Energy Efficient Load Balancing in Multiband Cellular Networks via Reinforcement Learning

Ahmed El Soukkary and Karim G. Seddik
Electronics and Communications Engineering Department, American University in Cairo, Egypt
Email: {ahmedelsoukkary, kseddik}@aucegypt.edu

***Abstract*—The exponential growth of mobile data traffic has intensified the need for energy-efficient and fair resource allocation in cellular networks. This paper addresses this challenge through two key contributions: a novel user association (UA) algorithm and a reinforcement learning (RL)-based dynamic power allocation framework employing Proximal Policy Optimization (PPO). The proposed UA algorithm dynamically assigns users to frequency bands to optimize energy efficiency, minimize dropped users, and enhance fairness. The RL agent dynamically adjusts power levels across high-frequency bands to further improve energy efficiency while maintaining Quality of Service (QoS).**

The simulation results demonstrate that the RL-based power allocation provides over a 15% improvement in energy efficiency compared to fixed full power configurations. Moreover, the proposed UA performs better than the Max-SINR baseline in terms of energy efficiency, load balancing fairness, and dropped users metrics. These findings underscore the potential of combining intelligent UA algorithms with RL-based power control to address the demands of next-generation cellular networks.

***Index Terms*—Reinforcement Learning, Cellular Networks, Machine Learning, Load Balancing,**

I. Introduction

The exponential growth of wireless communication technologies has transformed modern society, enabling seamless connectivity and unprecedented access to information. With the advent of 5G and the anticipated deployment of 6G networks, the demand for robust, high-speed, and energy-efficient cellular networks continues to increase [1]. As mobile data traffic increases, network operators face increasing pressure to manage scarce resources effectively while ensuring optimal user experiences.

Efficient resource management in cellular networks is a critical challenge that encompasses optimizing network performance, energy consumption, and user Quality of Service (QoS). Traditionally, network operators have relied on fixed and heuristic approaches to manage parameters such as user association, transmit power, and base station activity. However, these methods struggle to adapt to the dynamic and complex nature of modern networks.

Recent advances in machine learning (ML) have introduced a paradigm shift in network optimization [2]. Reinforcement learning (RL), in particular, has shown promise in addressing dynamic optimization problems by enabling agents to learn optimal strategies through interaction with their environments. RL-based approaches have been applied to various aspects of cellular network optimization, such as power control, user association, and energy efficiency [3]. Despite this progress, significant challenges remain.

One of the underexplored areas in cellular networks is traffic steering in homogeneous networks where base stations operate across multiple frequency bands. Current research primarily focuses on single-frequency band scenarios, overlooking the unique opportunities and challenges presented by multi-band operation. Recent efforts, such as [4], have explored multi-objective RL for load balancing; however, their approach does not explicitly address energy efficiency, which remains a critical challenge.

Our main contributions can be summarized as follows:

- We propose a novel user association (UA) algorithm that dynamically assigns users to frequency bands based on SINR metrics while prioritizing energy efficiency, load-balancing fairness, and minimizing dropped users.
- We develop a reinforcement learning (RL)-based power allocation framework using Proximal Policy Optimization (PPO) to dynamically adjust power levels in high-frequency bands, significantly improving energy efficiency.

The rest of the paper is organized as follows: Section II reviews the relevant literature on user association and RL-based power allocation in cellular networks. Section III outlines the system model and formulates the resource allocation problem. Section IV describes the proposed UA algorithm and RL-based power allocation framework in detail. Section V presents the simulation setup and evaluates the performance of the proposed methods. Finally, Section VI concludes the paper and outlines directions for future work.

II. Literature Review

Various analytical approaches have been proposed to improve energy efficiency and load balancing in cellular networks. Early research efforts focused on developing traditional optimization techniques to improve energy efficiency; for instance, the authors in [5] discussed the integration of millimeter wave (mmWave) technologies in ultra-dense networks (UDNs), highlighting the benefits of joint user association and power allocation. By employing mixed-integer programming and Lagrangian dual decomposition, the authors achieved significant improvements in energy efficiency and spectral efficiency. Meanwhile, in [6] the authors proposed long-term rate-based association strategies for load balancing in heterogeneous cellular networks (HCNs). The introduction of power control further reduced energy consumption by mitigating network interference.

