A novel Target-DQN model for handover control parameters optimization in ultra-dense cellular networks

by

Farhad EBRAHIMZADEH GONBADI

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT FOR A MASTER'S DEGREE
WITH THESIS IN ELECTRICAL ENGINEERING
M.A.Sc.

MONTREAL, AUGUST 12, 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Zbigniew Dziong, Thesis Supervisor
Department of Electrical Engineering, École de technologie supérieure

Mr. Kim Khoa Nguyen, President of the Board of Examiners
Department of Electrical Engineering, École de technologie supérieure

Mr. Michel Kadoch, Member of the jury
Department of Electrical Engineering, École de technologie supérieure

THIS THESIS WAS PRENSENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND PUBLIC

ON AUGUST 05, 2025

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGMENTS

First and foremost, I would like to express my heartfelt gratitude to my beloved wife for her endless support, patience, and encouragement throughout this journey. Her understanding and motivation have been my greatest source of strength. I am also deeply thankful to my supervisor, Prof. Zbigniew Dziong, for their invaluable guidance, support, and encouragement throughout the course of this program. Their expertise and insightful feedback have been instrumental in shaping both the technical depth and academic quality of this work. I would also like to thank the faculty and staff of the Department of Electrical Engineering, whose support and assistance have made this process significantly joyful.

Un nouveau modèle Target-DQN pour l'optimisation des paramètres de contrôle de transfert dans les réseaux cellulaires ultra-denses

Farhad EBRAHIMZADEH GONBADI

RÉSUMÉ

La gestion efficace de la mobilité dans les réseaux 5G ultra-denses demeure un défi majeur, notamment dans les environnements urbains à forte densité d'utilisateurs et aux conditions dynamiques. Cette thèse présente un cadre basé sur le Deep Q-Network (DQN) pour l'optimisation des paramètres de contrôle de transfert (HCP), notamment le temps de déclenchement (TTT), l'hystérésis (Hys) et l'A3Offset. Le modèle Target-DQN (TDQN) proposé intègre un rayon de recherche (RR) ajustable dans sa définition d'action afin de contrôler la granularité de l'espace d'action, permettant ainsi une optimisation multi-paramètre évolutive. Une représentation d'état personnalisée capture les conditions clés de l'utilisateur et du réseau, au moyen de la probabilité de panne, du profil de mobilité de l'utilisateur et de la charge cellulaire. Le modèle est entraîné par une approche d'apprentissage par renforcement qui exploite les mesures utilisateur en temps réel et les états du réseau pour améliorer les performances de transfert. Des simulations approfondies sous différents paramètres SR, combinaisons HCP et profils de mobilité utilisateur démontrent que le modèle proposé réduit significativement les transferts inutiles et les événements ping-pong, jusqu'à 53 % et 98 % respectivement, tout en améliorant le débit utilisateur moyen de 10 % et la probabilité de panne de 12 % par rapport à un benchmark de bonnes pratiques recommandé par les documents techniques de HUAWEI. L'étude examine également les compromis entre complexité de calcul et performances, soulignant l'adaptabilité du modèle à différentes granularités d'optimisation et vitesses utilisateur. Ces résultats suggèrent que le cadre TDQN proposé est une solution prometteuse pour l'optimisation dynamique et contextuelle de la mobilité dans les réseaux cellulaires de nouvelle génération.

Mots-clés : probabilité de couverture, deep Q-networks, réseau cible, DRL, optimisation du transfert, paramètres de contrôle du transfert, gestion de la mobilité, réseaux cellulaires ultra-denses, 5G/B5G

A novel Target-DQN model for handover control parameters optimization in ultra-dense cellular networks

Farhad EBRAHIMZADEH GONBADI

ABSTRACT

Efficient mobility management in ultra-dense 5G networks remains a critical challenge, especially in urban environments with high user densities and dynamic conditions. This thesis presents a Deep Q-Network (DQN)-based framework for optimizing handover control parameters (HCPs), including Time-to-Trigger (TTT), Hysteresis (Hys), and A3Offset. The proposed Target-DQN (TDQN) model incorporates a tunable Search Radius (SR) in its action definition to control the action space granularity, enabling scalable multi-parameter optimization. A custom State representation captures key user and network conditions, by means of outage probability, user mobility profile, and cell load. The model is trained using a reinforcement learning approach that leverages real-time user measurements and network states to improve handover performance. Extensive simulations under various SR settings, HCP combinations, and user mobility profiles demonstrate that the proposed model significantly reduces unnecessary handovers and ping-pong events up to 53% and 98% respectively while improving average user throughput by 10% and outage probability by 12% compared to a best practice benchmark recommended by HUAWEI technical documents. The study also investigates the trade-offs between computational complexity and performance, highlighting the model's adaptability across different optimization granularities and user speeds. These findings suggest that the proposed TDQN framework is a promising solution for dynamic and context-aware mobility optimization in next-generation cellular networks.

Keywords: coverage probability, deep Q-networks, target network, DRL, handover optimization, handover control parameters, mobility management, ultra-dense cellular networks, 5G/B5G

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABREVIATIONS

AI	Artificial Intelligence
BS	Base Station
CDR	Call Drop Ratio
CIO	Cell Individual Offset
CRAN	Cloud Radio Access Networks
DNN	Deep Neural Network
DQN	Deep Q-learning Network
DRL	Deep Reinforcement Learning
FLC	Fuzzy Logic Control
GBR	Guaranteed Bit Rate
HCP	Handover Control Parameter
HO	Handover
HOI	Handover Instance
HOM	Handover Margin
HOR	Handover Rate
Hys	Hysteresis
ISD	Inter-site Distance
KPI	Key Performance Indicator
LA	Location Update
MCS	Modulation and coding scheme
MG	Measurement Gap
MGRP	Measurement Gap Repeat Period
ML	Machine Learning
MM	Mobility Management
MRO	Mobility Robustness Optimization
NFV	Network Function Virtualization
PCI	Physical Cell ID
PPHO	Ping-pong Handover
PPP	Poisson Point Process

PRB	Physical Resource Block
QoS	Quality of Service
RL	Reinforcement Learning
RLF	Radio Link Failure
RSRP	Reference Signal Received Power
SARSA	State-Action-Reward-State-Action
SC	Small Cell
SDN	Software Defined Network
SR	Search Radius
TL	Transfer Learning
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TTT	Time To Trigger
UDN	Ultra-Dense Network
5G	Fifth Generation

LIST OF SYMBOLS

$\mathrm{A1}_{\mathrm{rsrp}}$	$rsrp$ threshold for triggering event $A1$
$A2_{rsrp}$	$rsrp$ threshold for triggering event $A2$
A_x	Maximum Index in value vector of parameter x
$\mathbb{C}$	Number of Base Stations
CHO	Cost of a Handover Instance
$indx_i(t)$	Action vector of indexes at time t
$indx_i^x(t)$	Value index of parameter x at time t
$\mathbb{M}$	Number of stationary users
MS_n	Measured $rsrp$ of the neighbor Cell
MS_p	Measured $rsrp$ of the serving Cell
$\mathbb{N}$	Number of Mobile users
N_o	Thermal Noise power per Hz
Oc_n	Neighbor Cell Individual Offset
Oc_p	Serving Cell Individual Offset
P_j	Output Power of BS j
P_{ij}^{outage}	Outage Probability of user i served by BS j
r_{ij}	User i's data rate served by BS j
$rsrp_i^j$	User i's $rsrp$ from Serving cell j
$S_i(t)$	State of the user i at time t
$s_i^{dir}(t)$	User i's direction relative to the serving cell
$s_i^{rsp}(t)$	User i's serving cell $rsrp$ distance from the $A2_{rsrp\_thr}$ at time t
$s_i^{load}(t)$	User i's serving cell $Load$ at time: t
$s_i^v(t)$	User i's velocity (mps) at time t
$\bar{S}_i^{hcp}(t)$	user i's HCPs vector at time t
$step_i^x(t)$	Change step of parameter x at time t
γ_i^{min}	User i's minimum required Throughput ($mbps$)
β	HO Failure penalty

Δ	Hysteresis
Ω	A3Off
δ	Handover Trigger Delay
θ	Single PRB Bandwidth
Θ_{ij}	Allocated Bandwidth to user i by BS j
ψ_{ij}	Channel power gain between BS j and user i
ϕ_{ij}	Large Scale Fading
φ_{ij}	Small Scale Fading
λ_{ij}	Channel Power gain Expected value

INTRODUCTION

0.1 Motivation

Over the past decade, the explosion in mobile data demand, driven by ubiquitous smart devices and emerging applications like autonomous driving, real-time video streaming, and smart city infrastructure, has revolutionized the architecture of cellular networks. To address the increasing need for higher capacity, lower latency, and seamless connectivity, the deployment of ultra-dense networks (UDNs) has become a defining feature of 5G and beyond (B5G) systems. UDNs consist of a large number of closely spaced small cells operating at high frequencies, which enhance spectral efficiency and capacity. However, these advantages come at a cost: the mobility management challenge becomes significantly more complex, with mobile users encountering frequent handover events. This is due to low dwelling time in cell coverage areas caused by the shrinked coverage radius of the cells as well as the user mobility speeds that can degrade the overall quality of service (QoS). A robust and intelligent handover control system is, therefore, necessary to support high-mobility users in dense network environments.

0.2 Problematic

Traditional handover procedures rely on statically configured parameters—such as Time-to-Trigger (TTT), Hysteresis (Hys), and A3 offset—that were originally designed for macro-cellular networks with limited handover frequency. In UDNs, however, these fixed parameters fail to adapt to the rapidly changing user and network contexts. As a result, users experience problems such as ping-pong handovers (PPHOs), radio link failures (RLFs), and unnecessary handover instances (HOIs), leading to throughput degradation, increased signaling overhead, and poor user experience. These issues are further intensified in high-speed vehicular or pedestrian scenarios where the handover decisions must be both rapid and accurate. Hence, static and rule-based handover strategies are insufficient for meeting the real-time demands of 5G UDN environments.

0.3 Existing Research Gap

While significant progress has been made using heuristic, fuzzy logic, and rule-based algorithms for handover optimization, most existing techniques either lack scalability, context-awareness, or the capacity for joint optimization of multiple handover control parameters (HCPs). Furthermore, many reinforcement learning (RL)-based methods in the literature use simplistic or insufficiently communicative state definition, failing to capture crucial network conditions or provide a context-aware and QoS-aware representation of the user. Another limitation lies in the action space design, where most models optimize only one or two HCPs due to computational limitations and action space size explosion, neglecting the possibility of multi-parameter joint optimization. Resulting in combining multiple parameters as single parameter, leading to inaccurate optimization. Finally, where most of the developed techniques and models focus on LTE networks, an adaptive and scalable model that can be used in a realistic 5G simulation environment is missing.

0.4 Objectives

This thesis aims to propose a Deep Q-Learning framework for optimizing multiple handover control parameters in an ultra-dense 5G networks and beyond. The main objectives are:

- To design a realistic, user- and network-aware state representation.
- To define a scalable, tunable action space for multi-parameter optimization.
- To implement a reward function that
  - Minimize HOI
  - Minimize PPHO
  - Reduce HOF
  - Maximize Average Throughput
  - Maximize Coverage Probability
- To evaluate the model under different mobility profiles, parameter combinations, and Search Radius (SR) values.

0.5 Methodology

The proposed approach is based on a Target Deep Q-Network (TDQN) that learns optimal HCP values through interaction with a simulated 5G NR UDN environment. A custom state space is designed to include user movement direction and speed, load status, and outage-based QoS indication. The action space is defined through a Search Radius (SR) mechanism, enabling flexible control over the number of actions. The reward function is also structured to minimize handover instances and maintain throughput. The model is evaluated using Python-based simulations across various deployment settings.

0.6 Achieved Results and Contributions

The proposed model significantly improves handover decision-making. Simulation results show reductions in unnecessary handovers and ping-pong effects up to 53% and 98% respectively while improving HOF up to 20%, as well as 10% improvements in throughput and OP. The main contributions include:

- A joint state–reward design centered on outage probability.
- A tunable action space enabling multi-HCP optimization.
- A novel RSRP-based formulation for OP estimation.
- A complexity-performance analysis for practical deployment.
- Extensive simulations in a 5G NR UDN scenario.

This thesis is based on a journal paper with the same title that has been submitted to the IEEE Transactions on Vehicular Technology. The Paper content is presented in the Chapter 1. Within the chapter, an introduction is presented in Section 1.1 while the related works are reviewed in Section 1.2. Section 1.3 presents the system model, including preliminary concepts and the adjusted outage probability formulation. The proposed Target-DQN framework, detailing the state space, action space, and reward function design is depicted in Section 1.4. Section 1.5 presents the simulation results and evaluates the model's performance through detailed analysis and discussion, while Section 1.6 concludes the paper by summarizing the key findings and disclosing some potential future paths.

CHAPTER 1

A NOVEL TARGET-DQN MODEL FOR HANDOVER CONTROL PARAMETERS OPTIMIZATION IN ULTRA-DENSE CELLULAR NETWORKS

1.1 Introduction

Over the past decade, the exponential growth in mobile data traffic has catalyzed a paradigm shift in cellular network design, giving rise to ultra-dense networks (UDNs). Unlike traditional macro-cellular systems, UDNs feature aggressive small cell densification operating at higher frequency bands. Densification has thus become a cornerstone of 5G and beyond (B5G) architectures, where the demand for high-speed connectivity, low latency, and reliable user experience continues to grow. While ultra-dense network (UDN) deployments enhance capacity and spectral efficiency, they also lead to more frequent handovers (HOs), resulting in increased signaling overhead, service interruptions, and a higher likelihood of radio link failures (RLFs). The latency involved in each handover—from initiation to execution—can be substantial, particularly in high-mobility scenarios. Moreover, UDNs introduce significant operational challenges in mobility management (MM), which is responsible for ensuring continuous user connectivity across densely deployed cells. Conventional HO mechanisms, which rely on fixed parameters such as TTT, handover margin (HOM), and cell individual offset (CIO), are typically designed under quasi-static conditions and lack the adaptability required to respond to dynamic user mobility and fluctuating network load (David Lopez-Perez 2012).

