Carrier Priority Control for Energy-Efficient Multi-Band Networks: A DRL-Based Approach

Anh-Khoa Dang[†‡], Hicham Khalifé[†], Stéphane Rovedakis[‡], Stefano Secci[‡]
Maxime Bouton[§], Jaeseong Jeong[§], Mathias Sintorn[†]
[†]Ericsson {anh.khoa.dang, hicham.khalife, mathias.sintorn}@ericsson.com
[§]Ericsson Research {maxime.bouton, jaeseong.jeong}@ericsson.com
[‡]Cnam, Paris, France {anh-khoa.dang, stephane.rovedakis, stefano.secci}@cnam.fr

***Abstract*—In mobile networks, base stations (BSs) typically consist of three sectors, each operating multiple carriers with distinct frequency bands to balance coverage and capacity demands. In this paper, we propose a joint carrier sleeping and traffic steering scheme tailored for carrier aggregation (CA)-enabled multi-band networks. First, we present a priority-based traffic steering strategy that uses carrier priorities not only to steer traffic toward the desired carriers but also to control carrier shutdown, thus optimizing the network energy performance. Then, we leverage Deep Reinforcement Learning (DRL) to dynamically adapt these priorities to spatio-temporal traffic variations. Results show that priority control enables more energy-efficient carrier utilization than binary sleep control, while maintaining user experience. Moreover, the proposed scheme achieves more than twice the energy savings compared to a legacy solution deployed in today's operational mobile networks.**

***Index Terms*—Multi-band networks, carrier shutdown, traffic steering, carrier aggregation, deep reinforcement learning.**

I. Introduction

Energy performance has emerged as a critical design pillar for future mobile networks. Although 5G systems achieve approximately 4 times higher energy efficiency than legacy 4G networks, their total network power demand may still increase by a factor of 12 due to wider spectrum utilization, massive antenna arrays, and ultra-dense base station layouts required to meet traffic capacity targets [1]. Consequently, energy management is of growing importance to network operators, both to reduce operational expenditures (OPEX) and to align with the industry-wide commitment toward carbon neutrality under the "Net Zero 2050" vision [2].

In practical mobile network deployments, a BS usually employs a three-sector configuration, where each sector operates multiple co-located carriers across distinct frequency bands (e.g., 0.8 and 2.6 GHz), each providing a unique trade-off between coverage and capacity (see Fig. 1). Indeed, while lower bands offer good propagation properties (i.e., long coverage distance), they provide lower bandwidth and throughput, in contrast to higher bands that offer higher capacities but smaller coverage zones [3]. In addition, such a multi-band setup enables flexible spectrum utilization through carrier aggregation (CA) to improve load distribution and enhance throughput performance [4]. However, while provisioning multiple carriers ensures robust network performance, operating all carriers during low-traffic periods (e.g., midnight hours) is often unnecessary and usually incurs considerable energy waste. Such inefficiency highlights the need to adapt carrier operation to temporal traffic dynamics [3]. For this reason, several energy management techniques [5] have been proposed and implemented in mobile networks, with carrier shutdown being one of the most prominent approaches. This technique consists of turning off carriers during low-load periods (e.g., off-peak hours) and reactivating them when traffic increases. Carrier shutdown can be timer-based, where capacity carriers are switched off during predefined periods such as nighttime, or rule-based, where a load threshold determines when a carrier is turned off [6].

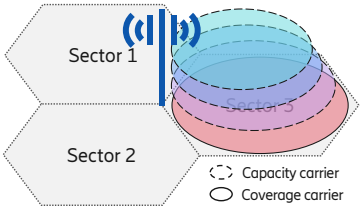


Fig. 1: Typical base station with a three-sector deployment, where each sector is served by multiple overlapping carriers: coverage carriers operate in sub-1 GHz frequency bands to provide wide-area coverage, while capacity carriers operate in frequency bands above 1 GHz to enhance network capacity.

More recently, machine learning (ML) based algorithms have been applied to adapt carrier operation in several ways, including traffic-aware planning [7], adaptive sleep thresholds [8], and direct carrier shutdown [9]. However, in carrier shutdown studies, traffic from deactivated carriers is usually handled implicitly by

This research work was supported by the French government, in the framework of France 2030 program (INTENTION-6G project).

ISBN 978-3-903176-82-9

the underlying network dynamics rather than explicitly controlled. As a result, the network has less flexibility to influence traffic redistribution and user experience, which may limit the achievable energy-saving gains.

In this paper, we leverage the *carrier priority* to empower direct carrier shutdown with the ability to influence traffic steering towards remaining active carriers in multi-band networks. The main contributions of this work are as follows:

- We present a system model for multi-band cellular networks with carrier aggregation, where users may aggregate multiple co-located carriers within the same sector. We introduce a *carrier priority* mechanism that steers user traffic toward aggregated carriers while jointly controlling carrier activation to improve energy efficiency.
- We implement a spatio-temporal traffic model that captures both spatial heterogeneity and temporal variations in user demand. The model employs a Weibull distribution to statistically capture daily traffic dynamics from recent live network datasets. This is essential for training carrier shutdown strategies, since their effectiveness depends on adaptation to time-varying traffic demand.
- We design a deep reinforcement learning-based energy-efficient traffic steering approach that adapts carrier priorities to traffic dynamics while preserving user experience and preventing cell saturation. Our solution has two variants, one using priority solely for binary carrier activation, while the other dynamically explores all priority levels for joint traffic steering and energy savings. Results show that priority control achieves more energy-efficient resource utilization than binary control, while both strictly outperform existing mobile network solution.

The rest of the paper is organized as follows. After discussing related work on Section II, we present the system model, priority-based traffic steering strategy, and spatio-temporal traffic model in Section III. Next, Section IV formulates the joint energy-saving and traffic-steering as a reinforcement learning problem and introduces its DRL-based solution framework. Section V presents performance evaluation leveraging the earlier presented traffic models. Finally, Section VI concludes the paper and outlines future directions.

