Deep Learning Applications in Visible Light Communications

Thesis submitted in partial fulfilment

for the Award of Degree

DOCTOR OF PHILOSOPHY

by

TANYA VERMA

RAJIV GANDHI INSTITUTE OF PETROLEUM TECHNOLOGY

JAIS-229304

ROLL NUMBER
21EE0003

YEAR OF SUBMISSION
2025

CERTIFICATE

It is certified that the work contained in the thesis titled **Deep Learning Applications in Visible Light Communications** by **Tanya Verma** has been carried out under our supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy, SOTA and Open Seminar.

Dr. Shivanshu Shrivastava
(Supervisor)

Dr. U.D.Dwivedi
(Co-Supervisor)

DECLARATION BY THE CANDIDATE

I, **Tanya Verma**, certify that the work embodied in this thesis is my own bona fide work and carried out by me under the supervision of **Dr. Shivanshu Shrivastava** and **Dr. U.D. Dwivedi** from **August 2021** to **March 2025**, at the **Department of Electrical and Electronics Engineering**, **Rajiv Gandhi Institute of Petroleum Technology, Jais**. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.*, reported in journals, books, magazines, reports dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date: **Tanya Verma**

Place: (21EE0003)

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of our knowledge.

Dr. Shivanshu Shrivastava	**Dr. U.D.Dwivedi**	**Head of Department**
(Supervisor)	(Co-Supervisor)	(Electrical and Electronics)

CERTIFICATE

CERTIFIED that the work contained in the thesis titled "Deep Learning Applications in Visible Light Communications" by Ms. **Tanya Verma** has been carried out under our supervision. It is also certified that she fulfilled the mandatory requirement of TWO quality publications that arose out of her thesis work.

It is further certified that the two publications (copies enclosed) of the aforesaid Ms. **Tanya Verma** have been published in the Journals indexed by –

(a) SCI

(b) SCI Extended

(c) SCOPUS

Dr. Shivanshu Shrivastava	**Dr. U.D.Dwivedi**	**Dr. Shivanshu Shrivastava**
(Supervisor)	(Co-Supervisor)	(Convener, DPGC)

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Deep Learning Applications in Visible Light Communications

Name of the Student: Tanya Verma

COPYRIGHT TRANSFER

The undersigned hereby assigns to the Rajiv Gandhi Institute of Petroleum Technology, Jais, all rights under copyright that may exist in and for the above thesis submitted for the award of the DOCTOR OF PHILOSOPHY.

Date: **Tanya Verma**

Place: (21EE0003)

ACKNOWLEDGMENT

Praises to the Divine Lord Mahadev, with whom blessings I got the courage to reach here. Also, thanks to my family for their constant support and motivation that allows me to reach here.

Next, my sincere gratitude goes to my thesis supervisor, **Dr. Shivanshu Shrivastava**, for his unwavering belief in me. I would like to express my heartfelt appreciation for his persistent support, invaluable guidance, and constant encouragement throughout my doctoral journey. His expertise and insightful feedback have been instrumental in shaping my work. I am truly grateful to Sir for the time and effort that Sir has dedicated in reviewing my work and for always being available to discuss ideas and address challenges. His patience, understanding, and confidence in my abilities have inspired me to push my limits and achieve my best. This thesis would not have been possible without his mentorship, and I am forever thankful for his invaluable role in my doctoral journey. I also express my heartfelt thanks to my co-supervisor, **Dr. Umakant Dhar Dwivedi**, for his motivation, guidance, and constructive feedback throughout my doctoral journey. My profound gratitude goes to **Prof. A.S.K. Sinha**, whose unwavering support has been a great source of inspiration to me.

I extend my heartfelt gratitude to my RPEC members, faculty, and administrative staff of the Department of Electrical and Electronics Engineering for their valuable suggestions and support.

Moreover, I would like to express my heartfelt gratitude to my senior and labmates for their constant support, encouragement, and camaraderie throughout this journey. The collaborative and friendly atmosphere you created made even the most challenging moments

more manageable. We will surely cherish the memories we have built and the bonds we have formed during this journey.

To my parents, thank you for instilling in me the values of hard work, perseverance, and education, and for always believing in my potential even when I doubted myself. Your encouragement and prayers have given me strength along the way. This achievement is as much yours as it is mine. To my brother Mayank, your guidance, support, unwavering faith in my abilities, and your encouragement have been my anchor. To my partner Amol, you have been incredible! Your patience and understanding during my long hours of work, as well as your belief in my abilities, have been my strengths. I am thankful to you all for standing by me throughout this doctoral journey and celebrating every milestone. This achievement would not have been possible without your unconditional love. I am forever grateful to share this life journey with you people. Your love and faith in me have been my greatest motivation throughout my life.

Tanya Verma

Dedicated to

Maa-Papa ji

and

Bhaiya

Publications from the Thesis

Journal publications

1. **Tanya Verma**, Arif Raza, Shivanshu Shrivastava, Dwarkadas Prahladadas Kothari, U.D. Dwivedi, A Novel On-Policy DRL Based Approach for Resource Allocation in Hybrid RF/VLC Systems, *IEEE Transactions on Consumer Electronics (Early Access)*, doi: 10.1109/TCE.2025.3529846, 2025.

2. **Tanya Verma**, Arif Raza, Shivanshu Shrivastava, Bin Chen, U.D. Dwivedi, Amarish Dubey, Intelligent resource allocation in Hybrid RF/LiFi networks via deep deterministic policy gradient based DRL mechanism, *AEU - International Journal of Electronics and Communications*, Volume 187, 2024, 155499, ISSN 1434-8411, https://doi.org/10.1016/j.aeue.2024.155499.

3. **Tanya Verma**, Shivanshu Shrivastava, Umakant Dhar Dwivedi, D. P. Kothari, Transfer learning for resource allotment in dynamic hybrid WiFi/ LiFi communication systems, *Optics Communications*, Volume 546, 2023, 129761, ISSN 0030-4018, https://doi.org/10.1016/j.optcom.2023.129761.

Conference publications

1. **Tanya Verma**, Arif Raza, Shivanshu Shrivastava, U.D. Dwivedi, Efficient DRL based technique for Resource Allocation in Heterogeneous WiFi/LiFi Systems, 31^{st} IEEE national conference on communications (NCC), IIT Delhi, $6-9$ March 2025.

2. **Tanya Verma**, Arif Raza, Shivanshu Shrivastava, U.D. Dwivedi, "DRL based Efficient Resource Allocation in Hybrid WiFi/LiFi Systems", 21^{st} IEEE INDICON

2024, IIT Kharagpur, $19 - 21$ December 2024.

3. **Tanya Verma**, Shivanshu Shrivastava, Arif Raza, U.D. Dwivedi, "Deep Reinforcement Learning based Q-networks for Efficient Resource Allocation in Hybrid Systems", 19^{th} EAI BODYNETS, IIT BHU, $15 - 16$ December 2024.

Contents

List of Figures

List of Tables

Abbreviation	Definition
LOS	Line-of-Sight
SNR	Signal-to-noise-ratio
SINR	Signal-to-interference-plus-noise-ratio
CU	Controlling unit
UE	User equipment
DC	Direct current
IM/DD	Intensity Modulation/Direct Detection
DNN	Deep neural network
DQN	Deep Q-network
AP	Access point
WiFi	Wireless fidelity
LiFi	Light fidelity
VLC	Visible light communication
RF	Radio frequency
PD	Photo diode
FOV	Field of view
NLOS	Non-Line-of-Sight
LED	Light emitting diode
PDF	Probability density function
CC	Channel capacity
AWGN	Additive white Gaussian noise
MLP	Multi layer perceptron
DRL	Deep reinforcement learning

DDPG	Deep deterministic policy gradient
QoS	Quality of service
PPO	Proximal policy optimization
TD3	Twin delayed deep deterministic policy gradient
DDQN	Double deep Q network
A2C	Advantage actor critic
HetNets	Heterogeneous networks
OFDM	Orthogonal frequency division multiplexing
PS1	Proposed scheme 1
PS2	Proposed scheme 2
TD	Temporal difference
FLOPs	Floating point operations
CSI	Channel state information
FCC	Federal communications commission
OWC	Optical wireless communication
EE	Energy efficiency
SE	Spectral efficiency
JEITA	Japan Electronics and Information Technology Industries Association
VLCC	Visible Light Communications Consortium
OMEGA	Home Gigabit Access project
AI	Artificial intelligence
D2D	Device to device
IoT	Internet-of-things
ES	Exhaustive search
SDR	Software defined radio
USRP	Universal software radio peripheral
UAV	Unmanned Aerial Vehicle
MSE	Mean square error

Preface

The fast development of 5G mobile communication systems has created a need for better communication solutions. Hybrid radio frequency (RF) / visible light communication (VLC) systems, which combine visible light and RF based technologies, offer a promising solution. These systems use the unique, non-overlapping spectrums of both technologies to improve data speeds and reliability, especially in changing environments with physical obstacles. However, managing resources in these hybrid systems is difficult because of the network's dynamic nature and the complexity of optimizing bandwidth, user associations, and power distribution.

This thesis explores improved deep reinforcement learning (DRL) methods for resource allocation in hybrid RF/VLC communication systems. VLC has a standardized subset called as light fidelity (LiFi) like RF has wireless fidelity (WiFi). Hence, the hybrid system formed is also known as hybrid WiFi/LiFi systems. We first propose a deep-Q-network (DQN) and transfer learning approach that enhances throughput in hybrid RF/VLC networks by adapting to changing network conditions and new mobile users entering into the indoor environment. Simulations reveal that this approach performs better than conventional optimization algorithms in maximizing data rates with fewer number of iterations with the help of transfer learning.

We extend the approach to dynamic hybrid networks that combine RF and LiFi. RF provides wide coverage, while LiFi offers high-speed data transmission. To manage resources in real-world scenarios with moving users and signal blockages, we use a model-free DRL technique named as deep deterministic policy gradient (DDPG). In this method, DRL

agent interacts directly with the environment to improve resource usage and network performance with the help of continuous state and action spaces. As a result, it achieves higher total data rates and better transmit power efficiency compared to traditional methods.

The last part of the thesis investigates the effect of random orientation of user equipment (UEs) on VLC, that is caused by dynamism of the users. It shows that combining VLC with RF ensures reliable connectivity across different network environments. It introduces two on-policy DRL methods advantage actor-critic (A2C) and proximal policy optimization (PPO) to improve resource allocation and load balancing in large, dynamic hybrid RF/VLC systems. Simulations reveal that A2C and PPO outperform existing methods, leading to significant improvements in data rates and overall system performance.

This thesis focuses on creating an efficient hybrid RF/VLC communication systems for 5G and future networks. It introduces new ideas on using DRL methods to optimize resources in real-time, even in complex scenarios.

