Mobility Conditioned RL for Handover and Energy Efficient Dual Carrier Allocation in 5G Urban Macro Networks

Chrysostomos-Athanasios Katsigiannis
Computer Engineering and Informatics Department
University of Patras
Patras, Greece
Email: up1072490@ac.upatras.gr

Konstantinos Tsachrelias
Computer Engineering and Informatics Department
University of Patras
Patras, Greece
Email: up1096511@upatras.gr

Vasileios Kokkinos
Computer Engineering and Informatics Department
University of Patras
Patras, Greece
Email: kokkinos@cti.gr

Apostolos Gkamas
Department of Chemistry
University of Ioannina
Ioannina, Greece
Email: gkamas@uoi.gr

Christos Bouras
Computer Engineering and Informatics Department
University of Patras
Patras, Greece
Email: bouras@upatras.gr

Philippos Pouyioutas
Computer Science Department
University of Nicosia
Nicosia, Cyprus
Email: pouyioutas.p@unic.ac.cy

Abstract- **Real time radio resource management in 5G and beyond networks has become increasingly challenging under heterogeneous user mobility and multi-carrier operation. High-mobility users increase channel variability and may trigger frequent handovers, while dual-carrier deployments combining sub-6 GHz and mmWave resources introduce asymmetric coverage, capacity and reliability conditions. This paper proposes a mobility-conditioned reinforcement learning framework for joint handover control and dual-carrier resource allocation in a 5G urban macro network. The problem is formulated as a constrained Markov decision process in which the controller observes radio-quality indicators, mobility-related features, queue states and cell-load information, and then selects association updates together with carrier-level allocation decisions. The proposed framework is evaluated in a paper-scale dual-carrier scenario with mixed mobility regimes and a 3GPP TR 38.901 inspired urban macro channel model, using conventional handover and scheduling baselines together with an ablated reinforcement learning variant. The evaluation focuses on the trade-off between network throughput, queue backlog, energy efficiency and mobility stability. The results indicate that PPO-based adaptive control can preserve high network capacity while reducing unnecessary handovers and ping-pong behavior compared with classical mobility-management baselines.**

Keywords- 5G/6G, Reinforcement Learning, Handover, Dual-carrier, Power Control, Scheduling.

I. INTRODUCTION

Fifth generation (5G) networks have enabled high-throughput and low-latency services, but their performance increasingly depends on real-time resource management decisions under rapidly varying conditions. Modern deployments serve heterogeneous traffic and device types, from low-mobility IoT terminals to high-mobility users in vehicular or transit scenarios, while simultaneously operating under more complex spectrum usage [1]. A representative example is dual-carrier operation, where a lower-frequency carrier (e.g., FR1) provides robust coverage and a higher-frequency carrier (e.g., FR2/mmWave) offers large bandwidth but suffers from stronger blockage sensitivity and faster link fluctuations. In such settings, a resource management policy must not only schedule users efficiently but also decide when to trigger handovers and how to distribute carrier resources in a way that remains stable, capacity-aware and energy-conscious [2][3].

Conventional threshold-based handover manually tuned carrier selection and proportional-fair scheduling remain simple and interpretable, but they may be suboptimal under mixed mobility. Aggressive use of mmWave resources can improve short-term capacity while increasing ping-pong handovers, signaling overhead and queue growth [4][5]. Reinforcement learning (RL) can address this sequential trade-off, provided that the action space remains feasible and the inference loop is lightweight enough for near-real-time RAN control [6][7].

Therefore, a useful RL design for this domain should explicitly address feasibility, avoid invalid allocations and expose the operational timescale used for control. It should also evaluate not only throughput, but also queue evolution and mobility stability, since aggressive association or carrier decisions may improve short-term capacity while producing excessive handovers or ping-pong behavior.

Recent studies have already applied reinforcement learning to handover management, radio resource allocation, and energy-aware RAN control, therefore the use of Proximal Policy Optimization (PPO) is not presented here as a new learning method. The contribution of this work is instead the joint control formulation, where sector association and dual-carrier allocation are considered together with queue state, sector load, carrier availability, and mobility stability in the same near-real-time decision loop. Handover frequency and ping-pong behavior are therefore treated together with capacity and queue evolution rather than as separate mobility metrics. The mobility-related observations are further examined through an ablated PPO model, although the present results do not establish an independent performance gain from these additional features.

The rest of the paper is organized as follows. Section II presents the system architecture, Section III introduces the mathematical formulation, Section IV describes the system model, Section V presents the RL formulation, Section VI summarizes the experimental setup, Section VII discusses the performance results, and Section VIII concludes the paper.

II. System Architecture

This paper proposes a mobility-conditioned PPO framework for joint handover control and dual-carrier allocation. The controller observes radio-quality indicators, queue states, cell-load information and mobility-related features, including speed class, recent SINR variation and handover history. It selects association updates and carrier-level allocations, while final UE scheduling is handled by a proportional-fair rule. Constrained action handling prevents invalid decisions, and the reward penalizes excessive handovers and short-lived associations. Fig. 1 summarizes the complete RL control loop, including the observations provided to the controller, feasibility handling, reward components, control actions, proportional-fair scheduling, and state feedback from the network environment.

