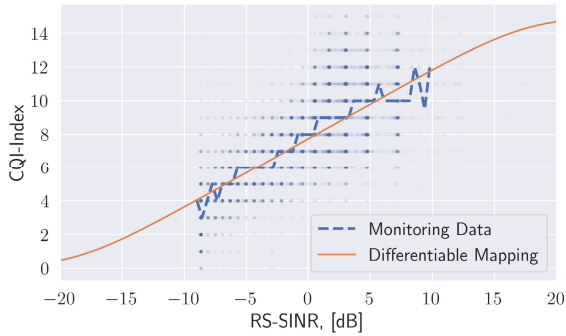
**FIGURE 18.** Obtained RS-SINR to CQI mapping from MDT, only considering UEs from the most common manufacturer.



**FIGURE 19.** Comparison of standard BO and the scalable variants. Results are computed over 100 runs in high demand.

**TABLE 7.** Comparison of GD and BO variants, showing steps per second and optimized outage ratios for high demand.

| | Steps | Only CPU | With GPU | $\mathcal{L}_{\text{outage}}$ |
|---|---|---|---|---|
| Gradient Descent | 50 | 0.740 step/s | 5.741 step/s | 0.14 |
| BO, Standard | 500 | 0.050 step/s | - | 0.15 |
| BO, TuRBO | 5 000 | 0.021 step/s | - | 0.14 |
| BO, SAASBO | 250 | 0.007 step/s | - | 0.19 |

reassignment of UEs. In contrast, the *load-aware* formulation does not suffer from this issue and adequately balances the interference while avoiding congested cells. For the interested reader, further analysis is provided in Appendix C, where the complete UE population is visualized during the optimization procedure.

## VIII. CONCLUSION

In this work, we derived a comprehensive, differentiable objective function for *load-aware* cellular network optimization that utilizes data-driven components while relying on model-based formulations for well-understood mechanisms in mobile communications. This enables direct calibration to MDT measurements and monitoring data, and still provides gradient information for efficient and scalable optimization.

In our large-scale evaluation, the proposed GD-based approach showed faster convergence and a lower risk of degrading end-user performance compared to the black-box BO baseline. Similarly, none of the considered global-optimization methods was able to identify a substantially better performing transmit power configuration, indicating that the GD-based approach guides the optimization towards well performing minima. We further observed a clear relation between cellular network optimization and load balancing, as the commonly used *load-unaware* SINR-based objective consistently yielded configurations with congested cells and thus insufficient throughput for connected UEs. This highlights the importance of accounting for limited resources and the respective UE demand instead of solely targeting interference management. Our results clearly show that both of these aspects can be combined in the proposed objective without leaving the differentiable framework.

Overall, our results suggest a prominent role for GD-based network optimization, in particular when considering recent advances for differentiable propagation modeling that enable the extension to further antenna parameters. In future work, we will incorporate such models into our framework while also studying potential side constraints, e.g., energy consumption. We further target an online implementation where the model interacts directly with the
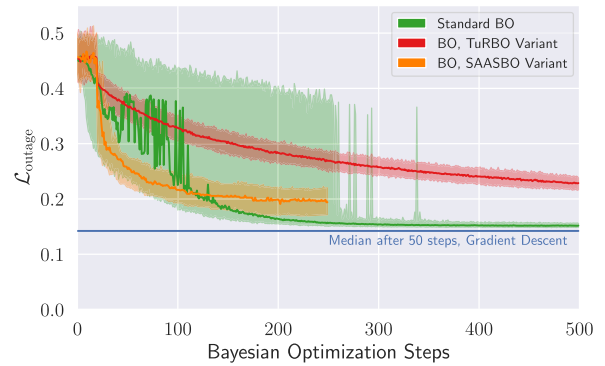
network deployment, receiving more immediate feedback from monitoring data.

