Handover Optimization in 6G Network using Reinforcement learning

Project Report Submitted in Partial Fulfilment of the Requirements for the Degree of

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

Jasmehar Singh Chadha: (Roll No. 2210110719)

Under the Supervision of

Dr. Shankar K. Ghosh
Assistant Professor

SHIV NADAR
INSTITUTION OF EMINENCE DEEMED TO BE
UNIVERSITY
DELHI NCR

Department of Computer Science and Engineering

October, 2025

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of the Student: Jasmehar Singh Chadha

Signature and Date:
October 15, 2025

Handover Optimization in 6G Network using Reinforcement Learning

Jasmehar Singh Chadha
Department of Computer Science and Engineering
Shiv Nadar University, Greater Noida, India
jc270@snu.edu.in

Dr .Shankar K. Ghosh
Department of Computer Science and Engineering
Shiv Nadar University, Greater Noida, India
shankar.ghosh@snu.edu.in

Abstract

Future 6G wireless networks must support ultra-dense, multi-tier heterogeneous deployments and deliver seamless mobility under highly dynamic radio environments. Traditional handover methods struggle to adapt to rapid fluctuations in signal quality, load, and beam alignment particularly in Sub-THz and mmWave bands. In this work, we model a realistic 6G three-tier heterogeneous network (macro, micro, and pico base stations) and formulate mobility management as a Contextual Multi-Armed Bandit (CMAB) problem. We compare the performance of two reinforcement learning approaches: standard Upper Confidence Bound (UCB), which is context-free, and LinUCB, which incorporates channel, mobility, and network state features such as SINR, distance, frequency band, beam alignment, and cell load. Simulation results using a full 6G propagation pipeline featuring 3D beamforming, Sub-THz molecular absorption, Rayleigh/Rician fading, and spatially correlated shadowing demonstrate that LinUCB significantly outperforms standard UCB. LinUCB achieves higher average SINR and throughput, reduces handovers by over 95 % , and lowers the ping-pong rate while providing superior cell-edge performance. These findings highlight the importance of context-aware learning for scalable, reliable, and low-latency mobility optimization in next-generation 6G networks.

Index Terms — 6G, heterogeneous networks, mobility management, handover optimization, Sub-THz/mmWave, reinforcement learning, contextual multi-armed bandit, LinUCB, beamforming, SINR.

I. Introduction

The handover problem refers to the difficulties that arise when a User Equipment (UE) transitions its connection from one base station (BS) to another while moving across a cellular network. Ideally, this transfer should be seamless, but in practice it often suffers from delays, poor timing, or incorrect cell selection. When the handover occurs too late or the target cell fails to acquire the UE, users experience disruptions such as dropped connections, frozen video playback, reduced data rates, or increased latency. These issues directly affect applications that rely on continuous connectivity, making handover reliability essential to maintaining overall quality of service (QoS).

Fifth-generation (5G) networks were developed to deliver higher data rates, lower latency, and improved efficiency compared to earlier wireless technologies. To achieve these goals, 5G introduced new features such as millimeter-wave (mmWave) communication, massive MIMO, and dense deployments of small cells. While these advancements significantly enhanced network capacity and user throughput, they also made mobility management more challenging. The short range and high sensitivity of mmWave signals mean that users frequently move in and out of coverage, requiring handovers to occur more often and with greater precision. However, traditional handover strategies in 5G still depend largely on fixed signal thresholds and rule-based decision-making. As a result, they often react too slowly to rapid changes in signal quality, struggle to cope with sudden drops caused by beam misalignment or blockage, and sometimes trigger unnecessary or premature handovers. These limitations reveal a clear mismatch between the speed and complexity of modern wireless environments and the static nature of conventional handover logic.

Sixth-generation (6G) networks pushes these requirements even further. By moving towards sub-THz frequencies, extremely dense deployments, and highly directional beams, 6G aims to support ultra-high data rates, real-time applications, and seamless connectivity in dynamic environments. However, the same characteristics that enable these capabilities also amplify the difficulty of performing reliable handovers. Narrow beams and fast-changing channel conditions make link quality highly sensitive to even small movements or environmental variations. In such settings, traditional mechanisms based on simple threshold comparisons are often inadequate, as they cannot adapt quickly enough to the pace of change or anticipate when a link is likely to deteriorate.

In this context, **Reinforcement Learning (RL)** provides a promising path forward for enhancing handover performance in next-generation networks. Rather than relying on pre-defined rules, an RL system learns from experience, observing how signal conditions evolve and how user movement affects link stability. The handover task can be naturally viewed through the lens of a **Contextual Multi-Armed Bandit (CMAB)** problem, where each base station represents a possible action and the agent must select the most suitable one based on the current context. By continuously learning from the outcomes of its decisions, the agent becomes increasingly capable of choosing the best cell at the right moment, reducing unnecessary handovers, avoiding ping-pong effects, and maintaining more stable connections even under rapidly changing conditions.