In this paper, a sector denotes one directional serving cell of a three-sector macro site and is treated as the serving entity for UE association, scheduling, and carrier-resource allocation.

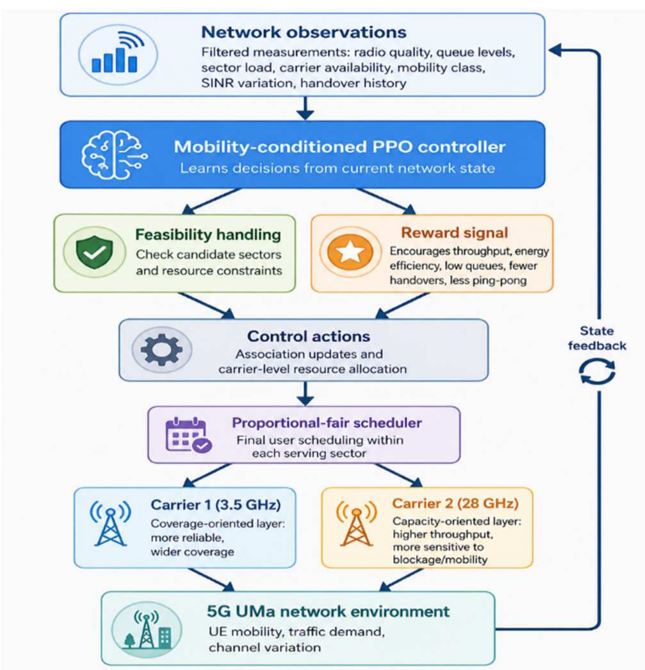

Fig. 1. Mobility-conditioned PPO control loop showing observations, reward, actions, feasibility handling, scheduling, and state feedback.

Evaluation uses a 3GPP TR 38.901-inspired 5G urban macro (UMa) scenario with FR1/FR2-like carriers and mixed UE mobility. The proposed controller is compared with Event-A3-based baselines and an ablated PPO variant.

III. Mathematical Model

This section defines the optimization framework used for joint handover control and dual-carrier allocation. The network is modeled in discrete time, where each control interval is indexed by $t \in \{0,1,2,\dots\}$ and has duration Δ. Let $\mathcal{B} = \{1,\dots,B\}$ denote the set of serving gNB sectors, $\mathcal{U}(t) = \{1,\dots,U_t\}$the set of active UEs at time t, and $\mathcal{C} = \{1,2\}$the set of available carriers. Carrier 1 represents the more reliable coverage-oriented carrier, while Carrier 2 represents the higher-capacity carrier with more sensitive propagation behavior.

For each UE u, a binary association variable $x_{u,b}(t)$ is defined as:

$$x_{(u,b)(t)} = \begin{cases} 1 b = b_{u(t)} \\ 0 b \neq b_{u(t)} \end{cases} \tag{1}$$

Since each UE is connected to one serving sector during a control interval, the association constraint is:

$$\sum_{b \in \mathcal{B}} x_{u,b}(t) = 1, \quad \forall u \in \mathcal{U}(t). \tag{2}$$

This assumption keeps the control process stable, while still allowing the serving gNB to allocate resources on one or both carriers.

Each serving sector operates under a fixed power budget and distributes carrier-level radio resources across the two carriers according to the current network state. The carrier-level power and resource-allocation constraints can be written as:

$$0 \le P_{b,c}(t), \quad \sum_{c \in \mathcal{C}} P_{b,c}(t) \le P_b^{max}, \quad \forall b \in \mathcal{B}. \tag{3}$$

$$0 \le \alpha_{u,b,c}(t) \le 1 \quad \sum_{u \in \mathcal{U}(t)} x_{u,b}(t) \alpha_{u,b,c}(t) \le 1, \quad \forall b \in \mathcal{B},\ c \in \mathcal{C} \tag{4}$$

Here, $P_{b,c}(t)$denotes the carrier-level transmit-power budget of sector $b$on carrier c, while $\alpha_{u,b,c}(t)$denotes the fraction of carrier c resources assigned to UE u by sector b. In the evaluation, the total sector power budget is fixed, so energy-efficiency differences mainly reflect the amount of useful data served under the same power-budget assumption.

Based on the carrier-level allocation, the instantaneous SINR of UE $u$from sector $b$on carrier $c$is expressed as:

$$\mathrm{SINR}_{\mathrm{u,b,c}}(t) = \frac{P_{b,c}(t)\, g_{u,b,c}(t)}{\sum_{j \in \mathcal{B},\, j \neq b} P_{j,c}(t)\, g_{u,j,c}(t) + N_0 W_c} \tag{5}$$

where $g_{u,b,c}(t)$is the effective channel gain between UE $u$and sector $b$on carrier c, $N_0$is the noise power spectral density, and $W_c$is the bandwidth of carrier c [8].