Similarly, [7] introduced an alternating optimization algorithm to enhance utility-energy efficiency in heterogeneous networks (HetNets). The algorithm leveraged Lagrangian dual

analysis and auxiliary variables to transform the original non-convex problem into a convex one, ensuring efficient solutions and convergence to a local optimum. In [8] a three-layer iterative algorithm was developed for ultra-dense heterogeneous networks (UDHNs). The algorithm combined base station (BS) on/off operations with user association to maximize long-term rates, demonstrating better performance than existing user association methods. However, such centralized approaches are often computationally intensive and require high signaling overhead, limiting their scalability in dense networks.

The advent of distributed learning methods has further advanced energy efficiency in cellular networks. The authors in [9] explored a distributed user association algorithm in HetNets, aiming to maximize network-wide energy efficiency. By turning off BSs with low user counts and offloading users to active BSs, the proposed method minimized energy consumption and maximized throughput, outperforming conventional load balancing strategies. A belief propagation-based message-passing approach for user association in HetNets was introduced in [10]. This method improved energy efficiency without sacrificing spectral efficiency, offering higher energy efficiency geometric mean compared to prior art.

The latest advancements in machine learning and deep learning have opened new avenues for enhancing energy efficiency. A joint cell activation and user association scheme using a Q-learning-based algorithm was introduced in [11] to address power consumption and backhaul load balancing in green dense heterogeneous networks. This approach showed substantial improvements in fairness, quality of service (QoS), and energy efficiency. The authors in [12] presented a multiagent Q-learning algorithm for user association and power allocation in UDHNs, focusing on load balancing and energy efficiency. Simulation results confirmed the convergence and effectiveness of the proposed scheme. The authors in [13] proposed a deep neural network (DNN)-based scheme for dynamic cell selection and power allocation in coordinated multi-point (CoMP) transmissions. The DNNs were trained to maximize spectral and energy efficiency, achieving similar performance to optimal algorithms with lower complexity. In [14], a decentralized user association technique based on multi-agent actor-critic (AC) networks was introduced for ultra-dense networks (UDNs). This approach utilized local observations and critic networks to inform energy-efficient decisions, resulting in a 50% average energy efficiency gain over conventional techniques. The authors in [15] proposed a multi-agent deep reinforcement learning (MA-DRL) scheme for channel assignment and power allocation in two-tier HetNets. The deep Q network (DQN) and deep deterministic policy gradient (DDPG) network worked together to optimize system capacity and reduce power consumption effectively.

Lastly, a recent work explored load balancing in homogeneous multi-band networks using a multi-objective reinforcement learning (MORL) framework [4]. Their approach integrates meta-reinforcement learning (meta-RL) to adapt to varying trade-offs between network KPIs. Specifically, they aim to maximize the minimum user throughput while minimizing its standard deviation, ensuring a more balanced distribution of network resources. Their control variables include handover (HO) and cell reselection (CRS) thresholds, which are adjusted dynamically to optimize load balancing. Additionally, they propose an extension to gradient-based meta-RL methods by incorporating policy distillation to enhance the meta-policy's performance. While this work demonstrates the potential of MORL for load balancing in multi-band networks, it primarily focuses on throughput fairness and does not explicitly address energy efficiency. In contrast, my research aims to enhance energy efficiency in multi-frequency band networks. By using reinforcement learning (RL), base stations dynamically switch higher frequency bands on or off to conserve energy and steer users to appropriate bands based on their volume and service requirements. Furthermore, my approach introduces a novel SINR-based user association metric that simplifies decision-making while maintaining load balancing and QoS. By combining RL-based power allocation with efficient user association, this work offers a complementary perspective to existing multi-band network optimization strategies, with a primary focus on energy-efficient load balancing