II. Related Work

Carrier shutdown is one of the energy-saving techniques specified by 3GPP [5], offering significant potential for energy efficiency in the BS by exploiting temporal variations in daily traffic patterns to deactivate underutilized carriers during periods of low demand. The currently adopted carrier sleeping in production networks is the fixed threshold-based strategy [6], due to its simplicity and ease of implementation. In this approach, two sets of ON/OFF thresholds based on Resource Block (RB) utilization, which are manually configured by the network operator, are employed to trigger carrier (de)activation. To improve adaptability, ML techniques have been applied to tune these thresholds while maintaining user experience. Joan *et al.* [10] apply DRL to adjust the thresholds, whereas Maggi *et al.* [8] employ Bayesian optimization. Since traffic patterns are generally predictable, carrier shutdown can also be proactively planned based on traffic forecasts. Dang *et al.* [7] aggregate RB usage across overlapping cells to construct a univariate time series and apply probabilistic traffic prediction to proactively keep only the necessary carriers operational, without incurring traffic loss. In addition, DRL can be used to make shutdown decisions directly. Attai *et al.* [9] employ a modified actor-critic algorithm to control cell sleeping in ultra-dense networks (UDNs). Gan *et al.* [11] propose a DRL-based strategy for joint sleep control and energy sharing with renewable energy sources in UDNs. In all previously mentioned works, traffic handling after carrier deactivation is not explicitly controlled, and is instead left to implicit load-balancing mechanisms. To address this, our previous work [12] proposes carrier shutdown with traffic offloading to co-located active carriers. However, it does not consider CA and limits offloading to lower-frequency bands due to the lack of propagation modeling. In contrast, this work jointly addresses carrier sleep control and traffic handling in CA-enabled multi-band networks.

In operational networks, CA is adopted to leverage the benefits of multi-band deployments by allowing users to aggregate resources from co-located carriers within the same BS [4]. In this context, several studies have explored energy-efficient resource management in CA-enabled settings [13]–[15]. Yu *et al.* [13] formulate an optimization problem for energy-efficient carrier aggregation; however, their model considers a single base station and focuses on transmit power tuning for each carrier rather than direct carrier shutdown, which may limit energy savings. Fahime *et al.* [15] use a Double Deep Q-Network (DDQN) to decide carrier activation in CA-enabled networks from the user perspective, while Elsayed *et al.* [14] address the same problem using a constrained multi-agent Markov decision process. However, both works mainly focus on reducing energy consumption at the user side. In contrast, this work targets direct carrier shutdown to optimize energy savings at the BS side.

III. System Model

In this section, we present the system model for a multi-band network, reflecting the realistic setup currently adopted in commercial networks [3]. In general, a BS consists of 3 sectors, each covering a number

of collocated carriers operating in different frequency bands (cf. Fig. 1). More precisely, we first present a realistic yet accurate model that characterizes carrier load on a multi-band mobile network, which is needed to estimate power consumption per band. Our modelling also accounts for the CA, where user can consume radio resources of multiple carriers within the same sector[1]. We then introduce *carrier priority* concept, that assigns a priority level to each carrier of the sector. This priority will then govern the splitting of traffic to the user across aggregated carriers, while also controlling carrier shutdown. Lastly, we detail a spatio-temporal traffic model that captures time-varying user demand across the day. This statistical model, derived from real network data, is essential for training and evaluating carrier shutdown strategies, whose effectiveness depends on their ability to adapt to time-varying traffic demand.

A. Network Model

We consider a downlink multi-carrier cellular network composed of N BSs and U users deployed in a hexagonal layout. Each BS follows a 3-sector deployment and is further divided into K carriers within each sector. We index each sector globally by $i \in [3N]$. Let $C_{i,k}$ denote the k-th carrier ($k \in [K]$)[2] of sector i. Each carrier $C_{i,k}$ operates on a distinct frequency band f_k with bandwidth B_k. Throughout this paper, simplified notations depending only on k (e.g., f_k) are homogeneous across all sectors.

Let $p^{(t)}_{i,k,u}$ denote the received power at user $u \in [U]$ from carrier $C_{i,k}$ at time t, which includes propagation modeling according to the NR Urban Macro (UMa) model [16]. We can then calculate signal-to-interference-plus-noise ratio (SINR) at user u from $C_{i,k}$ with noise power level N_0 as follows:

$$SINR^{(t)}_{i,k,u} = \frac{p^{(t)}_{i,k,u}}{\sum\limits_{i' \in [3N], i' \neq i} p^{(t)}_{i',k,u} + N_0}. \tag{1}$$

The spectral efficiency of the link between user u and carrier $C_{i,k}$ is denoted by $\delta(\mathrm{SINR}^{(t)}_{i,k,u})$, where $\delta(\cdot)$ is generally determined by the Shannon capacity formula or, as in our case, by mapping link-level results from our proprietary simulator [17]. Accordingly, the achievable data rate with full bandwidth utilization is expressed as $r^{(t)}_{i,k,u} = \delta(\mathrm{SINR}^{(t)}_{i,k,u}) B_k$.

We denote $V^{(t)}_u$ the total traffic demand (in bits/s) of user u at time t. Since CA is considered, this demand is split across multiple carriers within the same sector. Let $x^{(t)}_{i,k,u} \in [0,1]$ be traffic split ratio of carrier $C_{i,k}$ toward user u at time t. Accordingly, this split satisfies $\sum_k x^{(t)}_{i,k,u} = 1, \forall u \in \mathcal{U}^{(t)}_i$, where $\mathcal{U}^{(t)}_i$ denotes the set of users served by sector i at time t. Then, the amount of traffic served by carrier $C_{i,k}$ to user u is $V^{(t)}_u x^{(t)}_{i,k,u}$. The load of carrier $C_{i,k}$, i.e., the RB utilization, is denoted by $\ell^{(t)}_{i,k} \in [0,1]$ and is computed as follows [18]:

$$\ell^{(t)}_{i,k} = \sum_{u \in \mathcal{U}^{(t)}_i} \frac{V^{(t)}_u x^{(t)}_{i,k,u}}{r^{(t)}_{i,k,u}}. \tag{2}$$

Based on the cell load, we model the perceived throughput of user u from $C_{i,k}$ using processor-sharing queueing theory [19], where $\ell^{(t)}_{i,k}$ represents the load factor associated with carrier $C_{i,k}$. Thus, the throughput perceived by user u from carrier $C_{i,k}$, denoted by $T^{(t)}_{i,k,u}$, is modeled as a function of the carrier load and the achievable rate, i.e., $T^{(t)}_{i,k,u} = T(\ell^{(t)}_{i,k}, r^{(t)}_{i,k,u})$. Readers are referred to reference [19] for the derivation and closed-form expression of $T(\cdot)$. When user u is simultaneously served by multiple carriers in the same sector due to carrier aggregation, the total perceived throughput of user u at sector i is $T^{(t)}_{i,u} = \sum_k T^{(t)}_{i,k,u}$.