Chapter 1

Introduction

1.1 Motivations

In recent years, wireless internet data traffic has seen phenomenal growth across the radio frequency (RF) spectrum. As reported by Federal Communications Commission (FCC) in 2010, the total available licensed spectrum is less than the required spectrum to accommodate such huge growth in internet users leading to a serious spectrum scarcity [1]. Further, the growth is estimated to fully saturate the RF spectrum by 2035 [2], highlighting significant future challenges. As per the recent Cisco report [3], 66% of the global population are having internet access with nearly 30 billion connections. As shown in Fig. 1.1, approximately 2 billion users are using 5G communication networks. As per the recent mobility report of Ericsson, 80% of mobile data traffic will be from 5G communication devices by the end of 2030 [4]. Handling such upsurging amount of mobile data traffic on the scarce RF spectrum is highly challenging. Hence, exploration of alternatives to RF communication becomes significant. Optical wireless communication (OWC) has been identified as an efficient alternative to RF communication.

OWC uses light signals for performing communications. Employing light for communication is a historical concept, with OWC dating back over three centuries. Early methods such as ship flags, semaphore, and fire beacons laid the groundwork for it. Another primi-

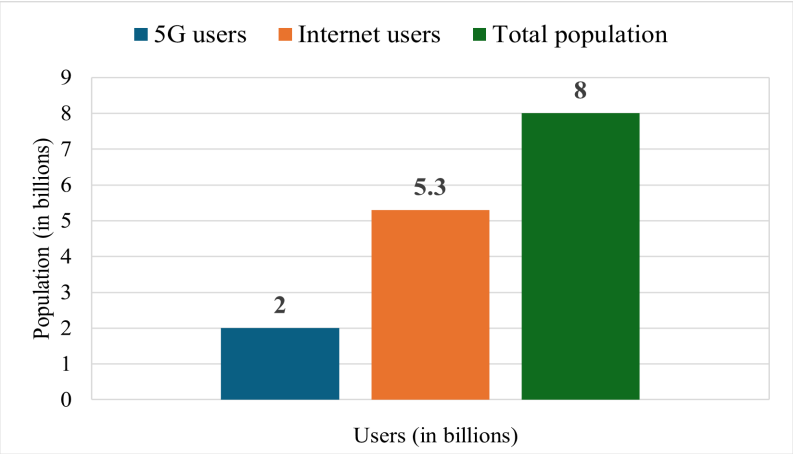


Figure 1.1: Global 5G users

tive technique involved reflecting sunlight using mirrors [5]. Alexander Graham Bell and Sumner Trainer pioneered the photophone in 1880, a revolutionary system of free space communication based on light. This groundbreaking invention enabled the transmission of sound through a beam of light [6]. Komine et al. [7] presented the visible light as a technology for wireless communication. Fig. 1.2 shows the visible light range in the electromagnetic spectrum [8].

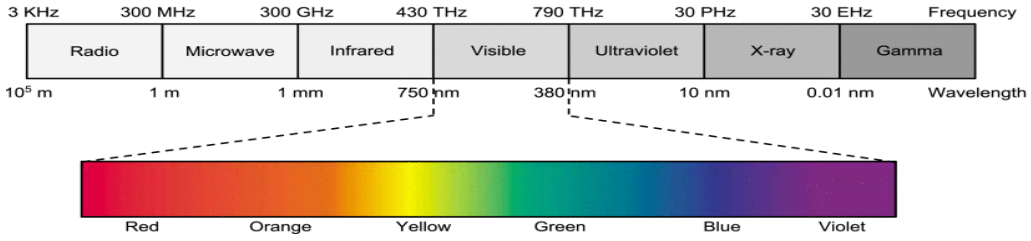


Figure 1.2: Visible light range in electromagnetic spectrum

The surge in demand for high data rate and license-free spectrum applications has prompted numerous researchers to explore visible light communication (VLC), a subset of OWC. VLC, sometimes also referred to as light fidelity (LiFi), is a broader term referring to any communication using visible light. It can be used for indoor position, and communication and follows IEEE 802.15.7 standards [9]. LiFi is a subclass of VLC only used for high speed internet access instead of WiFi and follows IEEE 802.11 standards [10]. It appears as a promising solution for indoor connectivity [7, 11]. However, standalone VLC network deployment is impractical due to its susceptibility to blockages. Thus, its co-deployment with RF network has been proposed, which creates a hybrid RF/VLC sys-

tem.

Hybrid RF/VLC is a type of heterogenous networks (HetNets) [12] that can provide higher data rates with broader coverage areas. HetNets are multi-tier networks in which several small cells spatially co-exist using the same spectrum bands [12, 13]. As a specialized class of HetNets, hybrid RF/VLC can make use of the large capacity of VLC links and the uninterrupted connectivity of RF links [14]. Thus, they enhance user rates and mobility on the one hand and optimize system's overall power and bandwidth usage on the other [15]. Synergy of both the technologies provides better data rates with uninterrupted communication [16]. Hybrid RF/VLC systems has emerged as a viable alternative within indoor environments. It has garnered considerable research attention due to its ability to provide traffic decongestion in densely populated RF environments. Apart from cumulating the advantages of both RF and VLC in a single network, hybrid RF/VLC also enhances mobility and energy efficiency in communications.

Similar to the joint resource allocation and association optimization in HetNets [17, 18, 19], optimal resource allocation and association optimization in hybrid RF/VLC systems remains a key research topic. In [20], the subject of resource allocation that maximizes feasible sum-rates in hybrid RF/VLC has attracted a significant attention. The downlink bandwidth and transmission power assigned to the access points (APs) for transmission of data, along with the affiliation of user equipments (UEs) with the APs to receive the downlink data, have a substantial impact on the system's possible sum-rate.

Optimizing resource allocation and association in hybrid RF/VLC involves non-concavity and integer optimization [21]. To address these dual challenges, several conventional optimization algorithms have been proposed in the existing literature [22, 20, 23, 24, 25, 26]. However, conventional optimization mechanisms often rely on assuming values for at least one of the optimization parameters and finding the best values for the remaining parameters [27]. Notably, as the optimization parameters jointly impact the data rate, their joint optimization must be free of any presumptions on their values. The assignment of downlink power and bandwidth has a direct impact on signal-to-interference-plus-noise-

ratio (SINR), and vice-versa. Presuming a value for the downlink bandwidth, transmission power, or association parameter results to suboptimal outcomes. A robust solution can be obtained by the inclusion of the interplay between each optimization parameter and the objective function in a comprehensive joint optimization problem.

To address above issues, deep learning, a flagship of machine learning, has been found to have high potential for resource optimization in hybrid RF/VLC. It can solve the optimization problem with higher efficiency and accuracy [17]. Instead of optimizing block-by-block from transmitter to receiver, as done by current technologies, deep learning facilitates the optimization of the whole system, incorporating the interplay between different optimization parameters [28]. When variables have a strong inter-relation, model-based mechanisms severely suffer as they involve optimization of one parameter at the cost of an other [27]. As deep learning is dependent on a moment-to-moment update, it can outperform sequential optimization methods.

Investigations on the application of deep learning and deep reinforcement learning (DRL) methods for optimal resource allocation in hybrid RF/VLC have been extensively carried out in [17, 29, 30, 31]. In [17], the authors have used deep Q-network (DQN) learning to solve the joint optimization problem of association parameter, bandwidth, and transmission power. Wang et al. in [29] use deep learning for seamless handover and increased downlink data rate in an ultra-dense deployment of VLC APs. Deep learning based solutions have been used against the heuristic methods to improve the stability and optimize the transmit power [30]. Along with transmission power, Yifei Wei et al. in [32] also considered user scheduling using deep learning techniques for hybrid energy supply. It considers renewable energy harvesting technique along with the conventional. Mohamad Azizi et al. in [33] uses deep learning mechanisms for enhanced energy efficiency (EE) and quality of service (QoS). Helin Yang et al. in [34] proposed deep learning based EE uplink and downlink resource management in HetNets. Parvez Shaik et al. in [35] considered outage probability and human blockers consideration in deep learning based hybrid RF/VLC dynamic communication systems. Liqiang Wang et al. in [29] proposed deep learning based seamless handover protocol in for hybrid network architecture. Duc M. T.

Hoang et al. in [36] and Danya A. Saifaldeen et al. in [37] uses deep learning based mechanisms for enhanced secrecy capacity and reliable data rate in hybrid VLC communication systems.

This thesis work is focused on joint optimization of the downlink bandwidth, transmission power of the APs, and the integer association parameter for achieving maximum achievable sum-rate for UEs. We have used state-of-the-art deep learning techniques to achieve this goal.

1.2 Contributions and Thesis Organisation

The thesis has been organized as follows:

- **Chapter 2**: Related works and Literature Review - A comprehensive review of existing research on hybrid RF/VLC systems and deep learning applications in communication systems is presented in this chapter. The review covers traditional optimization methods, recent developments in deep learning based approaches, and the evolution of hybrid RF/VLC systems. Gaps in the current literature are identified, highlighting the need for advanced, model-free DRL techniques to tackle resource allocation problems in dynamic and large-scale networks along with transfer learning.
- **Chapter 3**: Transfer Learning in Dynamic hybrid WiFi/ LiFi - This chapter presents a DQN combined with transfer learning based approach to optimize resource allocation in hybrid WiFi/LiFi systems. Here, the wireless fidelity (WiFi) network has been used as the RF network. The chapter explores the challenges in dynamic environments where users frequently enter and exit the network, requiring adaptive optimization of bandwidth, power, and user association. The proposed DQN-based method addresses these challenges by leveraging transfer learning to quickly adapt to new users without retraining, improving overall network throughput. Simulations validate the algorithm's efficiency, showing superior performance in dynamic conditions compared to existing optimization approaches.

- **Chapter 4**: Actor-critic deep deterministic policy gradient (DDPG) in hybrid RF/LiFi systems - This chapter focuses on a model-free DRL approach namely DDPG for efficient resource allocation in hybrid RF/LiFi networks. The method does not rely on predefined models, instead it learns from real-world interactions within the environment. It effectively handles challenges such as blockages and user mobility by dynamically optimizing resource allocation. The chapter highlights how the DRL model significantly improves network performance, enhancing data rates and power efficiency over traditional techniques, as demonstrated through simulations.

- **Chapter 5**: Advantage actor-critic (A2C) and proximal policy optimization (PPO) with random orientation in hybrid RF/VLC - This chapter investigates advanced on-policy DRL algorithms for resource allocation and load balancing including random orientation of UEs in hybrid RF/VLC systems. Two schemes namely A2C and PPO are developed to handle randomly oriented UEs and their demands of dynamic, and large scale networks. The chapter explores how these algorithms optimize data rates and improve load balancing efficiency, offering substantial performance gains over existing solutions. Simulation results demonstrate the superiority of these methods in achieving higher data rates and better resource allocation compared to other reinforcement learning techniques.

- **Chapter 6**: Conclusion and Future scope - This chapter summarizes the contributions of the thesis, emphasizing the impact of DRL techniques on resource allocation and load balancing in hybrid communication systems. It discusses the key findings, the improvements over traditional methods, and potential future directions for research in this field.