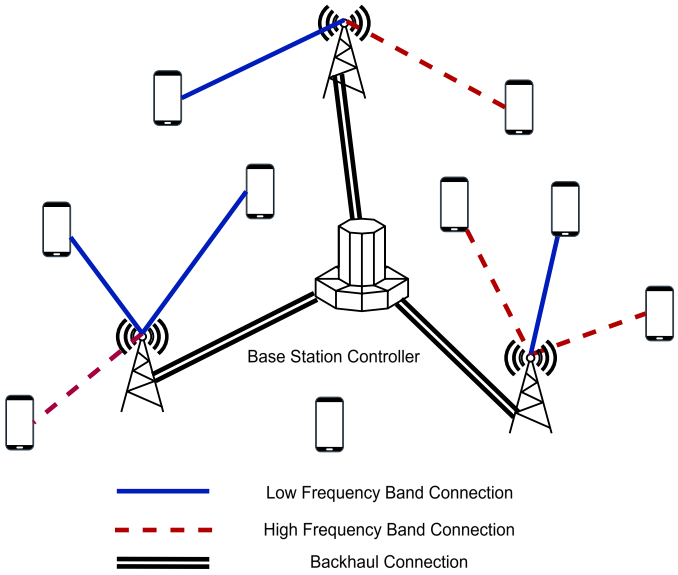

Fig. 1: System Model

III. System Model And Problem Formulation

Consider the downlink transmission of a homogeneous cellular network consisting of a set of base stations $\mathcal{N} = \{1, 2, ..., N\}$ and a set of mobile users $\mathcal{U} = \{1, 2, ..., U\}$. Each base station $n \in \mathcal{N}$ operates on two distinct frequency bands: a low-frequency band l_n, characterized by robust coverage and high penetration, and a high-frequency band h_n, which offers higher capacity but limited coverage. Moreover, let the sets $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}$ be defined as $\mathcal{B}_1 = \{l_n, n \in \mathcal{N}\}$, $\mathcal{B}_2 = \{h_n, n \in \mathcal{N}\}$, and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$. Moreover, all base stations are assumed to be connected to a centralized controller, which facilitates coordination and decision-making for resource allocation and user association. Figure 1 provides an illustrative example of the system model.

Each user $u \in \mathcal{U}$ is associated with a single base station $n \in \mathcal{N}$ and is assigned resources on either l_n or h_n based

on the user association policy. Let x_{bu} be a binary indicator variable such that $x_{bu} = 1$ if user $u \in \mathcal{U}$ is associated with band $b \in \mathcal{B}$; and 0 otherwise. Note that by the definitions of $\mathcal{B}$ and $\mathcal{N}$, any band $b \in \mathcal{B}$ uniquely determines a base station $n \in \mathcal{N}$.

The signal to interference and noise ratio (SINR) at user u when it is served by band $b \in \mathcal{B}_i$ is given by

$$\gamma_{bu} = \frac{P_b H_{bu}}{\sum_{b' \in \mathcal{B}_i \setminus b} P_{b'} H_{b'u} + \sigma^2}, \quad i \in \{1, 2\} \tag{1}$$

where P_b and H_{bu} represent the transmit power of band b and the channel gain between band b and user u, respectively, and σ^2 is the noise power.

Let the load of each band b be defined as $y_b = \sum_{u \in \mathcal{U}} x_{bu}$, thus the transmit power P_b of band b is set to zero whenever $y_b = 0$, ensuring that unused bands do not consume energy. We assume that the bandwidth of each band $b \in \mathcal{B}$, denoted W_b, is shared equally among all users connected to it. Consequently, the data rate of user u when connected to band b is given by $R_{bu} = \frac{W_b}{y_b} \log_2(1 + \gamma_{bu})$, and the overall system data rate is given by $R_{\text{total}} = \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} x_{bu} R_{bu}$. In this system, the transmit powers for the low-frequency bands $b \in \mathcal{B}_1$ are fixed to ensure robust coverage. In contrast, the transmit powers for the high-frequency bands $b \in \mathcal{B}_2$ can take values in the discrete set $\mathcal{P}_b$, enabling adaptive power control.

The total power consumption of the system is given by $P_{\text{total}} = \sum_{b \in \mathcal{B}} P_b$. The energy efficiency (EE) of the system, defined as the ratio of the total data rate to the total power consumption, is expressed as $\text{EE} = \frac{R_{\text{total}}}{P_{\text{total}}}$. Our objective is to maximize the system's energy efficiency by jointly optimizing the user association variables $\{x_{bu}\}$ and the transmit powers of the high-frequency bands $\{P_b : b \in \mathcal{B}_2\}$. The optimization problem can be formulated as:

$$\max_{\{x_{bu}\}, \{P_b : b \in \mathcal{B}_2\}} \quad \text{EE} = \frac{\sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} x_{bu} R_{bu}}{\sum_{b \in \mathcal{B}} P_b}, \tag{2a}$$