Sector-wide User eXperience (UX): Later, in Sec. IV-A, we formulate an energy-saving objective that ensures a throughput reliability target for all users within a sector. To do so, we define the ratio of users in sector i whose throughput falls below a target τ (e.g., $\tau = 5$ Mbps) as:

$$Q^{(\tau,t)}_i = \frac{1}{|\mathcal{U}^{(t)}_i|} \sum\nolimits_{u \in \mathcal{U}^{(t)}_i} \mathbf{1}\left(T^{(t)}_{i,u} < \tau\right), \tag{3}$$

where $\mathbf{1}(\cdot)$ is the indicator function, returning 1 if the condition is true and 0 otherwise. $Q^{(\tau,t)}_i$ represents the low-throughput user ratio of the sector.

Power consumption: We define the variable $\sigma^{(t)}_{i,k} \in \{0,1\}$ to indicate the state of carrier $C_{i,k}$ whether it is active (1) or inactive (0) at time t. In realistic deployments, the lowest-frequency layer $C_{i,1}$ is always active to ensure coverage, while $C_{i,k}$, $\forall k > 1$, may be shut down to save energy. We express the power consumption of $C_{i,k}$ at time t as $P^{(t)}_{i,k} = P(\ell^{(t)}_{i,k}, \sigma^{(t)}_{i,k})$, where it is a function of carrier load and state. In our simulator, $P(\cdot)$ interpolates real power consumption data from vendor products, along with radio characteristics such as frequency, number of transceivers N^{TR}_k, transmit power P^T_k. Fig. 6c shows how $P(\cdot)$ varies with the carrier load $\ell^{(t)}_{i,k}$. For reproducibility, readers can refer to the open-source $P(\cdot)$ in [20].

B. Priority-Based Traffic Steering

In our framework, traffic steering refers to the mechanism that governs radio resource utilization across multiple carriers (i.e., CA) within the same sector. The proposed traffic steering framework draws inspiration from the *cell reselection priority* standardized for idle-mode mobility in 3GPP [21]. In this mechanism, each

[1]Aggregating multiple carriers from different base stations is classified as dual connectivity (DC), rather than carrier aggregation (CA) [4].

[2]For a positive integer K, $[K]$ denotes the set $\{1, 2, \ldots, K\}$.

frequency is assigned a priority value that guides user to camp on highest-priority frequencies when signal quality permits, thus enabling implicit load balancing across frequency layers. Inspired by this design, we present a similar concept for connected-mode operation, whereby a priority value is assigned to each carrier and used internally by the base station to steer user traffic across multiple bands. This strategy remains simple yet practical and can be easily implemented within current network operations. Fig. 2 illustrates this priority-based traffic steering procedure. In general, traffic is steered toward the carrier with the highest priority. If multiple carriers share the same priority level, CA is enabled.

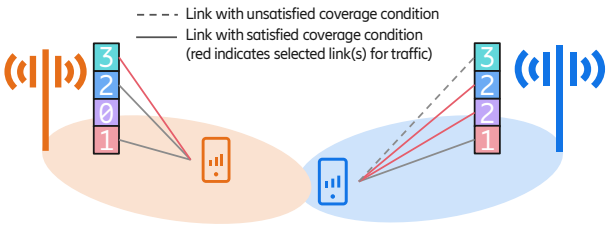


Fig. 2: **Priority-based traffic steering with two users:** Orange sector sends traffic to orange user on highest-priority (3) carrier. The blue sector aggregates traffic from two carriers of equal highest-priority (2) to serve blue user when priority-3 carrier's coverage condition is unsatisfied. The priority-0 carrier in orange sector is shut down, with no signaling present.

More formally, we denote by $\rho_{i,k}^{(t)} \in \{0, 1, \ldots, \rho_m\}$ the integer priority of carrier $C_{i,k}$ at time t, where ρ_m is the maximum priority level. We formulate that the priority not only guides traffic steering but also enables carrier shutdown control: $\rho_{i,k}^{(t)} = 0$ deactivates the carrier (i.e., $\sigma_{i,k}^{(t)} = 0$), while $\rho_{i,k}^{(t)} > 0$ keeps it active (i.e., $\sigma_{i,k}^{(t)} = 1$) and steers traffic accordingly. Recall that the lowest frequency carrier (i.e., $k = 1$) remains always ON ($\sigma_{i,1}^{(t)} = 1$).

First, the BS determine the serving sector i^* for each user u at time t by associating the user with the sector that provides the strongest received power on the always-on carrier ($k = 1$), as $i^* = \arg\max_i p_{i,1,u}^{(t)}$. Then, within sector i^*, the BS filter out carriers that do not satisfy the coverage condition $\frac{p_{i^*,k,u}^{(t)}}{N_0} > p_{\text{th}}$, where $\frac{p_{i^*,k,u}^{(t)}}{N_0}$ is effectively the signal-to-noise ratio (SNR), and p_{th} is the coverage threshold.

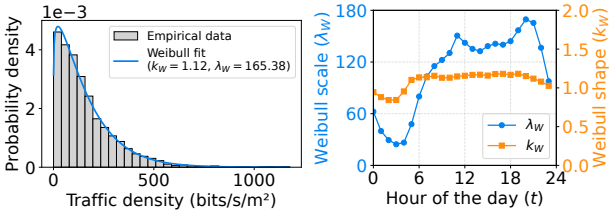
Finally, we denote by $\mathcal{K}_u^{(t)}$ the set of coverage-satisfied carriers with the highest priority in sector i^* for serving user u at time t. We model the traffic fraction assigned to user u on each selected carrier $C_{i^*,k}$, for all $k \in \mathcal{K}_u^{(t)}$, as [14]:

$$x_{i^*,k,u}^{(t)} = \frac{r_{i^*,k,u}^{(t)}}{\sum_{k' \in \mathcal{K}_u^{(t)}} r_{i^*,k',u}^{(t)}}. \tag{4}$$

This allocation distributes traffic proportionally to the carrier capacities. When $\mathcal{K}_u^{(t)}$ contains a single carrier k, all traffic is assigned to it ($x_{i^*,k,u}^{(t)} = 1$); otherwise, if $\mathcal{K}_u^{(t)} = \emptyset$, the traffic defaults to the always-on carrier ($x_{i^*,1,u}^{(t)} = 1$).