Using the above SINR, the achievable downlink rate of UE $u$is modeled as:

$$r_u(t) = \sum_{b \in \mathcal{B}} \sum_{c \in \mathcal{C}} x_{u,b}(t)\ \alpha_{u,b,c}(t)\, W_c \log_2 \left(1 + \mathrm{SINR}_{\mathrm{u,b,c}}(t)\right) \tag{6}$$

This expression represents the instantaneous achievable downlink data rate under the allocated radio resources using a Shannon-style capacity approximation. It should not be interpreted directly as application-level throughput, since the amount of traffic actually served during an interval also depends on the available queued data and the scheduling and allocation decisions reflected in the queue evolution. [9].

To account for non-full-buffer traffic, the serving sector maintains a per-UE downlink queue. Let $q_u(t)$ denote the backlog of traffic waiting for UE $u$at the beginning of interval t, and let $a_u(t)$denote the newly arriving downlink traffic during that interval. Since this queue is maintained at the network side, its backlog is directly available to the centralized controller through the gNB scheduling and buffer state, and no UE-side queue reporting is assumed:

$$q_u(t+1) = \max\{q_u(t) + a_u(t) - r_u(t)\Delta,\ 0\} \quad (7)$$

This queue model allows the controller to consider actual traffic demand instead of only maximizing instantaneous physical-layer rate.

Since mobility management is one of the main objectives of the proposed method, handover activity is explicitly considered. Let $b_u(t)$ denote the serving sector of UE u at interval t. A handover indicator is defined as:

$$h_{u(t)} = \begin{cases} 1 b_{u(t)} \neq b_{u(t-1)} \\ 0 b_{u(t)} = b_{u(t-1)} \end{cases} \quad (8)$$

Frequent handovers may increase signaling overhead and reduce association stability, especially in high-mobility scenarios. Therefore, the proposed controller does not aim only to maximize throughput, but also to reduce excessive handover behavior and short-lived association changes [10] [11].

In addition, the model includes the energy cost of network operation. The total network power consumption is modeled as the sum of static infrastructure power and carrier-level transmit-power terms:

$$P_{tot(t)} = \sum_{b\in\mathcal{B}}\left(P_b^{static} + \sum_{c\in\mathcal{C}} P_{(b,c)(t)}\right) \quad (9)$$

The instantaneous energy efficiency is then defined as:

$$\eta(t) = \frac{\sum_{u\in\mathcal{U}(t)} r_{u(t)}}{P_{tot(t)}} \quad (10)$$

In the present evaluation, $P_{tot}(t)$ follows a fixed sector power-budget assumption. Consequently, the reported energy-efficiency metric should be interpreted as served data per unit of assumed network-side power, rather than as evidence of dynamic power savings through sleep modes, power-amplifier adaptation, or carrier deactivation. UE energy consumption is outside the present model because the considered control decisions operate at the RAN side, through serving-sector association and carrier-resource allocation. A joint BS–UE energy model would be required to evaluate end-to-end energy consumption and is left for future work.

To combine these targets, the decision process is guided by an immediate utility (F(t)) that includes normalized served throughput, energy efficiency and Jain's fairness index, while penalties are applied for queue backlog, handovers, ping-pong events and outage. The numerical weights and normalization factors used in the PPO implementation are provided in Section V. The policy objective is:

$$\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \beta^t F(t)\right] \quad (11)$$

where $\pi$is the control policy, $\beta \in (0,1)$is the discount factor, and $F(t)$is the immediate utility at interval t.

This formulation provides the basis for the reinforcement learning framework developed in the following section. The system state is formed by radio measurements, mobility-related variables and traffic indicators, while the action space includes handover-control decisions together with carrier-level allocation choices. In this way, the model captures the central trade-off of the paper: maintaining high service capacity and energy efficiency while limiting queue growth, excessive handovers and ping-pong behavior.

IV. System Model

This section describes the network setting considered in the study. A 5G urban macro (UMa) downlink scenario is assumed, consisting of B serving sectors deployed over a fixed service area and a time-varying set of active UEs. The network operates over two carriers: Carrier 1 represents a lower-frequency coverage-oriented carrier, while Carrier 2 represents a higher-frequency capacity-oriented carrier. This dual-carrier structure captures the trade-off between robust coverage and high-throughput but less stable links.

For simplicity and control stability, each UE is associated with one serving sector during a given control interval. However, the serving sector may allocate resources to the UE on one carrier or on both carriers. This keeps the association decision discrete while still allowing the resource-allocation policy to exploit carrier diversity.