Chapter 2

Related works and Literature Review

In this chapter, a detailed literature review has been performed on the existing researches on VLC, hybrid RF/VLC, and deep learning approaches used in communications. From past few years, VLC has been significant area of research. The development and usage of light emitting diodes (LEDs) over the incandescent bulbs have given tremendous boost to the area. LED lights offer several advantages over incandescent and fluorescent lights in terms of energy efficiency, light density, longer lifetime span, reliability, low power consumption, and minimal heat generation [38]. LEDs are widely used in common life for general illumination, automotive headlights, traffic signals, displays, and smart lighting applications like automotive light management and VLC [39]. Komine et al. [7] from their laboratory proposed indoor VLC using white LED. They have shown the usage of visible white light as a medium for communication. Using LEDs that serve as a transmitter that emits both light and information signals to users serves the dual purpose. In a typical VLC system the receiver is equipped with a photodetector (PD) to receive the light signal and convert it into an electrical signal. The electrical signal is further processed to retrieve the transmitted data.

2.1 Overview

VLC is a form of short-range OWC that utilizes the visible light spectrum, ranging from 380 nm to 780 nm [8, 9]. VLC works by modulating the intensity of light sources known as intensity modulation/direct detection (IM/DD) technique, at speeds faster than the human eye can detect [40]. VLC uses dimming and flickering to transmit data which is invisible to eyes. Flicker refers to the variation in light brightness, which can lead to noticeable and harmful physiological effects in humans. Generally, frequency greater than 200Hz is considered safe [41]. Another important consideration is dimming support as it is used for transmission of data. As dimming leads to power saving but communication need to be maintained. The characterization of deterministic and stochastic VLC channels has been primarily investigated through simulation-based studies in various environments, including indoor settings [42, 43, 44], underground mines [45], and outdoor areas [46] .

Furthermore, numerous collaborative initiatives were launched worldwide, including the Japan Electronics and Information Technology Industries Association (JEITA) set standards for a "visible light ID system" in 2007. The following year, in 2008, the Visible Light Communications Consortium (VLCC) released a Specification Standard. Concurrently, in Europe, the Home Gigabit Access project (OMEGA) is advancing the development of VLC for home networks [47].

In indoor environments, a VLC cell typically covers only a few square meters due to the inherent properties of light. Generally, rooms can have multiple light sources, allowing for high spatial spectral efficiency with VLC. However, despite the dense deployment of APs, VLC does not offer uniform coverage because optical signals are prone to blockages. When light beams are obstructed, the data rate decreases due to low optical channel gain. Studies have shown that while VLC networks can deliver very high data rates, their outage rate performance in multiuser environments can be significantly low [16]. VLC is secure, license-free, and have no RF interference. They also have a huge bandwidth potential compare to RF counterparts [48]. However, VLC experiences issues when it goes non-line-of-sight (NLOS). Unlike RF networks, it operates effectively only when there is

a clear line-of-sight (LOS) between the transmitter and receiver. Therefore, standalone deployment of VLC may lead to obstruction in communication. To maintain ubiquitous connectivity for UEs, integrating an RF AP with the VLC network enhances coverage, maintain connectivity and boosts overall system capacity[49].

2.1.1 Hybrid RF/VLC

Hybrid RF/VLC system is consider to be more energy efficient than the standalone. Kashef et al. in [50], have discussed about the energy efficient hybrid RF/VLC systems in terms of bandwidth and power allocation in HetNets. Khreishah et al. in [51] have shown that the hybrid RF/VLC is more energy efficient particularly in terms of power than the standalone. Basnayaka et al. in [16] have shown that hybrid RF/VLC network can enhance both the average and outage data rate performance. Since this system operates on non-overlapping spectra, hence can harness the advantages of both the technologies. RF systems offer widespread coverage, ensuring consistent throughput across various locations. This integration can deliver a combined system throughput that surpasses the standalone VLC or RF networks can achieve, without causing mutual interference. Hence, combining advantages of both the network and forming a hybrid RF/VLC network which can significantly enhance both system's throughput and user experience. A hybrid RF/VLC systems belong to HetNets.

2.1.2 HetNets

One of the essential technologies within the diverse range of HetNets technologies is VLC [52][53]. HetNets have been proposed to integrate different wireless technologies, enhancing overall system capacity [54]. With huge surge in data demands, HetNets have been suggested to address these demands and are widely seen as a practical solution to accommodate this exponential growth [55, 56, 54]. These enhancements are possible due to the diversity in fading channels, propagation losses, and available resources across various networks. HetNets are also known as multi-tier networks. In these type of networks, assignment of APs is of utmost importance. In [55], the authors have introduced

user association to the AP with minimum distance and maximum SINR based condition. They have shown that the system achieves this with optimal balance load performance [57]. However, to realize these benefits, the critical challenge lies in developing resource allocation algorithms that effectively distribute power and bandwidth among HetNets to meet diverse service requirements [50]. Generally, the combined optimization of resource allocation and association continues to be a key research area in HetNets [14, 13, 58, 59].

2.1.3 Resource Allocation

Similarly, resource allocation is an important research area in hybrid RF/VLC systems. The affiliation of UEs with APs for downlink data receipt, as well as the allocation of downlink bandwidth and transmission power to APs for data transmission, all have a substantial impact on the system's possible sum-rate. As a result, the topic of optimal resource allocation to maximise the possible sum-rate in hybrid RF/VLC systems has received significant scientific interest [55, 60, 21, 61, 50]. Several studies have addressed the resource allocation in hybrid RF/VLC systems, targeting the various objectives such as sum-rate maximization [20, 62, 31, 18, 19] , spectral efficiency [55, 63] , power consumption [64, 65], and energy efficiency [66, 67]. Rui Jiang et al. in [27] have shown joint optimization of user association and power allocation for the sum-rate maximization and improved system performance in a cell-free VLC network. Mohanad Obeed et al. in [60] proposed an iterative algorithm for joint optimization of load balancing and allocation of power with focus on maximizing the achievable data rate. The common challenge faced in these joint optimization problem is non-concavity [21].

2.1.4 Non-concavity

The challenge of non-concavity in the downlink resource allocation problem is persistent difficulty faced in these studies, particularly when involving the joint optimization of downlink bandwidth, transmission power, and the association parameter. Since the association of UEs to APs may depend on various conditions such as minimum distance [55], maximum SINR [17], maximum power [68] depending upon the optimization technique

used. Typically, this issue is addressed by presuming values for at least one of these parameters, then using standard convex optimization techniques to determine the optimal values for the remaining parameters. However, presuming values for the association parameter, downlink bandwidth, or APs' transmit power might not be the optimal approach for maximizing the system's potential sum-rate. A robust solution can be achieved through a comprehensive joint optimization process that considers the interdependence of each parameter and its effect on the objective function. A combined optimization of association parameter, downlink bandwidth, and transmission power without any presumptions is necessary. Since presuming the value may not provide the optimal robust solution. Therefore, to get optimal solution, exploration of machine learning and deep learning techniques have been done in several studies.

2.1.5 Learning based Joint Optimization

Learning based solutions are capable of providing the robust solution for such kind of joint optimization problems. Given the challenge of non-concavity of the optimization problem, which compromises precision and accuracy, hence limiting the scope. Several studies have explored the machine learning and deep learning techniques in application of hybrid RF/VLC systems. Recently, machine learning techniques have been introduced for channel modeling to address the high complexity and site-specific constraints of deterministic methods, as well as the accuracy limitations inherent in stochastic models [69]. Machine learning based channel modeling seeks to create accurate, low-complexity models for complex channels by directly learning data patterns without relying on assumed analytical expressions. Additionally, machine learning models differentiate between scenarios by using physical parameters specific to each scenario as inputs [69, 70]. Deep learning techniques have demonstrated significant potential in handling a variety of intelligent tasks. In recent years, the fields of machine learning and, more specifically, deep learning have experienced tremendous growth, with their applications now spanning nearly every industry and research domain.

Reinforcement learning [71] has emerged as a crucial area of research in machine learn-

ing, significantly influencing the development of Artificial Intelligence (AI) over the past two decades. In reinforcement learning, an agent makes periodic decisions, observes the outcomes, and adjusts its strategy to achieve the optimal policy. Despite its proven convergence, this learning process can be time-consuming due to the need for extensive exploration and system understanding, making it impractical for large-scale networks. Consequently, the practical applications of reinforcement learning have been limited.

Recently, the advent of deep learning [72] has introduced a breakthrough technique that addresses these limitations. This advancement has led to the development of DRL, which leverages the power of Deep Neural Networks (DNNs) to enhance the learning process, improving both the speed and performance of reinforcement learning algorithms. As a result, DRL has been widely adopted in various practical applications, including robotics, computer vision, speech recognition, and natural language processing. In the fields of communications and networking, DRL has recently emerged as a powerful tool to tackle various issues and challenges. Specifically, contemporary networks like the Internet of Things (IoT), HetNets, and Unmanned Aerial Vehicle (UAV) networks are becoming increasingly decentralized, ad-hoc, and autonomous [73]. DRL is recognized as an efficient learning mechanism. It operates through interaction with the environment, requiring minimal prior information. As an online learning method, DRL has been extensively studied in the field of AI [74].

Q-learning is one of the most popular reinforcement learning technique, initially proposed in [75]. The convergence theorem for Q-learning was later established in [76]. In [77], an autonomous Q-learning algorithm was introduced for HetNets to optimize resource allocation for device-to-device (D2D) communication. This approach formulates a utility function as the difference between achievable throughput and power consumption cost, modeled as a stochastic non-cooperative game. Here, each D2D pair acts as a player and learning agent, tasked with determining its optimal strategy. Additionally, in [78], an online reinforcement learning approach was utilized to address the association problem in vehicular networks, leveraging the regularities in the network features. Ghadimi et al. introduced a reinforcement learning approach for rate adaptation in cellular networks in

[79]. However, it is important to note that achieving an optimal solution using the Q-learning method becomes challenging when the state and action vectors in the joint optimization problem are large. In this context, deep learning [80] has emerged as a promising technique for addressing issues involving large state and action vectors. Recently, deep learning-based methods have been applied to various areas, including dynamic channel access [81], power allocation [82], mobile offloading [83], cloud radio access networks [84], interference management [85], and mobile edge computing and caching [86]. We will first explore the usability of deep learning in communication networks.

The integration of machine intelligence into future mobile communication networks is garnering significant research interest. Deep learning, a flagship of machine learning, is particularly captivating the attention of communication network researchers. Studies like [87] and [88] have explored its potential to address challenges in the mobile networking domain, encouraging the use of deep learning in 5G mobile communication systems, which are predominantly HetNets. These systems generate highly heterogeneous data, originating from various formats with complex correlations [89]. Traditional machine learning tools struggle with these challenges due to their lack of performance improvement with increased data [90] and their inability to handle high-dimensional state/action spaces [80].