## APPENDIX

### A. RS-SINR TO CQI MAPPING FROM MDT DATA

The spectral efficiency mapping $\eta\left(s_{i,j}\right)$ used in Fig. 2 consists of the mapping from RS-SINR to the CQI, from which the spectral efficiency is computed following the 3GPP standard [40, p. 220]. As the RS-SINR does not uniquely[12] determine the selected CQI [2], [4], we retrieve an empirical mapping from MDT data, capturing the representative UE characteristics and average channel conditions. Further, the RS-SINR is not directly reported in MDT. Hence, we follow the approach in [2] and retrieve the RS-SINR from the RSRQ. With the definition of the received signal strength indicator (RSSI) we compute in linear scale:

$$
\begin{aligned}
\text{RS-SINR} &= \frac{12 \cdot N_{\text{RB}} \cdot \text{RSRP}}{\text{RSSI} - N_{\text{RB}} \cdot \text{RSRP} \cdot (2 + 10 \cdot \lambda)} \\
&= \frac{12 \cdot N_{\text{RB}} \cdot \text{RSRP}}{\frac{N_{\text{RB}} \cdot \text{RSRP}}{\text{RSRQ}} - N_{\text{RB}} \cdot \text{RSRP} \cdot (2 + 10 \cdot \lambda)} \\
&= \frac{12 \cdot \text{RSRQ}}{1 - \text{RSRQ} \cdot (2 + 10 \cdot \lambda)}.
\end{aligned}
\tag{14}
$$

Hence, (14) yields the respective RS-SINR from the RSRQ by discarding the servings cells contribution to the RSSI. Note, that the expression is a minor adaptation of [2], by distinguishing between the load-dependent (ten resource elements) and the load-independent (two resource elements) interference contributions. We refer the interested reader to [4] for a detailed discussion of RSRQ and RS-SINR.

---

[12]Due to different channel characteristics a given RS-SINR value can correspond to different CQIs [4] — see also [44]. The specific relation further depends on the considered UE [2].
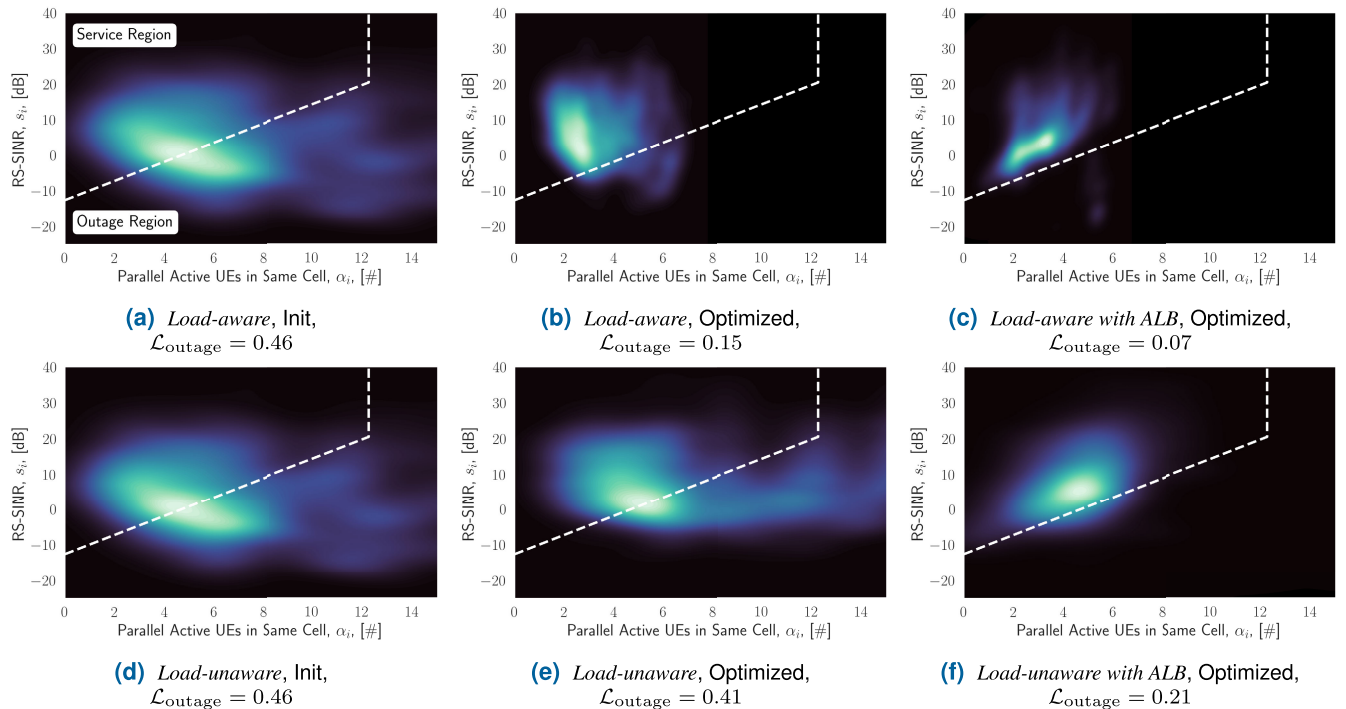
**FIGURE 20.** Density plots of the UE population under high demand. We show the RS-SINR and the number of active UEs in the same cell on a per-user level. The service and outage region refers to a threshold of $t^{(\text{thresh})} = 10$ MBit/s.