The propagation environment follows a 3GPP TR 38.901-inspired UMa modeling approach. For each UE–sector–carrier combination, the effective channel gain is expressed as:

$$g_{u,b,c}(t) = z_{u,b,c}(t)\, 10^{-PL_{u,b,c}(t)/10} \left|h_{u,b,c}(t)\right|^2 (12)$$

where $PL_{u,b,c}(t)$ denotes the large-scale path-loss and shadowing component in dB, $h_{u,b,c}(t)$represents the small-scale fading term, and $z_{u,b,c}(t)$is an availability factor. The use of $z_{u,b,c}(t)$is especially important for the higher-frequency carrier, where blockage or limited coverage may reduce or remove carrier usability. This representation is intentionally modular: path loss, fading and blockage terms can be replaced by more detailed models without changing the RL formulation.

Mobility is a key part of the proposed framework. Each UE is described by a time-varying mobility state. Let $p_u(t) \in \mathbb{R}^2$denote the UE position, $v_u(t)$the UE speed, and $\theta_u(t)$the direction of movement. A simple two-dimensional kinematic update is used to evolve the UE position over time:

$$p_u(t+1) = p_u(t) + v_u(t)\Delta \begin{bmatrix} \cos\theta_u(t) \\ \sin\theta_u(t) \end{bmatrix} \quad (13)$$

This mobility evolution affects path loss, cell ranking, carrier usability and the likelihood that a handover becomes beneficial. In addition to basic motion variables, the controller may also use mobility-related indicators such as speed class, recent handover history and short-term SINR variation. These features allow the policy to be conditioned on mobility behavior rather than relying only on instantaneous channel measurements.

Traffic is represented through per-UE downlink queues maintained at the serving sectors. At each interval, new traffic arrivals are added to the UE queues, while served data are removed according to the rate achieved through the selected serving sector and carrier allocation. This allows the model to evaluate not only instantaneous capacity, but also whether traffic demand accumulates over time. Within each serving sector, the final user-level scheduling is handled by a proportional fair rule, while the RL controller operates at the higher carrier-control and mobility-control level.

At each control interval, the controller observes a compact state description. At the modeling level, the network state combines UE-related observations with cell-level information such as aggregate load and carrier utilization. A representative state vector can include serving sector information, candidate-sector SINR values, queue state, mobility class, previous handover indicators, sector load and carrier availability. In the implemented controller, these variables are represented through compact features so that the RL policy remains computationally feasible in a dense scenario with many UEs.

The control interval $\Delta$is assumed to lie in the near-real-time control range rather than at the PHY-symbol timescale. This means that the controller acts on filtered radio measurements and short-term network summaries, not on raw physical-layer samples. Such a timescale is more appropriate for learning-based RRM because it reduces action frequency and keeps policy evaluation computationally tractable. The concrete interval value used in the experiments is specified in Section VI.

For performance evaluation, conventional measurement-based mobility baselines are defined using Event-A3-style handover triggering with hysteresis and time-to-trigger (TTT), since these remain standard parameters in 5G handover configuration. These baselines provide realistic non-learning reference points for comparing the learned policy. The detailed baseline variants used in the evaluation are presented in Section VI.

V. RL FORMULATION

The proposed control problem is modeled as a constrained Markov decision process with a centralized reinforcement learning controller. The centralized design allows the policy to jointly consider handover behavior, carrier usage, traffic demand and sector load instead of treating each UE independently. At every control interval, the controller receives a compact representation of the current network state. This state includes UE-level features, such as radio-quality indicators, queue backlog, mobility class, recent SINR variation, handover history and time spent in the serving sector, together with sector-level information, including aggregate load, carrier utilization and carrier availability. These inputs allow the controller to consider both the current radio conditions and the recent mobility behavior of each UE.

The action space includes two main types of decisions. First, the controller determines whether selected UEs should remain connected to their current serving sector or be handed over to a feasible neighboring sector. Second, it adjusts the carrier-level resource allocation between the two available carriers. The controller does not perform detailed scheduling for every UE, as this would significantly increase the action-space complexity in a dense network. Instead, the final UE-level resource distribution within each serving sector is handled by a proportional-fair scheduler. Before any action is applied, feasibility handling ensures that handovers are restricted to valid candidate sectors and that carrier-resource constraints are satisfied.

The reward combines normalized served throughput and energy efficiency with Jain's fairness index, while penalizing queue backlog, handovers, ping-pong events, and outage. The corresponding weights are 1.0 for throughput, 0.8 for energy efficiency, 0.5 for fairness, 0.7 for queue backlog, 0.6 for handovers, 0.5 for ping-pong events, and 0.9 for outage. Throughput, energy efficiency, and queue backlog are normalized using hyperbolic tangent functions with scales of 2500 Mbps, 5 Mbit/J, and 1500 Mbits, respectively.

The PPO controller uses separate actor and critic multilayer perceptrons, each with two hidden layers of 512 and 256 neurons and ReLU activations. Training is performed for 120,000 environment steps using a learning rate of 3×10^{-4}, discount factor $\gamma = 0.99$, GAE parameter $\lambda = 0.95$, clipping ratio 0.2, minibatch size 128, and 10 optimization epochs per update. The value and entropy coefficients are 0.5 and 0.01, respectively, and the maximum gradient norm is 0.5.