In contrast, deep learning thrives on big data, eliminating the need for domain expertise and utilizing hierarchical feature extraction. Consequently, it has become a highly effective solution for addressing problems in communication networks, especially in HetNets. A comprehensive applications of deep learning in communication systems are discussed in following works such as authors in [91], discuss about deep learning for network cybersecurity; in [92], which reviews approaches for network traffic control; in [93], which presents methods for physical layer modulation, resource allocation, and network routing; and [94], which explores emerging issues like edge caching and computing, multiple radio access, and interference management.

A significant advantage of DRL is its ability to solve complex network optimization

problems. This capability allows network controllers, such as base stations, to address non-convex and intricate issues like joint user association, computation, and transmission scheduling, achieving optimal solutions even without complete and accurate network information.

Chapter 3

Transfer Learning in Dynamic Hybrid WiFi/ LiFi

As discussed in the previous chapter, the demand of newer technologies for fifth generation mobile communication systems has been continuously growing in the current arena. A hybrid form of RF and VLC has emerged as a promising candidate to fulfill this demand. In this chapter, we discuss standardized subsets of RF and VLC technologies, termed as WiFi and LiFi, respectively. Such a system is thus termed as a hybrid WiFi/LiFi communication system. The joint optimization problem of bandwidth, user association parameter and transmission power for sum-rate maximization in these hybrid systems is non-concave. DQN learning based algorithms offer solution to non-concavity. However, existing DQN learning based solutions are often restricted to static networks. They face complexity issues in dynamic networks. In this chapter, we address the dynamic hybrid WiFi/LiFi communication system with DQN transfer learning algorithm. Transfer learning is used to gather information about a newly entering UE in the network, thereby improving the overall throughput of the network. Simulations show that the proposed algorithms perform well than the existing optimization algorithms in throughput maximization.

3.1 Overview

The demand for wireless data has been increasing exponentially, particularly in education and industry. As mentioned in Chapter 2, the present conventional WiFi communication based technology may fail to fulfil the desirable data requirement in near future [8]. LiFi can be a powerful supplement to conventional WiFi based systems. It uses the indoor deployed LEDs for communication. The deployed LED lights are used for data transfer using dimming of light. They provide various advantages like high data rate, lesser interference, high energy efficiency and better security, unregulated bandwidth in visible spectrum range, illumination and communication simultaneously [38]. However, certain limitations like inefficiency of NLOS components prevent it's stand-alone deployment [8]. The proposal of hybrid WiFi/LiFi system has been found as a solution to the problem [95, 96].

A typical hybrid WiFi/LiFi is a merger of WiFi and LiFi systems. The light sources present in the indoor setup act as multiple LiFi APs. Generally, one or more WiFi APs are also present. An UE present in the set up can connect to any of the APs. When connected to a LiFi AP, it gets a high data rate. However, when LOS components are unavailable, the minimum required signal-to-noise ratio (SNR) is not maintained. In this situation, it gets connected to WiFi AP. In this way, both the networks compensate for the limitations of each other. Further, practical hybrid WiFi/LiFi observe a continuous change in the number of UEs in the set-up. For instance, consider an airport waiting area where UEs are entering or exiting the room continuously. Such situations demand efficient handling of the dynamism in the network, to quickly associate a newly entering UE with the AP that is offering the highest data-rate to it.

Another significant issue in hybrid WiFi/LiFi is resource allocation. Resource allocation primarily implies the allocation of bandwidth, transmission power, and association on the AP-UE links. The parameters significantly affect the achievable sum-rate [62, 97, 98, 99, 42, 100, 101, 102]. A crucial issue encountered in these works is non-concavity of the achievable sum-rate maximization problem. A common way of addressing this problem

is by presuming values of one of the resource parameters. However, the problem becomes more challenging and complex when new UEs are entering or leaving the system. Such a system can be termed as a dynamic hybrid WiFi/LiFi system. The association of UEs depends on SINR, which depends on bandwidth, transmission power and association parameter. Presuming a value for them significantly affects the system performance. Hence, getting optimal solution without such presumptions is required.

To solve the above problem, we aim to jointly optimize the transmission power, bandwidth and association parameter with *DQN transfer learning* to maximize achievable sum-rate for a dynamic hybrid WiFi/LiFi systems.

3.1.1 DQN transfer learning

DQN learning was proposed by Mnih et al. in 2015 [80]. It is a combination of reinforcement learning with DNN for Q-networks. Mnih et al. proposed DQN as an extension of classical Q-networks that uses DNN without reinforcement learning [76]. Reinforcement learning uses were limited to low-dimensional state spaces. However, DQN can learn successfully from high-dimensional inputs using reinforcement learning. The other challenges faced in using reinforcement learning were instabilities due to correlations and divergence due to non-linear function approximators in neural network. DQN provides two key ideas to solve these challenges [103]. First, experience replay that randomizes data, thereby removing correlations in the observation sequence and smoothing over changes in the data distribution. Second, we used an iterative update that adjusts the Q-function towards target values periodically, thereby reducing correlations with the target.

In reinforcement learning, an agent interacts with its environment. This interaction occurs through a sequence of observations, actions, and rewards. The agent observes the environment to gather information. Based on these observations, it selects and performs actions. These actions lead to certain outcomes, which the agent evaluates as rewards. The main goal of the agent is to choose actions strategically. By doing so, it aims to maximize the total reward it will receive over time.

The notion of transfer learning in neural network evolved between 1970 to early 1980s. In this method, a model created for one specific task is not discarded after completing that task. Instead, the model is reused to perform another related task. The knowledge gained from the initial model (first task) serves as a starting point for the second task. The first work on transfer learning was published by Bozinovski et al. in 1976 [104]. It was also termed as *learning by learning*, *sequential learning*, *adaptive generalization*, and *lifelong learning* [104]. Subsequent researches in 1977 [105] and 1978 [106] have shown experimental assessment of transfer learning. The authors in [105] and [106] mentioned two types of transfer learning, namely positive transfer learning and negative transfer learning. Positive transfer learning occurs when the first task helps the model learn the second task more efficiently. In this case, the second task requires fewer iterations or a shorter training sequence to achieve the same level of performance. This happens when the two tasks share similar features, allowing the model to generalize well. Since the model has already learned a useful representation of the data, it does not need extensive retraining. Thus, it is performed when the second task is showing shorter sequence for learning after the first task. On the other hand, negative transfer learning happens when the first task makes it harder for the model to learn the second task. This results in longer training sequences or degraded performance. Negative transfer occurs when the tasks have different or conflicting features, causing the model to overfit to irrelevant patterns from the first task. Thus it is performed when the second task is showing longer sequence for learning after the first task.

In 1981, research was performed for application of transfer learning in alphabets letter recognition from computer terminals [107]. In 1985 [108], this work was extended for modeling of supervised learning in neural network. As synaptic weights are not observable in biological systems, they can be observed in neural network. Bozinovski et al. discussed supervised learning problem representation without representing synaptic weights as primary concept. It was termed as teaching space. Multilayer neural network gained interest in 1986 with the publication of a book " Parallel Distributed Processing " by David E. Rumelhart et al. [109]. Transfer learning regained research focus in early 1990s, when

Pratt et al. [110] made additional developments to it by adding multi-task learning to it.

The most significant advantage of transfer learning is that it saves time and effort because the model does not need to be built entirely from scratch for the new task [111]. The previous task knowledge acts as a source to the new task or target, making it somewhat analogous to human behavior. Human beings generally learn from their past experience and apply that experience while performing that same task or another related task in present or future. Similarly, transfer learning uses neural network already trained for one task to speed-up the training process for a new task. This approach with transfer of learning or knowledge is widely used in modern DNN to improve learning efficiency.

The joint optimization and resource allocation in 5G and beyond systems are complex and having large-scale links. Several researchers have explored transfer learning for resource management in 5G wireless communication systems for making them faster, more secure and energy efficient [112]. Wang et al. discuss about varied application of transfer learning in wireless communication systems such as indoor wireless localization, APs switching efficiency, spectrum allocation, etc [113]. Du et al. [43] uses WiFi with a knowledge transfer approach to choose between the RF and LiFi networks in hybrid WiFi/LiFi systems. Yang et al. [114] proposed reinforcement learning based transfer learning for low latency and high reliability in an energy efficient hybrid WiFi/LiFi systems.

3.2 Chapter Contributions

In this chapter, we aim to jointly optimize the transmission power, bandwidth and association parameter with *DQN transfer learning* to maximize achievable sum-rate for a dynamic hybrid WiFi/LiFi systems. In the existing literature, DQN learning has been used for static hybrid WiFi/LiFi system. However, DQN learning fails to perform optimally in dynamic hybrid WiFi/LiFi environment. First we consider a static hybrid WiFi/LiFi set-up with a fixed number of UEs and APs. The controlling unit (CU) of this system optimizes the bandwidth on each of the AP-UE links, the transmit power of the AP and the association parameter between the APs and the UEs with the help of DQN learning algorithm.

However, when a new UE enters, the static system transforms into a dynamic system. To optimize the bandwidth, transmit power and association parameter of this dynamic system, the DQN learning information at the UE nearest to the newly entered UE is transferred to the newly entering UE with the help of DQN transfer learning algorithm.

The main contributions made in this chapter can be enlisted as follows:

- *Comprehensive assessment of the resource allocation problem:* Addressing the joint optimization problem comprehensively by incorporating the bandwidth, transmission power, and association parameter. The resource allocation problem for hybrid WiFi/LiFi system is neither concave nor convex. DQN learning technique is applied to solve this problem. The optimal solution is obtained with moment-to-moment update and without modeling errors.

- *Practicality of the scheme:* We have considered the practicality of the system and considered idle APs. Idle APs are APs which are not taking part in communication due to hardware malfunctioning. The consideration of idle APs has been made in SINR expression to make the system more comprehensive and practical.

- *Dynamicity with transfer learning:* DQN transfer learning has been introduced in the system. When a new UE joins the set-up, it learns from the experience of already existing UE. This makes the system more efficient in quick convergence, as the newly entered UE requires 54% less iterations.

The rest of the chapter is organized as follows: Section 3.3 discuss the system model, resource allocation algorithm for DQN based hybrid WiFi/LiFi system is discussed in section 3.4. Results of simulation are shown in 3.5 and finally section 3.6 concludes the chapter.