In contrast to macro networks where users traverse few cells while connected to the network, the reduced coverage area in UDNs causes frequent handovers. Additionally, in 5G New Radio (NR), relocating control-plane processing to the network edge (i.e., base stations) reduces latency but increases signaling overhead during HO due to frequent UE context transfers. In ultra-dense deployments, excessive signaling can strain base station CPUs and degrade spectrum efficiency by consuming valuable air interface resources (Syed Muhammad Asad Zaidi 2020). Interference management also becomes more complex, as neighboring cells often operate on overlapping frequencies, reducing modulation and coding scheme (MCS) indices and degrading user throughput (TS38.214 2025).

To address these challenges, recent research has focused on AI-driven solutions. The massive data generated in UDNs presents new opportunities for intelligent mobility management using machine learning (ML). Conventional heuristic and optimization-based algorithms often fall short in scalability or responsiveness, particularly in dynamic 5G environments. For example, mobility robustness optimization (MRO) techniques, based on RSRP, velocity, weight functions, and fuzzy logic, lack adaptability and require extensive manual tuning (Waheeb Tashan, et al 2022). In contrast, ML-based models can autonomously learn optimal handover configurations, providing predictive and data-driven optimization capabilities without explicit programming (Yaohua Sun, et al 2019), (Mingzhe Chen, et al 2019).

Among ML techniques, reinforcement learning (RL) and its deep variants (DRL) have emerged as promising approaches for mobility management due to their interactive decision-making framework. Unlike supervised or unsupervised learning, RL agents learn optimal policies by directly interacting with the environment through state-action-reward dynamics. However, the success of RL in UDNs depends heavily on how the environment is modeled, especially the definition of states, actions, and reward functions (Waheeb Tashan, et al 2022). Poorly defined states may fail to capture user behavior and network dynamics, while basic action definitions often restrict the number of HCPs that can be jointly optimized. Commonly used HCPs include TTT, Hys, A3 offset, and CIO. A naive expansion of the action space by considering multiple parameters can lead to combinatorial explosion, making learning inefficient. As these parameters directly influence HO performance, improper configurations can result in well-known issues such as ping-pong HOs, late HOs, or HO failures (David Lopez-Perez 2012), (Wasan Kadhim Saad, et al 2021).

To address the limitations in existing ML-based models, we propose a deep Q-learning framework that employs a carefully designed state–action–reward mechanism. The model aims to represent a more realistic user state by integrating multiple contextual factors, including outage probability, movement direction and speed as well as network load. A scalable action space is defined in alignment with the state space, enabling the optimization of multiple HCPs simultaneously with a configurable degree of freedom. The reward function is crafted to minimize unnecessary handover instances (HOIs), reduce ping-pong handovers (PPHOs), and improve overall handover performance and maintain user throughput.

Additionally, the concept of outage probability is incorporated into the reward structure to ensure quality of service (QoS) guarantees for users with a specified service requirement.
The contributions of this paper can be classified as follows:

- **Outage-Aware Joint State-Reward Design:** A unified state and reward formulation centered on outage probability is introduced, enabling a user-centric, real-time representation of network dynamics. It allows the model to distinguish between users under similar signal strength or BS assignment, improving HO decisions by capturing both signal quality and cell load with respect to service requirement of the user, impacting QoS.
- **Scalable and Flexible Action Space with Tunable Parameter Selection:** The proposed model introduces a scalable action space design controlled by a tunable Search Radius, enabling fine-grained selection of handover control parameter values. This structure supports multi-parameter optimization while providing a balance between performance and computational complexity. Moreover, the model demonstrates robust adaptability when optimizing different subsets of HCPs (e.g., TTT and Hys) and under varying user mobility profiles, confirming its flexibility across diverse network conditions and deployment scenarios.
- **RSRP-based Outage Probability derivation:** A novel outage probability formulation is derived using RSRP values and PRB allocation, enabling a realistic and analytically tractable estimate of QoS satisfaction. This allows seamless integration of outage dynamics into the learning framework for more effective handover parameter optimization
- **Complexity-Performance Analysis of TDQN:** The complexity analysis is presented for both inference and training phases, showing how the model scales with the number of tuned HCPs and SR values, which provides insights into feasible configurations for practical deployment.
- **Extensive Simulation Evaluation in 5G NR UDN Environment:** A comprehensive set of simulations using a Python-based 5G NR UDN simulator demonstrates the effectiveness of the TDQN model under various scenarios.

The remainder of this paper is structured as follows. Section 1.2 reviews related work. Section 1.3 presents the system model, including preliminary concepts and the adjusted outage probability formulation. Section 1.4 describes the proposed Target-DQN framework, detailing the state space, action space, and reward function design. Section 1.5 evaluates the model's performance through simulation results and discussions. Section 1.6 concludes the paper. Prior to these sections, the list of acronyms and notations used throughout the paper is provided in the List of Abbreviations and the List of Symbols included at the beginning of this thesis.

1.2 Related works

Significant research has been conducted in the handover optimization domain using conventional MRO techniques and ML-based technique, as summarized in (Syed Muhammad Asad Zaidi 2020) (Waheeb Tashan, et al 2022) (Yaohua Sun, et al 2019) (Mingzhe Chen, et al 2019) (Waheeb Tashan, et al 2022). Among ML approaches, all three learning paradigms, supervised, unsupervised, and reinforcement learning (RL), have been explored across various deployment scenarios, HCPs, and key performance indicators (KPIs). RL models have shown particular promise for HO optimization due to their adaptability and online learning capabilities. For instance, (Rana D. Hegazy 2018) addresses two major handover challenges in LTE networks: RLF and ping-pong HO (PPHO). To tackle these issues, the authors classify users into four distinct groups based on their mobility speed and traffic load. A fuzzy Q-learning algorithm is then employed for each category to dynamically tune handover parameters, particularly the handover margin (HOM).

In (Jin Wu 2015) the authors propose a dynamic fuzzy Q-learning algorithm that begins without predefined fuzzy rules and incrementally develops them through continuous system learning. This article uses call drop ratio (CDR) and handover rate (HOR) to define system state as well as reward function to dynamically adjust HOM by means of Q-learning and FLC models. The model’s performance was tested for different TTT and UE speeds which showed improved network performance and user satisfaction for a dense small cell environment.

The proposed methodology in (Tanu Goyal 2019) operates in two main phases: first, AHP and TOPSIS (a type of multi-objective decision making methods) are employed to select the optimal target cell for handover by evaluating multiple criteria such as signal strength, load

conditions, and user direction and location to rank the UEs. Subsequently, Q-learning is utilized to dynamically adjust handover control parameters, specifically the TTT and HOM, based on the varying speeds of user equipment. These adaptive tunings aim to minimize handover failures and reduce the ping-pong effect.

Authors in (Raja Karmakar 2023) used a Kalman filter to predict the RSRP values for the serving and target cells that would be used in a state-action-reward-state-action (SARSA) RL model to find the Target Cell. The Serving cell was considered as the system state while the outputs of the filter were used to generate the related reward. Then an epsilon greedy policy was deployed to tune the HCPs (TTT, Hys). The proposed model improved throughput and packet loss rate. Another widely used model in the domain of RL are deep learning models, such as DQN or deep neural networks (DNN).

(Mubashir Murshed 2024) introduces HMUD-H, a SARSA reinforcement learning model designed to manage handovers in ultra-dense, high-mobility 5G vehicular environments. Frequent handovers in such settings can cause disruptions, including failures and ping-pong effects. HMUD-H uses real-time signal metrics (RSSI, RSRQ, SINR), mobility factors (speed, distance), and network density to make adaptive and intelligent handover decisions. It dynamically adjusts learning rates based on vehicle speeds to improve decision accuracy. Evaluated using simulations in urban and highway scenarios, HMUD-H significantly reduces handover frequency, failure rates, and cumulative handover time compared to two baseline methods. The results demonstrate that HMUD-H provides more stable and efficient connectivity in 5G vehicular networks, especially under varying speed and density conditions.

In (Wei Huang, et al 2022), an SDN- enabled handover optimization strategy is proposed, combining the TOPSIS method with a DQN. In this approach, TOPSIS is used to preselect the optimal target BS based on key metrics such as RSRP, SINR, and traffic load. Once the target BS is selected, the SDN controller employs deep reinforcement learning to automatically fine-tune the TTT and HOM. In the deep model, RSRP and load of all the BSs as well as the TOPSIS output BS index define the state of the system while the action space is all the combinations of TTT and HOM values. The reward function is also a function of HOF and PPHO. The proposed model showed improvement in throughput, HOF and PPHO performance.

Authors in (Kang Tan, et al 2022) has considered a double-deep Q-learning model in which they considered the received RSRP of the neighboring BS and a Hot-vector representing the Serving BS ID as the state of the system. Action space in this model is choosing the target BS among the neighbor BS as well as the serving cell. To define the Reward function, they considered the RSRP gain considering the maximum RSRP received after HO and the RSRP of the Target BS when taking the related action. They also considered a penalty constant for a HO instance. The important point is that they first trained their model offline without any interaction with the environment, using the data generated from some UE trajectories. Then, the trained model was used to find the optimal target BS with no exploration. Their results showed improvement in HO delay time and packet loss compared to baseline event A3-based HO.

Despite significant progress in RL–based handover optimization, several critical gaps remain unaddressed in the existing literature. Most prior works rely on limited state representations that lack a fine-grained and context-aware state representations. Additionally, traditional action spaces are typically restricted to the tuning of only one or two HCPs, as the action space size grows exponentially with the number of HCPs, making multi-parameter optimization computationally prohibitive without careful design. Many existing models also lack a direct connection between handover decisions and user-perceived performance, often ignoring key QoS metrics such as throughput satisfaction. Furthermore, few studies evaluate their solutions under a realistic 5G system model that includes high carrier frequencies and dense base station deployments. These gaps motivate the development of a more flexible, scalable, and context-aware handover optimization framework, as proposed in this work.

1.3 Preliminaries and system model

In this thesis an ultra-dense 5G cellular network composed of $\mathbb{C}$ small cells (SCs) deployed in an urban area, serving $\mathbb{N}$ mobile user equipment (UEs), and $\mathbb{M}$ stationary UEs is considered. a uniform hexagonal grid with a predefined inter-site distance (ISD) is adopted for BSs positioning to represent an idealized scenario with minimal coverage gaps. BS c concurrently serves $\mathbb{N}_c$ mobile and $\mathbb{M}_c$ stationary UEs satisfying the constraints $\sum_{c=1}^{\mathbb{C}} \mathbb{N}_c = \mathbb{N}$; $\sum_{c=1}^{\mathbb{C}} \mathbb{M}_c = \mathbb{M}$.

1.3.1 5G NR handover mechanisms

5G NR handover scheme follows a mechanism similar to that of LTE. According to 3GPP specifications (TS38.300 2025), in intra-frequency handovers for example, the UE continuously performs measurements of neighboring cells at regular intervals defined by the measurement gap repetition period (MGRP). Each measurement instance occurs within a measurement gap (MG) of 6 ms duration. In the simulation framework, with no loss of generality, the scheduling interval aligns with the MGRP interval, meaning resource allocation to UEs is carried out based on the modulation and coding scheme (MCS) index and the current cell load, concurrently with physical layer measurements. These measurements are subsequently processed by the Layer 3 (L3) filtering mechanism, to produce the final filtered signal metric ($F[n]$) used in the handover decision process (David Lopez-Perez 2012) which is formulated as follows:

$$F[n] = (1-\alpha) * F[n-1] + \alpha * 10log_{10}(M[n])$$
$$\alpha = {}^{1}/_{2^{ki/4}} \tag{1.1}$$

where $M[n]$ is the latest received measurement result from physical layer and $F[n]$ is the filtered measurement result used for evaluation and reporting. The L3 filtering is governed by a recursive filter, where the filter coefficient α determines the smoothing level, meaning that how much the previous samples of the measurement are going to affect the current measurement output. α value itself is determined by ki where the default value is 4 resulting in $\alpha = 0.5$ (TS38.331 2025). The previous filtered value $F[n-1]$ is used to compute the current one, with $F[0] = M[1]$ at initialization.

Before diving deeper in the HO mechanism lets first check the common events in mobility management in 5G NR:

- Event $A1$: Serving cell's signal quality becomes better than configured threshold $A1_{rsrp}$ (lower bound that defines when the serving cell's signal quality is considered good).
- Event $A2$: Serving cell's signal quality becomes worse than configured threshold $A2_{rsrp}$ (upper limit below which the serving cell's signal quality is considered bad).