We then apply (14) to MDT measurements from the area of interest and extract the relation for the median CQI over bins of the RS-SINR. Fig. 18 highlights the obtained relation for around 50 000 measurements, filtered for the most common device manufacturer in our dataset. In line with the lab measurements in [4, p. 73], the observed trend from the MDT data is approximately linear. For our mapping, we thus fit a linear model of the form $\tilde{\kappa}_{i,j} = w_0 \cdot s_{i,j} + w_1$, with $\tilde{\kappa}$ denoting the CQI index and the fitted parameters given by $w_0 = 0.41$ and $w_1 = 7.74$. Subsequently, we ensure that the output adheres to the valid interval by introducing squashing functions through $\kappa_{i,j} = \text{softplus}\left(\tilde{\kappa}_{i,j}\right) - \text{softplus}\left(\tilde{\kappa}_{i,j} - 15\right)$.

Alternatively, the mapping can also be computed over all device categories. In our case, this resulted in a lower average CQI, as we found a substantial number of internet-of-things devices with less advanced hardware.

### B. BO VARIANTS & COMPUTATIONAL COMPLEXITY
In Sec. V-A, the standard BO implementation from [43] shows a substantially prolonged optimization phase compared to the proposed GD-based approach. Still, standard BO performs surprisingly well given the high dimensional search space. For a more comprehensive picture of the given loss function and the BO performance, we additionally evaluate two state-of-the-art BO variants, which specifically address the issue of scaling to higher dimensions.

In particular, we consider the TuRBO variant proposed in [26], which tackles the high dimensionality by fitting a collection of local models. Additionally, we evaluate the SAASBO variant from [27], which utilizes surrogate models defined on sparse axis-aligned subspaces. Both of these

methods showed state-of-the-art performance on different benchmarking problems. The obtained optimization history[13] of TuRBO and SAASBO is shown in Fig. 19, with the standard BO results from Fig. 8 included for reference. For both methods, we note a much narrower outage ratio interval during optimization. This can be explained by the tendency of SAASBO and TuRBO to limit exploration, which standard BO often prioritizes over exploitation in high dimensional spaces [26], [27]. Still, SAASBO is not able to improve upon the obtained GD and standard BO configuration in terms of the outage ratio. Similarly, also TuRBO, with much slower convergence, shows poorer performance after the first 500 steps. In our evaluation, TuRBO does, however, overtake standard BO and SAASBO, achieving similar performance as GD after 5 000 function evaluations. We summarize these results in Tab. 7, where we also report the required steps until convergence, which differ drastically between the methods. While the relative performance of the different BO variants seems counterintuitive at first, it seems that the underlying assumptions of TuRBO and SAASBO are not directly applicable to the considered problem. We also suspect that the tendency of standard BO to extensively explore the domain boundary (see [27]) might play a certain role. From Fig. 10 we know that sufficiently well performing configurations can be found in that proximity. Hence, standard BO can find a relatively well performing starting point for further

---

[13]We use the reduced sampling budget configuration for SAASBO as described in [27]. For TuRBO we only consider TuRBO-1, as TuRBO-M is unsuitable for such high-dimensional problems [26]. We further select a batch size of one, which showed faster convergence in our analysis. For both methods, the optimization starts with 20 randomly drawn samples.

optimization rather quickly, while TuRBO and SAASBO struggle to identify this region. All in all, we would argue that the results in Fig. 19 and Tab. 7 again highlight the challenge of exploring the high dimensional search space in a black-box manner. In contrast, the GD approach reliable guides the optimization towards well performing minima.

For the BO variants, the drastically increased computational complexity is another limiting factor. Tab. 7 summarizes the iterations per second achievable on a mid-range gaming PC with a *AMD Ryzen 5 3.9 GHz* CPU. Hereby, an iteration refers to a GD step or a function evaluation for BO, respectively. Clearly, the GD-based approach is substantially more efficient per iteration as compared to the BO variants. In practice, the improvement is even larger, as the GD requires a much smaller number of iterations in the first place. As it is implemented in *tensorflow*, we can further make use of a *NVIDIA GeForce RTX 3060* GPU, which again drastically speeds up the optimization. For the BO variants, we also found that the required computation time varies from step to step and generally increases throughout the optimization procedure.[14] The results in Tab. 7 thus refer to the average over the required steps for the respective models.