C. Spatio-Temporal Traffic Model

In preparation of the evaluation in the next part, we present the spatio-temporal traffic model, inspired by [22], to capture both the spatial variation of traffic volume across the deployment area and its temporal evolution throughout the day.



(a) Weibull PDF fitting at 21:00 (b) $(\lambda_W^{(t)}, k_W^{(t)})$ over 24 hours

Fig. 3: (a) Fitted Weibull probability density at 9 p.m. (b) Temporal evolution throughout the day of Weibull parameters

It has been shown that spatial traffic density is typically highly skewed and can be well approximated by a log-normal or Weibull distribution [22]. For this reason, we first collect hourly traffic measurements from 136 sectors of an anonymous operator between February 3 and March 2, 2025. The data is then processed and subsequently fitted with a Weibull distribution to statistically capture its spatial heterogeneity. The Weibull fitting results on the collected dataset are depicted in Fig. 3, where the estimated parameters $k_{\text{W}}^{(t)}$ (shape) and $\lambda_{\text{W}}^{(t)}$ (scale) characterize the distribution of traffic density for each hour of the day $t \in \{0, 1, \ldots, 23\}$.

Given the fitted Weibull distribution, one can now generate a *spatial* traffic density distribution. The process begins by generating a Gaussian random field $R_G(m, n)$ over the 2D grid $(m, n) \in [M_{\text{g}}] \times [N_{\text{g}}]$, using a sum of cosine functions:

$$R_G(m, n) = \frac{2}{\sqrt{L}} \sum_{l=1}^{L} \cos(i_l m + \phi_l) \cos(j_l n + \psi_l), \tag{5}$$

where L is the number of sinusoids (set to 10), i_l and j_l are uniformly distributed in $[0, \omega_{\max}]$ with $\omega_{\max} = 0.011592$, and ϕ_l and ψ_l are uniformly distributed in $[0, 2\pi]$, following the configuration in [22]. The generated field is normalized to $[0, 1]$ to obtain $R'_G(m, n)$, subsequently transformed via the Weibull percent point function [23]:

$$V^{(t)}(m, n) = F^{-1}\big(R'_G(m, n);\, k_W^{(t)}, \lambda_W^{(t)}\big), \tag{6}$$

where F^{-1} denotes the inverse cumulative distribution function (CDF) of the Weibull distribution, yielding the traffic density $V^{(t)}$ at each grid point (m, n) and time t.

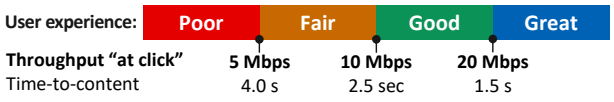


Fig. 4: Correlation between TTC and DL throughput [24]

The resulting continuous distribution $V^{(t)}$ is quantized into ten discrete levels, creating hotzones $h \in [10]$ with relative volumes $V_h^{(t)}$. A total of U users are uniformly deployed over the map, each located in a hotzone $h(u)$ (cf. Fig. 5). The per-user traffic demand $V_u^{(t)}$ introduced earlier can then be computed as $V_u^{(t)} = V_{h(u)}^{(t)}/|\mathcal{U}_{h(u)}|$, where $\mathcal{U}_{h(u)}$ denotes the set of users within hotzone $h(u)$, such that $\sum_{u\in[U]} V_u^{(t)} = \sum_{h\in[10]} V_h^{(t)}$.

IV. Joint Energy Saving and Traffic Steering

A. Problem Formulation

The goal of joint energy saving and traffic steering is to reduce the energy consumption of BS while maintaining acceptable UX and preventing cell saturation. In this work, we use the **Time-To-Content (TTC)** [24] metric proposed by Ericsson as a practical proxy for user experience, reflecting the responsiveness perceived by end-users (Fig. 4). Our UX target is to minimize the occurrence of `Poor` UX by constraining the proportion of users with throughput below 5 Mbps to less than 5% (i.e., $Q_i^{(5,t)} < \xi$, with $\xi = 0.05$). We formulate the problem at the sector level and define the objective function for each sector i as follows:

$$\min_{\rho_{i,k}^{(t)}} \quad \sum_{t=0}^{T-1}\sum_{k=1}^{K} P_{i,k}^{(t)}, \quad \forall i, \tag{7a}$$

$$\text{s.t.} \quad \ell_{i,k}^{(t)} < 1, \quad \forall t, i, k, \tag{7b}$$

$$Q_i^{(5,t)} < \xi, \quad \forall t, i, \tag{7c}$$

The objective (7a) minimizes the total power consumption of sector i over the horizon $T = 24$ hours by controlling priority variables $\rho_{i,k}^{(t)}$, while (7b) prevents cell saturation and (7c) maintains the UX target. This formulation yields a non-convex mixed-integer non-linear programming (MINLP) problem, which is NP-hard due to the discrete priority variables and the non-linear coupling between power usage, cell load, and throughput through the traffic-split variable $x_{i,k,u}^{(t)}$ (Sec. III-B) and the traffic demand $V_u^{(t)}$ (Sec. III-C). Moreover, as $V_u^{(t)}$ evolves spatio-temporally and is unknown *a priori*, optimization methods would require additional forecasting or repeated re-optimization to track these variations. The MINLP mainly serves as a conceptual formulation of the problem, while direct optimization is beyond the scope of this work because the objective and constraints are simulator-driven, making tractable approximation difficult without further simplifying assumptions. In contrast, reinforcement learning (RL) can learn adaptive policies directly from interactions with the environment, enabling instantaneous adaptation to traffic dynamics while optimizing long-term energy efficiency, which makes it a suitable framework for this problem.

B. Deep Reinforcement Learning (DRL) formulation

DRL is a decision-making framework where an agent leverages deep neural networks to learn a policy that maximizes cumulative rewards through interaction with an unknown environment. It is modeled as a Markov Decision Process (MDP) $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R} \rangle$. At each time step t, the agent observes the current state $s^{(t)} \in \mathcal{S}$ and selects an action $a^{(t)} \in \mathcal{A}$. This action causes a transition to a new state $s^{(t+1)} \in \mathcal{S}$ according to the transition probability function $\mathcal{P} : \mathcal{S}\times\mathcal{A}\times\mathcal{S} \to [0,1]$, and yields a reward $r^{(t+1)}$ from the function $\mathcal{R} : \mathcal{S}\times\mathcal{A} \to \mathbb{R}$. In order for the base station to set the priority for each carrier, we consider each sector i as an independent agent that operates autonomously. The MDP is defined as follows:

1) State: At each time t, the agent observes $s^{(t)}$ with the following features from sector i: the total number of users connected to sector i, denoted by $|\mathcal{U}_i^{(t)}|$; the average throughput of each carrier $\overline{T}_{i,k}^{(t)} = \mathbb{E}_u[T_{i,k,u}^{(t)}], \forall k$; the load of each carrier $\ell_{i,k}^{(t)}, \forall k$; the previous action $a^{(t-1)}$; and encoding of the time step t to capture daily periodic patterns. The dimension of the state is then a vector $s^{(t)} \in \mathbb{R}^{3K+3}$.