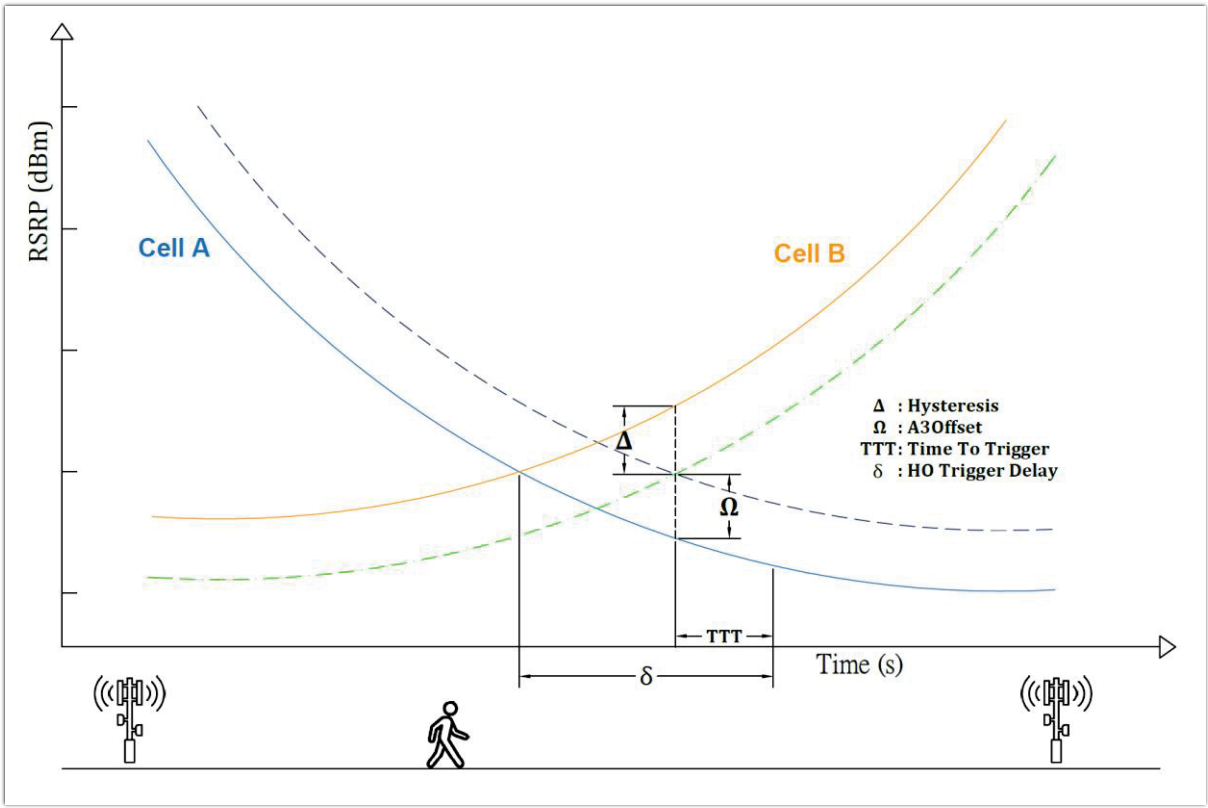

Figure 1.1 Event A3-based handover mechanism with the related HCPs and their effect on the user's perception of the received signals

- Event $A3$: Neighbor cell's signal quality becomes offset better than the serving cell (explained further on).
- Event $A4$: Neighbor cell's signal quality becomes better than a configured threshold (absolute threshold that defines when the neighbor cell's signal quality is considered good).
- Event $A5$: Serving cell's signal quality becomes worse than threshold-1 (below which the signal quality is considered bad) AND neighbor cell's signal quality becomes better than threshold-2 (above which the signal quality is considered good) simultaneously.

The most common handover mechanism in real scenarios is the event $A3$-based handover. The reason is that, in this type of the handover the neighbor cells and serving cell are compared with each other and it makes sure that user is always connected to the best BS. However, event A5-based is also used but the neighbor cells and serving cells are compared to separate thresholds and usually are used for inter-RAT HO. Other event are used within the HO

mechanism to start or stop actions such as measurement. Event $A3$-based handover is visualized in Figure 1.1, along with the main handover control parameters. Δ and Ω are the hysteresis and offset of event $A3$-based HO scheme. Δ is a buffer value added to avoid unnecessary handovers due to small or temporary signal fluctuations between cells while Ω is a threshold used to trigger handover when a neighboring cell's signal becomes better than the serving cell's signal by a certain margin. The entering condition should hold for the whole period of TTT to trigger $A3$ event. δ is the handover trigger delay which is the time difference between the ideal handover time with zero HCP values and the actual handover time, triggered under the HCPs. How they impact the signal strength perception, and the HO timing can be seen clearly. The corresponding entering and leaving conditions of the $A3$-event are defined as follows:

$$\begin{aligned} Enter&: MS_n + Oc_n - Hys > MS_p + Oc_p + A3Off \\ Leave&: MS_n + Oc_n + Hys < MS_p + Oc_p + A3Off \end{aligned} \tag{1.2}$$

where MS_n and MS_p are the measured $RSRP$ of the neighboring and serving cells, respectively. Oc_n and Oc_p are the cell individual offsets (CIO) of the neighbor and serving cell, respectively. These values are used to bias handover decisions, making a cell more or less attractive for handovers in order to improve load balancing. As previously mentioned, value of each parameter can have different impact on the handover performance metrics such as the HOI, PPHO, $RSRP$ and $SINR$ as it will affect the triggering time of the $A3$ event. The impact of different values of the HCPs on different KPIs such as $RSRP$, OP and $Data\ Rate$ is investigated in the simulation results section. However, what TTT does, is either delaying or expediting the HO time. While, $A3Off$ and Hys can alter the actual signal interpretation of the BS (negatively or positively as illustrated in Figure 1.1) and are more influential specially in cell edge area (Yasir Ullah, et al 2024) (Adnan Farooq Bhat 2024). Consequently, improper configuration of these parameters can lead to various handover issues, as summarized in Table 1.1.

Table 1.1 Handover Problems and Parameters Relation

HO Problem	**TTT**	**δ**	Δ
Too late HO	High	High	High
Early HO	Low	Low	Low
Wrong HO	Unsuitable	Unsuitable	Unsuitable

1.3.2 SINR, RSRP and outage probability

Regarding the users' service request, each user acquires a non-GBR (Nonguaranteed bit rate) bearer while trying to reach a minimum throughput of γ_i^{min}. Different service types with minimum requirements can be defined, which are assigned randomly to users during the simulation.

Considering P_j as the output power of BS j while ψ_{ij} is the channel power gain between BS j and UE i, then received power at i from j would be $\psi_{ij}P_j$. The channel gain ψ_{ij} is composed of two multiplicative components $\phi_{ij}\varphi_{ij}$: I) large-scale fading (ϕ_{ij}) including distance related attenuation, lognormal shadowing, antenna gain, and II) small-scale fading (φ_{ij}) meaning Rayleigh fading (Sunil Kandukuri 2002). Accordingly, Signal-to-Interference-plus-Noise Ratio (Goldsmith 2005) for user i connected to BS j ($SINR_{ij}$) is defined as the ratio of the received signal power from j to the combined power of interference from other BSs and the noise that can be mathematically expressed as:

$$SINR_{ij} = \frac{P_j \phi_{ij} \varphi_{ij}}{\Theta_{ij} N_o + \sum_{k \neq j} P_k \phi_{ik} \varphi_{ik}} \tag{1.3}$$

where:

$$\Theta_{ij} = PRB_{ij} * \theta$$

N_o is the thermal noise power per Hz (-174dBm (Goldsmith 2005)), Θ_{ij} is the user i's total bandwidth, which is the number of allocated PRBs by BS j multiplied by single PRB bandwidth (θ). To fulfill the QoS requirement of a user, the system should be able to provide the user with at least the corresponding SINR threshold ($SINR_i^{th}$) or respectively, the minimum

throughput (γ_i^{min}). Based on the Shannon channel capacity formula, there is a direct relation between the maximum achievable throughput and the SINR value:

$$r_{ij}^{max} = \Theta_{ij} log_2(1 + SINR_{ij}) \tag{1.4}$$

where r_{ij} is the achievable data rate of i while been served by j. The logarithmic term represents the spectral efficiency of the user based on the SINR which is used to find the proper MCS index for the users in the simulations. That been said, the $outage\ probability$ for a user i and BS j pair can be defined as:

$$P_{ij}^{outage} = \Pr\left(SINR_{ij} \leq SINR_i^{th}\right)\ or\ \Pr\left(r_{ij} \leq \gamma_i^{min}\right) \tag{1.5}$$

Which can be interpreted as the fraction of time the ij pair experiences an outage due to fading (Sunil Kandukuri 2002) and BS fail to provide the user with the required QoS. Inspired by the approaches in (Sunil Kandukuri 2002) and considering the expected value of channel power gain (ψ_{ij}):

$$E[\psi_{ij}] = E[\phi_{ij}\varphi_{ij}] = \phi_{ij} = \frac{1}{\lambda_{ij}} \tag{1.6}$$

which indicates that the small-scale fading can be averaged out in channel measurement and the large-scale fading is fixed over one channel association. We can calculate the outage probability for an ij pair as:

$$P_{ij}^{outage} = \Pr\left(\Theta_{ij} log_2\left(1 + SINR_{ij}\right) \leq \gamma_i^{min}\right) \tag{1.7}$$

$$= 1 - \Pr\left(\Theta_{ij} log_2\left(1 + SINR_{ij}\right) > \gamma_i^{min}\right)$$

$$= 1 - \Pr\left(1 + SINR_{ij} > 2^{\frac{\gamma_i^{min}}{\Theta_{ij}}}\right)$$

Using the $SINR$ formula and the fact that the fading terms are independent exponentially distributed random variables with mean $1/\lambda_{ij}$:

$$P_{ij}^{outage} = 1 - \Pr\left(P_j \phi_{ij} \varphi_{ij} > \Gamma\left(\sum_{k \neq j} P_k \phi_{ik} \varphi_{ik} + \Theta_{ij} N_o\right)\right) \tag{1.8}$$

$$= 1 - e^{-\frac{\lambda_{ij} \Theta_{ij} N_o}{P_j} \Gamma_{ij}} \prod_{k \neq j} \frac{1}{1 + \frac{P_k \lambda_{ij}}{P_j \lambda_{ik}} \Gamma_{ij}}$$

where:

$$\Gamma_{ij} = 2^{\frac{\gamma_i^{min}}{\Theta_{ij}}} - 1$$

Outage probability is a decreasing function of the bandwidth allocated to the user by the BS. As more bandwidth is allocated, the likelihood that the user achieves its minimum required data rate γ_i^{min} increases, thereby reducing the outage probability. In cellular networks the signal strength is represented as $rsrp$ which is the linear average over the power contributions (in watt) of the resource elements that carry secondary synchronization signals while $SINR$ is the linear average over the power contribution (in watt) of the resource elements carrying secondary synchronisation signals divided by the linear average of the noise and interference power contribution (in watt) (TS38.215 2024). So, without loss of generality, we can represent the $SINR$ as a function of $rsrp$s which is also known as G-factor metric which is in fact the worst case $SINR$ where the received signal from other BS are the only interference sources (Marvin Manalastas, et al 2022). So, we can rewrite the outage probability as a function of $rsrp$s and BW as below:

$$P_{ij}^{outage} = 1 - e^{-\frac{\Theta_{ij} N_o}{rsrp_i^j} \Gamma_{ij}} \prod_{k \neq j} \frac{1}{1 + \frac{rsrp_i^k}{rsrp_i^j} \Gamma_{ij}} \tag{1.9}$$

1.3.3 User throughput

User throughput is a critical performance metric in 5G NR mobile networks, reflecting the effective data rate experienced by an individual user. It is influenced by several factors,

including signal quality, allocated bandwidth, scheduling policies, and network load. The theoretical throughput based on the Shannon-Hartley theorem was already mentioned in (4), which provides an upper bound on achievable throughput, assuming ideal conditions and no overheads. But the realistic value is calculated based on the physical layer resources and MCS index as follows:

$$r_{ij} = prb_{ij} * N_{sc} * N_{Syb} * N_{slot} * Q * R * N_{ch} \quad (1.10)$$

where prb_{ij} is the number of allocated PRBs to user i by BS j that can be affected by the cell load and the scheduling method, N_{sc} is the number of subcarrier per PRB (12 in LTE/5G NR), N_{Syb} represents the number of symbols per slot (14 in 5G NR, 7 in LTE), N_{slot} is the number of slot per second which depends on the numerology (μ) of the 5G NR and can be formulated as $2^{\mu} * 1000$, Q is the modulation order (2: QPSK,…, 8: 256QAM), R is the coding rate and N_{ch} represents the number of MIMO layer (in this paper a single layer MIMO is considered). Moreover, the resulting data rate using (10) assumes ideal conditions, where all available resources are dedicated solely to data transmission, neglecting any signaling overhead. However, in practical deployments, a portion of resources is reserved for control signaling and protocol procedures.

In the context of HO, the same limitation factor persists and is commonly referred to as normalized HO delay or the HO cost (Rabe Arshad, et al 2016), which accounts for the resource consumption and performance degradation associated with mobility events:

$$HOC = \min(1, H_t \times d_h) \quad (1.11)$$

where H_t is the handover rate per unit time and d_h is the service delay in seconds per HO. So, the actual data rate would be:

$$Throughput = r_{ij} * (1 - HOC) \quad (1.12)$$

which indicates that high HOIs or PPHOs can degrade the user throughput due to the overhead and corresponding interruption time.