To examine the independent contribution of explicit mobility information, an ablated PPO variant is also evaluated. In this version, mobility-related inputs, such as speed class, movement direction and handover-history indicators, are removed from the observation space. The remaining radio-quality, load and queue-related features are retained. This comparison makes it possible to determine whether the performance improvement is mainly produced by the PPO-based control structure or whether explicit mobility conditioning provides an additional benefit.

VI. EXPERIMENTAL SETUP

The proposed method is evaluated in a system-level 5G UMa downlink simulation comprising 19 macro sites, three sectors per site, and a total of 57 serving sectors. The two carriers operate at 3.5 GHz with 100 MHz bandwidth and at 28 GHz with 400 MHz bandwidth. The scenario contains 2400 UEs divided into pedestrian (50%, 0.6–1.8 m/s), moderate-mobility (30%, 2–6 m/s) and vehicular (20%, 8–15 m/s) classes, with randomized initial positions and trajectories. Per-UE downlink queues are initialized between 0 and 10 Mbits, traffic arrival rates have a mean of 6 Mbps and a standard deviation of 2 Mbps, and the queue size is limited to 200 Mbits. A 1 s near-real-time control interval and 300-interval episodes are used.

The radio model follows the TR 38.901 UMa LOS/NLOS path-loss formulation [3]. The inter-site distance is 500 m, with BS and UE heights of 25 m and 1.5 m. Log-normal shadowing standard deviations are 4/6 dB for FR1 and 5/8 dB for FR2 under

LOS/NLOS conditions, while small-scale fading is represented by a Rayleigh power gain. Co-channel interference is accumulated from all serving sectors. The receiver noise figure is 7 dB, and the BS/UE antenna gains are 17/0 dBi at 3.5 GHz and 24/5 dBi at 28 GHz. FR2 blockage is represented through a time-varying availability state with a 20 dB blockage loss, while the sector transmit-power ceiling is 46 dBm.

The mobility-conditioned PPO controller is compared with Event-A3 + PF, Load-aware A3 and an ablated PPO variant without explicit mobility inputs. The evaluation reports throughput, energy efficiency, final queue backlog, handovers per UE per minute and ping-pong rate under common system-level assumptions; it is not intended as a full 3GPP conformance-level simulator. The selected baselines represent three complementary reference points: conventional measurement-driven handover with PF scheduling, a more conservative load-aware mobility policy, and an ablated PPO controller that preserves the learning structure while removing explicit mobility features. This comparison separates conventional mobility control, load-aware conservative behavior, and the contribution of the PPO observation design. A broader comparison with additional learning-based handover methods remains outside the scope of the present evaluation.

The main evaluation metrics, as seen in Table 1 are average network throughput, average UE throughput, energy efficiency, final mean queue backlog, handovers per UE per minute and ping-pong rate. These metrics are chosen so that the comparison is not limited to raw throughput alone. The goal is to examine whether the proposed controller can maintain high service capacity while also improving mobility stability and avoiding excessive queue growth under mixed mobility conditions.

TABLE 1. MAIN SIMULATION PARAMETERS

Parameter	Value
Scenario	5G UMa downlink
Macro sites	19
Sectors per site	3
Total sectors	57
Carrier 1	3.5 GHz, 100 MHz
Carrier 2	28 GHz, 400 MHz
Number of UEs	2400
Mobility classes	pedestrian 0.6–1.8 m/s (50%), moderate 2–6 m/s (30%), vehicular 8–15 m/s (20%)
Control interval	1 s
Episode length	300 intervals
Simulated time per episode	5 minutes
Traffic arrivals	6 Mbps mean, 2 Mbps standard deviation
Main scheduler	Proportional fair
Compared schemes	Event-A3 + PF, Load-aware A3, PPO, ablated PPO
Maximum UE queue	200 Mbits

The PPO training was performed using a fixed random seed of 7, while the evaluation configuration used five 300-interval episodes. Independent retraining across multiple random seeds was not performed; therefore, confidence intervals across training runs are not reported, and the small numerical differences between the two PPO variants should not be interpreted as statistically significant.

VII. PERFORMANCE EVALUATION

This section evaluates the proposed mobility-conditioned PPO controller against the reference schemes described in Section VI. The comparison focuses on both capacity-related and mobility-stability metrics, since the objective is not only to maximize throughput, but also to avoid excessive handovers, ping-pong behavior and queue growth. The evaluated schemes are Event-A3 with proportional fair scheduling, Load-aware A3, the proposed mobility-conditioned PPO controller and an ablated PPO variant without explicit mobility-related inputs as seen in Table 2. The queue values reported in Table II are final mean per-UE backlogs after traffic arrivals and service, rather than fixed queue sizes; the configured per-UE queue limit is 200 Mbits.