Motivated by these challenges and opportunities, our work focuses on developing an improved RL-based handover strategy that leverages contextual information more effectively while remaining simple, practical, and efficient. Using a controlled 6G-like simulation environment that reflects realistic mobility patterns and signal variations, we aim to demonstrate that a context-driven RL approach can provide more stable and reliable handover behaviour than conventional methods. By addressing the shortcomings observed in 5G and aligning with the demands of emerging 6G technologies, the proposed approach contributes to building more intelligent and adaptive mobility management systems for future wireless networks.

II. Literature Review

A. Evolution of Cellular Networks and Handover Challenges

The handover problem has evolved significantly across generations of cellular networks, with each advancement introducing new complexities in mobility management. Traditional handover mechanisms in 4G and early 5G networks primarily relied on fixed signal thresholds and rule-based decision-making, where handover decisions were triggered when signal quality metrics such as Reference Signal Received Power (RSRP) or Signal-to-Interference-plus-Noise Ratio (SINR) crossed predefined thresholds. While these approaches provided adequate performance in relatively stable environments, they exhibit significant limitations in modern wireless networks characterized by high mobility, dense deployments, and rapidly changing channel conditions.

Fifth-generation networks have introduced advanced features including millimeter-wave (mmWave) communication, massive MIMO, and dense small cell deployments to achieve higher data rates and improved network efficiency [5]. However, these enhancements have simultaneously amplified the complexity of mobility management. Rappaport et al. [4] demonstrated that the short range and high sensitivity of mmWave signals require more frequent handovers with greater precision, while narrow pencil beams can be easily disrupted by blockages or small changes in user equipment orientation.

B. 5G Deployment Challenges and Network Densification

The deployment of ultra-dense heterogeneous networks (HetNets) in 5G has created new challenges for handover optimization. Uusitalo et al. [7] analyzed 5G network coverage planning and identified deployment challenges related to inter-tier handover management, load balancing, and interference coordination across macro, micro, and pico cell layers. The work by Martín-Vega et al. [6] developed system models for average downlink SINR in multi-beam 5G networks, revealing that beam-based communication introduces additional complexity in predicting handover timing and target cell selection.

Large-scale propagation measurements conducted by Rappaport et al. [8] provided critical insights into distance-dependent path loss models for outdoor and indoor 5G systems. Their findings indicate that mmWave frequencies experience significantly higher path loss and shadowing effects compared to sub-6 GHz bands, necessitating more intelligent handover strategies that can anticipate channel quality degradation before complete link failure occurs.

C. 6G Vision and Sub-THz Communication

Sixth-generation networks aim to push beyond the capabilities of 5G by targeting sub-THz frequencies (100 GHz to 1 THz) to achieve ultra-high data rates exceeding 100 Gbps, sub-millisecond latency, and massive connectivity [3]. Rappaport et al. [11] provided comprehensive analysis of sub-THz channel propagation, highlighting that these frequencies experience severe atmospheric absorption, requiring highly directional beamforming with extreme angular precision. The authors demonstrated that path loss at 142 GHz can exceed that at 28 GHz by more than 20 dB over typical cell ranges, dramatically reducing coverage areas and necessitating more frequent handovers.

The transition to sub-THz frequencies introduces several fundamental challenges for mobility management. First, the extremely narrow beamwidths required for adequate link budget (often less than 10 degrees) make beam alignment highly sensitive to user movement and orientation changes. Second, blockage events become more frequent and severe, as sub-THz signals exhibit poor penetration through common obstacles. Third, the increased Doppler spread at higher frequencies complicates channel estimation and beam tracking, particularly for high-velocity users.

D. Reinforcement Learning for Handover Optimization

Recognizing the limitations of traditional threshold-based approaches, researchers have increasingly explored reinforcement learning (RL) techniques for adaptive handover management. Yajnanarayana et al. [1] pioneered the application of RL to 5G handover control, proposing a centralized agent that processes radio measurement reports and selects handover actions to maximize long-term utility. Their simulation results demonstrated significant improvements in handover success rate and throughput compared to conventional 3GPP handover mechanisms.

Polese et al. [9] provided a comprehensive survey of RL applications in 5G and beyond networks, categorizing approaches into model-free methods (Q-learning, SARSA, policy gradient) and model-based techniques. They identified key challenges including sample efficiency, scalability to large state spaces, and the need for safe exploration during the learning phase.

The survey emphasized that contextual bandit formulations offer a promising middle ground between full Markov Decision Processes and simple multi-armed bandits, providing sufficient learning capability while maintaining computational tractability.

E. Multi-Armed Bandit Approaches

Multi-armed bandit (MAB) frameworks have emerged as particularly suitable for handover optimization due to their ability to balance exploration and exploitation in sequential decision-making. Chen et al. [10] applied MAB techniques to mobility management in ultra-dense networks, demonstrating that Upper Confidence Bound (UCB) algorithms can effectively reduce handover frequency while maintaining quality of service. Their work showed that UCB algorithms can effectively reduce handover frequency while maintaining quality of service. However, their context-free approach does not exploit real-time channel state information, potentially limiting performance in highly dynamic environments.