1.4 Proposed Target-DQN model

The DQN is a reinforcement learning architecture that combines Q-learning with deep neural networks to approximate the optimal action-value function $Q^*(s,a)$. Unlike traditional tabular Q-learning, where a Q-value is stored for each state-action pair, DQN uses a neural network parameterized by weights ω to estimate the Q-values: $Q(s,a;\omega)$. The input to the network is the current state s, and the output is a vector of Q-values for all possible actions a in that state. The agent selects actions using an ϵ-greedy policy to balance exploration and exploitation. The model calculates the target Q-values (y) to train the Q-network as mentioned below:

$$y = R + \xi max_{a'} Q(s', a'; \omega) \tag{1.13}$$

where R is the received reward and ξ is the discount factor while ω are the Q-network parameters. In reinforcement learning, especially in non-stationary environments such as user behavior in mobile networks, a common phenomenon known as catastrophic forgetting poses a significant challenge to maintaining performance over time, where a neural network's parameters are updated globally during training, and new information can overwrite the weights that encoded older knowledge. To deal with this problem, experience reply has been incorporated where experience samples (s,a,r,s') are stored in a buffer and then batches are randomly chosen from the buffer to train the model. This way we are letting the Q-network to take actions in different states before updating the network parameters to stabilize the network. However, the Q-network can still be very noisy. To rectify this, a Target network $(\hat{Q};\hat{\omega})$ can be added where its weights are a copy of the Q-network in the beginning $\hat{\omega} = \omega$. Afterward, it will be used to predict the target Q-values (y) in order to train the main Q-network as follow:

$$y = R + \xi max_{a'} Q(s', a'; \hat{\omega}) \tag{1.14}$$

Eventually, with a frequency of the choice (predefined number of iterations), it will update its parameters from the trained Q-network $(\widehat{\omega'} = \omega')$. The architecture and data flow of the described TDQN is illustrated in Figure 1.2. As it is shown, in each instance of the algorithm, the Q-networks receives the state of the system (s) and output the action $max_a Q(s,a;\omega)$. Then, based on the received reward (r) and the next state (s'), the experience is stored in the

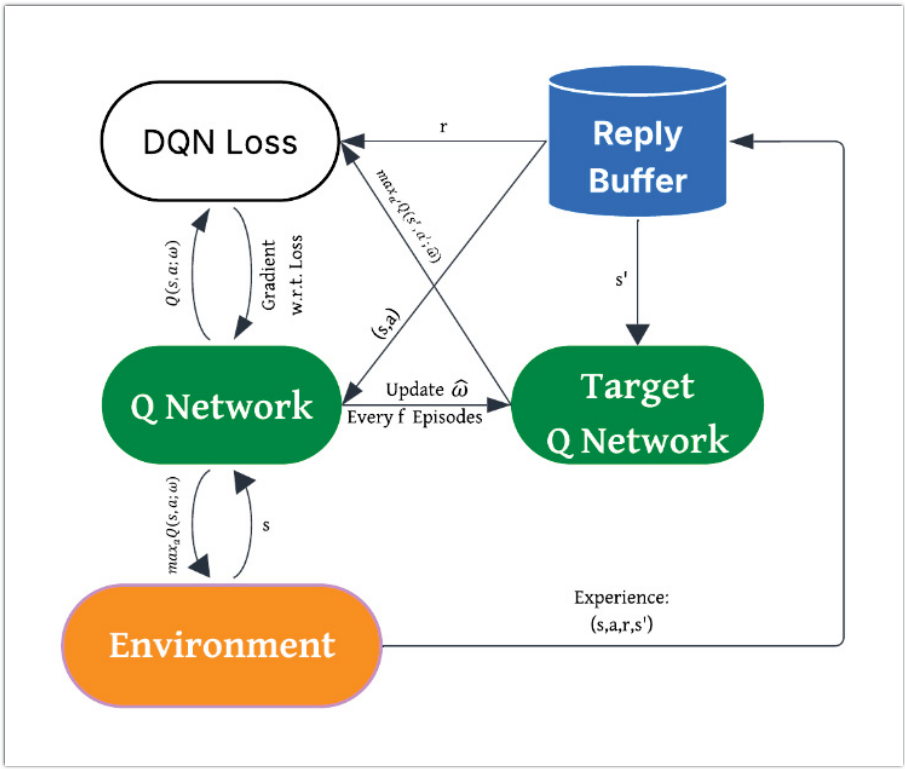

Figure 1.2 Target-DQN with Experience reply buffer

Reply buffer. Finally, after enough experiences are accumulated, a batch is randomly selected from the buffer and based on the target values calculated using Equation (1.14) and recorded reward (r), Q-network weights are optimized and after couple of iterations they are copied to Target network as new weights.

1.4.1 State space

In reinforcement learning, the definition of state space is critical, as it directly influences the agent's ability to distinguish between different user conditions and make optimal decisions in a dynamic environment. The state space serves as the agent's perception of the environment, encapsulating all relevant contextual information required to select appropriate actions. If the state representation is incomplete or overly simplified, it may lead to suboptimal policies and degraded system performance. Radio signal strength/quality ($rsrp$ and its derivation), user mobility patterns (speed, trajectory), network congestion levels, and the HCP values of the system are some of the factors that can be considered.

In the proposed model, $S_i(t)$ is defined as the state of the user i at time t which consists of several elements as below:

$$\begin{cases} S_i(t) = \{P_{ij}^{outage}, s_i^{load}(t), s_i^d(t), s_i^v(t), \bar{S}_i^{hcp}(t)\} \\ \quad \bar{S}_i^{hcp}(t) \in [s_i^{TTT}(t), s_i^{Hys}(t), s_i^{Off}(t)] \end{cases} \quad (1.15)$$

Where:

- P_{ij}^{outage} is the outage probability of user i and BS j pair, representing the likelihood that user i fails to meet its required quality of service (QoS) given the current signal quality and the load of the BS j.
- $s_i^{load}(t) \in [-1,1]$ represent the load factor of the user i at time t. It is defined as the difference between the load of the serving cell and the average load of the neighboring cells where the cell load is calculated based on the minimum required PRB for users' across the BSs. The load of the serving cell is directly affected by the number of served UEs and their minimum required throughput γ_i^{min}.
- $s_i^d(t) \in [-1,1]$ represents the UE's direction relative to its serving cell at time t. Positive values indicate that the user is moving toward the BS with $s_i^d(t) = +1$ denoting direct movement toward the BS. Conversely, negative values indicate movement away from the BS.
- $s_i^v(t)$ is the speed of the UE in "$meter\ per\ second$". In the scope of this paper, constant speeds during each simulation run ($5\ to\ 80\ Kmph$) is considered.
- $\bar{s}_i^{hcp}(t)$ is a vector indicating the handover control parameters' central value at time t. The size of this vector depends on the number of HCPs that are considered in the optimization model. For instance, considering TTT and Hys: $\bar{s}_i^{hcp}(t) = \{s_i^{TTT}(t), s_i^{Hys}(t)\}$.

Instead of using the actual discrete values of the parameters that can be of very different scales and metrics (ms, dB), which can create dominant features, we will use the corresponding index in the parameter vectors as the normalization of HCP values. For example, $s_i^{TTT}(t) = 2$ corresponds to $TTT = 40ms$ while $s_i^{Hys}(t) = 2$ corresponds to $Hys = 1dB$. These values are further normalized by dividing them by the maximum number of values of the corresponding

parameter. As it is clear, the size of the state space can grow with the number of HCPs as well as the size of the value vector of each parameter. But, based on the action mechanism defined, not all the states are reachable, limiting the state space size.

1.4.2 Action space

In RL optimization models concerning HO procedure, the most common approaches focus on optimizing the handover control parameters such as TTT and HOM. However, choosing proper parameters' value as the state action has its own drawbacks such as state-action size which can grow dramatically if more parameters are factored in. For example, if only TTT, Hys are considered, $a_i(t) = \left[a_i^{TTT}(t),\ a_i^{Hys}(t)\right]$, then there will be 976 different possible action combinations. To address this problem, in this paper we have implemented a local search mechanism for the action space where, instead of considering the whole possible action space, just an area with a defined search radius (SR) around the current values of the parameters is considered.

Let's first clarify what do we mean by search radius. To simplify, let's just consider one parameter such as Hys. If the current value of the Hys is 3dB and the $SR = 1$, it means that we can choose either 3dB meaning no change or decrease/increase the Hys by one step (2.5dB or 3.5dB). Now, if we consider $SR = 2$ then our possible choices would be among one of these values: 2dB, 2.5dB, 3dB, 3.5dB, 4dB. This method can reduce the action size from 31 to 3 or 5 actions in each state for $SR = 1$ and $SR = 2$ respectively. A 2D and 3D visualization of action search space for two HCPs with $SR = 2$ and 3 HCPs with $SR = 1$ is illustrated in Figure 1.3. The first one will have 25 actions at each state while the other one will only have 27 action possibilities. Moreover, the SR can be parameter specific, meaning that each parameter can have a different degree of freedom in the search procedure (SR_x). In the scope of this paper same SR values are assumed for different parameters. So, with a flexible tuning capability, the state action is defined as increasing or decreasing the central HCPs with SR_x degree of freedom: $x \in \{TTT, Hys, A3off, \dots\}$. As previously mentioned, the HCPs have different units and scales, and to simplify the notations and implementation of the algorithm, the

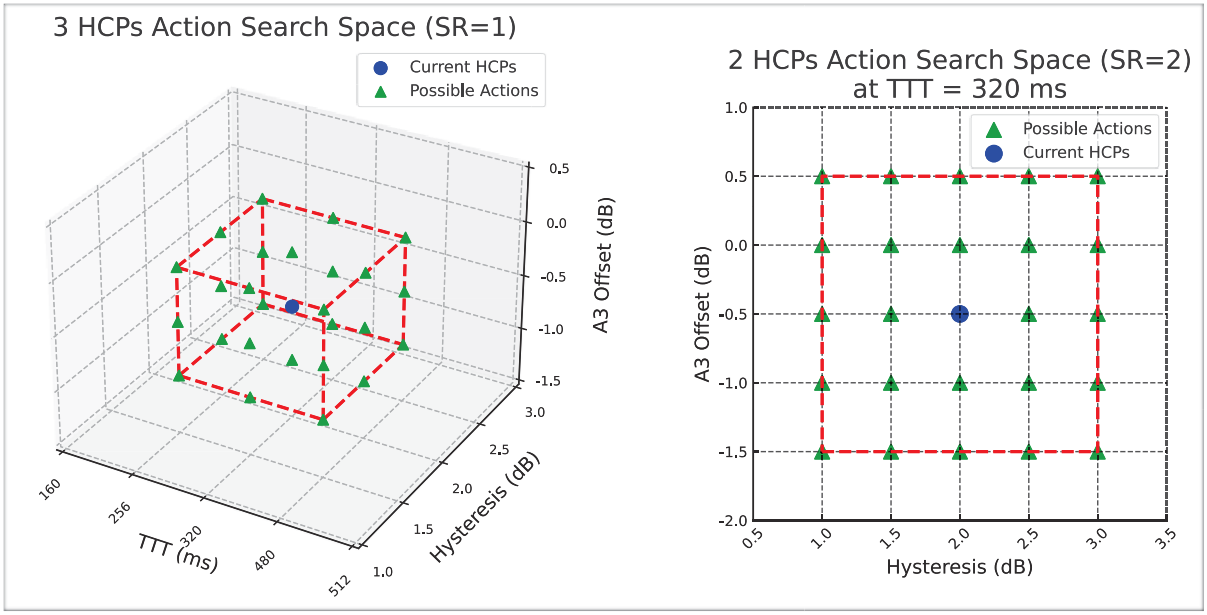

Figure 1.3 3D and 2D visualization of the action search space: Left is for 3 HCPs with SR = 1 while the Right is for 2 HCPs with SR = 2

corresponding indexes are used ($a_i(t) \approx indx_i(t)$), similar to the state definition for the $\overline{s}_i^{hcp}(t)$, as follow:

$$a_i(t) = \left[a_i^{TTT}(t), a_i^{Hys}(t), a_i^{A3off}(t), \dots\right] \tag{1.16}$$

$$\downarrow$$

$$indx_i(t) = \left[indx_i^{TTT}(t), indx_i^{Hys}(t), indx_i^{A3off}(t), \dots\right] \tag{1.17}$$

$$\begin{gathered} indx_i^{TTT}(t) = Indx_i^{TTT} + step_i^{TTT}(t) \\ indx_i^{Hys}(t) = Indx_i^{Hys} + step_i^{Hys}(t) \\ indx_i^{A3off}(t) = Indx_i^{A3off} + step_i^{A3off}(t) \\ \downarrow \\ step_i^x(t) \in \{0, \pm 1, \dots, \pm SR_x\} \\ x \in \{TTT, Hys, A3off, \dots\} \end{gathered} \tag{1.18}$$

where $indx_i(t)$ is the action vector containing the indexes of the parameters in their corresponding value vectors. $indx_i^x(t) \in \{0, 1, \dots, A_x\}$ is the new index of the handover control parameter x for user i at time t, where A_x is the length of the value vector of the parameter x. $Indx_i^x$ are the central values of parameter x which are fixed in the scope of this

paper and $step_i^x(t)$ is the chosen steps for the parameter x at time t. Table 1.2 presents the value vector for the most important HCPs used in literature. Lower values of the SR can reduce the algorithm's complexity by limiting the size of the action space, thereby enabling the inclusion of additional HCPs without incurring excessive computational overhead. But on the other hand, it limits the algorithm's degree of freedom and restrict the model to local optimal values. This can be improved by using higher SR values at the cost of complexity but finding the proper values of the SR_x based on the environment's level of dynamicity can be an interesting topic to further explore. Finally, at each state, the algorithm chooses a random action with probability ε and the action with the highest Q-value with probability $(1-\varepsilon)$.

Table 1.2 HCP Values Vectors

HCP ($\boldsymbol{x}$)	Value Vector	$\boldsymbol{A_x}$
TimeToTrigger (TTT)-ms	$\{0, 40, 64, 80, \dots, 2560, 5120\}$	16
A3Offset (A3Off)-0.5dB	$\{0, \pm1, \pm2 \ \dots, \pm 29, \pm30\}$	61
Hysteresis (Hys)-0.5dB	$\{0, 1, 2, 3, \dots, 28, 29, 30\}$	31
Cell Specific Offset (CIO) -dB	$\{0, \pm1, \pm2 \dots, \pm6, \pm8 \ \dots, \pm24\}$	31

1.4.3 Reward function

There are different KPIs that can be considered when designing the *Reward Function*. KPIs such as Average Throughput, HOR, PPHOR, HOSR, HOFR can be targeted to tune the DQN weights such that the targeted KPI is improved. In 5G/B5G networks, throughput is one of the key performance metrics which guarantees the eMBB(+) usage scenario realization. Similarly, *Outage Probability* has a direct impact on this KPI, because the higher the SINR status and allocated PRBs, the lower the OP and consequently the higher average throughput of the user. The reward function is structured such that it pushes the algorithm to aim for actions with better coverage probability, better load status of the serving cell and lower HOIs and HOFs to consequently achieve better throughput and user experience. It is defined as follows:

$$R(t) = P^{coverage} + load^{score} - CHO - \beta$$
$$\beta = \{1 \; \boldsymbol{if} \; HO \; fail \; \boldsymbol{else} \; 0\} \tag{1.19}$$
$$CHO = c$$

where $P^{coverage}$ is the user's coverage probability, $load^{score}$ is the load score of the user which is the gain that user has achieved in terms of cell load. CHO is the cost of a handover instance to control the HOI and β is the penalty due to HO failure to address the wrong HO instances.