2) Action: After observing the state $s^{(t)}$, the agent determines the discrete priority level $\rho_{i,k}^{(t)}$ for each carrier k in sector i. As introduced earlier in Sec. III-B, priority variable $\rho_{i,k}^{(t)}$ can control energy saving, where $\rho_{i,k}^{(t)} = 0$ deactivates a carrier, and $\rho_{i,k}^{(t)} > 0$ governs CA traffic steering by determining how user traffic is split among the active carriers. The number of possible actions is $|\mathcal{A}| = (\rho_m + 1)^K$.

3) Reward: The reward signal is aligned with the optimization objective (7) and consists of three components:

- **Power consumption:** Encourages energy efficiency by penalizing high power usage, defined as $r_P^{(t)} = -\tilde{P}_i^{(t)}$, where $\tilde{P}_i^{(t)} \in [-1, 1]$ is the min–max scaled version of the total power of sector i, with $P_i^{(t)} = \sum_{k=1}^{K} P_{i,k}^{(t)}$.
- **User eXperience (UX):** Penalizes `Poor` service quality when the proportion of `Poor`-throughput users exceeds the target ξ, expressed as $r_Q^{(t)} = -\mathbf{1}\left(Q_i^{(5,t)} \geq \xi\right)$.
- **Cell saturation:** Penalizes cell overload when any carrier in sector i operates at or above full capacity, defined as $r_\ell^{(t)} = -\mathbf{1}\left(\ell_{i,k}^{(t)} \geq 1, \forall k\right)$.

The overall reward is computed as a weighted sum of the individual components:

$$r^{(t)} = w_P\, r_P^{(t)} + w_Q\, r_Q^{(t)} + w_\ell\, r_\ell^{(t)}, \tag{8}$$

where w_P, w_Q, and w_ℓ control the trade-off among energy efficiency, user experience, and saturation avoidance, respectively. In this work, after operational and empirical assessments, we set $w_\ell = 1$, $w_Q = 2$, and $w_P = 1$.

C. Proximal Policy Optimization

In our MDP formulation, the state–transition dynamics $\mathcal{P}$ are unknown, requiring the agent to learn an adaptive policy $\pi_\theta(a^{(t)}|s^{(t)})$, parameterized by θ. This policy defines the probability of selecting an action $a^{(t)}$ given a state $s^{(t)}$ and is optimized to maximize the expected discounted return $\mathbb{E}_{\tau\sim\pi_\theta}\left[\sum_{t=0}^{T-1}\gamma^t r^{(t)}\right]$, where $\gamma \in [0,1]$ controls the trade-off between short- and long-term rewards.

We adopt the Proximal Policy Optimization (PPO) algorithm [25] to learn the policy π_θ, as it provides stable learning by preventing large policy updates. Importantly, PPO is particularly suitable for our multi-carrier control problem, in which the agent operates at the sector level and jointly controls multiple carriers. By natively supporting multi-discrete action spaces, PPO enables each action dimension to correspond to an individual carrier-level decision within the sector. This factorized action representation avoids enumerating all carrier combinations in a single discrete action space, thereby reducing learning complexity and improving scalability.

Formally, PPO maximizes the clipped surrogate objective that constrains policy updates:

$$\mathcal{L}(\theta) = \mathbb{E}_\theta\left[\min\left(\Upsilon_t(\theta)\hat{A}_t, \hat{A}_t + \varepsilon|\hat{A}_t|\right)\right], \quad (9)$$

where $\Upsilon_t(\theta) = \frac{\pi_\theta(a^{(t)}|s^{(t)})}{\pi_{\theta_{\text{old}}}(a^{(t)}|s^{(t)})}$ denotes the probability ratio between the updated and old policies, and the clipping operation constrains the policy update within ϵ to ensure stable learning. $\hat{A}_t$ is the advantage function used in Generalized Advantage Estimation (GAE) method [26] to reduce the variance of policy gradient estimates while maintaining low bias, expressed as:

$$\begin{aligned}\hat{A}_t &= \sum_{l=0}^{T-t-1}(\gamma\lambda)^l\,\delta_{t+l},\\ \delta_t &= r^{(t)} + \gamma V_\phi(s^{(t+1)}) - V_\phi(s^{(t)}).\end{aligned} \quad (10)$$

where $\lambda, \gamma \in [0,1]$ are the GAE parameter and the discount factor, respectively. V_ϕ is a value function (a neural network parameterized by ϕ) used to guide policy updates. In addition, an entropy bonus $c_e\,\mathbb{E}_{a\sim\pi_\theta}\left[-\log\pi_\theta(a \mid s^{(t)})\right]$ is added to $\mathcal{L}(\theta)$ to encourage exploration, where c_e denotes the entropy coefficient. Finally, the policy and value networks are updated with learning rate α by stochastic gradient ascent and descent:

$$\theta \leftarrow \theta + \alpha\nabla_\theta\mathcal{L}(\theta), \quad (11)$$

$$\phi \leftarrow \phi - \alpha\nabla_\phi\mathbb{E}\left[V_\phi(s^{(t)}) - \textstyle\sum_{k=t}^{T}\gamma^{k-t}r^{(k)}\right]^2. \quad (12)$$

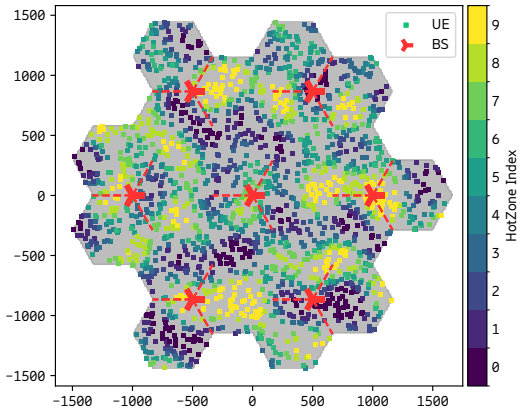


Fig. 5: Hexagonal network layout. The user (UE) is color-coded by HotZone, each corresponding to a different level of traffic demand (Recall sec. III-C).