Note, that in addition to the BO variants in Fig. 19, we also evaluated other common optimization methods such as Nelder-Mead and L-BFGS-B from [45]. We omit the details, as both of them were not able to identify a configuration below an outage ratio of 0.35 after 10 000 function evaluations.

### C. VISUALIZATION OF THE UE POPULATION

Throughout the discussion of the exemplary optimization procedure in Fig. 9 and 15, we consider median values to characterize the improvement or degradation of the end-user performance. To obtain a more comprehensive picture, we can visualize the relation between interference management and load-balancing on a per-user level. For this, we rely on the RS-SINR $s_i$ from Sec. IV-A and additionally represent the congestion for UE $i$ via the number of UEs in the same cell, again given by an expectation over the assignment $u_i = \sum_{j=1}^{C} a_{i,j} \cdot u_j$. By mapping $u_i$ to the number of active UEs $\alpha_i = \alpha\,(u_i)$, we then resolve the two contributing factors $s_i$ and $\alpha_i$, which determine the throughput for UE $i$.

The density plots in Fig. 20, now visualize the optimization procedure over the complete UE population for the exemplary initialization considered throughout this work. For reference, we also include the approximate boundary, which splits the service and outage region for the throughput threshold of $t^{(\text{thresh})} = 10$ MBit/s. Again, Fig. 20 shows the same trend already discussed throughout this work, with the *load-aware* objective in Fig. 20b improving upon the random initialization in 20a, both in terms of congestion and interference management. In contrast, the *load-unaware* objective in Fig. 20e only improves the RS-SINR, with

a substantial ratio of UEs still suffering from congestion. Finally, Fig. 20c and 20f shed light on the *active load balancing (ALB)* procedure, where reassignment of UEs allows for the reduction of congestion, such that the throughput target can be achieved for a higher number of UEs. Again, this is accompanied by a minor reduction of the RS-SINR, as the RSRP-based assignment, which generally yields the highest RS-SINR, is extended by load balancing considerations. Hence, UEs are reassigned among neighboring cells and the two frequency layers to reduce congestion, as long as a sufficient but potentially also lower RS-SINR is ensured.

### REFERENCES

[1] Z. A. Dahi and J. Á. Morell, "Models and solvers for coverage optimisation in cellular networks: Review and analysis," in *Proc. IEEE 9th Int. Conf. Sci. Electron., Technol. Inf. Telecommun. (SETIT)*, May 2022, pp. 312–319.

[2] M. Skocaj, L. M. Amorosa, G. Ghinamo, G. Muratore, D. Micheli, F. Zabini, and R. Verdone, "Cellular network capacity and coverage enhancement with MDT data and deep reinforcement learning," *Comput. Commun.*, vol. 195, pp. 403–415, Nov. 2022.

[3] R. M. Dreifuerst, S. Daulton, Y. Qian, P. Varkey, M. Balandat, S. Kasturia, A. Tomar, A. Yazdan, V. Ponnampalam, and R. W. Heath, "Optimizing coverage and capacity in cellular networks using machine learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 8138–8142.

[4] V. Raida, "Data-driven estimation of spatiotemporal performance maps in cellular networks," Ph.D. thesis, Inst. Telecommun., TU Wien, Vienna, Austria, 2021, doi: 10.34726/hss.2021.80644.

[5] F. Vannella, G. Iakovidis, E. A. Hakim, E. Aumayr, and S. Feghhi, "Remote electrical tilt optimization via safe reinforcement learning," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–7.

[6] Y. Liu, W. Huangfu, H. Zhang, and K. Long, "An efficient stochastic gradient descent algorithm to maximize the coverage of cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3424–3436, Jul. 2019.

[7] L. Eller, P. Svoboda, and M. Rupp, "A deep learning network planner: Propagation modeling using real-world measurements and a 3D city model," *IEEE Access*, vol. 10, pp. 122182–122196, 2022.

[8] J. Hoydis, F. A. Aoudia, S. Cammerer, M. Nimier-David, N. Binder, G. Marcus, and A. Keller, "Sionna RT: Differentiable ray tracing for radio propagation modeling," 2023, *arXiv:2303.11103*.