The average network throughput results are shown in Fig. 2. Event-A3 + PF achieves 4479.55 Mbps, while the proposed mobility-conditioned PPO achieves 4474.37 Mbps. The difference is only about 0.12%, which shows that the proposed controller preserves almost the same network capacity as the conventional high-throughput baseline.

TABLE 2. PERFORMANCE COMPARISON OF THE EVALUATED SCHEMES

Metric	**Event-A3 + PF**	**Load-aware A3**	**PPO mobility-conditioned**	**PPO ablated**
Avg. network throughput (Mbps)	**4479.55**	**4203.60**	**4474.37**	**4480.30**
Avg. UE throughput (Mbps)	**1.866**	**1.752**	**1.864**	**1.867**
Energy efficiency (Mbit/J)	**0.4560**	**0.4279**	**0.4555**	**0.4561**
Final mean queue backlog (Mbits)	**9.89**	**14.73**	**10.64**	**10.01**
Handovers / UE / min	**2.670**	**1.427**	**1.967**	**1.938**
Ping-pong rate	**0.1779**	**0.0205**	**0.0711**	**0.0694**

The Load-aware A3 baseline achieves 4203.60 Mbps, which is lower than both Event-A3 + PF and the PPO-based schemes. Compared with Load-aware A3, the mobility-conditioned PPO improves average network throughput by approximately 6.4%. This indicates that the proposed controller avoids the excessive capacity loss caused by the more conservative load-aware heuristic.

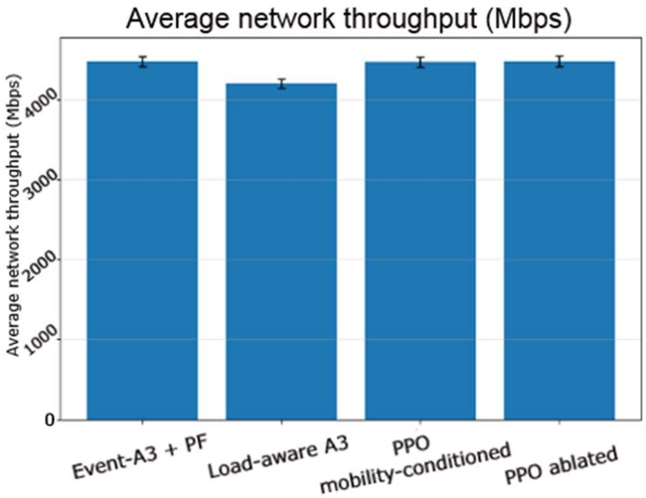

Fig. 2. Average Network Throughput of the Evaluated Schemes.

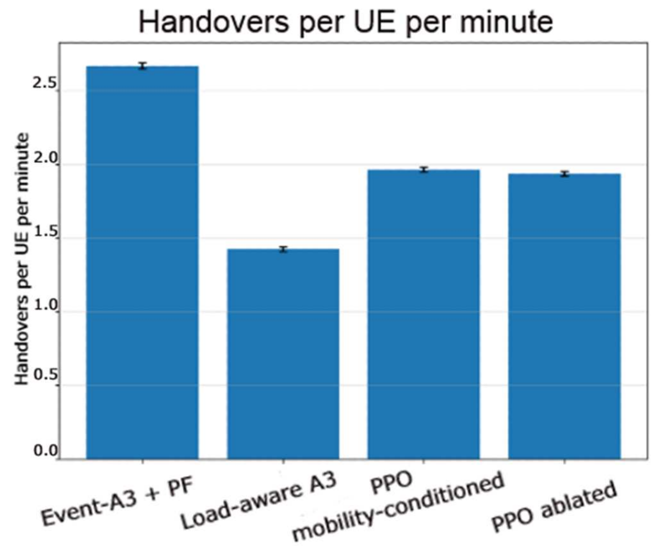

Fig. 4. Handovers per UE per Minute.

Queue results are shown in Fig. 3. Event-A3 + PF gives the lowest final mean queue backlog among the non-ablated schemes, with 9.89 Mbits, while the mobility-conditioned PPO gives 10.64 Mbits. This shows a small queue increase compared with Event-A3 + PF. However, Load-aware A3 reaches 14.73 Mbits, meaning that its conservative behavior leaves more traffic unserved by the end of the episode. Compared with Load-aware A3, the mobility-conditioned PPO reduces the final mean queue backlog by approximately 27.8%.