$$\text{subject to:} \quad \sum_{b \in \mathcal{B}} x_{bu} \leq 1, \quad \forall u \in \mathcal{U}, \tag{2b}$$

$$P_b \in \mathcal{P}_b, \quad \forall b \in \mathcal{B}_2, \tag{2c}$$

$$\sum_{b \in \mathcal{B}} x_{bu} R_{bu} > R^u_{min}, \quad \forall u \in \mathcal{U}, \tag{2d}$$

$$x_{bu} \in \{0, 1\}, \quad \forall b \in \mathcal{B}, \forall u \in \mathcal{U} \tag{2e}$$

where (2b) ensures that each user u is associated with at most one band b across all base stations, (2c) restricts the power levels of the high-frequency bands to a discrete set of values, including the option to turn off the band ($P_b = 0$), and (2d) guarantees that each user u receives a minimum required data rate $R^u_{\min}$, ensuring quality of service. The formulated problem is a Mixed-Integer Nonlinear Programming (MINLP) problem, which is known to be NP-hard and computationally challenging to solve directly for practical network sizes. To address this complexity, we decompose the problem into two subproblems: user association and power allocation. This decomposition allows for a tractable and efficient solution to the overall problem, as detailed in the following section.

IV. Proposed Solution

To address the complexity of the formulated MINLP problem, we decompose it into two subproblems: user association (UA) and power allocation (PA). In this section, we detail the two user association algorithms considered in this work: the newly proposed algorithm and a baseline algorithm based on maximum SINR. The power allocation subproblem will be addressed in subsequent sections using reinforcement learning.

A. Proposed User Association Algorithm

The proposed user association algorithm is designed to optimize system performance by prioritizing users based on their SINR ratios and ensuring that constraints are respected during the assignment process. The key steps of the algorithm are as follows:

- Compute the metric for each user as the ratio of the two highest SINRs they experience.
- Order users in descending order of their metrics, prioritizing those with a higher disparity in SINR. This prioritization ensures that users with the highest ratio, indicating a significant performance difference between their best and second-best SINR, are assigned to their optimal base station first. By doing so, the algorithm minimizes the likelihood of severe performance degradation for these users if they are forced to connect to a suboptimal band.
- Assign each user to the band providing the highest SINR, provided it does not violate any constraints. If a violation occurs, the band is closed to new assignments.
- Repeat the process for remaining users and open bands until all users are assigned or all bands are closed.

The pseudocode for the proposed algorithm is given in Algorithm 1.

B. Baseline User Association Algorithm (Max SINR)

The baseline algorithm, referred to as Max SINR, serves as a comparative approach to the proposed method. This algorithm assigns each user to the band providing the highest SINR, resolving violations iteratively. The main steps are:

- Assign each user to the band with the highest SINR.
- Remove users from bands where rate constraints are violated and reassign them to the next highest SINR band, excluding previously banned bands.
- Repeat until all users are either assigned or banned from all bands.

Algorithm 2 represents the pseudocode for the baseline algorithm.

C. Reinforcement Learning for Power Allocation

To address the power allocation subproblem, we employ Proximal Policy Optimization (PPO), a state-of-the-art reinforcement learning (RL) algorithm well-suited for continuous or discrete action spaces [16]. PPO is known for its sample efficiency and stability, achieved by restricting policy updates to a small region via a clipped objective function. The RL framework for power allocation in this work is designed with the following components:

Algorithm 1: Proposed User Association Algorithm

Input: SINR values γ_{bu} for all $u \in \mathcal{U}$ and $b \in \mathcal{B}$.
Output: User association $\{x_{bu}\}$.

Initialize $x_{bu} \leftarrow 0$ for all $b \in \mathcal{B}$ and $u \in \mathcal{U}$.
Set $\mathcal{U}_{\text{unassigned}} \leftarrow \mathcal{U}$ and $\mathcal{B}_{\text{open}} \leftarrow \mathcal{B}$.
while $\mathcal{U}_{unassigned} \neq \emptyset$ *and* $\mathcal{B}_{open} \neq \emptyset$ **do**
  foreach $u \in \mathcal{U}_{unassigned}$ **do**
    Compute the two highest SINRs in $\mathcal{B}_{\text{open}}$: $\gamma_{b_1 u}$ and $\gamma_{b_2 u}$.
    Compute the metric $M_u = \frac{\gamma_{b_1 u}}{\gamma_{b_2 u}}$.
  end
  Order users in $\mathcal{U}_{\text{unassigned}}$ in descending order of M_u.
  foreach *user* u *in the ordered list* **do**
    if *assigning* u *to* b_1 *does not violate the rate constraints of users connected to* b_1 **then**
      Assign u to b_1: $x_{b_1 u} \leftarrow 1$.
      Remove u from $\mathcal{U}_{\text{unassigned}}$.
    else
      Remove b_1 from $\mathcal{B}_{\text{open}}$ (close the band).
    end
  end
end
return $\{x_{bu}\}$.