Sun et al. [12] extended MAB approaches by incorporating spatial and temporal context into handover decisions for ultra-dense mmWave networks. Their Contextual Multi-Armed Bandit (CMAB) formulation uses features such as received signal strength, user location, and cell load to guide base station selection. The key innovation lies in exploiting the empirical distribution of users' post-handover trajectories and line-of-sight blockage patterns, learned online through the MAB framework. By formulating two different MAB problems focusing on spatial and space-time contexts respectively, their simulation results demonstrated that contextual handover mechanisms significantly outperform existing counterparts in reducing unnecessary handovers across all scenarios. The Linear UCB (LinUCB) algorithm, which assumes a linear relationship between context features and expected rewards, demonstrated particularly strong performance due to its sample efficiency and convergence guarantees.

F. Deep Reinforcement Learning Approaches

More recent work has explored deep reinforcement learning (DRL) methods that can handle high-dimensional state spaces and complex reward structures. Machumilane et al. [2] developed a DRL-based approach for handover and SINR-aware path optimization in 5G unmanned aerial vehicle (UAV) mmWave communication. Their Deep Q-Network (DQN) agent learns optimal handover policies while simultaneously optimizing UAV trajectory to maintain connectivity. Results demonstrated significant improvements in both handover efficiency and end-to-end throughput for mobile aerial platforms.

Despite their promise, DRL approaches face several practical challenges. First, they typically require extensive training periods and large amounts of data, which may not be available in newly deployed networks. Second, the black-box nature of deep neural networks makes it difficult to interpret decisions and ensure robust performance across diverse scenarios. Third, DRL agents may exhibit instability during online learning, potentially causing temporary service degradation.

G. Research Gaps and Motivation

While existing literature has made significant progress in applying learning-based techniques to handover optimization, several critical gaps remain, particularly in the context of 6G networks:

Limited 6G-specific modeling: Most existing RL approaches focus on 5G mmWave scenarios and do not adequately address sub-THz propagation characteristics, including molecular absorption, beam split effects, and ultra-dense beam management requirements.

Context feature engineering: Previous CMAB approaches have not systematically investigated which contextual features are most informative for 6G handover decisions, particularly in three-tier heterogeneous networks with multiple frequency bands.

Handover cost modeling: The reward functions used in existing work often oversimplify the true cost of handovers, failing to capture the trade-off between throughput gain and signaling overhead in realistic 6G scenarios.

Scalability and interpretability: While DRL methods offer high capacity, they sacrifice the interpretability and convergence guarantees of simpler linear models. There is a need for approaches that balance learning capability with practical deployment constraints.

Motivated by these gaps, this work develops a CMAB-based handover optimization framework specifically designed for 6G heterogeneous networks. We employ LinUCB as a context-aware yet tractable solution that exploits real-time channel, mobility, and network state information. Our approach is evaluated using a comprehensive 6G propagation model incorporating sub-THz molecular absorption, 3D beamforming, and spatially correlated fading. By comparing LinUCB against context-free UCB, we demonstrate the value of contextual learning for achieving superior signal quality, throughput, and handover efficiency in next-generation wireless networks.

III. Experimental Setup

A. 6G Grid Conditions Considered

To evaluate learning-based handover strategies in realistic sixth-generation (6G) conditions, we construct a multi-layer heterogeneous cellular grid that models the spatial, spectral, and propagation characteristics expected in future networks. The simulation includes macro, micro, and pico base stations, each operating on sub-6, mmWave, and sub-THz carrier frequencies. This layered architecture enables both wide-area and highly localised coverage, reflecting the density, beam dependence, and ultra-high-frequency operation anticipated in 6G deployments.

1) ***Macro cells****:* constitute the highest-power layer and provide the widest coverage region. They ensure global connectivity across the grid, especially in areas lacking dense deployment. Due to their large coverage footprint, macro cells deliver lower beam precision and experience slower variations in link quality but are essential for mobility continuity.

2) ***Micro cells****:* occupy the intermediate tier, operating at moderate transmit power and covering smaller regions within the macro layer. They relieve macro-cell load and create smoother handover transitions by offering higher capacity in moderately dense areas. The reduced coverage radius results in more distance-sensitive path loss and more frequent beam fluctuations than in macro-tier links.

3) ***Pico cells****:* form the densest layer and are deployed in hotspots requiring extremely high capacity. Their low transmit power and short-range operation greatly amplify the impact of blockage and small-scale fading, especially at mmWave and sub-THz frequencies. Pico cells naturally induce frequent handovers because users traverse their coverage areas rapidly. Their dense placement makes them ideal for evaluating mobility performance in beam-based 6G environments.

Together, the three tiers create a hierarchical HetNet in which coverage responsibility is distributed across wide-area, mid-range, and ultra-dense access points. This arrangement generates highly heterogeneous link conditions and realistic handover triggers, forming a natural testbed for reinforcement learning-based mobility management.

B. Path-loss and Fading models

The simulator implements a 3D multi-scale radio channel model tailored for 6G bands (Sub-6, mmWave, Sub-THz). The implementation and parameters follow the simulation files used in this work.