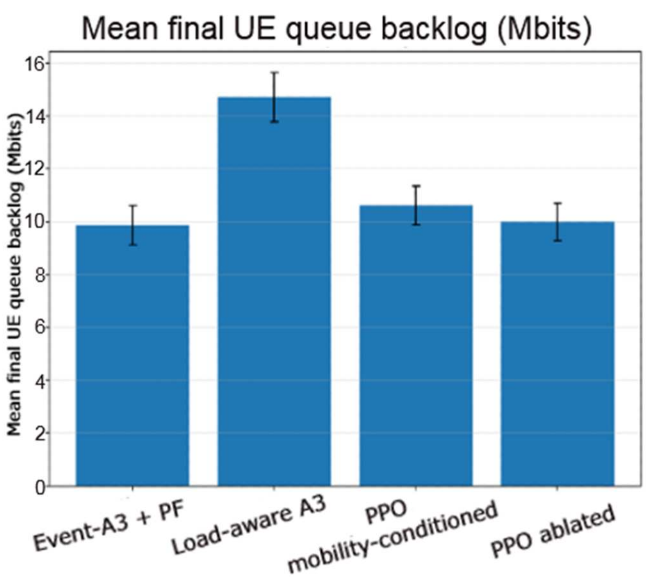

Fig. 3. Final Mean Queue Backlog of the Evaluated Schemes.

The main advantage of the proposed controller appears in the mobility-stability metrics. As shown in Fig. 4, Event-A3 + PF produces 2.670 handovers per UE per minute, while the mobility-conditioned PPO reduces this value to 1.967. This corresponds to a reduction of approximately 26.3%. Therefore, the proposed controller maintains almost the same throughput as Event-A3 + PF while reducing unnecessary mobility activity.

The ping-pong results in Fig. 5 confirm the same trend. Event-A3 + PF has a ping-pong rate of 0.1779, while the mobility-conditioned PPO reduces it to 0.0711. This is a reduction of approximately 60.0%. This result is important because it shows that the PPO controller does not only reduce the total number of handovers but also limits unstable short-term association changes.

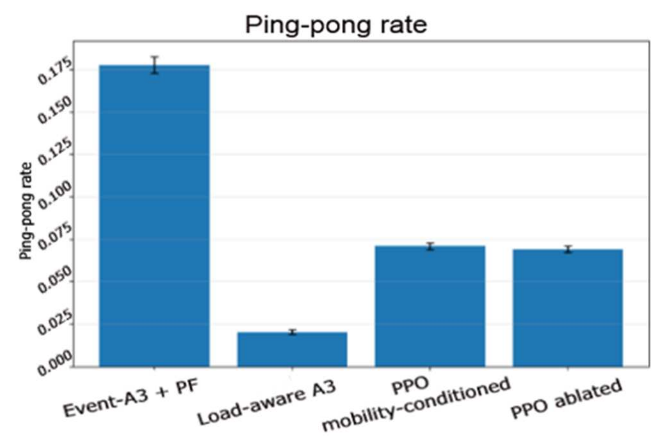

Fig. 5. Ping-Pong Rate of the Evaluated Schemes.

Load-aware A3 achieves the lowest handover and ping-pong values, but this comes with a clear throughput and queue penalty. Therefore, Load-aware A3 represents a conservative operating point, while the mobility-conditioned PPO provides a more balanced trade-off between capacity and mobility stability.

These results are better interpreted as a throughput–mobility trade-off than as a throughput improvement. Relative to Event-A3 + PF, the mobility-conditioned PPO reduces average network throughput by only about 0.12%, while reducing handovers by 26.3% and ping-pong events by approximately 60.0%. Load-aware A3 is more conservative and produces fewer mobility events, but this is accompanied by lower throughput and a larger final queue backlog. The PPO-based schemes therefore operate between the aggressive capacity-oriented and conservative mobility-oriented reference points.

Energy efficiency follows the same general behavior as network throughput, as seen in Fig. 6. The mobility-conditioned PPO achieves 0.4555 Mbit/J, which is very close to the 0.4560 Mbit/J obtained by Event-A3 + PF and higher than the 0.4279 Mbit/J obtained by Load-aware A3. Since the evaluation uses a fixed sector power budget, energy efficiency should be interpreted as served throughput per unit of assumed network power, rather than as dynamic power-saving behavior.

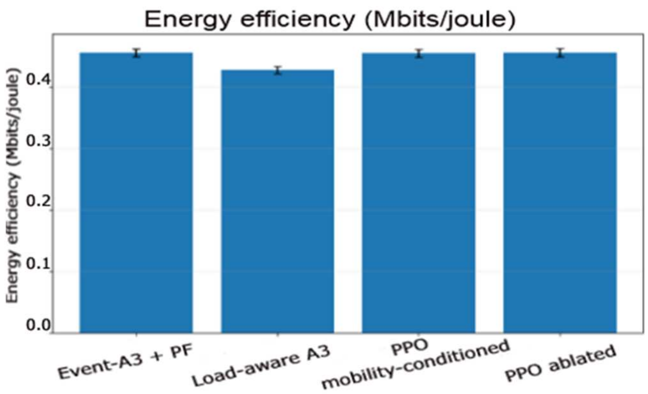

Fig. 6. Energy Efficiency under Fixed Sector Power Budget.