1.4.4 Target-DQN algorithm (TDQN)

Based on the system model of this paper and with no loss of generality, a transmission time interval (TTI) equal to MGRP $= 80ms$ has been considered, over which the resources are scheduled based on a fair distribution of resources according to the required BW by each user. The proposed TDQN algorithm for optimizing HCPs is presented in Algorithm I. This method extends the standard Deep Q-Network (DQN) framework by incorporating a replay buffer and a target network to enhance training stability in the dynamic environment of cellular networks. The algorithm begins with parameter initialization, including the Q-network, target network, and replay buffer. At each simulation step, physical layer measurements are collected, and an initial measurement report is generated. The algorithm then checks whether the handover potential list (HO-Potential-list) includes at least one neighboring cell offering better or comparable signal quality than the current serving cell. Only under this condition does the algorithm proceed to evaluate the user's current state $S_i(t)$, which is then used for action selection and learning (similar to event $A3$ with $TTT = 0, Hys = 0, A3off = 0$). Based on the evaluated state, the corresponding action $a_i(t)$ is selected from the Q-network and applied to configure the user's measurement parameters. A timer is then initiated to monitor the outcome. If a handover instance is observed before the timer expires, or if the timer expires without triggering a HOI (whichever occurs first), the algorithm proceeds to compute the reward $R(t)$ and the subsequent user state $S_i(t+1)$. If a HOI is detected, the next state $S_i(t+1)$ is treated as a terminal state that will affect the target value computation in the training phase using the terminal-state formulation in Equation (1.14). This experience tuple ($S_i(t)$, $a_i(t)$, $R(t)$, $S_i(t+1)$, $Done$) is stored in the replay buffer. Once the buffer size reaches a certain

value (usually the batch size), a mini-batch of samples is randomly drawn and used to train the Q-network. Finally, the Target-Network updates its parameters at a synchronization frequency (ℓ) based on Q-Network parameters.

Algorithm 1.1 Proposed Target-DQN Algorithm

Proposed Target-DQN Algorithm

TDQN Parameters Initialization:
- Q-Network (ω) & Target-Network ($\widehat{\omega}$)
- ξ, Reply Buffer, Batch-size, learning rate, ε, Epsilon decay
- E, $Simulation_{time}$, TTI, timer

for epoch $\boldsymbol{e} = 1$: E **do:**
  for $t = 0: Simulation_{time}, tti = 80ms$ **do:**
    for user $i = 1: \mathbb{N}$ **do**:
      Measure the $rsrp_i^j$ from the neighbor BS list
      Generate Measurement Report
      If HO-Potential-list **not** Empty **do:**
        Get the user state $S_i(t)$
        $\begin{cases} a_i(t) = argmax_a Q(s_i, a; \omega) & with\ Prob.\ \varepsilon \\ Random\ a_i(t) & with\ Prob.\ 1-\varepsilon \end{cases}$
        Apply Action $a_i(t)$
        Start the DQN timer
        Reassess the Measurement Report
      If HO-Candidate-list **not** Empty **do**:
        Execute HO
      If timer expired **or** HO Executed **do**:
        Calculate reward $\boldsymbol{R}$
        Get the user state $S_i(t+1) = s_i'$
        Record the Experience $(S_i(t), a_i(t), R, S_i(t+1))$
      If Reply Buffer size > Batch Size **do**:
        Randomly Sample a **Batch** from Buffer
        Calculate $y = \begin{cases} \boldsymbol{R} & HO\ Complete \\ \boldsymbol{R} + \xi Q(s_i', a_i; \widehat{\omega}) & HO\ not\ Complete \end{cases}$
        Train the Q-Network (ω)
    Every ℓ training rounds **Copy** Q-Network parameters (ω) to Target-Network ($\widehat{\omega}$)

1.4.5 Computational complexity as a function of search radius and number of HCPs

The output layer of the reinforcement learning agent's neural network is designed to produce Q-values for each discrete action in the space. Given that the total number of actions is defined as $|A| = (2 \cdot SR + 1)^K$, where SR is the search radius and K is the number of HCPs, the output layer must contain one neuron per action. Consequently, the computational complexity of each forward pass through the network is:

$$O(H \cdot (2 \cdot SR + 1)^K) \tag{1.20}$$

where H is the number of neurons in the final hidden layer. This formulation reveals that complexity scales exponentially with the number of tuned parameters K, and polynomially with the SR. Thus, SR provides a direct trade-off between action space granularity and computational cost. Similar complexity orders hold for the training phase and backpropagation.

1.5 Simulation results

A comprehensive set of simulations was conducted to evaluate the performance of the proposed model across various deployment scenarios and parameter configurations. The results are benchmarked against a best-practice baseline where $TTT = 320ms$, $Hys = 1$ and $A3Off = 0$, reflecting handover parameter settings recommended by Huawei. Also, some worst-case parameter settings are examined to somehow illustrate the performance of the system under value settings that are not properly chosen.

Table 1.3 General Simulation Parameter Settings

Parameter	Value
Simulation area	$2Km^2$
Mobility model	Self-Developed using OSM graph
HO type	NR Intra-Frequency HO
Cell type	Micro
ISD	350m
Cell antenna gain	10dB
Deployment scenario	Uniform Hexagonal
Area type	Urban
Central frequency	25GHz
Numerology	$\mu = 2$
Bandwidth	100MHz
# of small cells	12
Small cell height	10m
User height	1.5m
User antenna gain	4dB
Small cell TX power	46dBm
Thermal noise density	$-174\text{dBm}(\text{dBm/Hz})$
Pass loss model	ETSI_TR_138_901
Shadow fading std	4 (LOS), 7.8 (NLOS)
# of stationary users	120
# of mobile users	40
Possible users' speeds	[5,10,20,40,80]Km/h
MGRP/TTI	80ms
Measurement reporting	Periodical
Simulation RUN time	30min

Table 1.4 TDQN Parameter Settings

Parameter	Value
Model	TDQN
Hidden Layers	2 Layer
Discount Factor (ξ)	0.9
Epsilon Decay	0.999
Learning Rate	10^{-4}
Buffer Size	512
Batch Size	32
Activation Function	Leaky RELU
Loss Function	Smooth L1Loss

1.5.1 Parameters and metrics

The performance of the proposed model was evaluated under various parameter configurations, as summarized in Table 1.3 and Table 1.4. Stationary users are distributed across the network to emulate varying levels of cell congestion, allowing mobile UEs to experience different load conditions as they move. In general, higher cell load results in lower achievable data rates, even under similar signal quality conditions. To illustrate this setup, we consider a representative deployment consisting of $\mathbb{M} = 132$ randomly distributed stationary users and a single mobile user ($\mathbb{N} = 1$) within a network of $\mathbb{C} = 12$ small cells with 350m inter-site distance (ISD) operating at 25GHz. Figure 1.4 depicts the received signal reference power ($RSRP$) where darker green presents stronger $RSRP$ while red samples on the graph represent weaker $RSRP$. Physical cell ID (PCI) footprint of the mobile user along its trajectory is also added to the graph, highlighting the serving cell PCI, and HO instances (HOI), especially at cell boundaries. The corresponding $SINR$ and achievable theoretical throughput of the mobile user during this trajectory, and the $RSRP$ values received from all BSs along its path, is also illustrated in Figure 1.5. The results show that despite better $SINR$ under certain BSs, the actual throughput can vary significantly depending on the BS load. This observation demonstrates the impact of cell congestion on throughput, even under favorable radio conditions. For example, when the user is connected to BS7 (PCI = 64), which serves 12 users with a 37% load factor, the achieved throughput is considerably lower than when it is served by BS8 (PCI = 72), which has only 5 users and a 5% load and provide user with a worse signal quality.

It also worth mentioning that 20% reserved resources have been considered for signaling overhead. While a frequency reuse factor of 1 is assumed in the scope of this paper (meaning all BSs operate on the same carrier frequency), other reuse factors are also valid and applicable. The Only difference will be in the type of HO and not the HO procedure. A list of sample service types and their minimum throughput requirements (values are approximations) used in the simulation runs can be found in Table 1.5. Key handover metrics, including the HOI, PPHO, and RLF, were assessed alongside average user throughput and outage probability. These performance indicators are influenced by several factors, such as the action space search radius (SR), the number of HCPs (K), user mobility speed $s_i^v(t)$. The simulations were carried

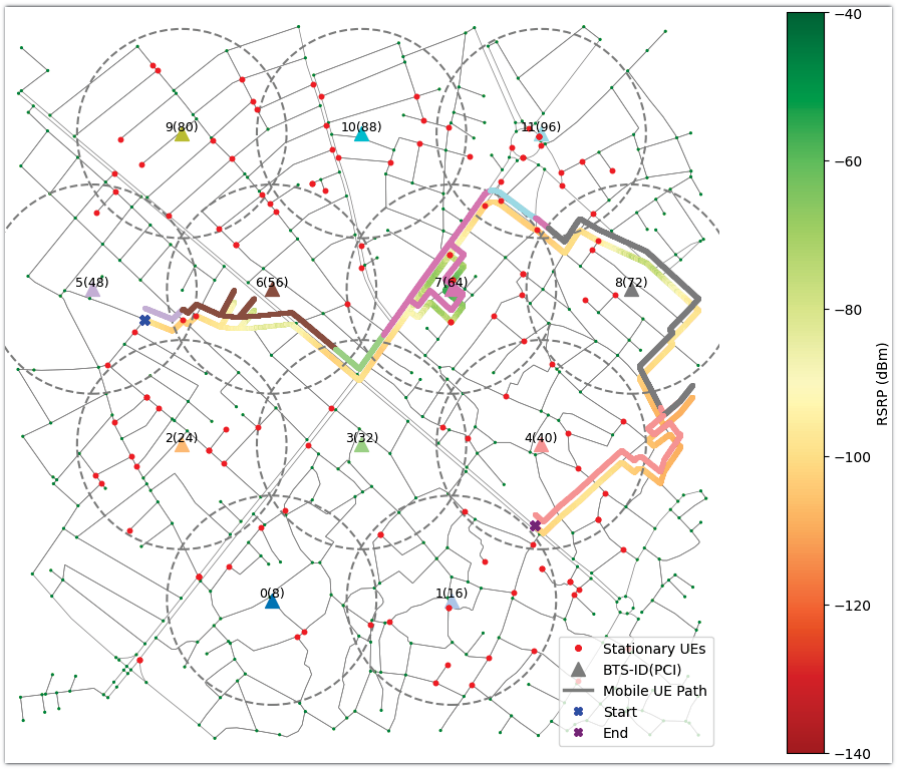

Figure 1.4 Sample Stationary UEs (red dots) deployment with a mobile User's trajectory route within the network navigating through Cells with different load levels (ISD=350m, $v = 5Km/h$)

out using a uniform hexagonal deployment of 12 5G gNodeBs with a predefined inter-site distance (ISD). In the scope of this study the evaluated KPIs are list as follows:

- HOI refer to the total number of handovers performed during a simulation run.
- PPHO are defined as handovers where a user equipment switches back to the previous cell within a short duration—specifically: within 1 second following the last handover.
- RLF are counted when a user's data rate falls below its minimum required throughput (γ_i^{min}) or the $rsrp_i^j$ falls below minimum required values for a predefined number of consecutive transmission time intervals (TTIs).

Additionally, average throughput and outage probability (OP) are calculated as the mean throughput and mean OP of the users within a 5-sec time window centered around each HO event.

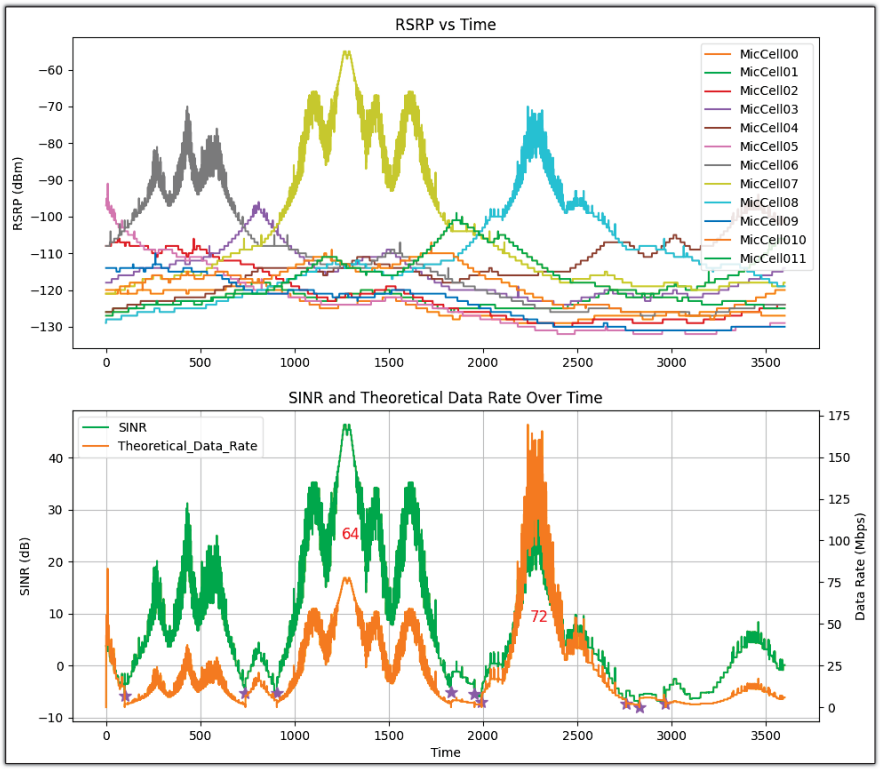

Figure 1.5 a) Upper graph representing the $RSRP$ of all the 12 SCs at user's location throughout the trajectory route without fading

b) Lower graph depicting the user's SINR and achievable theoretical data rate. Purple stars indicate the HO instances

Table 1.5 Throughput Requirement of Different Service Types

Service Type	$\boldsymbol{\gamma_i^{min}}$(Mbps)
Voice over IP	0.256
Web Browsing	1
Cloud Gaming	5
4K Video Streaming	15

1.5.2 Outage probability as a unified indication of throughput and cell load

As previously mentioned, the outage probability is inversely related to both the $SINR$ (signal quality) and the amount of bandwidth allocated to the user (serving cell load). For a fixed BW allocation, higher values of the minimum required data rate γ_i^{min} result in increased outage probability. Additionally, since throughput is directly proportional to both SINR and the number of PRBs, increased cell load leads to fewer available PRBs per user and, consequently,

higher outage probability. This establishes a logical relationship among outage probability, throughput, and cell load. Specifically, lower outage probabilities typically correspond to higher achievable data rates. Conversely, high outage probability under good SINR conditions may indicate either significant cell congestion or high target throughput requirement, or both. These relationships are illustrated better in Figure 1.6 which is based on the data of previous test simulation for 40 mobile users, and the relationship between outage probability and Throughput and the impact of $SINR$ and PRB allocation can be observed. Whenever $SINR$ is weak and lower PRBs are allocated, the OP is high, consequently the achievable data rate is low.