Algorithm 2: Baseline User Association Algorithm: Max SINR

Input: SINR values γ_{bu} for all $u \in \mathcal{U}$ and $b \in \mathcal{B}$
Output: User association $\{x_{bu}\}$

Initialize $x_{bu} \leftarrow 0$ for all $b \in \mathcal{B}$ and $u \in \mathcal{U}$ Initialize $\mathcal{U}_{\text{unassigned}} \leftarrow \mathcal{U}$, $\mathcal{B}_u \leftarrow \mathcal{B}$ for all $u \in \mathcal{U}$
while $\mathcal{U}_{unassigned} \neq \emptyset$ *and* $\mathcal{B}_u \neq \emptyset$ *for all* $u \in \mathcal{U}_{unassigned}$ **do**
  foreach $u \in \mathcal{U}_{unassigned}$ **do**
    Assign u to band $b_u^\star = \arg\max_{b \in \mathcal{B}_u} \gamma_{bu}$: $x_{b_u^\star u} \leftarrow 1$
  end
  Identify users violating rate constraints: $\mathcal{U}_{\text{violating}}$
  foreach $u \in \mathcal{U}_{violating}$ **do**
    $x_{b_u^\star u} \leftarrow 0$
    Remove $b_u^\star$ from $\mathcal{B}_u$
  end
  $\mathcal{U}_{\text{unassigned}} \leftarrow \mathcal{U}_{\text{violating}}$
end
return $\{x_{bu}\}$

a) States: For each frequency band $b \in \mathcal{B}$, define the percentage of "good users" as: $g_b = \frac{\sum_{u \in \mathcal{U}} \mathbb{1}(\text{RSRP}_{bu} > \text{RSRP}_{\text{threshold}})}{U}$, i.e., those users whose Reference Signal Received Power (RSRP) is more than a predetermined threshold. Using this definition, the state vector $\mathbf{s}$ at each decision step is expressed as:

$$\mathbf{s} = [U, g_{b_1}, g_{b_2}, \ldots, g_{b_{2N}}], \tag{3}$$

b) Actions: The actions are defined as the power allocation across all high-frequency bands in the network. Specifically, the state vector $\mathbf{a}$ at each decision step can be expressed as:

$$\mathbf{a} = [P_b | P_b \in \mathcal{P}_b, b \in \mathcal{B}_2], \tag{4}$$

At each step, the agent selects one power level for each high-frequency band from the set of predefined power levels. This decision impacts the signal strength and coverage area of the corresponding bands, influencing the overall network performance in terms of energy efficiency, quality of service, and fairness in load balancing.

c) Reward: The reward function is designed to balance energy efficiency, user satisfaction, and fairness in load distribution among the bands. At each step, the reward is computed as:

$$r = \hat{\text{EE}} - D + \alpha J, \tag{5}$$

where $\hat{\text{EE}}$ is the normalized energy efficiency defined as

$$\hat{\text{EE}} = \frac{\text{EE} - \text{EE}_{min}}{\text{EE}_{max} - \text{EE}_{min}}, \tag{6}$$

where $\text{EE}_{min}, \text{EE}_{max}$ are the minimum and maximum energy efficiency for a given fixed sum rate R, respectively; i.e., $\text{EE}_{min} = \frac{R}{\sum_{b \in \mathcal{B}_1} P_b + \sum_{b \in \mathcal{B}_2} P_b^{max}}$, and $\text{EE}_{max} = \frac{R}{\sum_{b \in \mathcal{B}_1} P_b}$. The variable D represents the percentage of users who cannot be connected to any band without violating the rate constraints of already connected users., i.e.,

$$D = \frac{\sum_{u \in \mathcal{U}} \mathbb{1}(\sum_{b \in \mathcal{B}} x_{bu} = 0)}{U}, \tag{7}$$

and J is Jain's Fairness Index [17], which is a metric quantifying the fairness of resource allocation across bands, defined as:

$$J = \frac{\left(\sum_{b \in \mathcal{B}} y_b\right)^2}{|\mathcal{B}| \cdot \sum_{b \in \mathcal{B}} y_b^2} \tag{8}$$

where y_b is the load of band b. Finally, α is a tunable scalar that governs the emphasis placed on load balancing in the reward function.