Complexity analysis: The overall computational complexity of PPO is mainly determined by the size of its neural networks, with complexity approximately $\mathcal{O}(\sum_{l=1}^{L} n_{l-1}n_l)$, where n_l is the number of neurons in the l-th layer, and L is the total number of layers. The number of neurons in the input layer of the policy and value networks are $3K+3$. The size of the output layer of the policy network is $(\rho_m+1)K$, owing to the multi-discrete action formulation, instead of $(\rho_m+1)^K$ as in a flat discrete action space. In contrast, the output layer of the value network has only one neuron for the state value.

V. Performance Evaluation

A. Experiments setup

Our experiments use a proprietary time-static system-level simulator [17] to model multi-carrier 3-sector BSs with a hexagonal layout shown in Fig. 5. The policy is trained jointly across all sectors, and then applied individually at each sector. Aggregating the diverse experiences from all sectors enables the shared model to better learn and generalize to varying traffic demands across the network. Table I summarizes the system configuration and algorithm (PPO) hyperparameters.

TABLE I: Configured system parameters and PPO hyperparameters

System parameter	Value	PPO hyperparameter in Ray RLlib [27]	Value
Number of BSs (B)	7	Training time steps	96 000
Number of users (U)	2 000	Learning rate (α)	2×10^{-4}
Max priority level (ρ_m)	2	Hidden layers	[256, 256]
Frequency band (f_k)	(0.8, 1.8, 2.6, 3.5) GHz	Discount factor (γ), λ	0.99, 0.95
Bandwidth (b_k)	(10, 20, 20, 100) MHz	Clip parameter (ϵ)	0.25
N° of transceivers (N_k^{TR})	(2, 4, 4, 32)	Train batch size	480
Transmit power (P_k^T)	(40, 80, 80, 200) W	Minibatch size	240
Modulation scheme	(64, 64, 64, 256) QAM	Iterations per batch	6
Inter-site distance	1 000 m	Entropy coefficient (c_e)	0.06
Propagation model	NR Urban Macro [16]	N° of {Learners, Workers}	{1, 20}

1) Baselines: To assess the performance of our proposed solutions we implement the following baseline strategies for comparison with our DRL approaches as follows.

Always-On: All carriers remain active. The 3.5 GHz carrier with the highest capacity (100 MHz, 256QAM) is assigned priority 2, while the rest 0.8–2.6 GHz carriers (10–20 MHz, 64QAM) share priority 1, enabling CA among them. This setup reflects best practice in today's mobile networks throughput-oriented configuration.

Rule-Based [6]: A production network solution that pairs a capacity carrier with a coverage carrier and manually configures their cell load thresholds (t_{on}, t_{off}) to turn on/off the capacity carrier. In this baseline, the capacity and coverage carriers are set to 2.6 GHz and 1.8 GHz, respectively, with $(t_{\text{off}}, t_{\text{on}}) = (0.1, 0.25)$.

2) Our strategy: We consider two variants of our DRL-based approach to optimize priority configuration, as follows.

DRL-OnOff: The agent makes binary priority decisions for the 1.8–3.5 GHz carriers, where each carrier is either inactive ($\rho_{i,k}^{(t)} = 0$) or follows the Always-On priority. In practice, this policy replaces the thresholds in Rule-Based strategy with DRL-based on/off decisions.

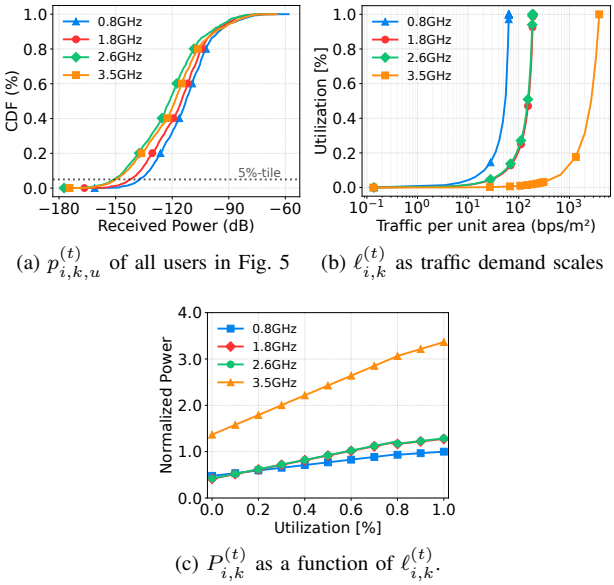
DRL-Priority: Each sector i acts as an agent that sets priority values $\rho_{i,k}^{(t)} \in \{0, 1, \ldots, \rho_m\}$ for each carrier to adapt traffic steering to spatio-temporal dynamics.

Note that, although simpler and aligned with current implemented priority settings, DRL-OnOff offers less flexibility than DRL-Priority. In fact, the latter strategy in addition to turning off carriers ($\rho_{i,k}^{(t)} = 0$) it offers the capability to decide traffic splitting to remaining active carriers.

B. Experimental results

1) Independent carrier-level evaluation: Before presenting the DRL-based energy-saving performance, we first conduct system-level simulations to characterize the behavior of each frequency carrier in the considered wireless environment. Each carrier is evaluated independently, without joint multi-carrier modeling. These results aim to provide a clear understanding on how our simulator realistically models each carrier et the parameters impacting its performance. Such analysis provide useful insights for analyzing the performance of the proposed DRL-based framework in the next sections of the paper.

Fig. 6a presents the received power distribution for the considered carriers. As expected, the 0.8 GHz carrier achieves the highest received power due to its favorable propagation characteristics and lower penetration loss. This is followed by the 1.8 and 2.6 GHz carriers, which exhibit reduced performance as frequency increases. The 3.5 GHz carrier achieves improved cell-center (median) received power than 2.6 GHz due to its 32 TRs massive MIMO array gains. At cell-edge (5%-tile), performance is comparable between the two carriers. Fig. 6b depicts the capacity of each carrier in terms of carrier load (utilization) as the traffic demand per area increases. The



(a) $p_{i,k,u}^{(t)}$ of all users in Fig. 5 (b) $\ell_{i,k}^{(t)}$ as traffic demand scales

(c) $P_{i,k}^{(t)}$ as a function of $\ell_{i,k}^{(t)}$.