3.3 System Model

Fig. 3.1 shows the set-up considered for investigation. It has a single WiFi AP (shown as RF AP in figure) and multiple LiFi APs (shown as VLC AP in figure) deployed on

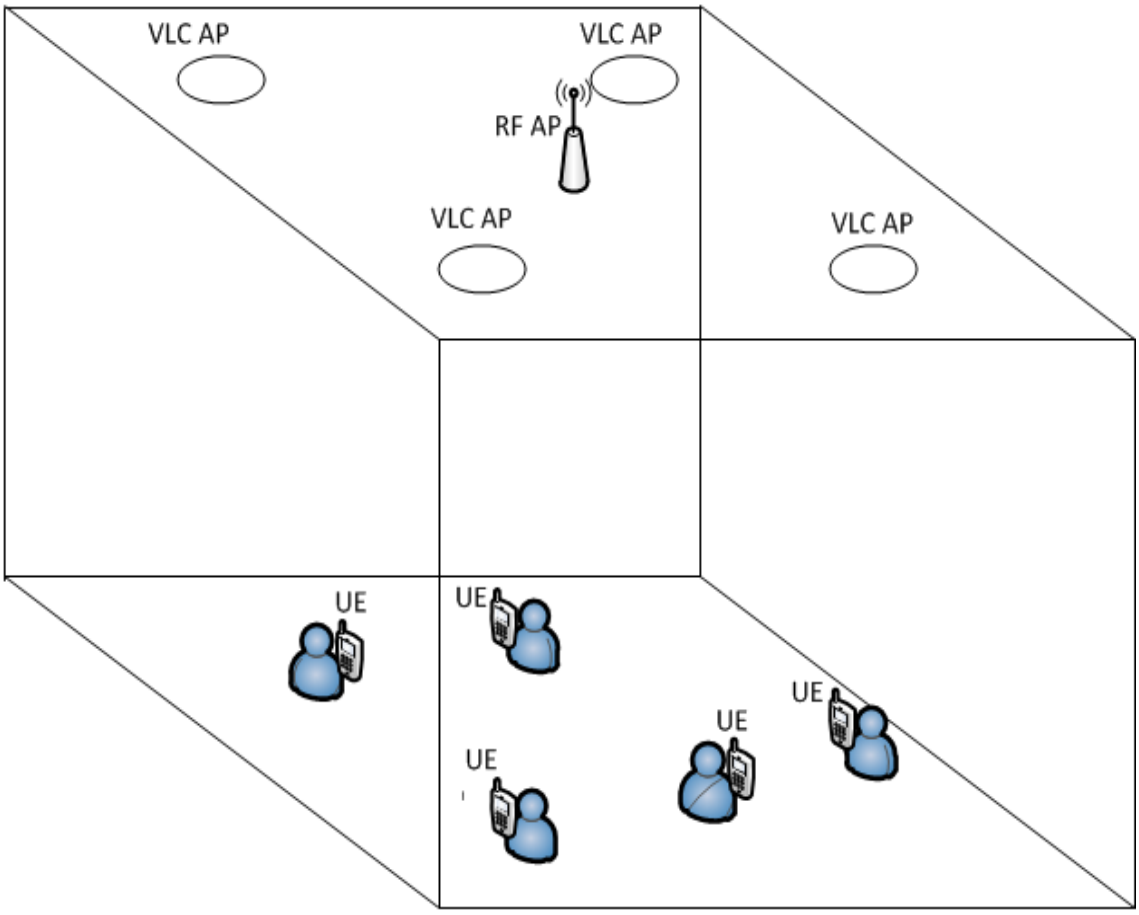


Figure 3.1: Hybrid WiFi/LiFi system

the ceiling of a room. The UEs are present on the floor. A newly entering UE is also considered in the set-up. Consider the set of APs as $\mathcal{D}$, where the APs are indexed as $d = 0, 1, 2, \ldots, |\mathcal{D}|$. The WiFi AP is indexed as $d = 0$ while the LiFi APs are denoted by indices $d = 1, 2, \ldots, |\mathcal{D}| - 1$. LiFi APs are light sources deployed in the room. The set of UEs $\mathcal{E}$ indexed as $e = 1, 2, \ldots, |\mathcal{E}|$ present in the room. The UEs are assumed at a constant height h from the floor.

3.3.1 Channel Model for Light propagation

The propagation of light is modeled with a Lambertian law model [50, 115] as most of the surfaces found in typical indoor set-ups like plaster walls and ceilings are ideal Lambertian reflector. In a typical indoor set-ups, plaster walls and ceilings are present. These surfaces provide excellent reflective properties for diffuse links. These diffused reflections from different indoor surfaces are well approximated by ideal Lambertian reflector. Diffused

links are the non-directed links that do not require any alignment between the transmitter and receiver. These paths can be categorized as either LOS or NLOS. Typical LOS path links requires direct path between transmitter and receiver. LOS links are unobstructed paths and it gets highly affected by shadowing. However, NLOS components are obtained from the diffused reflections from plaster walls, ceiling and other reflectors. As diffuse links do not require any direct paths, they are more robust to shadowing [115, 116].

LiFi systems assume diffuse reflection. Diffused reflections come from multiple directions, hence cannot be accurately modeled by simple ray-tracing laws that assume perfect linear propagation. Simple ray optics does not consider wave-like behaviors such as diffraction and interference, especially when encountering obstacles. Simple ray laws are insufficient for LiFi because they neglect critical aspects such as diffuse reflection, wavelength-specific behavior, wave phenomena, and real-world complexities such as multi-path propagation and receiver properties [117].

Consider LiFi APs as transmitters and UEs equipped with PD as receivers, in an indoor environment without reflectors. If the distance between the LiFi AP and UE is significantly larger than the area of photodiode, the received irradiance will be nearly uniform across the detector's surface. Hence, all the signal energy will reach the UE almost simultaneously. Therefore, the impulse response for this system can be approximated as a scaled and delayed Dirac delta function [115]. The LiFi APs transmit data to UEs on the downlink. In this process, light propagation in LiFi is modeled using diffused reflection, where an incident light ray on a surface is scattered at various angles.

The optical power of light after diffused reflection is described by the Lambertian law [50]. According to this model, the LOS direct current (DC) channel gain CG^v_{de} is given as

$$CG^v_{de} = \frac{(m+1)A_{\text{pd}}\cos^m\phi_{de}\cos\psi_{de}T_{\text{opt}}(\psi_{de})g(\psi_{de})}{2\pi d^2_{de}}, \tag{3.1}$$

where $T_{\text{opt}}(\psi_{de})$ is the optical receiver filter gain which is constant or unity within field-of-view (FOV) of receiver, ϕ_{de} is the angle of incidence at UE e from AP d, ψ_{de} is the

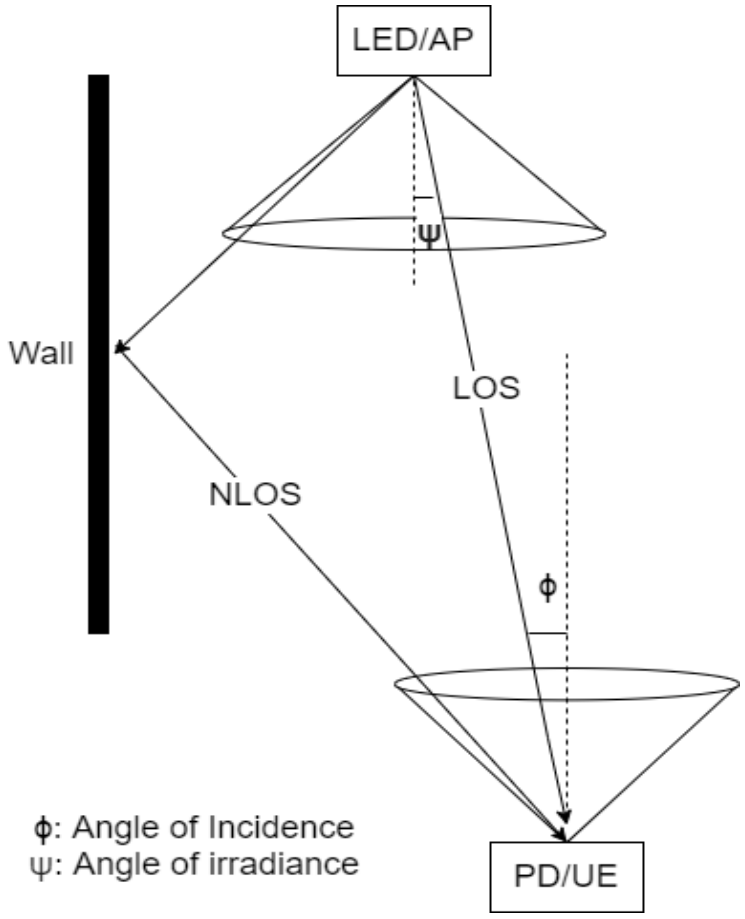


Figure 3.2: Representation of LOS and NLOS link in hybrid WiFi/LiFi system

angle of irradiance at AP d, d_{de} is the Euclidean distance between the dth UE and the eth AP, A_{pd} is the area of photodiode, and $g(\psi_{de})$ is the concentrator gain given as

$$g(\psi_{de}) = \begin{cases} \frac{n^2}{\sin^2 \psi_{\text{FOV}}} & \text{if } 0 \leq \psi_{de} \leq \psi_{\text{FOV}} \\ 0 & \text{if } \psi_{de} >_{\text{FOV}}, \end{cases} \tag{3.2}$$

ψ_{FOV} is the angle of receiver FOV of UE, the refractive index n is given as

$$n = \frac{\text{speed of light in vaccum}}{\text{speed of light in that optical material}}, \tag{3.3}$$

and m is the mode number of radiating lobe, also known as *order of Lambertian radiation*, and is given as

$$m = -\frac{\ln 2}{\ln \cos \psi_{1/2}}, \tag{3.4}$$

Table 3.1: Important notations and their meanings

Notation	**Meaning**
d	Index of APs
e	Index of UEs
k	The interferer AP index
r_{de}	The achievable data rate between the dth AP and the eth UE
r_d	Downlink data rate of the dth AP
$B_{\text{max}}^{\text{LiFi}}$	Maximum BW allotted to LiFi AP
$B_{\text{max}}^{\text{WiFi}}$	Maximum BW allotted to WiFi AP
P_d	Transmit power of the dth AP
P_{eff}	Total transmit power from an AP
$\mathcal{D}$	Set of APs
$\mathcal{E}$	Set of UEs
m	Lambertian coefficient
ϕ	Angle of incidence
ψ	Angle of irradiance
$CG^{(p)}$	DC Channel gain after pth reflection
A_{pd}	Area of PD
CG_{EffRef}	Effective channel gain after reflection
$P_q^{(p)}$	The optical power of the reflected light wave at the pth reflecting point emitted from the qth transmitting AP
CC	Channel capacity
B_{de}	BW allotted to the eth UE by the dth AP
CG_{de}	Channel gain between dth AP and eth UE
ρ_e	Responsivity of the receiver PD
a_{de}	Indicator function showing association of AP d-UE e
N_0^{WiFi}	WiFi noise power
N_0^{LiFi}	LiFi noise power
$P_{\text{max}}^{\text{WiFi}}$	Maximum transmit power of the WiFi AP
$P_{\text{max}}^{\text{LiFi}}$	Maximum transmit power of the LiFi AP

where $\psi_{1/2}$ is half power semi-angle at half illuminance. $\psi_{1/2}$ represents the angle within which half power of the incoming light is concentrated. It can be seen that is inversely proportional to $\psi_{1/2}$. Thus, as $\psi_{1/2}$ increases, the directionality and the value of m decreases and vice-versa. It can be seen that the parameter m tells about the directionality of light emitted. Larger the value of m, higher is the directionality. For a traditional Lambertian source, $m = 1$.