V. Simulation Results and Analysis

The simulation environment was implemented using the QuaDRiGa (QUAsi Deterministic RadIo channel GenerAtor) tool in MATLAB. QuaDRiGa is a comprehensive channel modeling framework widely employed to simulate realistic radio propagation environments [18]. The simulated network consists of three base stations deployed in a $1\,\text{km} \times 1\,\text{km}$ area, forming an equilateral triangle centered around the origin with an inter-site distance (ISD) of 500 m. The mobile users are randomly distributed to emulate realistic traffic conditions. Each base station operates on two distinct frequency bands: a low-frequency band (800 MHz) and a high-frequency band (2.6 GHz). All base stations and mobile users are equipped with omni-directional antennas. Both low frequency bands and high frequency bands are modeled using the `3GPP_3D_UMa_LOS` scenario.

Additional simulation parameters are as follows:

- Low-frequency bands operate with a bandwidth of 5 MHz, while high-frequency bands use a bandwidth of 10 MHz.
- The Reference Signal Received Power threshold value, $\text{RSRP}_{\text{threshold}}$, is set to 10^{-8} mW (-80 dBm).
- $R^u_{min} = 0.1$ Mbps for all $u \in \mathcal{U}$.
- The power levels for high-frequency bands ($b \in \mathcal{B}_2$) are the same for all bands and consist of 10 discrete levels: $\mathcal{P}_b = \{0, 20, 22.5, \ldots, 37.5, 40\}$ W. These power levels increase in steps of 2.5 W, starting from 20 W and ending at 40 W, with an additional zero-power (OFF) state.
- Fairness Weight (α): Set to 2.1 to strike a balance between load balancing and the other performance metrics. This value was selected through parameter tuning to avoid overemphasis on any single objective.
- The training parameters for the PPO-based reinforcement learning algorithm are detailed in Table I.

Parameter	**Value**	**Parameter**	**Value**
Number of Episodes	600	Clip Factor	0.2
Steps per Episode	100	Mini Batch Size	25
Experience Horizon	50	Advantage Estimate Method	GAE
Number of Epochs	3	GAE Factor	0.95
Entropy Loss Weight	0.1	Discount Factor	0.995

TABLE I: PPO Training Parameters

To evaluate the performance of the proposed User Association (UA) algorithm, it is compared with the Max-SINR algorithm under two power allocation strategies: reinforcement learning (RL)-based dynamic power allocation and fixed full power. The four configurations analyzed are:

1) **Proposed + RL:** The PPO agent is trained with the proposed UA algorithm.
2) **Max-SINR + RL:** The PPO agent is trained with the Max-SINR UA algorithm.
3) **Proposed + Full Power:** Fixed power configuration with the proposed UA.
4) **Max-SINR + Full Power:** Fixed power configuration with the Max-SINR algorithm.

For fixed power configurations, we have $P_b = 20\text{W}, b \in \mathcal{B}_1$ and $P_b = 40\text{W}, b \in \mathcal{B}_2$.

The four configurations are evaluated according to three performance criteria: energy efficiency, as shown in Figure 2, percentage of dropped users, as shown in Figure 3, and the Jain's Fairness Index, as shown in Figure 4. All simulation results are averaged over 100 independent trials.

A. Comparison Between Proposed and Max-SINR UA Algorithms

The results demonstrate the superior performance of the proposed UA algorithm compared to Max-SINR in terms of energy efficiency (EE), percentage of dropped users, and fairness index:

- **Energy Efficiency (EE):** The proposed UA achieves an average improvement of **4.3%** over Max-SINR in the fixed full power case and **5.3%** in the RL-based dynamic power allocation case.