Fig. 6: Characteristics of each frequency carrier within a sector. Simulations are performed independently for each carrier (single-carrier analysis), not jointly across carriers. Detailed carrier configurations are listed in Table I.

1.8 and 2.6 GHz carriers achieve higher capacity than 0.8 GHz, while 3.5 GHz provides roughly an order-of-magnitude greater capacity than 2.6 GHz. This is mainly attributed to differences in available bandwidth, modulation schemes. Lastly, Fig. 6c shows the power consumption of each carrier as the load increases. As the power model $P(\cdot)$ accounts for carrier-specific characteristics (Sec. III-A), the 3.5 GHz carrier exhibits significantly higher power consumption than lower-band 0.8–2.6 GHz carriers, primarily due to its higher transmit power (200 W) and larger number of transceivers (32 TRs).

2) DRL-based performance: We show in Fig. 7 the learning curves for both DRL strategies. One can easily see that both training trajectories reach convergence after a number of training steps. In fact, DRL-OnOff, with its smaller action space (on/off over 3 carriers), converges faster, whereas the DRL-Priority approach explores a wider range of control strategies and thus converges more gradually.

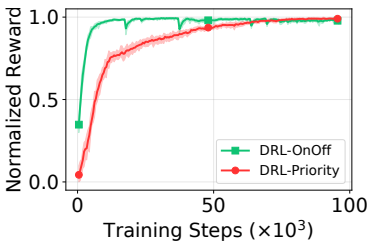


Fig. 7: Learning curves of the proposed strategy using PPO.

Fig. 8 further compares the performance of the pro-

posed DRL models in terms of energy saving (gains) and User eXperience guarantees to the classic Rule-Based and Always-On strategies. These results are obtained for a simulation run of 24 hours over the 21 sectors of Fig. 5. Looking at UX satisfaction, which reflects compliance with the throughput reliability objective, the DRL-OnOff achieves a performance close to the Always-On solution, while DRL-Priority attains slightly lower UX satisfaction. This is likely because DRL-Priority learns a more complex policy over a larger control space. Nevertheless, both DRL approaches obtain an acceptable UX satisfaction exceeding 99%. The strength of our strategies lies in their energy-saving gains, with DRL-Priority achieving slightly higher savings than DRL-OnOff. More interestingly, both of our proposed strategies outperform the legacy Rule-Based approach by more than doubling the energy saved during 24 hours. While the daily energy gain may seem moderate, it scales to substantial yearly savings across thousands of base stations nationwide [3].

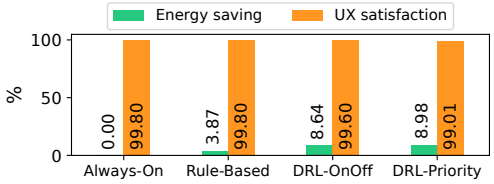


Fig. 8: Comparison of different energy-savings strategies.

3) Hourly and carrier-based analysis: We now compare the energy-saving behavior of the two DRL strategies by examining the power consumption trajectory over the day and contrast it with selected baselines in Fig. 9. The midnight period (0-6 a.m.) corresponds to low-traffic hours, during which energy-saving strategies are most effective. In this interval, DRL-Priority and DRL-OnOff exhibit similarly low power consumption, with DRL-Priority being opportunistically lower. After 6 a.m., when traffic demand increases, DRL-Priority consistently maintains lower power consumption than DRL-OnOff. This indicates the effect of priority adjustments on traffic steering, enabling more efficient carrier utilization.

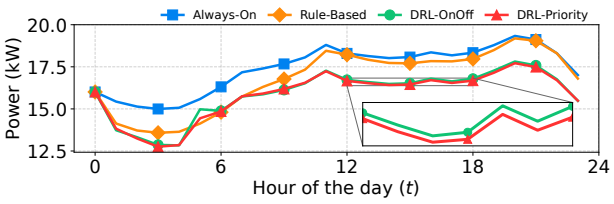


Fig. 9: Network power consumption throughout the day

For further investigation, we examine the active time of each carrier per sector in Fig. 10. During the midnight hours (0-6 am), all strategies (except the Always-On) shut down 2.6 GHz carriers (roughly 0% uptime). Furthermore, in this period, both DRL-Priority and DRL-OnOff generally shuts down the 3.5 GHz carrier (0% active time). After midnight, both DRL-based strategies keep the 1.8 and 3.5 GHz carriers active to satisfy UX constraints, while turning off the 2.6 GHz carrier. The Rule-Based approach, in contrast, conservatively adapts the 2.6 GHz activation according to carrier-load demand.

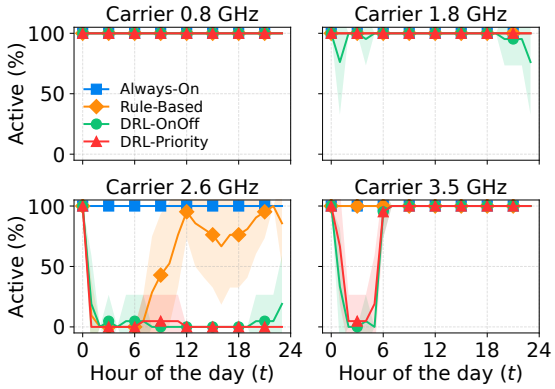


Fig. 10: The average percentage of active time for each carrier per sector (mean $\pm$ std).

While both DRL-based strategies exhibit similar behavior, DRL-Priority shows a tendency toward higher energy savings. Fig. 9 suggests that dynamic priority adjustment enables more efficient carrier utilization. Indeed, we observe in Fig. 11 that DRL-Priority steers more traffic to the 1.8 GHz carrier instead of the 0.8 GHz one. This is achieved by assigning the 0.8 GHz carrier the lowest priority. By steering less traffic toward the low-band 0.8 GHz carrier, energy consumption is effectively reduced. As shown in Fig. 6b, we illustrate the utilization (i.e., capacity usage) of each carrier as traffic demand scales. It can be observed that the 0.8 GHz carrier is the most sensitive to traffic load: even a smaller traffic demand results in higher utilization compared to higher-frequency bands. This behavior is expected, since low-band carriers typically have limited bandwidth despite their advantage of wider coverage. Consequently, this higher utilization translates into increased power consumption, as power scales with carrier load (see Fig. 6c). Therefore, steering traffic toward higher-frequency bands is more efficient in terms of carrier utilization and energy consumption.