1) LOS probability: A distance-, frequency- and BS-type-dependent LOS probability is used. The implemented functional form is

$$P_{\mathrm{LOS}}(d) = \min\left(\frac{d_0}{\max(d,1)}, 1\right)\left(1 - e^{-d/d_1}\right) + e^{-d/d_1},$$

where the constants d_0, d_1 depend on frequency band and BS type (macro/micro/pico), with more restrictive values at Sub-THz bands.

2) Path-loss: The simulator uses a 3GPP-inspired baseline path-loss with separate LOS and NLOS expressions, augmented with atmospheric absorption and height correction.

a) LOS path-loss:

$$PL_{\mathrm{LOS}}(d, f_c) = 32.4 + 21\log_{10}(d) + 20\log_{10}(f_c),$$

where d is in meters adn f_c is in GHz.

b) NLOS path-loss:

$$PL_{\mathrm{NLOS}}(d, f_c) = PL_{\mathrm{LOS}}(d, f_c) + \Delta_{\mathrm{NLOS}}(f_c),$$
$$\Delta_{\mathrm{NLOS}}(f_c) = 25 + 0.6\,(f_c - 24).$$

.

3) Shadow fading (large-scale): Spatially correlated log-normal shadow fading is applied to model large-scale environmental variations. The standard deviation σ is band-dependent:

- $\sigma \approx 4$ dB for Sub-6 and mmWave
- $\sigma \approx 8$ dB for Sub-THz

The sampled shadowing $X_\sigma \sim \mathcal{N}(0, \sigma^2)$ is clipped to a practical interval and added to the received power.

4) Small-scale fading: Small-scale fading selection depends on the LOS state.

a) Rician fading (LOS): The Rician K-factor is distance-adaptive:

$$K_{\mathrm{dB}} = \max\big(8,\ 20 - d/50\big), \qquad K = 10^{K_{\mathrm{dB}}/10}.$$

The complex channel gain is expressed as

$$h = \sqrt{\frac{K}{K+1}} + \sqrt{\frac{1}{K+1}}(X + jY),$$

and the resulting fading is expressed in dB and clipped to a practical range.

b) Rayleigh fading (NLOS): For NLOS links,

$$h = X + jY, \quad X, Y \sim \mathcal{N}(0,1),$$

and the magnitude is converted to dB and clipped to a practical interval.

5) Beamforming gain and received power: A 3-D directional beamforming gain G_{bf} is applied based on the azimuth and elevation alignment between the BS and UE. The received power is calculated as

$$P_r = P_t + G_{\mathrm{bf}} - PL_{\mathrm{total}} + X_\sigma + F_{\mathrm{small}},$$

where

- P_t = BS transmit power
- X_σ = shadow fading, and
- F_{small} = small-scale fading in dB.

6) Interference and SINR: Interference is computed by summing the linear received powers from all simultaneously active neighboring BSs operating on interfering frequencies. The SINR is given by

$$\mathrm{SINR} = 10\log_{10}\left(\frac{10^{P_r/10}}{\sum_{\ell\neq 0} 10^{P_{r,\ell}/10} + N_0}\right),$$

where N_0 = thermal noise power.

IV. CMAB Problem Formulation

The problem is formulated as a Contextual Multi-Armed Bandit (CMAB) instead of a deterministic or static optimization framework because the wireless environment is highly dynamic, uncertain, and time-varying. The achievable throughput, interference, and handover impact are not known a priori and change continuously due to user mobility, fading, and varying network load. CMAB enables online learning directly from real-time observations while exploiting contextual information, without requiring an accurate analytical model of the environment. This makes the framework well suited for adaptive user association and handover management in dynamic 6G heterogeneous networks.

At every time step, each user equipment (UE) observes the wireless environment, selects one serving base station (BS), and receives a scalar reward that reflects both throughput performance and mobility cost.

A. Arms

Let the set of candidate BSs available to a UE at time t be

$$\mathcal{A}_t = \{1, 2, \ldots, N_t\},$$

where each arm $a \in \mathcal{A}_t$ represents a possible serving BS. Selecting an arm corresponds to associating the UE with that BS for the current time step.

B. Context

For each UE–BS pair (t, a), a context vector

$$\mathbf{x}_{t,a} \in \mathbb{R}^d$$

is observed, containing real-time radio and network state information such as:

- instantaneous throughput,
- SINR,
- BS load,
- distance to BS,
- fading and shadowing conditions.

The context captures the time-varying nature of the wireless environment and guides the learning process.

C. Reward Variables

The reward at each time step is defined using the following observable variables:

$$\text{Rewards: } \{\text{Throughput},\ \text{Handover (Boolean)},\ \text{Previous Throughput}\}.$$

The base reward is computed from the current achieved throughput as

$$R_{\text{base}} = \frac{\text{Throughput}}{1000}.$$

A fixed penalty is assigned for any handover event:

$$H_o = 0.2.$$

The relative improvement in throughput is defined as

$$\Delta = \frac{\text{Throughput} - \text{Previous Throughput}}{1000}.$$

D. Final Reward Function

The final reward used by the CMAB agent is constructed as

$$R_t = \begin{cases} R_{\text{base}} - H_o, & \text{if } \Delta < 0.05, \\ R_{\text{base}}, & \text{otherwise.} \end{cases}$$

This formulation explicitly penalizes unnecessary handovers that do not provide a sufficient throughput improvement, thereby discouraging frequent or low-gain handovers while still promoting high data-rate links.