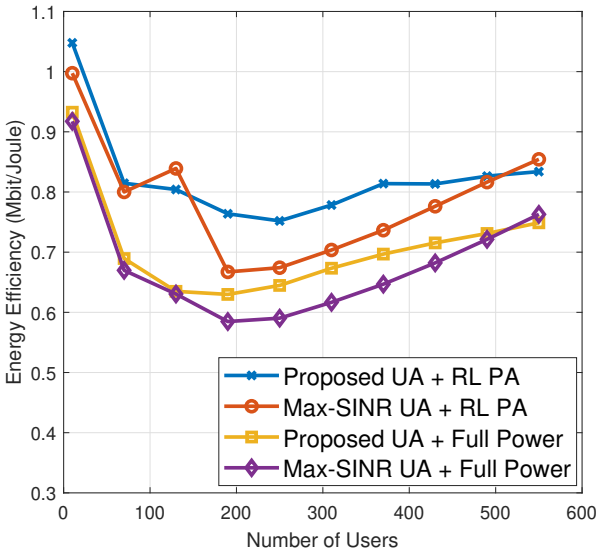

Fig. 2: Energy Efficiency Comparison

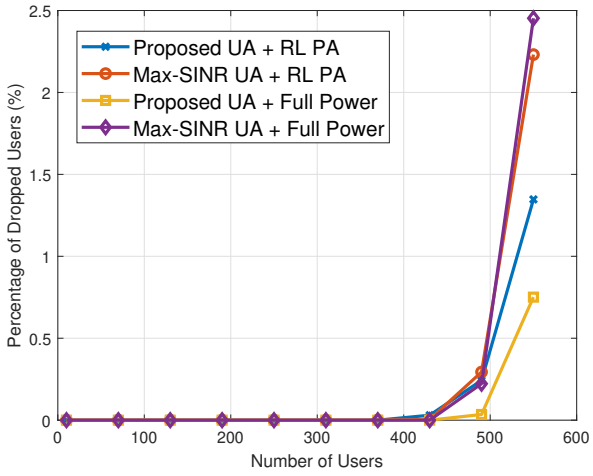

Fig. 3: Dropped Users Comparison

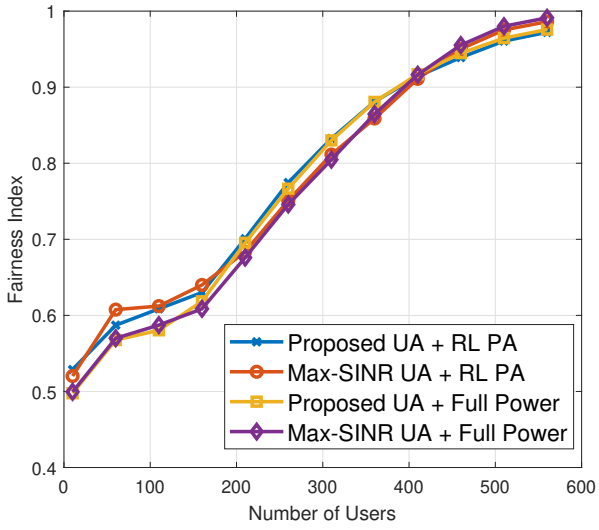

Fig. 4: Fairness Index Comparison

- **Dropped Users:** The proposed UA maintains a slightly better (lower) percentage of dropped users in both power allocation strategies.
- **Load Balancing Fairness:** The proposed UA exhibits a marginally better fairness index than Max-SINR in both cases for almost all user densities, indicating a more balanced load distribution between base stations.

B. Effect of RL-Based Power Allocation

The impact of RL-based dynamic power allocation is evident in the significant improvements in energy efficiency and its effect on other performance metrics:

- **Energy Efficiency (EE):** RL provides more than a **15% improvement** on average in energy efficiency compared to the corresponding fixed full power configurations, irrespective of the UA algorithm used (*Proposed vs. Proposed* and *Max-SINR vs. Max-SINR*).
- **Dropped Users:** RL has a negligible impact on the percentage of dropped users. For Max-SINR, the dropped users metric remains nearly identical between RL and full power configurations. For the proposed UA, there is a slight (negligible) increase in the dropped users percentage when RL is applied.
- **Load Balancing Fairness:** RL does not significantly affect the fairness index in either UA algorithm, indicating that load balancing fairness is primarily determined by the user association strategy rather than the power allocation approach.