[9] M. Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/

[10] S. R. Samal, K. Swain, S. Bandopadhaya, N. Dandanov, V. Poulkov, S. Routray, and G. Palai, "Dynamic coverage optimization for 5G ultradense cellular networks based on their user densities," *Wireless Pers. Commun.*, vol. 128, no. 1, pp. 605–620, Jan. 2023.

[11] F. Vannella, A. Proutiere, Y. Jedra, and J. Jeong, "Learning optimal antenna tilt control policies: A contextual linear bandit approach," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, May 2022, pp. 740–749.

[12] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "A learning-based approach for autonomous outage detection and coverage optimization," *Trans. Emerg. Telecommun. Technol.*, vol. 27, no. 3, pp. 439–450, Mar. 2016.

[13] Y. Zhao, K. Zhang, and R. Han, "Multi-antenna tuning simulation platform by deep reinforcement learning," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP)*, Dec. 2019, pp. 1–6.

[14] E. Balevi and J. G. Andrews, "Online antenna tuning in heterogeneous cellular networks with deep reinforcement learning," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 4, pp. 1113–1124, Dec. 2019.

[15] X. Liu, G. Chuai, X. Wang, Z. Xu, W. Gao, K. Zhang, Q. Liu, S. Maimaiti, and P. Zuo, "QoE-driven antenna tuning in cellular networks with cooperative multi-agent reinforcement learning," *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1186—1199, 2024, doi: 10.1109/TMC.2022.3230711.

[16] A. Marinescu, Z. Jiang, S. Zhou, L. A. DaSilva, and Z. Niu, "Deep learning–based coverage and capacity optimization," *Machine Learning for Future Wireless Communications*. Hoboken, NJ, USA: Wiley, 2020, pp. 63–83. [Online]. Available: https://ieeexplore.ieee.org/document/8958870

---

[14]This is especially true for SAASBO, where further optimization after the plateau at step 250 became more and more unfeasible.

[17] E. Tekgul, T. Novlan, S. Akoum, and J. G. Andrews, "Joint uplink-downlink capacity and coverage optimization via site-specific learning of antenna settings," *IEEE Trans. Wireless Commun.*, p. 1, 2023, doi: 10.1109/TWC.2023.3313916.

[18] E. Tekgul, T. Novlan, S. Akoum, and J. G. Andrews, "Sample-efficient learning of cellular antenna parameter settings," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2021, pp. 1–6.

[19] M. Benzaghta, G. Geraci, D. Lopez-Perez, and A. Valcarce, "Designing cellular networks for UAV corridors via Bayesian optimization," 2023, *arXiv:2308.05052*.

[20] S. Bandopadhaya, S. R. Samal, S. K. Dora, and V. Poulkov, "Base station transmission power optimization in interference-limited cellular networks for maximum energy efficiency," in *Proc. 13th Int. Conf. Adv. Technol., Syst. Services Telecommun. (TELSIKS)*, Oct. 2017, pp. 228–231.

[21] V. Buenestado, M. Toril, S. Luna-Ramírez, J. M. Ruiz-Avilés, and A. Mendo, "Self-tuning of remote electrical tilts based on call traces for coverage and capacity optimization in LTE," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4315–4326, May 2017.

[22] V. Raida, P. Svoboda, M. Lerch, and M. Rupp, "Crowdsensed performance benchmarking of mobile networks," *IEEE Access*, vol. 7, pp. 154899–154911, 2019.

[23] S. Tripkovic, L. Eller, P. Svoboda, and M. Rupp, "Unbiased benchmarking in mobile networks: The role of sampling and stratification," *IEEE Access*, vol. 11, pp. 53772–53787, 2023.

[24] J. Buckman, D. Hafner, G. Tucker, E. Brevdo, and H. Lee, "Sample-efficient reinforcement learning with stochastic ensemble value expansion," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 8234–8244.

[25] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.

[26] D. Eriksson, M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek, "Scalable global optimization via local Bayesian optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 5496–5507.

[27] D. Eriksson and M. Jankowiak, "High-dimensional Bayesian optimization with sparse axis-aligned subspaces," in *Proc. Uncertainty Artif. Intell.*, 2021, pp. 493–503.

[28] D. Romero and S.-J. Kim, "Radio map estimation: A data-driven approach to spectrum cartography," *IEEE Signal Process. Mag.*, vol. 39, no. 6, pp. 53–72, Nov. 2022.