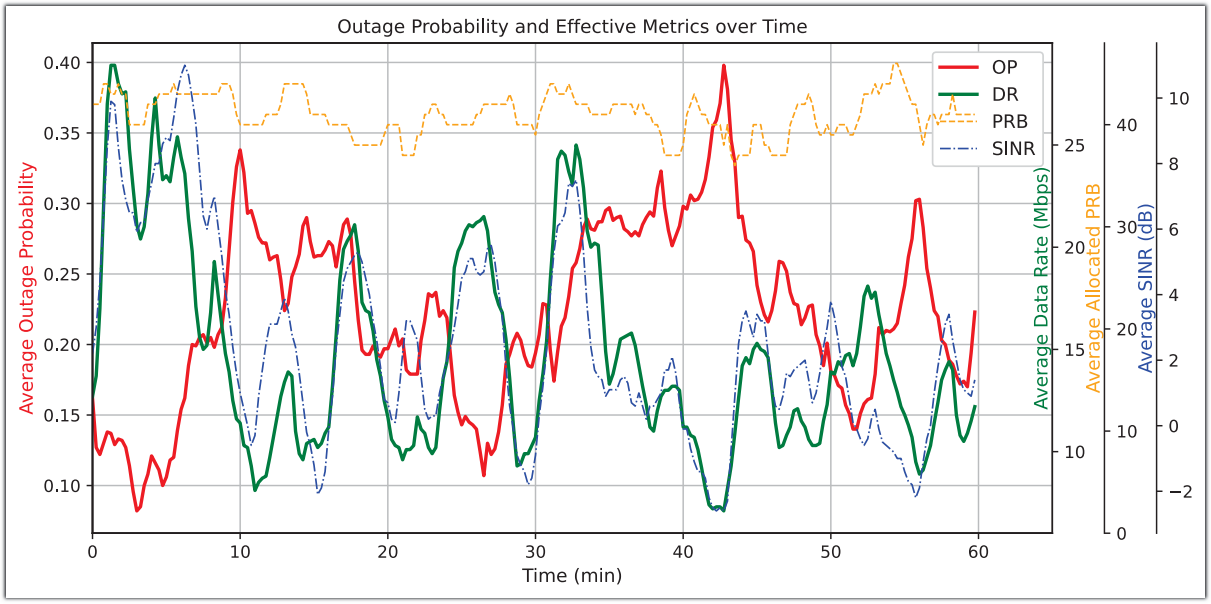

Figure 1.6 Relation of Outage Probability and Data Rate with respect to $SINR$ and allocated PRB values

1.5.3 TDQN and various SR settings

Based on the action space formulation of the proposed TDQN model, different search radius (SR) values can be configured, enabling the model to control the size of the action space. This flexibility allows the incorporation of more HCPs while potentially enhancing model performance. Under the general simulation settings described in Table 1.3, Table 1.4 and Table 1.5, a series of simulations were conducted to evaluate the optimization of three HCPs with varying SR values ranging from 1 to 10 and user speed of 5 Km/h. The results, illustrated in Figure 1.7, demonstrate the impact of SR on key handover performance metric introduced in the previous section. The proposed TDQN model yielded significant improvements across several metrics. When configured with $SR = 1$, the model achieved improvements of 33% in HOI, 55% in PPHO, and 9% in RLF compared to the benchmark. Further gains of up to 65%, 96%, and 20% in HOI, PPHO, and RLF respectively, were observed as SR increased to 10. However, larger SR values were associated with degradation in $RSRP$ and $SINR$ performance, as depicted in Figure 1.9. In this figure the probability density function (PDF) of the users'

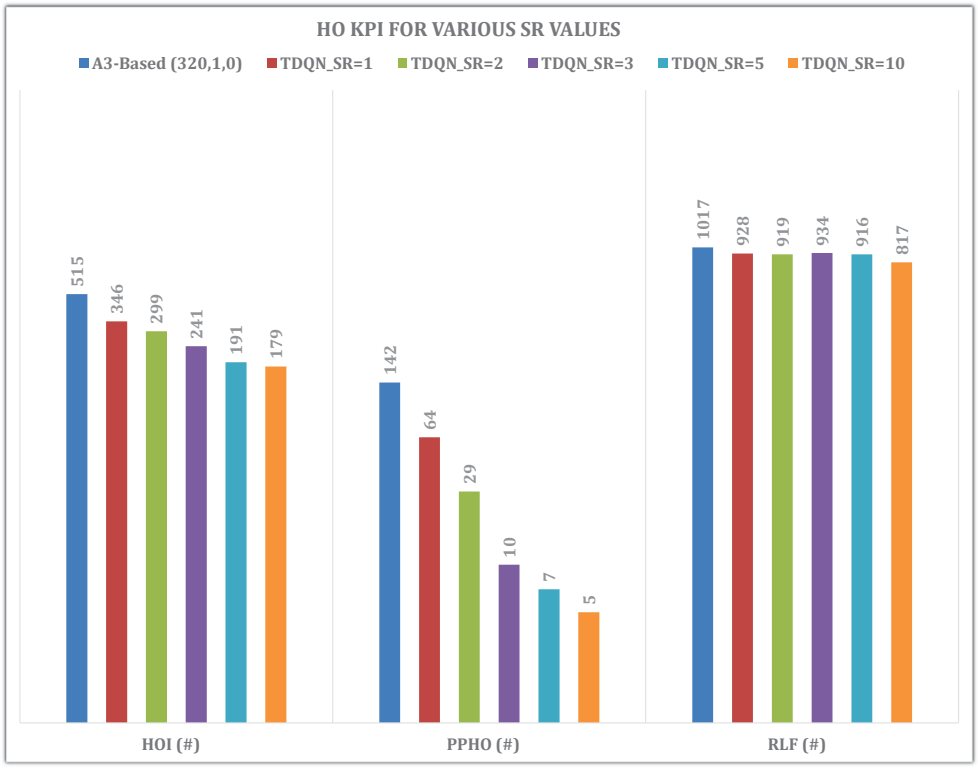

Figure 1.7 Handover KPI vs various SR values

$RSRP$ and $SINR$ is presented. Notably, $SR = 3$ provided a balanced trade-off, offering 53%, 93%, and 8% improvements in HOI, PPHO, and RLF respectively—while maintaining comparable $RSRP$ and $SINR$ levels relative to the benchmark configuration. If we consider the PDFs for $SR = 5$ and $SR = 10$, it is clear that the users start to experience worse $RSRP$ and $SINR$ conditions due to aggressive parameter settings by the algorithm. This is because the model needs more samples and training to fully adjust to the new SR setting. Moreover, the proposed TDQN as depicted in Figure 1.8, enhanced the average user throughput by 3% to 10% for SR values between 1 and 5 while improving the OP up to 10% for $SR = 3$. At $SR = 10$, however, throughput gains diminished, indicating an inflection point in performance. A similar trend was observed for OP. These findings emphasize the existence of a trade-off between performance, complexity and SR size, as increasing SR leads to polynomial growth in action space size and neural network computational complexity as formulated in Equation (3.8), as well as performance metrics to some orders But, after certain values the system

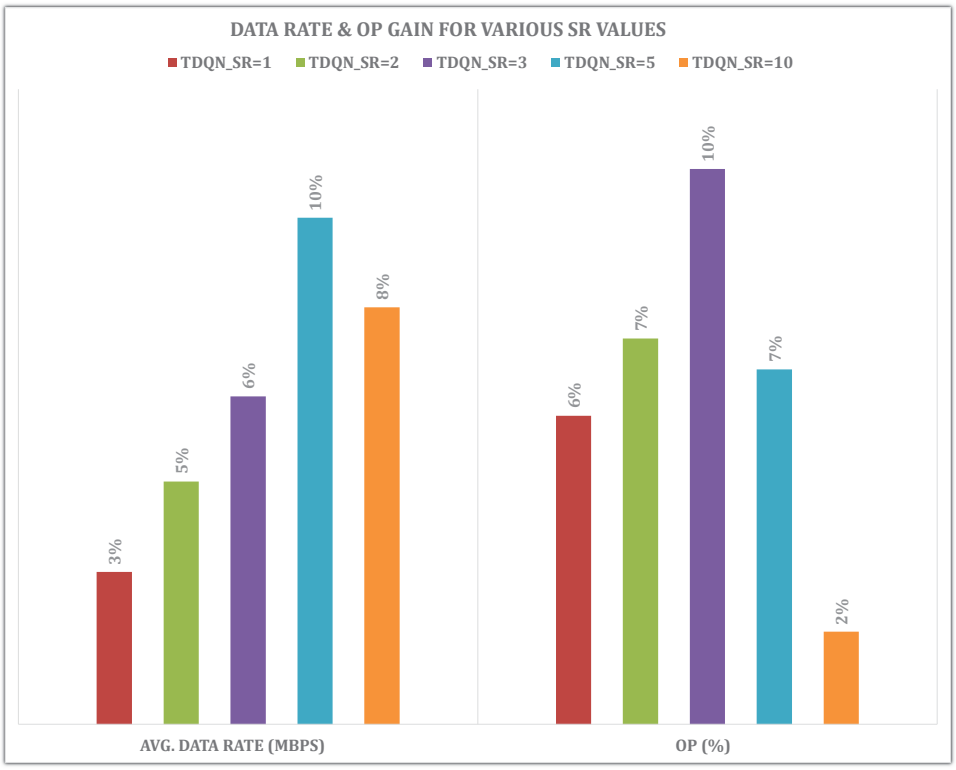

Figure 1.8 Data Rate and OP gains for various SR values

performance starts to degrade.

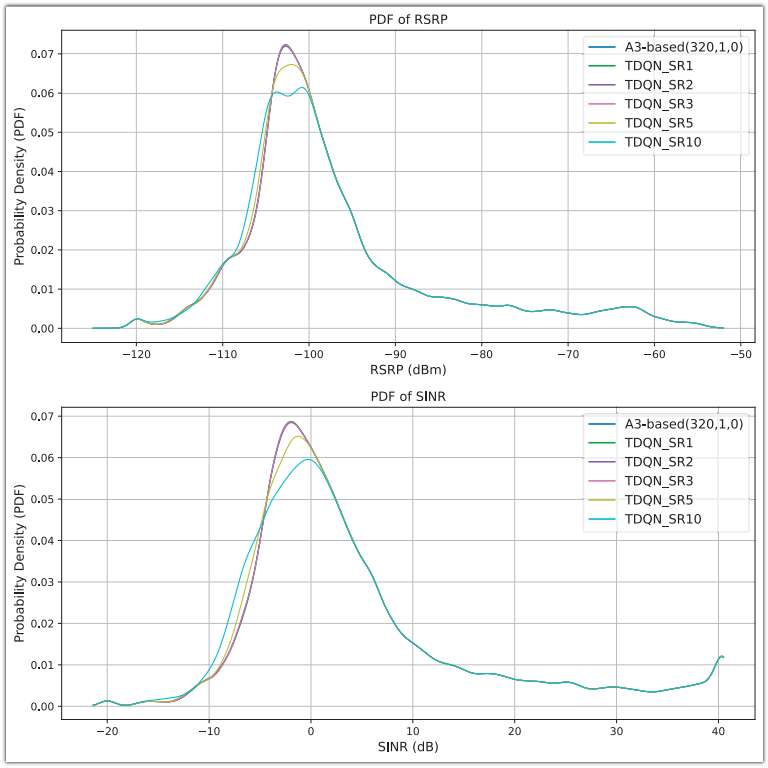

Figure 1.9 RSRP and SINR PDF for various SR values

1.5.4 Proposed TDQN performance for various number of incorporated HCPs

In this section, we evaluate the proposed TDQN model as a function of the HCPs being optimized. Specifically, we assess the model's performance when optimizing a reduced set of HCPs, while in the previous scenario three HCPs were jointly optimized. Based on the insights from the previous simulation analysis, an SR value of 3 and a user speed of 5 Km/h has been considered for this set of simulations. As illustrated in the bar graph in Figure 1.10, the proposed TDQN model continues to deliver substantial performance gains, particularly in reducing PPHO and overall HOIs. When optimizing TTT and Hys simultaneously, the model achieves a 98% reduction in PPHOs, whereas a 73% reduction is observed when optimizing TTT and $A3Off$. In both scenarios, a 43% decrease in HOI is observed compared to the benchmark configuration. It is worth noting that optimizing a single HCP yields less improvement than optimizing two parameters simultaneously. However, the results indicate that optimizing only TTT provides performance levels nearly comparable to those achieved