E. Learning Objective

The CMAB agent aims to learn an optimal association policy that maximizes the expected long-term cumulative reward:

$$\max\ \mathbb{E}\left[\sum_{t=1}^{T} R_t\right].$$

This objective jointly optimizes:

- long-term throughput,
- mobility stability,
- handover efficiency,
- reduction of unnecessary handovers.

V. Problem Solution Using Bandit-Based Learning

The CMAB-formulated user association and handover problem is solved using two learning-based approaches. First, a classical Upper Confidence Bound (UCB) algorithm is applied as a non-contextual baseline. Then, a Linear UCB (LinUCB) approach is employed as a context-aware solution that exploits real-time network information for improved decision-making.

A. Upper Confidence Bound (UCB) Based Solution

In the UCB approach, each candidate base station (BS) is treated as an arm with an unknown reward distribution. The UE does not use explicit contextual information and relies solely on historical reward observations to guide its association decisions.

Let

- $\bar{R}_a(t)$ = empirical mean reward of arm a upto time t,
- $n_a(t)$ = the number of times arm a has been selected

At each decision time t, the UE selects the arm according to

$$a_t = \arg\max_{a \in \mathcal{A}_t} \left[\bar{R}_a(t) + \sqrt{\frac{2 \ln t}{n_a(t)}} \right].$$

The first term promotes exploitation of the best-performing BS based on past rewards, while the second term enforces exploration of less frequently selected BSs. After selecting a BS and observing the reward R_t, the empirical mean and selection count are updated accordingly.

Algorithm 1 UCB-Based UE Association

Initialize $\bar{R}_a = 0$, $n_a = 0$ for all arms a
for $t = 1$ to T **do**
  for each arm $a \in \mathcal{A}_t$ **do**
    if $n_a = 0$ **then**
      Select arm a once
    else
      Compute $U_a(t) = \bar{R}_a + \sqrt{\frac{2 \ln t}{n_a}}$
    end if
  end for
  Select $a_t = \arg\max_a U_a(t)$
  Associate UE with BS a_t
  Observe reward R_t
  Update:

$$n_{a_t} \leftarrow n_{a_t} + 1, \quad \bar{R}_{a_t} \leftarrow \bar{R}_{a_t} + \frac{R_t - \bar{R}_{a_t}}{n_{a_t}}$$

end for

B. LinUCB as a Context-Aware Solution

To incorporate instantaneous network conditions into the learning process, a context-aware LinUCB framework is adopted. In this approach, each UE observes a feature vector $\mathbf{x}_{t,a}$ for every UE–BS pair, which captures real-time radio and network conditions.

The expected reward is assumed to follow a linear model:

$$\mathbb{E}[R_t \mid \mathbf{x}_{t,a}] = \mathbf{x}_{t,a}^\top \boldsymbol{\theta}_a,$$

where $\boldsymbol{\theta}_a$ is an unknown parameter vector associated with BS a.

At each time step, the UE selects the BS that maximizes the following upper confidence bound:

$$a_t = \arg\max_{a \in \mathcal{A}_t} \left(\mathbf{x}_{t,a}^\top \hat{\boldsymbol{\theta}}_a + \alpha \sqrt{\mathbf{x}_{t,a}^\top \mathbf{A}_a^{-1} \mathbf{x}_{t,a}} \right).$$

After observing the reward R_t, model parameters are updated as:

$$\mathbf{A}_{a_t} \leftarrow \mathbf{A}_{a_t} + \mathbf{x}_{t,a_t}\mathbf{x}_{t,a_t}^\top,$$
$$\mathbf{b}_{a_t} \leftarrow \mathbf{b}_{a_t} + R_t\mathbf{x}_{t,a_t},$$
$$\hat{\boldsymbol{\theta}}_{a_t} \leftarrow \mathbf{A}_{a_t}^{-1}\mathbf{b}_{a_t}.$$