[29] S. Bakirtzis, K. Qiu, J. Zhang, and I. Wassell, "DeepRay: Deep learning meets ray-tracing," in *Proc. 16th Eur. Conf. Antennas Propag. (EuCAP)*, Mar. 2022, pp. 1–5.

[30] T. Orekondy, P. Kumar, S. Kadambi, H. Ye, J. Soriaga, and A. Behboodi, "WineRT: Towards neural ray tracing for wireless channel modelling and differentiable simulations," in *Proc. 11th Int. Conf. Learn. Represent.*, 2023.

[31] M. Gupta, R. M. Dreifuerst, A. Yazdan, P.-H. Huang, S. Kasturia, and J. G. Andrews, "Load balancing and handover optimization in multi-band networks using deep reinforcement learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.

[32] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[33] H.-H. Chang, H. Chen, J. Zhang, and L. Liu, "Decentralized deep reinforcement learning meets mobility load balancing," *IEEE/ACM Trans. Netw.*, vol. 31, no. 2, pp. 473–484, Apr. 2023.

[34] A. J. Fehske, H. Klessig, J. Voigt, and G. P. Fettweis, "Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1974–1988, Jun. 2013.

[35] H. Klessig, A. Fehske, G. Fettweis, and J. Voigt, "Cell load-aware energy saving management in self-organizing networks," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–6.

[36] W. Wang, J. Zhao, H. Qu, Y. Zhang, and G. Ren, "Joint optimization algorithm of user association and power control based on spatial traffic distribution," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 12, 2022, Art. no. e4608.

[37] H. Zhuang, J. Chen, and R. Gilimyanov, "Hierarchical energy optimization with more realistic power consumption and interference models for ultra-dense networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4507–4518, Jul. 2020.

[38] D. F. P. Rojas and A. Mitschele-Thiel, "A data driven coordination between load balancing and interference cancellation," in *Proc. NOMS IEEE/IFIP Netw. Operations Manage. Symp.*, Apr. 2022, pp. 1–6.

[39] A. Asghar, H. Farooq, and A. Imran, "Concurrent optimization of coverage, capacity, and load balance in HetNets through soft and hard cell association parameters," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8781–8795, Sep. 2018.

[40] *LTE; E-UTRA; Physical Layer Procedures*, Standard (TS), TS 36.213, Release 16, 3GPP, 2020.

[41] M. Dianati, X. Shen, and S. Naik, "A new fairness index for radio resource allocation in wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2005, pp. 712–717.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[43] F. Nogueira. (2014). *Bayesian Optimization: Open Source Constrained Global Optimization Tool for Python*. [Online]. Available: https://github.com/fmfn/BayesianOptimization

[44] S. Schwarz, C. Mehlführer, and M. Rupp, "Low complexity approximate maximum throughput scheduling for LTE," in *Proc. Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1563–1569.

[45] P. Virtanen, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 261–272, Feb. 2020.

**LUKAS ELLER** (Graduate Student Member, IEEE) received the Dipl. (Ing.) degree in telecommunications from Technische Universität Wien (TU Wien), in 2020. He is currently a Project Assistant with the Institute of Telecommunications, TU Wien, focusing on end-user performance estimates for 4G and 5G networks. His research interest includes assessing the deployment of data-driven deep learning methods for the generation of performance models based on crowdsourced measurements.

**PHILIPP SVOBODA** (Senior Member, IEEE) received the Dr. (Ing.) degree in electrical engineering from Technische Universität Wien (TU Wien). He is currently a Senior Scientist with TU Wien, with his research focusing on the performance aspects of mobile cellular technologies. He is also examining the feasibility of using crowdsourcing to conduct performance measurements on 4G and 5G mobile networks. His research interests include common framework for evaluating the performance of mobile networks, guaranteeing reliable, and fair connectivity for end-users.

**MARKUS RUPP** (Fellow, IEEE) received the Dipl. (Ing.) degree from the University of Saarbrcken, Germany, in 1988, and the Dr. (Ing.) degree from Technische Universität Darmstadt, Germany, in 1993. Until 1995, he held a post-doctoral research position with the University of California at Santa Barbara, Santa Barbara, CA, USA. From 1995 to 2001, he was with the Nokia Bell Laboratories, Wireless Technology Research Department, Holmdel, NJ, USA. Since 2001, he has been a Full Professor in digital signal processing in mobile communications with TU Wien.

• • •