These results validate the effectiveness of the proposed UA algorithm, both independently and when combined with RL, in achieving a balanced trade-off between energy efficiency, load-balancing fairness, and connectivity reliability.

VI. Conclusions and Future Work

This paper presents a novel approach to optimizing resource allocation in multi-band cellular networks by combining a novel user association (UA) algorithm with reinforcement learning (RL)-based power allocation using Proximal Policy Optimization (PPO). The proposed UA algorithm improves energy efficiency, minimizes dropped users, and enhances load-balancing fairness by dynamically prioritizing user assignments while respecting network constraints. Additionally, the PPO-based RL agent adjusts power levels across high-frequency bands to further maximize energy efficiency without compromising Quality of Service (QoS). Simulation results show that the proposed UA consistently outperforms the Max-SINR baseline across energy efficiency, load balancing fairness, and dropped user metrics. Furthermore, the RL-based dynamic power allocation shows significant improvements in energy efficiency, exceeding 15% compared to fixed full power configurations, demonstrating the value of learning-based power control strategies.

Future work will explore enhancements to the proposed framework by incorporating user mobility, heterogeneous rate demands, and a decentralized approach to decision making. These extensions aim to improve the scalability and applicability of the proposed methods in real-world scenarios while addressing challenges such as signaling overhead and network latency.

References

[1] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5g networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, 2016.

[2] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine learning for resource management in cellular and iot networks: Potentials, current solutions, and open challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1251–1275, 2020.

[3] A. Alwarafy, M. Abdallah, B. S. Çiftler, A. Al-Fuqaha, and M. Hamdi, "The frontiers of deep reinforcement learning for resource management in future wireless hetnets: Techniques, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 322–365, 2022.

[4] A. Feriani, D. Wu, Y. T. Xu, J. Li, S. Jang, E. Hossain, X. Liu, and G. Dudek, "Multiobjective load balancing for multiband downlink cellular networks: A meta- reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2614–2629, 2022.

[5] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. M. Leung, and H. V. Poor, "Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1936–1947, 2017.

[6] T. Zhou, Z. Liu, J. Zhao, C. Li, and L. Yang, "Joint user association and power control for load balancing in downlink heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2582–2593, 2018.

[7] X. Huang, W. Xu, H. Shen, H. Zhang, and X. You, "Utility-energy efficiency oriented user association with power control in heterogeneous networks," *IEEE Wireless Communications Letters*, vol. 7, no. 4, pp. 526–529, 2018.

[8] T. Zhou, Y. Fu, D. Qin, X. Li, and C. Li, "Joint user association and bs operation for green communications in ultra-dense heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 2, pp. 2305–2319, 2024.

[9] S. H. Lee, M. Kim, H. Shin, and I. Lee, "Belief propagation for energy efficiency maximization in wireless heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 56–68, 2021.

[10] R. Chatzigeorgiou and A. Bletsas, "Distributed, inference-based, energy efficient user association with convergence guarantees," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 3228–3233.

[11] Y. L. Lee, W. L. Tan, S. B. Y. Lau, T. C. Chuah, A. A. El-Saleh, and D. Qin, "Joint cell activation and user association for backhaul load balancing in green hetnets," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1486–1490, 2020.

[12] D. Li, H. Zhang, K. Long, W. Huangfu, J. Dong, and A. Nallanathan, "User association and power allocation based on q-learning in ultra dense heterogeneous networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–5.

[13] D. Kim, H. Jung, and I.-H. Lee, "Deep learning-based spectral and energy efficiency optimization for comp in hetnets," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 1970–1975.

[14] J. Moon, S. Kim, H. Ju, and B. Shim, "Energy-efficient user association in mmwave/thz ultra-dense network via multi-agent deep reinforcement learning," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 2, pp. 692–706, 2023.

[15] X. Zhao, Y. Cao, H. Chen, Z. Huang, and D. Wang, "Multi-objective resource allocation based on deep reinforcement learning in hetnets," in *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, 2022, pp. 574–578.

[16] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[17] R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Digital Equipment Corporation, Tech. Rep. DEC-TR-301, September 1984.

[18] S. Jaeckel, L. Raschkowski, K. BÃ¶rner, L. Thiele, F. Burkhardt, and E. Eberlein, "Quadriga- quasi deterministic radio channel generator, user manual and documentation," Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2.8.1, 2023.