Algorithm 2 Context-Aware LinUCB-Based UE Association

Initialize for each arm a:

$$\mathbf{A}_a = \mathbf{I}_d, \quad \mathbf{b}_a = \mathbf{0}_d$$

for $t = 1$ to T **do**
  for each arm $a \in \mathcal{A}_t$ **do**
    Observe context $\mathbf{x}_{t,a}$
    Compute:

$$\hat{\boldsymbol{\theta}}_a = \mathbf{A}_a^{-1}\mathbf{b}_a$$
$$p_a(t) = \mathbf{x}_{t,a}^\top\hat{\boldsymbol{\theta}}_a + \alpha\sqrt{\mathbf{x}_{t,a}^\top\mathbf{A}_a^{-1}\mathbf{x}_{t,a}}$$

  end for
  Select $a_t = \arg\max_a p_a(t)$
  Associate UE with BS a_t
  Observe reward R_t
  Update:

$$\mathbf{A}_{a_t} \leftarrow \mathbf{A}_{a_t} + \mathbf{x}_{t,a_t}\mathbf{x}_{t,a_t}^\top$$
$$\mathbf{b}_{a_t} \leftarrow \mathbf{b}_{a_t} + R_t\mathbf{x}_{t,a_t}$$

end for

VI. Experiment Parameters

A. Network Topology

The simulation is conducted over a 3000×3000 m^2 area, consisting of a three-tier heterogeneous network with 4 macro, 8 micro, and 12 pico base stations, totaling 24 BSs. The network serves 300 user equipments (UEs) positioned at a height of 1.5 m.

TABLE I
Network Topology Parameters

Parameter	Value
Simulation Area	3000×3000 m^2
Macro BSs	4
Micro BSs	8
Pico BSs	12
Total BSs	24
Number of UEs	300
UE Height	1.5 m

B. Base Station Parameters

Each BS tier has specific technical features to reflect its role in the heterogeneous architecture. Macro BSs provide wide-area coverage, micro BSs enhance capacity in medium-density areas, and pico BSs target high-density hotspots.

TABLE II
BASE STATION TECHNICAL FEATURES BY TIER

BS Tier	Height (m)	Antennas	Max Beams	BF Gain (dB)
Macro	30	128	256	25
Micro	15	256	128	30
Pico	8	512	64	35
UE Antennas	1.5	4 / 8 / 16	–	–

VII. RESULTS AND ANALYSIS

This section presents the performance evaluation of the LinUCB and UCB algorithms based on extensive simulation experiments. The analysis focuses on key performance metrics including signal quality, throughput, handover efficiency, and user experience at cell edges.

A. Signal Quality Performance

The signal quality analysis reveals significant differences between the two algorithms. LinUCB achieved an average SINR of 27.41 dB, outperforming UCB which recorded 24.68 dB—representing an 11.1% improvement. This enhancement in signal quality is further evidenced by the 5th percentile SINR values, where LinUCB maintained -6.1 dB compared to UCB's -9.2 dB, indicating superior performance even in challenging signal conditions.

The improved SINR performance of LinUCB can be attributed to its contextual learning approach, which enables more intelligent cell selection decisions by considering user-specific features and environmental conditions. This contextual awareness allows the algorithm to better predict optimal serving cells, resulting in consistently higher signal quality across diverse network conditions.

B. Throughput Analysis

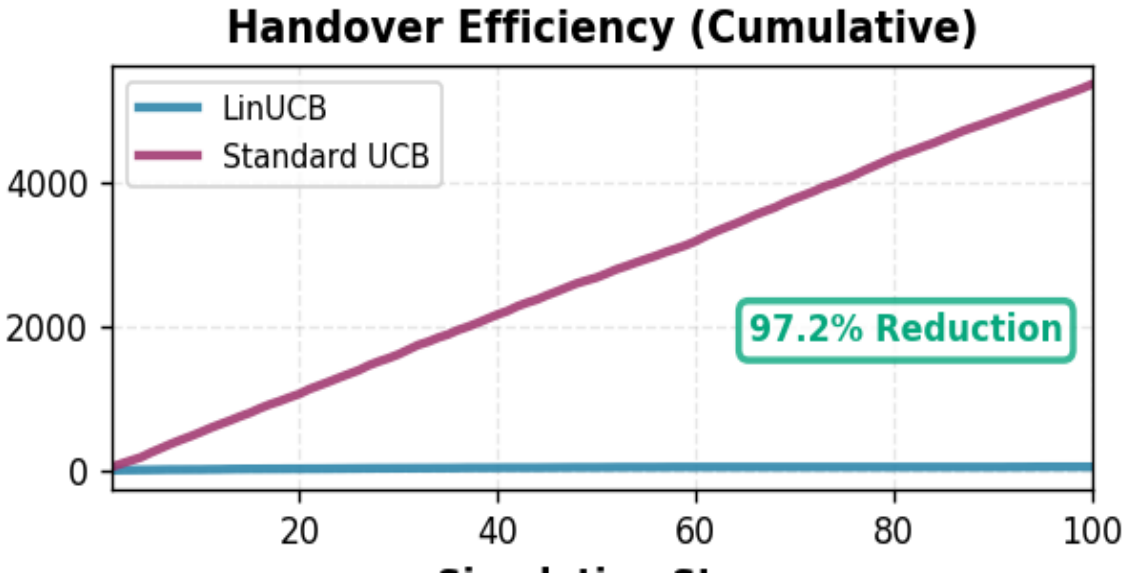

Fig. 1. Handover performance comparison between LinUCB and UCB algorithms. LinUCB demonstrates significantly reduced handover frequency with 153 total handovers compared to UCB's 5,548 handovers, indicating more stable cell selection decisions and reduced signaling overhead.