The ablated PPO variant performs very close to the mobility-conditioned PPO and is slightly better in all reported metrics. It achieves 4480.30 Mbps average network throughput, 10.01 Mbits final mean queue backlog, 1.938 handovers per UE per minute, and a 0.0694 ping-pong rate. The differences between the two PPO variants are small, but the present results do not provide evidence that explicit mobility features independently improve performance. The more appropriate conclusion is that the PPO-based joint control structure provides the observed throughput–mobility trade-off, while the specific benefit of mobility conditioning remains to be established.

Overall, the results show that the proposed mobility-conditioned PPO controller maintains almost the same throughput as Event-A3 + PF, while reducing handovers by approximately 26.3% and ping-pong rate by approximately 60.0%. Compared with Load-aware A3, it improves throughput by approximately 6.4% and reduces final mean queue backlog by approximately 27.8%. These results support the use of PPO-based adaptive control for balancing capacity and mobility stability in dual-carrier 5G networks.

VIII. Conclusion and Future Work

This paper presented a PPO-based framework for joint handover control and dual-carrier resource allocation in a 5G urban macro network. The mobility-conditioned controller preserved nearly the same average network throughput as Event-A3 + PF (4474.37 versus 4479.55 Mbps), while reducing handovers per UE per minute from 2.670 to 1.967 and the ping-pong rate from 0.1779 to 0.0711. Compared with Load-aware A3, it improved throughput by approximately 6.4% and reduced final mean queue backlog by approximately 27.8%. The ablated PPO variant performed similarly and was slightly better in all reported metrics. Therefore, the present results support PPO-based joint control for balancing capacity and mobility stability, but they do not establish an independent performance gain from explicit mobility conditioning.

Future work will refine mobility-specific reward terms and incorporate richer temporal features, such as dwell time, trajectory history, predicted cell residence time and short-term SINR evolution. Longer and larger evaluation campaigns are also needed to assess statistical significance against the ablated variant. The present evaluation also uses a fixed reward configuration, traffic setting, mobility configuration, and control interval, and therefore does not isolate the sensitivity of the results to individual design parameters. Future evaluation will examine handover and ping-pong penalties, traffic load, mobility speed, and control-interval duration together with repeated runs under different random seeds. Extending the simulator with a more detailed power-consumption model, dynamic sleep states, multi-node connectivity and more realistic Open RAN or system-level validation would further strengthen the practical assessment.

References

[1] ITU-R, “Minimum requirements related to technical performance for IMT-2020 radio interface(s),” Report ITU-R M.2410-0, Nov. 2017.

[2] ETSI, “5G; NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (3GPP TS 38.101-1 Release 18),” ETSI TS 138 101-1 V18.8.0, Feb. 2025; and ETSI, “Part 2: Range 2 Standalone (3GPP TS 38.101-2 Release 18),” ETSI TS 138 101-2 V18.8.0, Jan. 2025.

[3] ETSI, “Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 Release 16),” ETSI TR 138 901 V16.1.0, Nov. 2020.

[4] ETSI, “NR; Radio Resource Control (RRC); Protocol specification (3GPP TS 38.331 Release 17),” ETSI TS 138 331 V17.12.0, Apr. 2025.

[5] S. C. Borst and P. A. Whiting, “User-level performance of channel-aware scheduling algorithms in wireless data networks,” IEEE/ACM Trans. Netw., vol. 13, no. 3, pp. 636–647, Jun. 2005.

[6] M. Zangooei, N. Saha, M. Golkarifard, and R. Boutaba, “Reinforcement learning for radio resource management in RAN slicing: A survey,” IEEE Commun. Mag., vol. 61, no. 2, pp. 118–124, Feb. 2023.

[7] O-RAN Alliance, “O-RAN: Towards an Open and Smart RAN,” White Paper, Oct. 2018.

[8] Y. Azimi, S. Yousefi, H. Kalbkhani, and T. Kunz, “Mobility aware and energy-efficient federated deep reinforcement learning assisted resource allocation for 5G-RAN slicing,” Computer Communications, vol. 217, pp. 166–182, 2024.

[9] H. Choi, T. Kim, S. Lee, H.-S. Choi, and N. Yoo, “Energy-Efficient Dynamic Enhanced Inter-Cell Interference Coordination Scheme Based on Deep Reinforcement Learning in H-CRAN,” Sensors, vol. 24, no. 24, Art. no. 7980, 2024.

[10] H. Wang and B. Li, “Double-deep Q-learning-based handover management in mmWave heterogeneous networks with dual connectivity,” *Transactions on Emerging Telecommunications Technologies*, vol. 35, no. 1, Art. no. e4907, 2024.

[11] C. F. Kwong, C. Shi, Q. Liu, S. Yang, D. Chieng, and P. Kar, “Autonomous handover parameter optimisation for 5G cellular networks using deep deterministic policy gradient,” *Expert Systems with Applications*, vol. 246, Art. no. 122871, 2024.