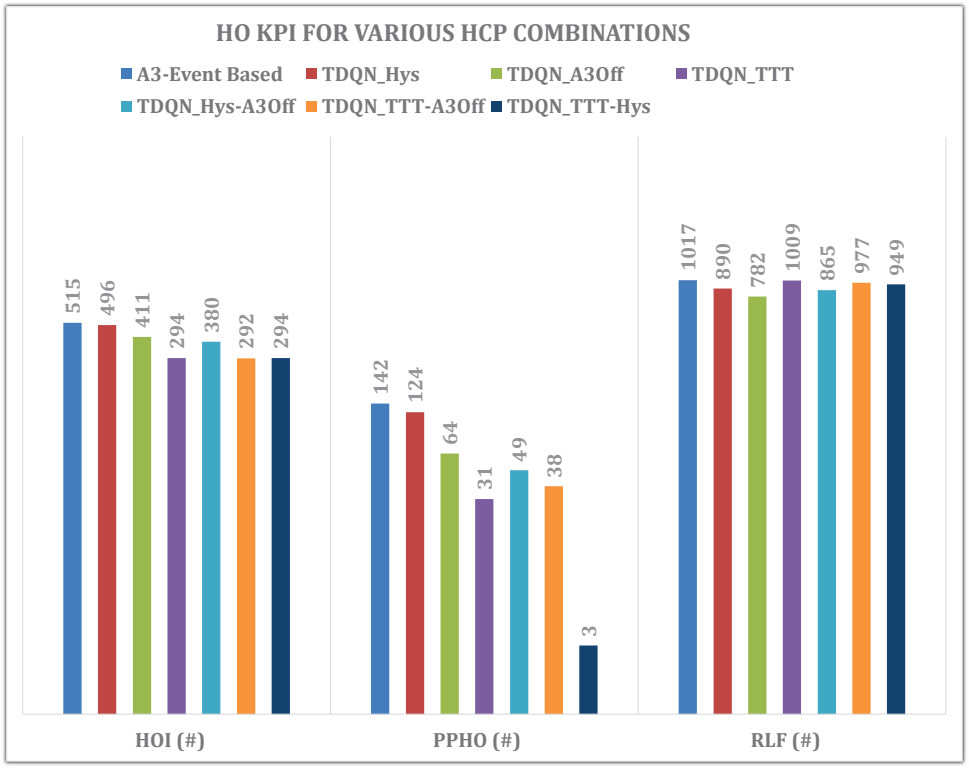

Figure 1.10 Handover KPI vs various HCP combinations

with two-HCP configurations. Additionally, average user data rates improve by approximately 5–6%, while outage probability is reduced by 7–10% across these scenarios as illustrated in Figure 1.11. It can be concluded that TTT plays a significant role in improving handover performance, particularly in reducing the number of HOI and PPHO. A comparison of the $RSRP$ and $SINR$ across different optimization scenarios, as illustrated in Figure 1.12, indicates that the impact on signal strength and quality is generally limited. However, scenarios involving the optimization of TTT alone or in combination with Hys demonstrate relatively better $RSRP$ and $SINR$ performance compared to others. While $A3off$ usually affect these KPIs negatively as the SR value increases. Although optimizing only one or two HCPs yields relatively lower performance gains compared to joint optimization of three parameters, it offers a favorable trade-off in terms of computational complexity considering Equation (3.8) as there is exponential relationship between the computational complexity and the number of HCPs.

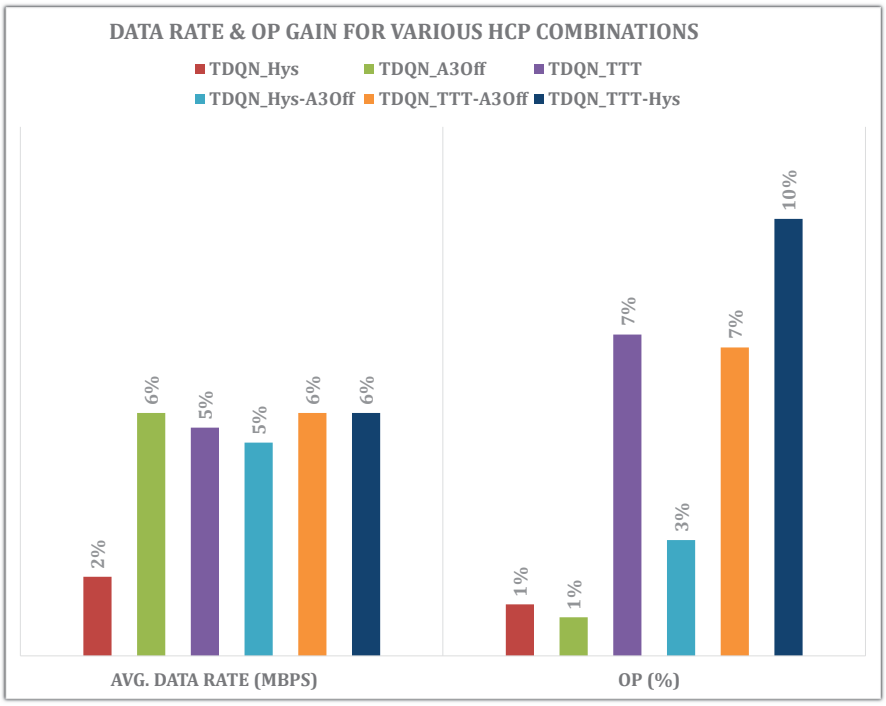

Figure 1.11 Data Rate and OP gains for various HCP combinations

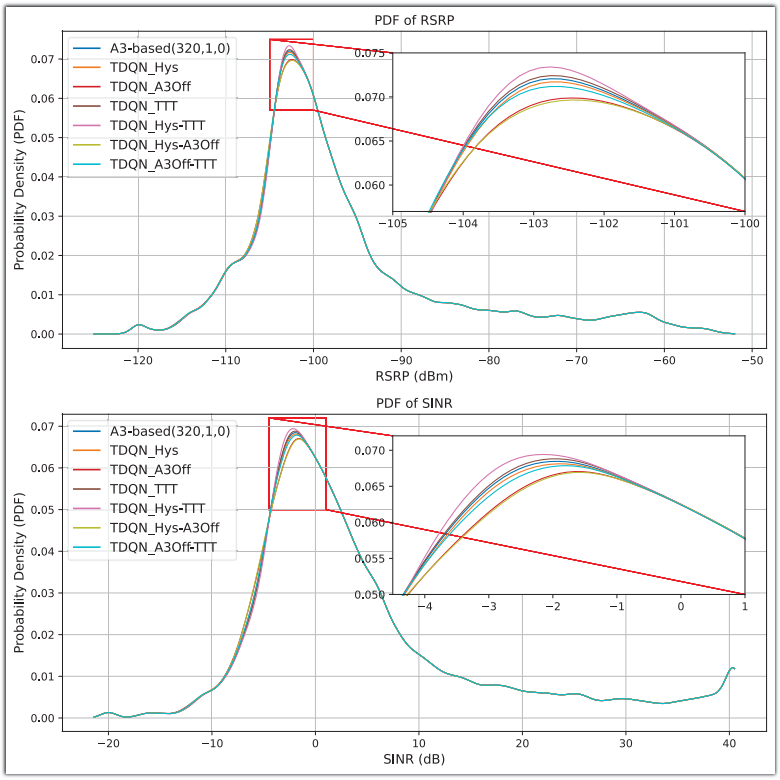

Figure 1.12 RSRP and SINR PDF for various HCP combinations

1.5.5 Proposed TDQN performance under different user speeds

In this section, we evaluate the effectiveness of the proposed TDQN model with $SR = 3$ while considering 3 HCPs under varying user speed scenarios. In this study, we focus on urban micro-BS deployments, considering user speeds ranging from 5 km/h to 80 km/h. Figure 1.13 presents key system KPIs across various speed profiles. As expected, higher user speeds lead to an increased number of HOI due to faster variations in signal quality and less dueling time of the user in a cell coverage area. One notable observation is that the performance gains of the proposed TDQN model diminish with increasing speed. At higher mobility, more frequent handovers are necessary to maintain connectivity, and skipping too many HOs may adversely affect signal quality and overall performance. For low-speed scenarios (5 km/h), the TDQN achieves a 53% reduction in HOI and 87% improvement in PPHO as previously outlined. These gains decline to 7% and 41%, respectively, at speeds of 40–80 km/h. At moderate speeds (10–20 km/h), HOI and PPHO reductions are approximately 20% and 62%, respectively. The impact on RLF is minimal, with a maximum degradation of only 3% observed. Moreover, as

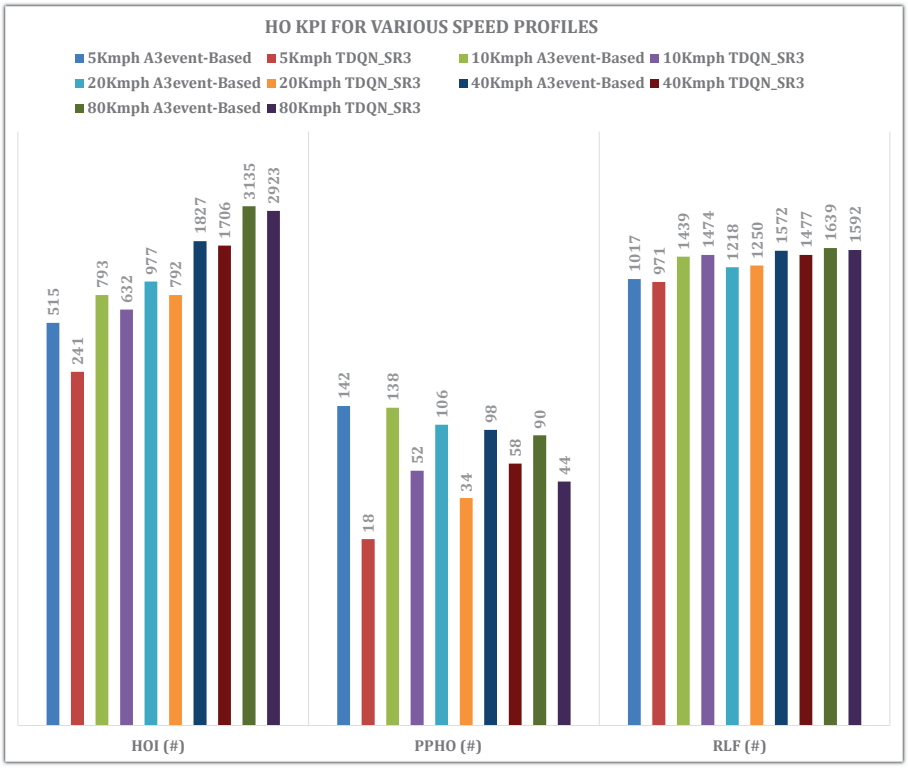

Figure 1.13 Handover KPI vs various speed profiles

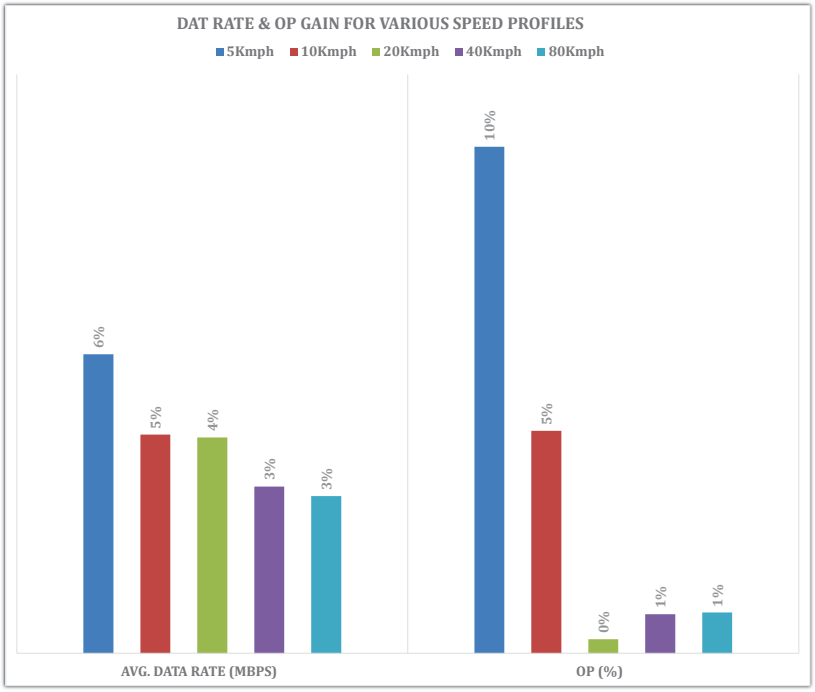

Figure 1.14 Data Rate and OP gains for various speed profiles

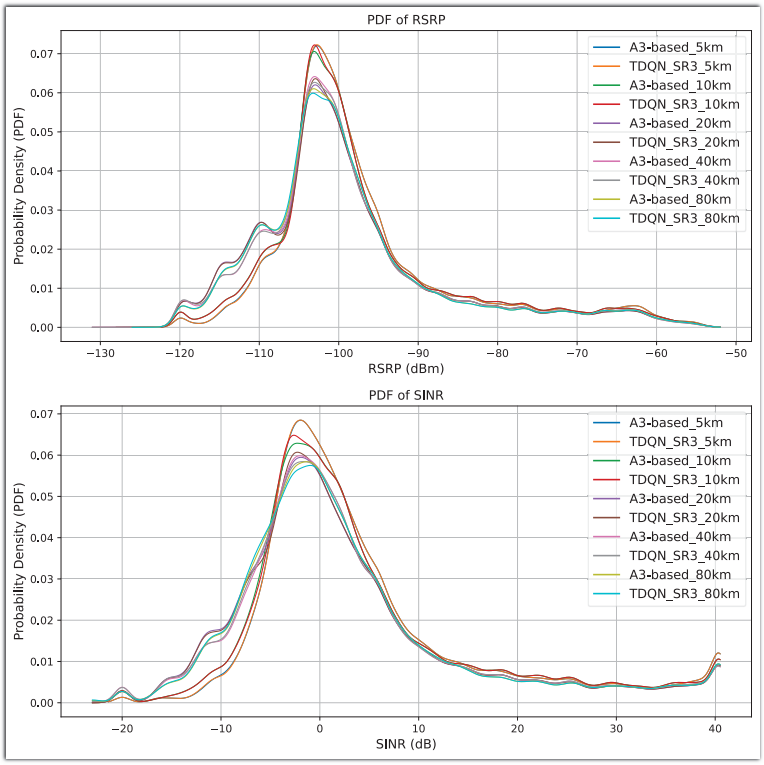

Figure 1.15 RSRP and SINR PDF for various speed profiles

Figure 1.14 highlights, improvements in outage probability are also modest, and TDQN reduces HOI and PPHO while preserving signal coverage. Additionally, the average gains in throughput decreased from 6% at 5 km/h to about 3% at 80 km/h. However, it is also worth noting that the total average system throughput under the benchmark configuration, with no TDQN active, decreases from 9.5 Mb/s at 5 km/h to 8.5 Mb/s at 80 km/h, highlighting the impact of user mobility on achievable data rates. As illustrated in Figure 1.15, while system-wide performance deteriorates with increasing user speed, the TDQN effectively maintains $RSRP$ and $SINR$ levels close to the benchmark configuration while still achieving meaningful KPI improvements presented in previous bar graphs.