Throughput performance demonstrates substantial differences between the algorithms across multiple metrics. LinUCB achieved an average throughput of 2,235,433.4 bps, exceeding UCB's 1,980,123.5 bps by approximately 12.9%. The median throughput values further confirm this trend, with LinUCB reaching 4,650.9 kbps compared to UCB's 3,850.2 kbps—a 20.8% improvement.

At the 95th percentile, LinUCB delivered 25,181.9 kbps while UCB achieved 19,890.7 kbps, representing a 26.6% advantage for LinUCB. This indicates that LinUCB not only provides better average performance but also delivers superior throughput for high-demand users. The consistent performance gains across different percentiles suggest that LinUCB's contextual learning mechanism effectively optimizes throughput for diverse user scenarios.

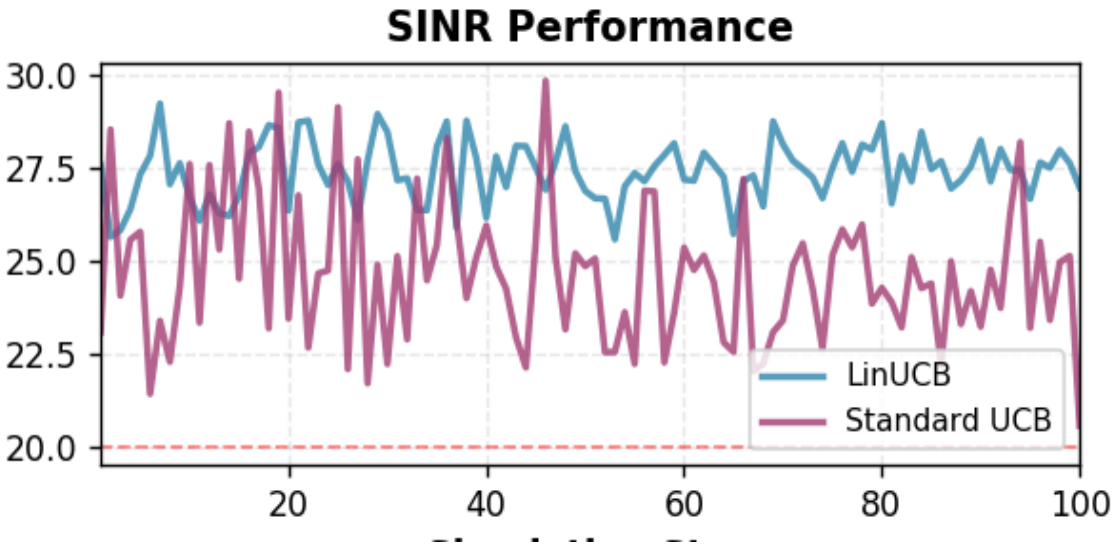

Fig. 2. Signal-to-Interference-plus-Noise Ratio (SINR) distribution for LinUCB and UCB algorithms. LinUCB achieves an average SINR of 27.41 dB compared to UCB's 24.68 dB, demonstrating superior signal quality across the network coverage area.

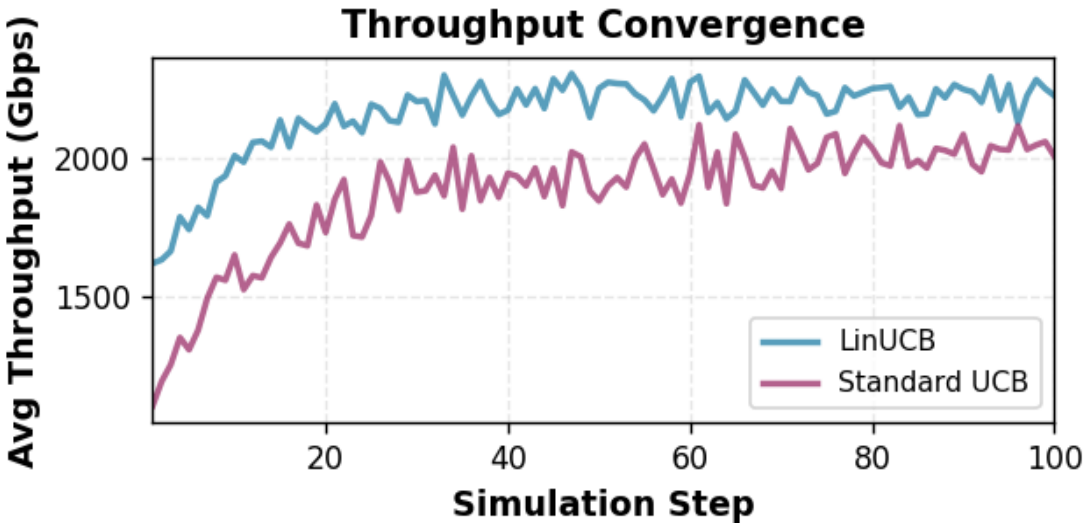

Fig. 3. Throughput performance comparison showing cumulative distribution of user data rates. LinUCB consistently outperforms UCB across all percentiles, with median throughput of 4,650.9 kbps versus 3,850.2 kbps for UCB, and demonstrates particularly strong performance for high-demand users at the 95th percentile.