Next, the channel gains of the NLOS light components received by the PD at a UE are calculated. The $(l-1)$th reflection point acts as the source of the lth reflected light ray. The lth reflection point is treated as a virtual receiver, while the $(l-1)$th reflection point acts

as a virtual light source. The authors in [116] show that the total channel gains between pairs of reflecting sites determine the effective DC channel gain, CG_{EffRef}, for a light ray traveling through multiple reflections. Mathematically,

$$CG_{\text{EffRef}} = \sum_{p=0}^{\infty} CG^{(p)}, \tag{3.5}$$

where $CG^{(p)}$ represents the DC channel gain after the pth reflection from the source LED. The variable p refers to the reflection index. This can also be expressed as

$$CG^{(p)} = \int_S CG_1 CG_2 \dots CG_{p+1} P_q^{(p)} dA_s, \tag{3.6}$$

$P_q^{(p)}$ represents the optical power of the reflected light ray after p reflections from the qth transmitting LiFi AP. The term dA_s refers to a very small reflection surface area. The integration considers the tiny areas of all wall surfaces as variables. $CG_1, CG_2, \dots, CG_{p+1}$ are the DC channel gains for the path of each reflected component, as described in [116], and are given as

$$\begin{aligned}
CG_1 &= \tfrac{(m+1)A_s}{2\pi d_1^2} \cos^m(\phi_1) \cos(\psi_1), \\
CG_2 &= \tfrac{A_s}{\pi d_2^2} \cos^m(\phi_2) \cos(\psi_2), \\
&. \\
&. \\
&. \\
CG_{p+1} &= \tfrac{A_s}{\pi d_{p+1}^2} \cos^m(\phi_{p+1}) \cos(\psi_{p+1}) T_{\text{opt}}(\psi_{p+1}) g(\psi_{p+1}),
\end{aligned} \tag{3.7}$$

for $b = 1, 2, \dots, p+1$, the irradiance angle is ψ_b and the incidence angle is ϕ_b at the p reflections. The variable A_s represents the incident surface area. Here, b is a dummy variable. The DC channel gain between the eth PD receiver and the dth LiFi AP is CG_{de}^v (3.1). The channel gain between the dth LiFi AP and the first reflection point is CG_1 (3.7). Similarly, the channel gain between the second and third reflection points is CG_2, and the gain between the pth reflection point and the receiver PD is CG_{p+1}. The channel gains have a similar shape at all the reflecting sites. T_{opt} and $g(\psi_{p+1})$ are properties of

the receiving PD. The function CG_{p+1} depends on these variables. It represents the gain between the receiving PD and the last reflection point. If the spectral reflectance of the material at the pth reflecting point is $\Gamma_p(\lambda)$, then $P_q^{(p)}$ is given by

$$P_q^p = \int_\lambda P_d(\lambda)\Gamma_1(\lambda)\Gamma_2(\lambda)\ldots\Gamma_p(\lambda)d\lambda. \tag{3.8}$$

We assume that the surface of each reflecting points is made of the same material. We also assume that for all d, $\Gamma_d(\lambda) = \Gamma$ for $d = 1, 2, \ldots, p$. This is because Γ_d depends on λ.

The total of the LOS and NLOS components gives the effective received optical power, P_{eff}, from a single LED. It is expressed as

$$P_{eff} = CG_{\text{EffRef}}P_d + CG_{de}^v P_d = CG_{de}P_d \text{ for } d \in \mathcal{D}\backslash\{0\}. \tag{3.9}$$

The effective channel gain between AP d and UE e is given by $CG_{de} = CG_{\text{EffRef}} + CG_{de}^v$ for $d \in \mathcal{D} \setminus \{0\}$. Here, P_d is the power transmitted by LiFi AP d.

3.3.2 Channel Model for WiFi signal propagation

The signal received by UE e from the WiFi AP d indexed as $d = 0$ adheres to the WiFi signal propagation model, incorporating both fading and path loss into the power channel gain. We describe the received signal power using the WINNER-II channel model. WINNER II model supports a wide range of propagation scenarios, including indoor, outdoor, urban, rural, and suburban environments. LOS and NLOS conditions make it more versatile and adaptable for several modern 5G and beyond wireless systems. Also, it comprehensively and meticulously considers advanced propagation phenomena like large-scale fading, small-scale fading, clustered propagation effects where signals are reflected, scattered, or diffracted in clusters. Several other features like Doppler spread and time evolution of clusters allow it to handle dynamic environment efficiently. The channel model for WiFi signal propagation from WiFi AP ($d = 0$) to UE e is given as [118]

$$G_{0e} = Ldt_{0e}^{-a_{0e}}\chi_{0e}, \tag{3.10}$$

where χ_{0e} denotes independent and identically distributed Nakagami fading channel parameter for WiFi links, dt_{0e} is the distance between UE e and WiFi AP ($d = 0$), and a_{0e} is the path-loss exponent. The value of L here is $L = 10^{X/10}$, where $X = M + N\log_{10}\left(\frac{f_c}{5}\right)$ describes the relationship, where f_c represents the carrier frequency in GHz, and M and N are propagation constants determined by the propagation model. For LOS environment, $M = 46.8$ and $N = 20$, where as for an NLOS environment, $M = 43.8$ and $N = 20$.

The channel parameter models the amplitude variations of a wireless signal due to multipath propagation. The probability density function (PDF) of χ_{0e} is characterized by the Nakagami-κ distribution as

$$PDF(r) = \frac{2\kappa^{\kappa}}{\Gamma(\kappa)\Omega^{\kappa}} r^{2\kappa-1}\exp\left(-\frac{\kappa}{\Omega}r^2\right), \quad r \geq 0, \tag{3.11}$$

where $\kappa \geq 0.5$ is the fading parameter, indicating the severity of fading and $r = \sqrt{\|y_{0e}\|^2}$ is amplitude of received signal. Ω is the average power of the received signal, and $\Gamma(\kappa)$ is the Gamma function. This versatile Gamma distribution encompassing various fading conditions. For $\kappa = 1$, it approximates to the Rayleigh distribution, representing severe fading with no LOS component. For $\kappa > 1$ to ∞, it approximates to the Rician fading distribution.

3.3.3 Achievable Data Rate

As mentioned earlier, our focus in this chapter is on maximizing the sum-rate of hybrid WiFi/LiFi systems. It is essential to understand how channel capacity (CC) for the achievable sum-rate works for a UE when connected to either a WiFi or LiFi AP. To determine the feasible data rate for a UE connected to a WiFi AP, we use the Shannon's capacity formula. However, when a UE connects to a LiFi AP, the application of Shannon's capacity has to be investigated. Though exact capacity calculations are yet under investi-

gation [15], the upper and lower bounds on achievable capacity have been derived [119]. When connected to a LiFi AP, a UE communicates through IM/DD of light, where the signal amplitude represents instantaneous optical power. Intensity modulation can be easily achieved by changing the bias current of an LED. In a direct-detection receiver, the photodiode generates a photocurrent that is directly proportional to the optical power it receives. Consequently, the signal must be real-valued and non-negative. Due to these constraints, directly applying the Shannon's capacity formula may not yield accurate results.

Investigations on the IM/DD CC impaired by Gaussian noise show that the lower bound serves as a useful approximation for CC in LiFi networks. Hranilovic et al. [120] approximated lower and upper bounds of CC for bandwidth and power constrained Gaussian noise corrupted intensity modulated channels. Lapidoth et al. [119] investigated CC for upper and lower bounds in optical channel. It considers additive white Gaussian noise (AWGN) corrupted output with non-negative channel inputs. Farid et al. [121] investigated the CC using pulse amplitude modulation and given lower and upper bounds. As investigated by authors in [120, 119, 121, 122], the CC of IM/DD can be approximated with a lower bound as

$$CC = \frac{1}{2} B \log_2 \left(1 + w \frac{\rho^2 P_{eff}^2}{\sigma^2} \right), \tag{3.12}$$

where ρ is the responsivity, $w = e/2\pi$ is a constant (e is the Euler's number), B is the modulation bandwidth, P_{eff} is optical power received and σ^2 is the Gaussian noise power. A factor of $1/2$ appears as a result of various constraints in LiFi [119]. It was found that at high SNR, (3.12) is true and in concurrence with upper bound.

3.3.4 Communication Model

Let $X = [x_0, x_1, \ldots, x_{|\mathcal{D}|}]$ be the signal vector transmitted by the LiFi APs and WiFi AP. For communication, the UEs receive signal either from LiFi AP or from WiFi AP. The received signal y_{0e} at UE e from WiFi AP, indexed as $d = 0$, is expressed as

$$y_{0e} = \sqrt{G_{0e} P_0} \times x_0 + N_0^{\text{WiFi}}, \tag{3.13}$$

where N_0^{WiFi} is AWGN. When the UE e is connected to a LiFi AP $d = 1, 2, \ldots, |\mathcal{D}| - 1$, y_{de} will be expressed as

$$y_{de} = \rho_e G_{de} P_d x_d + \sum_{k \in \mathcal{D} \setminus \{d\}} \rho_e G_{ke} P_k x_k D_k(\alpha_{ke'}) + N_0^{\text{LiFi}}, \tag{3.14}$$

where ρ_e is responsivity and N_0^{LiFi} includes thermal noise and shot noise. Shot noise is the ambient light present in indoor system. In the receiver circuitry, thermal noise occurs due to thermal agitation of electrons in the resistors. To accommodate the case of idle APs, a term $D_k(\alpha_{ke'})$ is included in (3.14) and is given as

$$D_k(\alpha_{ke'}) = \left(1 - \prod_{e' \in \mathcal{E} \setminus \{e\}} (1 - \alpha_{ke'})\right), \tag{3.15}$$

where $\alpha_{ke'}$ is indicator function showing association of UE e' with AP k such that

$$\alpha_{ke'} = \begin{cases} 1 \text{ if UE } e' \text{ is associated to AP } k \\ 0 \text{ otherwise.} \end{cases} \tag{3.16}$$

The term $D_k(\alpha_{ke'})$ is included after considering a practical scenario of idle APs. It is possible that a LiFi AP gets switched off due to hardware failure and it is not transmitting. Let AP d - UE e be the desired AP-UE pair, and AP k is the interferer AP. The term $\alpha_{ke'}$ is the indicator function showing whether interferer AP k is connected to a UE $e' \neq e$ or not. In case AP k is connected to some UE $e' \neq e$, $D_k(\alpha_{ke'}) = 1$ and the interference caused to UE e will be taken into account. If AP k is lying idle, it will not cause interference to UE e and $D_k(\alpha_{ke'}) = 0$.