1.5.6 System performance under aggressive HCP settings

Finally, it is instructive to examine system behavior under aggressive and worst-case settings of HCPs. Figure 1.16 illustrates the cumulative distribution function (CDF) and the PDF of the users' $RSRP$ and OP for various HCP configurations. As shown, applying extremely high or low values to HCPs can significantly degrade system performance. These scenarios effectively define the performance boundaries of the system. In the figure, the light blue curve represents the benchmark configuration, while the dark blue curve denotes an idealized case in which no handover delay or margin is applied. The two red curves correspond to two extreme HCP settings. In these extreme cases, users suffer from degraded signal conditions pushing the PDF and CDF toward low RSRP values as well as worse outage probabilities which indicate low achievable data rates. Because users either remain connected to the serving cell for too long, or they are prematurely handed over to neighboring cells before entering an area of sufficient signal coverage. These results underscore that simply reducing the number of HOIs by increasing HCP values (e.g. Hys or A3Off), without any consideration, is not a viable solution leading to performance decline in other key metrics, such as those observed in previous simulations. So, the process of assigning these parameters under dynamic states of the system should be executed carefully and intelligently. This highlights the importance and necessity of a well-developed model for assignment of proper HCPs for the users under different states.

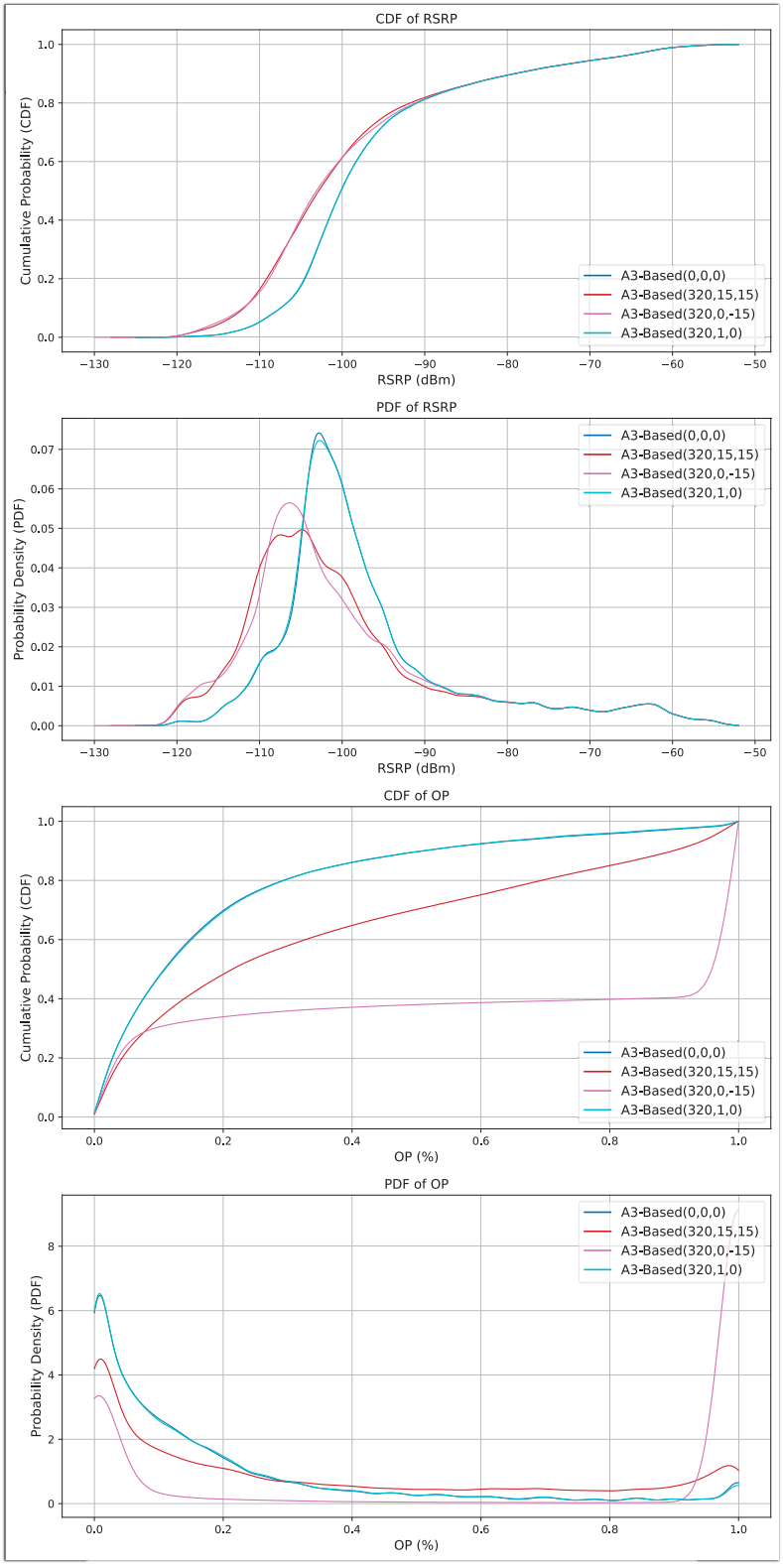

Figure 1.16 CDF and PDF of $RSRP$ & OP for different HCP settings. The last setting is HUAWEI recommendation

CONCLUSION AND FUTURE WORKS

This thesis proposed a TDQN-based framework for optimizing handover control parameters in ultra-dense 5G networks. The model introduces a carefully structured state space that incorporates essential user-level and network-level information—including mobility, signal quality and load in form of outage probability—enabling more accurate and efficient handover decisions. More importantly, the action space is defined through a search radius (SR) mechanism that offers flexibility in selecting more HCPs and controls the trade-off between decision granularity and computational complexity. Simulation studies across a range of deployment conditions, user mobility profiles, and optimization settings demonstrated that the TDQN model consistently enhances handover performance while maintaining stable signal quality and user experience.

Beyond the numerical improvements, the proposed approach addresses long-standing limitations in mobility management by jointly optimizing multiple parameters in a scalable manner, making it adaptable to the complex and dynamic conditions of next-generation networks. The results confirm that the framework is effective not only in reducing unnecessary handovers, ping-pong events, and radio link failures, but also in maintaining high throughput and coverage probability across diverse scenarios. Its ability to balance performance with computational feasibility makes it a promising candidate for integration into practical network management systems.

The adaptability of the model across different speeds, cell densities, and parameter configurations underscores its robustness and real-world viability, including applications in high-mobility vehicular environments and dense urban deployments. By providing an intelligent, context-aware, and scalable solution, this research contributes a significant step toward achieving seamless connectivity and enhanced quality of service in 5G/B5G networks. Overall, the TDQN framework offers a forward-looking alternative to traditional handover strategies, paving the way for more autonomous, efficient, and resilient mobility management in future wireless systems.

Future works: while the proposed TDQN model has demonstrated strong potential in optimizing mobility management, several promising directions remain for future research. One valuable extension is leveraging the model's capacity to incorporate additional parameters for load balancing by dynamically optimizing Cell Individual Offsets (CIOs). Exploring more realistic mobility profiles, including non-constant speeds and rural area scenarios, would further validate the model's robustness. Additionally, hyperparameter tuning and investigation into neural network architectures and sizes could help enhance learning efficiency and performance. A decentralized approach, where each user operates an independent TDQN agent while sharing knowledge through Transfer Learning (TL), presents another compelling research avenue, paving the way for realization of user-managed MM. Lastly, adapting the model to support multi-tier and multi-frequency networks would extend its applicability to more heterogeneous and complex cellular environments.

BIBLIOGRAPHY

Adnan Farooq Bhat, Shahid Mehraj Shah. 2024. "SIGMAML: SNR-Guided 5G Mobility Management using Machine Learning Algorithms." *2024 IEEE Space, Aerospace and Defence Conference (SPACE).* Bangalore, India: IEEE. 474-478. doi:10.1109/SPACE63117.2024.10667970.

David Lopez-Perez, Ismail Guvenc, and Xiaoli Chu. 2012. "Mobility management challenges in 3GPP heterogeneous networks." *IEEE Communications Magazine* 50: 70-78. doi:10.1109/MCOM.2012.6384454.

Goldsmith, Andrea. 2005. *Wireless Communications.* Cambridge university press.

Jin Wu, Jing Liu, Zhangpeng Huang, and Shuqiang Zheng. 2015. "Dynamic fuzzy Q-learning for handover parameters optimization in 5G multi-tier networks." *International conference on wireless communications & signal processing (WCSP)* 1-5. doi:10.1109/WCSP.2015.7341220.

Kang Tan, et al. 2022. "Intelligent Handover Algorithm for Vehicle-to-Network Communications with Double-Deep Q-Learning." *IEEE Transactions on Vehicular Technology* 71: 7848-7862. doi:10.1109/TVT.2022.3169804.

Marvin Manalastas, et al. 2022. "A Data-Driven Framework for Inter-Frequency Handover Failure Prediction and Mitigation." *IEEE Transactions on Vehicular Technology* 71: 6158-6172. doi:10.1109/TVT.2022.3157802.

Mingzhe Chen, et al. 2019. "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial." *IEEE Communications Surveys & Tutorials* 21: 3039-3071. doi:10.1109/COMST.2019.2926625.

Mubashir Murshed, Glaucio HS Carvalho and Robson E. De Grande. 2024. "Ultra-Density Aware Learning-Based Handover Management in High-Mobility 5G Vehicular Networks." *ICC 2024.* Denver: IEEE International Conference on Communications. 2324-2329. doi:10.1109/ICC51166.2024.10623074.

Rabe Arshad, et al. 2016. "Handover Management in 5G and Beyond: A Topology Aware Skipping Approach." *IEEE Access* 4: 9073-9081. doi:10.1109/ACCESS.2016.2642538.

Raja Karmakar, Georges Kaddoum, and Samiran Chattopadhyay. 2023. "Mobility management in 5G and beyond: A novel smart handover with adaptive Time-To-Trigger and Hysteresis Margin." *IEEE Transactions on Mobile Computing* 5995-6010. doi:10.1109/TMC.2022.3188212.

Rana D. Hegazy, O. A. Nasr, and H. A. Kamal. 2018. "Optimization of user behavior based handover using fuzzy Q-learning for LTE networks." *Wireless Networks* 24: 481-495. doi:10.1007/s11276-016-1348-2.

Sunil Kandukuri, Stephen Boyd. 2002. "Optimal power control in interference-limited fading wireless channels with outage-probability specifications." *IEEE Transactions on Wireless Communications* 1: 46-55. doi:10.1109/7693.975444.

Syed Muhammad Asad Zaidi, Marvin Manalastas, Hasan Farooq and Ali Imran. 2020. "Mobility Management in Emerging Ultra-Dense Cellular Networks: A Survey, Outlook, and Future Research Directions." *IEEE Access* 8: 183505-183533. doi:10.1109/ACCESS.2020.3027258.

Tanu Goyal, Kaushal Sakshi. 2019. "Handover optimization scheme for LTE-advance networks based on AHP-TOPSIS and Q-learning." *Computer communications* 133: 67-76.

TS38.214, 3GGP. 2025. *5G; NR; Physical layer procedures for data (Release 18).* France: 3rd Generation Partnership Project. https://www.3gpp.org/dynareport/38214.htm.

TS38.215, 3GGP. 2024. *5G; NR; Physical layer measurements (Release 18).* France: 3rd Generation Partnership Project. https://www.3gpp.org/dynareport/38215.htm.

TS38.300, 3GPP. 2025. *NR and NG-RAN Overall description; Stage-2.* France: 3GPP. https://www.3gpp.org/dynareport/38300.htm.

TS38.331, 3GPP. 2025. *NR; Radio Resource Control (RRC) protocol specification (Release 18).* France: 3rd Generation Partnership Project. https://www.3gpp.org/dynareport/38331.htm.

Waheeb Tashan, et al. 2022. "Advanced Mobility Robustness Optimization Models in Future Mobile Networks Based on Machine Learning Solutions." *IEEE Access* 10: 111134-111152. doi:10.1109/ACCESS.2022.3215684.

Waheeb Tashan, et al. 2022. "Mobility Robustness Optimization in Future Mobile Heterogeneous Networks: A Survey." *IEEE Access* (IEEE Access) 10: 45522-45541. doi:10.1109/ACCESS.2022.3168717.

Wasan Kadhim Saad, et al. 2021. "Handover parameters optimisation techniques in 5G networks." *Sensors* 21: 5202. doi:10.3390/s21155202.

Wei Huang, et al. 2022. "Self-adapting handover parameters optimization for SDN-enabled UDN." *IEEE Transactions on Wireless Communications* 21: 6434-6447. doi:10.1109/TWC.2022.3149415.

Yaohua Sun, et al. 2019. "Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues." *IEEE Communications Surveys & Tutorials* 21: 3072-3108. doi:10.1109/COMST.2019.2924243.

Yasir Ullah, et al. 2024. "Evaluation Study of Handover Control Parameters in Mobile HetNets." *2024 Multimedia University Engineering Conference (MECON).* Cyberjaya: IEEE. 1-6. doi:10.1109/MECON62796.2024.10776041.