C. Handover Performance

The handover behavior exhibits a striking contrast between the two algorithms. LinUCB executed a total of 153 handovers throughout the simulation period, while UCB performed 5,548 handovers—representing a 97.2% reduction in handover frequency for LinUCB. This dramatic difference demonstrates LinUCB's ability to make more stable cell selection decisions, reducing unnecessary handovers that can degrade user experience and increase signaling overhead.

The ping-pong handover rate, which measures the frequency of repetitive handovers between the same cells, was 16.2% for LinUCB compared to 20.5% for UCB. While LinUCB shows improvement in this metric, the relatively modest difference suggests that when handovers do occur, both algorithms face similar challenges in certain edge-case scenarios. However, the substantially lower absolute number of handovers in LinUCB means that the overall impact of ping-pong effects is significantly reduced.

D. Cell Edge User Performance

Cell edge performance is a critical indicator of network quality, as users at cell boundaries typically experience the poorest conditions. LinUCB achieved a cell edge rate of 1,879.7 kbps, outperforming UCB's 1,650.4 kbps by 13.9%. This improvement is particularly significant as it demonstrates LinUCB's capability to enhance service quality for the most vulnerable users in the network.

The superior cell edge performance can be attributed to LinUCB's contextual learning approach, which considers user location and channel conditions to make optimal handover decisions. By reducing unnecessary handovers and maintaining better signal quality, LinUCB ensures that even users at cell edges experience more stable and higher-quality connections.

E. Comparative Summary

Table III summarizes the key performance metrics for both algorithms, providing a comprehensive view of their relative strengths.

TABLE III
PERFORMANCE COMPARISON: LINUCB VS. UCB

Metric	**LinUCB**	**UCB**	**Improvement**
Average SINR (dB)	27.41	24.68	+11.1%
5th Percentile SINR (dB)	-6.1	-9.2	+3.1 dB
Average Throughput (bps)	2,235,433.4	1,980,123.5	+12.9%
Median Throughput (kbps)	4,650.9	3,850.2	+20.8%
95th Percentile Throughput (kbps)	25,181.9	19,890.7	+26.6%
Total Handovers	153	5,548	-97.2%
Ping-pong Rate (%)	16.2	20.5	-21.0%
Cell Edge Rate (kbps)	1,879.7	1,650.4	+13.9%

VIII. CONCLUSION

The results conclusively demonstrate that LinUCB provides superior performance across all evaluated metrics. The contextual learning capability enables LinUCB to make more informed decisions, resulting in better signal quality, higher throughput, dramatically reduced handover frequency, and improved cell edge performance. These improvements translate to enhanced user experience and more efficient network operation.

REFERENCES

[1] V. Yajnanarayana, H. Rydén, L. Hévizi, A. Jauhari, and M. Cirkic, "5G handover using reinforcement learning," *arXiv preprint arXiv:1904.02572v2*, 2019.

[2] A. K. Machumilane, A. Gotta, and P. Cassarà, "Handover and SINR-aware path optimization in 5G-UAV mmWave communication using DRL," *arXiv preprint arXiv:2504.02688*, 2025.

[3] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[4] T. S. Rappaport, R. W. Heath, and R. C. Daniels, "An Introduction to Millimeter-Wave Broadband Systems," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 40–50, 2015.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[6] F. J. Martín-Vega, S. Lagen, L. Giupponi, and D. López-Pérez, "System Model for Average Downlink SINR in 5G Multi-Beam Networks," *IEEE Access*, vol. 8, pp. 21481–21495, 2020.

[7] M. A. Uusitalo et al., "5G Network Coverage Planning and Analysis of the Deployment Challenges," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 40–47, Dec. 2018.

[8] T. S. Rappaport, Y. Xing, G. R. MacCartney Jr., A. F. Molisch, E. Mellios, and J. Zhang, "Millimeter-Wave Distance-Dependent Large-Scale Propagation Measurements and Path Loss Models for Outdoor and Indoor 5G Systems," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 4, pp. 2293–2310, Apr. 2017.

[9] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Rao, and M. Zorzi, "Reinforcement Learning for Handover and Mobility Management in 5G and Beyond Networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1772–1800, 2020.

[10] H. Chen, Y. Ye, and L. Wang, "Mobility Management via Multi-Armed Bandits in Ultra-Dense Networks," *IEEE Transactions on Mobile Computing*, vol. 20, no. 12, pp. 3380–3393, Dec. 2021.

[11] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless Communications and Sensing in 6G: Sub-THz Channel Propagation, Models, and Key Technologies," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 24–30, Nov. 2021.

[12] L. Sun, J. Hou, and T. Shu, "Spatial and Temporal Contextual Multi-Armed Bandit Handovers in Ultra-Dense mmWave Cellular Networks," *IEEE Transactions on Mobile Computing*, vol. 20, no. 12, pp. 3423–3438, Dec. 2020.