Furthermore, all strategies assign the highest priority to the 3.5 GHz carrier, attracting about 90% of total traffic as shown in Fig. 11. We acknowledge that such behavior is energy-efficient, as the 3.5 GHz carrier's high capacity yields lower carrier load than the 0.8–2.6 GHz bands under the same demand. Fig. 6b shows that the 3.5 GHz carrier can handle roughly ten times the traffic demand of the 0.8–2.6 GHz bands while maintaining a low carrier utilization of approximately 0.2. This is

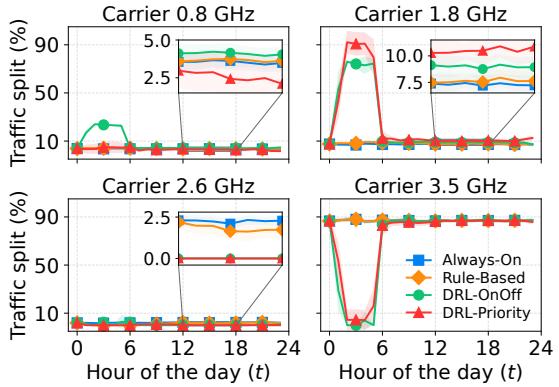


Fig. 11: The ratio of total traffic distributed among all carriers per sector (mean ± std).

attributed to the use of Massive MIMO antennas and the wide available bandwidth of 100 MHz (cf. Table I).

Our results underscore the practical potential of dynamic priority control, a mechanism that can be easily implemented in existing Radio Access Networks (RAN) architectures to jointly adapt carrier activation and traffic steering to varying traffic demands. As a consequence, energy efficiency can be enhanced without affecting user experience.

VI. Concluding Remarks

In this paper, we leverage the concept of *carrier priority* as a unified parameter to control both carrier activation and traffic splitting in CA-enabled multi-band mobile networks. To do so, we train a DRL agent to dynamically adapt these priorities across spatial and temporal traffic variations, optimizing energy savings while satisfying user experience constraints. Numerical results show that the proposed method can intelligently steer traffic for more efficient carrier utilization, achieving additional energy savings compared to baseline approaches. This approach is particularly promising, as the priority mechanism can be easily implemented in existing RAN systems. For future work, we plan to assess the generalization of our solution under diverse load scenarios and extend the framework to multi-agent cooperation among neighboring sectors.

References

[1] S. Han, S. Bian *et al.*, "Energy-efficient 5G for a greener future," *Nature Electronics*, vol. 3, no. 4, pp. 182–184, 2020.

[2] GSMA. Mobile Net Zero: State of the Industry on Climate Action 2023.

[3] Ericsson. Sustainable Networks: The RAN modernization handbook. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/further-insights/sustainable-networks-the-ran-modernization-guide

[4] C. Pupiales *et al.*, "Multi-Connectivity in Mobile Networks: Challenges and Benefits," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 116–122, 2021.

[5] 3GPP, "Study on network energy savings for NR," version 18.1.0.

[6] D. Lopez-Perez *et al.*, "Data-driven energy efficiency modeling in large-scale networks: An expert knowledge and ml-based approach," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 780–804, 2024.

[7] A.-K. Dang *et al.*, "Energy optimization for multi-band cellular networks: A traffic prediction-based strategy," in *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 2025, pp. 1–6.

[8] L. Maggi *et al.*, "Energy savings under performance constraints via carrier shutdown with bayesian learning," in *2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2023, pp. 1–6.

[9] A. I. Abubakar *et al.*, "A secured energy saving with federated assisted modified actor-critic framework for 6g networks," *IEEE Transactions on Vehicular Technology*, 2025.

[10] J. S. Pujol-Roigl *et al.*, "Deep Reinforcement Learning for cell on/off energy saving on Wireless Networks," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 01–07.

[11] J. Gan *et al.*, "Joint Sleep Control and Energy Sharing Strategy With Deep Reinforcement Learning in Green Ultra-Dense Networks," *IEEE Transactions on Green Communications and Networking*, 2025.

[12] A.-K. Dang *et al.*, "Data-driven energy optimization in mobile networks with user experience guarantees," in *IEEE INFOCOM 2025 - IEEE Conference on Computer Communications*, 2025, pp. 1–10.

[13] G. Yu *et al.*, "Joint downlink and uplink resource allocation for energy-efficient carrier aggregation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3207–3218, 2015.

[14] M. Elsayed *et al.*, "Energy-efficient carrier aggregation in 5g using constrained multi-agent mdp," *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 4, pp. 1595–1606, 2024.

[15] F. Khoramnejad *et al.*, "Delay-aware and energy-efficient carrier aggregation in 5g using double deep q-networks," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6615–6629, 2022.

[16] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 38.901, July 2020, release 16.

[17] H. Asplund *et al.*, "A set of propagation models for site-specific predictions," in *12th European Conference on Antennas and Propagation (EuCAP 2018)*. IET, 2018, pp. 1–5.

[18] I. Viering, M. Dottling, and A. Lobinger, "A mathematical perspective of self-optimizing wireless networks," in *2009 IEEE International Conference on Communications*. IEEE, 2009, pp. 1–6.

[19] N. Chen and S. Jordan, "Throughput in processor-sharing queues," *IEEE Transactions on Automatic Control*, vol. 52, no. 2, pp. 299–305, 2007.

[20] G. Vallero *et al.*, "A New Explainable Power Demand Model for 4G LTE and 5G NR Base Stations," in *ICC 2025-IEEE International Conference on Communications*. IEEE, 2025, pp. 4173–4178.

[21] 3GPP TS 38.304, "NR; User Equipment (UE) Procedures in Idle Mode and Inactive Mode," version 17.6.0, Release 17.

[22] D. Lee *et al.*, "Spatial modeling of the traffic density in cellular networks," *IEEE Wireless Communications*, vol. 21, no. 1, pp. 80–88, 2014.

[23] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[24] Ericsson. Time-to-content: Benchmarking network performance. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/articles/time-to-content

[25] J. Schulman *et al.*, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[26] J. Schulman *et al.*, "High-dimensional continuous control using generalized advantage estimation," *arXiv preprint arXiv:1506.02438*, 2015.

[27] E. Liang *et al.*, "RLlib: Abstractions for distributed reinforcement learning," in *International conference on machine learning*. PMLR, 2018, pp. 3053–3062.