The instantaneous achievable data rate is expressed as

$$r_{de} = \begin{cases} B_{0e} \log_2 (1 + SINR_{0e}), & \text{for } d = 0 \text{ and} \\ \frac{1}{2} B_{de} \log_2 (1 + wSINR_{de}), & \text{for } d \in \mathcal{D} \setminus \{0\}, \end{cases} \tag{3.17}$$

where $SINR_{0e}$ and $SINR_{de}$ are lower and upper bounds and given as

$$\begin{aligned} SINR_{0e} &= \frac{P_0 G_{0e}}{N_0^{\text{WiFi}} B_{0e}}, \text{ and} \\ SINR_{de} &= \frac{CG_{de}^2 P_d^2}{N_0^{\text{LiFi}} B_{de} + \sum_{k \in \mathcal{D}\setminus\{d\}} \rho_e C G_{ke}^2 P_k^2 \left(1 - \prod_{e' \in \mathcal{E}\setminus\{e\}} (1-\alpha_{ke'})\right)^2}, \end{aligned} \tag{3.18}$$

where B_{0e} and B_{de} are bandwidths of WiFi AP-UE pair and LiFi AP-UE pair respectively. Based on above expression the throughput of AP d can be given as

$$r_d = \sum_{e \in \mathcal{E}} a_{de} r_{de}. \tag{3.19}$$

3.3.5 The Resource Allocation Problem

In this section, we define the joint optimization problem of resource allocation with several constraints. The desirable objective in resource allocation is achieving maximum total data rate. However, the maximization of data rate is subject to constraints on bandwidth, transmission power, and SINR. The resource allocation problem is thus formulated as

$$\mathcal{P}: \qquad \max_{B_{de}, P_d, a_{de}} r_d, \qquad \text{for } d \in \mathcal{D}, e \in \mathcal{E}. \tag{3.20}$$

The constraint on bandwidth subjected to LiFi APs is given as

$$\mathcal{C}_1: \qquad \sum_{e \in \mathcal{E}} a_{de} B_{de} \leq B_{\text{max}}^{\text{LiFi}}, \quad \text{for } d \in \mathcal{D}\setminus\{0\}. \tag{3.21}$$

The constraint on bandwidth subjected to WiFi AP is given as

$$\mathcal{C}_2: \qquad \sum_{e \in \mathcal{E}} a_{0e} B_{0e} \leq B_{\text{max}}^{\text{WiFi}}, \quad \text{for } d = 0, \tag{3.22}$$

where $B_{\text{max}}^{\text{LiFi}}$ and $B_{\text{max}}^{\text{WiFi}}$ are the maximum bandwidths allotted to LiFi AP and WiFi AP respectively. The constraint to power transmission of LiFi AP is given as

$$\mathcal{C}_3: \qquad 0 \leq P_d \leq P_{\text{max}}^{\text{LiFi}}, \qquad \text{for } d \in \mathcal{D}\setminus\{0\}, \tag{3.23}$$

similarly for WiFi AP, the constraint on the transmission power $P_{\text{max}}^{\text{WiFi}}$ is given as

$$\mathcal{C}_4 : \qquad 0 \leq P_0 \leq P_{\text{max}}^{\text{WiFi}}, \quad \text{for } d = 0, \tag{3.24}$$

where $P_{\text{max}}^{\text{LiFi}}$ and $P_{\text{max}}^{\text{WiFi}}$ are the maximum transmission power allotted to LiFi AP and WiFi AP respectively. In addition the constraint on $SINR_{de}$ to have reliable communication is given as

$$\mathcal{C}_5 : \qquad SINR_{de} \;\; \geq \;\; \gamma_{de}, \text{ for } d \in \mathcal{D}, e \in \mathcal{E}. \tag{3.25}$$

In (3.25), γ_{de} is the minimum threshold for $SINR_{de}$ and when it holds the equality following conditions must be satisfied to prevent SINR constraint violation [101, 123]

$$\begin{aligned} &1 - \textstyle\sum_{d\in\mathcal{D}} \sum_{e\in\mathcal{E}} \xi_{de} > 0, \text{ and} \\ &\textstyle\sum_{d\in\mathcal{D}} \sum_{e\in\mathcal{E}} \beta_d \xi_e \leq 1, \end{aligned} \tag{3.26}$$

where,

$$\xi_{de} = \left(1 + \frac{1}{\gamma_{de}}\right)^{-1}, \text{ and} \tag{3.27}$$

$$\beta_{de} = \frac{N_0 B_{de}}{(C G_{de} P_d / \gamma_{de}) - N_0 B_{de}} + 1. \tag{3.28}$$

It can be seen that the maximization problem in (3.20) -(3.25) is non-concave. It also involves integer optimization due to the presence of association parameter (a_{de} and $a_{ke'}$).

Non-concave functions can have multiple local optima, hence it is difficult to identify global optimum. Algorithms may converge to a local minimum rather than the global minimum. For a non-concave objective function $f(z)$, there may exist several points $z_1, z_2, \ldots, z_k$ such that $f(z_1) \geq f(z_2) \geq \cdots \geq f(z_k)$, where z_1 may not correspond to the global maximum. Non-concave problems may involve complex, non-convex constraints, which can complicate the optimization process and may lead to suboptimal solutions. For a constraint $g(z) \geq 0$, where $g(z)$ is non-concave, the feasible region $S_z = \{z \mid g(z) \geq 0\}$ can be non-convex, possibly consisting of disconnected components, complicating the search for an optimal solution.

Integer optimization problems are generally NP-hard, meaning there is no known polynomial-time algorithm for solving them. They are challenging due to their non-convex and discrete nature. As the number of variables and constraints grows, their parameters and computational requirement increases exponentially.

To address the issue of non-concavity and integer optimization, we propose a DQN learning based solution.

3.4 Resource allocation Algorithm for DQN based Hybrid WiFi/LiFi System

In this section, a DQN-based learning algorithm has been proposed to address the resource allocation problem formulated in (3.20). The proposed method aims to maximize the achievable data rate of AP d while satisfying the constraints mentioned in (3.21)-(3.26). It works on the three fundamental components: state, action, and reward. The *state vector* represents the current status of the environment, the *action vector* defines the decision taken in response to the observed state, and the *reward vector* quantifies the system's performance based on the executed action.

Let the state vector be denoted as $\mathcal{S}_{de} = \{s_{de}^1, s_{de}^2, \ldots, s_{de}^l\}$ and the action vector as $\mathcal{A}_{de} = \{a_{de}^1, a_{de}^2, \ldots, a_{de}^m\}$. The values of l and m depend on the specific formulations of $\mathcal{S}_{de}$ and $\mathcal{A}_{de}$. At a given time step t, the system is in state $s_{de}(t) \in \mathcal{S}_{de}$ and receives a corresponding reward $R_d(s, a)$. When an action $a_{de}(t) \in \mathcal{A}_{de}$ is executed, the system transitions to a new state $s_{de}(t+1) \in \mathcal{S}_{de}$. The action $a_{de}(t)$ directly influences the reward obtained.

The CU is responsible for training the learning algorithm to optimize association parameter, bandwidth and power allocation to the APs. This process is executed iteratively, to achieve the maximum reward.

3.4.1 Framework for Learning

The maximization of achievable data rate while satisfying the constraints in (3.21)-(3.26) has been carried out with the help of DQN-transfer learning. It works with the help of state, action and reward vectors formulated as follows.

3.4.1.1 Action Space

After observing the present environment, the player (CU) takes an action. The action space defines the set of actions that can be taken by the player. Let $\mathbb{B}_{de}$ and $\mathbb{P}_d$ are the discretized sets of B_{de} and P_d. The following formulations made are

$$\mathbb{B}_{de} = \left\{ 0, B_{\text{min}}^{\text{WiFi/LiFi}} \left(\frac{B_{\text{max}}^{\text{WiFi/LiFi}}}{B_{\text{min}}^{\text{WiFi/LiFi}}} \right)^{\frac{u}{(|\mathbb{B}_{de}|-2)}} , u = 0, 1, 2, ..., |\mathbb{B}_{de}| - 2, \right. \tag{3.29}$$

where $B_{\text{min}}^{\text{WiFi/LiFi}}$ and $B_{\text{max}}^{\text{WiFi/LiFi}}$ are the minimum and maximum values of B_{de} for WiFi AP and LiFi APs respectively. Similarly,

$\mathbb{P}_d$ is obtained as

$$\mathbb{P}_d = \left\{ 0, P_{\text{min}}^{\text{WiFi/LiFi}} \left(\frac{P_{\text{max}}^{\text{WiFi/LiFi}}}{P_{\text{min}}^{\text{WiFi/LiFi}}} \right)^{\frac{u}{(|\mathbb{P}_d|-2)}} , u = 0, 1, 2, ..., |\mathbb{P}_d| - 2, \right. \tag{3.30}$$

where $P_{\text{max}}^{\text{WiFi/LiFi}}$ and $P_{\text{min}}^{\text{WiFi/LiFi}}$ are the maximum and minimum levels of the transmit power for WiFi AP and LiFi APs respectively. The discretized parameters $\mathbb{B}_{de}$ and $\mathbb{P}_d$, along with a_{de}, are utilized to obtain the threshold γ_{de}, as defined in (3.18), for each communication APd- UEe link. The state action vector is given as

$$\mathcal{A}_{de} = \{\gamma_{de}^1, \gamma_{de}^2, \ldots, \gamma_{de}^{|\mathbb{B}_{de}|}\}. \tag{3.31}$$

During each iteration, the CU selects a value from the set $\mathcal{A}_{de}$ for each AP. It is important to emphasize that the selection process does not directly involve choosing a threshold value. Instead, selecting the appropriate values of $\mathbb{B}_{de}$ and $\mathbb{P}_d$ to achieve the required minimum SINR, γ_{de}. Next, we define the state vectors.

3.4.1.2 State Space

We define a state space $\mathcal{S}_{de}$ with binary variables given as $\mathcal{S}_{de} = \{I_1^{de}, I_2^{de}, \ldots, I_6^{de}\}$, where I_1 to I_6 are indicators variables that assume a value 0 or 1. Each of these indicator variables respectively take a value 0 if the constraint in (3.21) - (3.28) are satisfied. On the other hand, each of these indicator variables take value 1 if these constraints are not satisfied. They are written as follows:

$$
\begin{aligned}
I_1^{de} &= \begin{cases} 0, \text{ if } \sum_{e\in\mathcal{E}} a_{de} B_{de} \leq B_{\text{max}}^{\text{LiFi}}, \text{ for } d \in \mathcal{D}\backslash\{0\}, e \in \mathcal{E}, \\ 1, \text{ otherwise.} \end{cases} \\
I_2^{de} &= \begin{cases} 0, \text{ if } \sum_{e\in\mathcal{E}} a_{0e} B_{0e} \leq B_{\text{max}}^{\text{WiFi}}, \quad \text{for } d = 0, e \in \mathcal{E}, \\ 1, \text{ otherwise.} \end{cases} \\
I_3^{de} &= \begin{cases} 0, \text{ if } 0 \leq P_d \leq P_{\text{max}}^{\text{LiFi}}, \quad \text{for } d \in \mathcal{D}\backslash\{0\}, e \in \mathcal{E}, \\ 1, \text{ otherwise.} \end{cases} \\
I_4^{de} &= \begin{cases} 0, \text{ if } 0 \leq P_0 \leq P_{\text{max}}^{\text{WiFi}}, \quad \text{for } d = 0, e \in \mathcal{E}, \\ 1, \text{ otherwise.} \end{cases} \\
I_5^{de} &= \begin{cases} 0, \text{ if } \sum_{d\in\mathcal{D}, e\in\mathcal{E}} \xi_{de}(\gamma_{de}) < 1, \quad \text{for } d \in \mathcal{D}, e \in \mathcal{E}, \\ 1, \text{ otherwise.} \end{cases} \\
I_6^{de} &= \begin{cases} 0, \text{ if } \sum_{d\in\mathcal{D}, e\in\mathcal{E}} \beta_{de}\xi_{de}(\gamma_{de}) < 1, \quad \text{for } d \in \mathcal{D}, e \in \mathcal{E}, \\ 1, \text{ otherwise.} \end{cases}
\end{aligned}
\tag{3.32}
$$

3.4.1.3 Reward

When an action is taken, the reward is received by the system. The immediate reward received after a particular action is

$$
R_d(s,a) = \begin{cases} r_{\text{fix}}, & \text{if } \sum_{c=1}^{6} I_c^d > 0, \\ r_d, & \text{otherwise,} \end{cases}
\tag{3.33}
$$

where r_{fix} is the reward smaller than reward obtained for action violating the interference constraints and r_d is reward received when constraint are satisfied.

3.4.2 DQN Transfer Learning for a newly entered UE

When a set-up has fixed number of UEs, it is static in nature. A UE entering or leaving the set-up makes it dynamic. Let us consider the former case. Whenever a new UE enters the room, data collection for the new UE and re-execution of the DQN algorithm are needed. It limits the application of DQN learning in hybrid WiFi/LiFi systems. Thus, DQN learning application for a dynamic hybrid WiFi/LiFi system is complex.

We investigate the application of transfer learning [124] to address this issue. Transfer learning has been found as an efficient tool to mitigate the data insufficiency problem. We use the deep transfer learning for efficient transfer of the knowledge gathered by the static hybrid WiFi/LiFi network when a new UE enters it. Our proposed transfer learning algorithm is an example of positive transfer learning. It obtains an optimal policy for transferring the knowledge obtained from the UEs present in the environment to the new UE entering the environment. With the entrance of the new UE, the static network converts to the dynamic network. With the help of the transfer of knowledge, a new UE achieves higher data rates with fewer numbers of iterations. For the second task, i.e., when the new UE enters into the environment, it receives the data from the nearest UE present in the environment. The number of iterations required to achieve a similar level of data rates for the new UE is far less than that required for the UE previously present in the environment. It is also evident from the simulation results plotted in Fig. 3.3 in Section 3.5. Also, Table 3.2 compares the number of iterations required for convergence. It shows positive transfer learning because the number of iterations required is much less. Positive transfer learning occurs when knowledge gained from the first task improves the learning performance in the second task, hence improving convergence speed. In a practical scenario, the system is dynamic, i.e, new UEs will be entering and existing UEs will be exiting the environment randomly. A newly entered UE needs to get connected to one of the APs. Most part of the information it uses is already learned while doing the previous task. First, the information

for the UE nearest to the new UE will be transferred to it. Now, the new UE will execute the algorithm based on this information. In this way, it speeds up the performance of the system. This happens as the DQN network estimates the new Q-function based on the reward of every action for each AP. The CU learns the environment with each AP and then it takes the highest reward action. That means associating the UE to an AP in a way that that AP gets the highest data rate as a reward. The parameters of Q-function are updated immediately as the AP receives the reward. When a new UE joins the setup, making use of already gathered information from the environment will be an efficient procedure. This process improves the learning performance. The knowledge retained from the environment while doing the previous task will be used as the new UE enters the scenario and the information will follow the algorithm. For DQN based learning algorithm the optimal policy π for immediate reward $R_d(s,a)$ over a long span of time has to be followed. Mathematically, the value of state function $V^{\pi}(s,a)$ is given as

$$V^{\pi}(s,a) = \max_{\pi} \{\sum_{t=0}^{\infty} \zeta^t E(R(s,a))_t | s_t = s, a_t = a, \pi\}, \tag{3.34}$$

where $Q^*(s,a)$ is the optimal action-value function $\stackrel{\Delta}{=} \max_{\pi} V^{\pi}(s,a)$. It is obtained with the help of Bellman's equation and is given as

$$Q^*(s,a) = \max_{a \in \mathcal{A}} \{r(s,a) + \zeta Q^*(s',a')\}, \tag{3.35}$$

where ζ is the learning rate update at $Q^*(s,a)$. In equation (3.34), the function $Q^*(s,a)$ iteratively converges to its optimal value as $t \rightarrow \infty$. However, for large-dimensional state-action spaces, obtaining the optimal action-value function $Q^*(s,a)$ becomes challenging with (3.35). To address this, a function estimator is utilized to approximate the optimal action-value function. As proposed in [80], a neural network can be employed for this estimation, where $Q(s,a;\theta) \approx Q^*(s,a)$. In this chapter, we used a multilayer perceptron (MLP) network for this purpose. Specifically, a fully connected feed-forward MLP network is implemented. The DQN-based methodology uses neural network as an action-value function approximator, incorporating the *experience replay* mechanism to enhance

learning efficiency.

The CU records experiences at each time step t, and is defined as $e_d(t) = \{a_{de}(t), s_{de}(t), r_d(t), s_{de}(t+1)\}$, and these experiences are stored in a replay memory $D_d(t) = \{e_d(1), e_d(2), \ldots, e_d(t)\}$.

To approximate the Q-value function, two MLP networks are employed: the current Q-network $Q(s, a)$ and the target Q-network $Q(s, a; \theta)$. Here, θ and θ^- denote the parameters of the current and target networks, respectively. At each iteration, the parameter θ of the action-value function is updated using experiences sampled from the replay memory D_d, where a random sample $(a, s, r, \widehat{s})$ is selected. After fixed number of iterations, the target network parameters θ^- are updated by setting them equal to the current network parameters ϕ. The update process employs the gradient descent method to optimize the parameters effectively. This update procedure is based on a gradient descent algorithm as

$$L(\theta_d) = E\left[\left(r_d(s,a) + \zeta \max_{\widehat{a}\in\mathcal{A}} \left(\widehat{Q_d}(\widehat{s}, \widehat{a}, \theta_d^-)\right) - Q_d(s, a, \theta_d)\right)^2\right]. \tag{3.36}$$

Algorithm 1 presents the DQN-based transfer learning approach to maximize the achievable data rate. We considered newly entering UEs into the system. Firstly, individual rate r_d is optimized. Since r_d values are non-negative and their summation constitutes the total system sum-rate r, hence optimizing each r_d ensures the overall system optimization.

3.5 Simulation Results

The set-up has been considered as having 4 LiFi APs, 1 WiFi AP and 4 UEs in a room. One AP can be connected to multiple UEs, but an UE can be connected to one AP only at a time. The data has been taken from [50] for our experimental simulations. The noise at LiFi AP N_0^{LiFi} is $10^{-21}\text{A}^2/\text{Hz}$, for each LiFi AP (LED lamp) the average optical power is 9.2 W, PD has an area of $A_{\text{pd}} = 1$ cm^2, and has a responsivity ρ of 0.28 A/W. The receiver that has been considered is of FOV $60°$, and its maximum illuminous intensity is considered as 30 cd. Discount factor of 0.9 and learning rate ζ of 0.01 are used for

Algorithm 1 Proposed DQN Transfer Learning Algorithm

for $d = 0, 1, 2, \ldots, |\mathcal{D}|$ **do**
  Start
  Start replay memory
  Start the parameter θ_d for policy $\pi(a_{de}|s_{de}; \theta_d)$
  Start neural network for function Q_d with random θ_d
  Start target $\widehat{Q}_d$ with $\theta_d^- = \theta_d$
end for
for Itr =1:KK **do**
  Receive the initial state
  for Episode = 1: EP_D **do**
    for $t < T_{itr}$ **do**
      for $i = 0, 1, 2, \ldots, |\mathcal{D}|$ **do**
        Select $a_{de}^*(t)$ as shown below for $e \in \mathcal{E}$
        Chose from action space by finding

$$a_{de}(t) = \arg\max_{a_{de}(t)} Q(s_{de}(t), a_{de}(t); \theta_d) \tag{3.37}$$

        If maximization unsuccessful, chose arbitrary action with probability ϵ
        Amend $s_{de}(t+1)$ and $r_d(t)$ according to (3.32) and (3.33)
        Save replay memory D_d created for AP d with $e_d(t) = (a_{de}(t), s_{de}(t), r_d(t), s_{de}(t+1)$
        Amend the present θ_d of $Q(s_{de}(t), a_{de}(t); \theta_d)$, by taking specimen mini-batch of transitions from $D_d(t)$
        After regular intervals, amend $\theta_d^- = \theta_d$
        Obtain mini batch specimens from D_d
      end for
    end for
  end for
end for
Execute $r = \sum_{d \in \mathcal{D}} r_d$
The optimal r_ds are obtained for all APs. Note that every $r_d > 0$, thus maximizing r will optimize the overall system
Arrival of a new UE
Newly arrived UE is indexed as $|\mathcal{E}| + 1$
Start Q for AP d with $a_{de}(t)$ parameters related to the UE closest to the newly arrived $|\mathcal{E}| + 1$th UE {The UE closest to the $|\mathcal{E}| + 1$th UE data is recorded by the CU (transfer learning)}
for $d = 0, 1, 2, \ldots, |\mathcal{D}|$ **do**
  Algorithm 1 is initiated with the persisting $a_{de}(t)$ for $|\mathcal{E}| + 1$ UEs. The iterations are carried on further.
end for
Perform $r = \sum_{d \in \mathcal{D}} \log_2 R_d$

all the APs. The exponent of path-loss a_{0e} is considered as 2.8, the order of Lambertian emission $m = 1.2$ has been considered, the dimensions of room are taken as 9 m $\times 9$

m $\times 5$ m, the separation between the UE and the floor is 0.9 m. The center of the room ceiling has the WiFi AP. The LiFi APs are located at room coordinates $\left[\pm\frac{10}{\sqrt{8}}, \pm\frac{10}{\sqrt{8}}\right]$. We consider a replay memory and buffer mini-batch of sizes 100 and 10 respectively. The algorithm is run for 1000 monte-carlo simulations. In the input layer, neural networks has 7 nodes : 6 state nodes and 1 action node. The DQN structure consists of two hidden layers with 3 and 2 neurons respectively. The association parameter, downlink bandwidth, and transmission power are state and action vector functions. Therefore, passing through input means passing through these three parameters. As iterations proceed, algorithm converges and the sum-rate achieved is maximized.

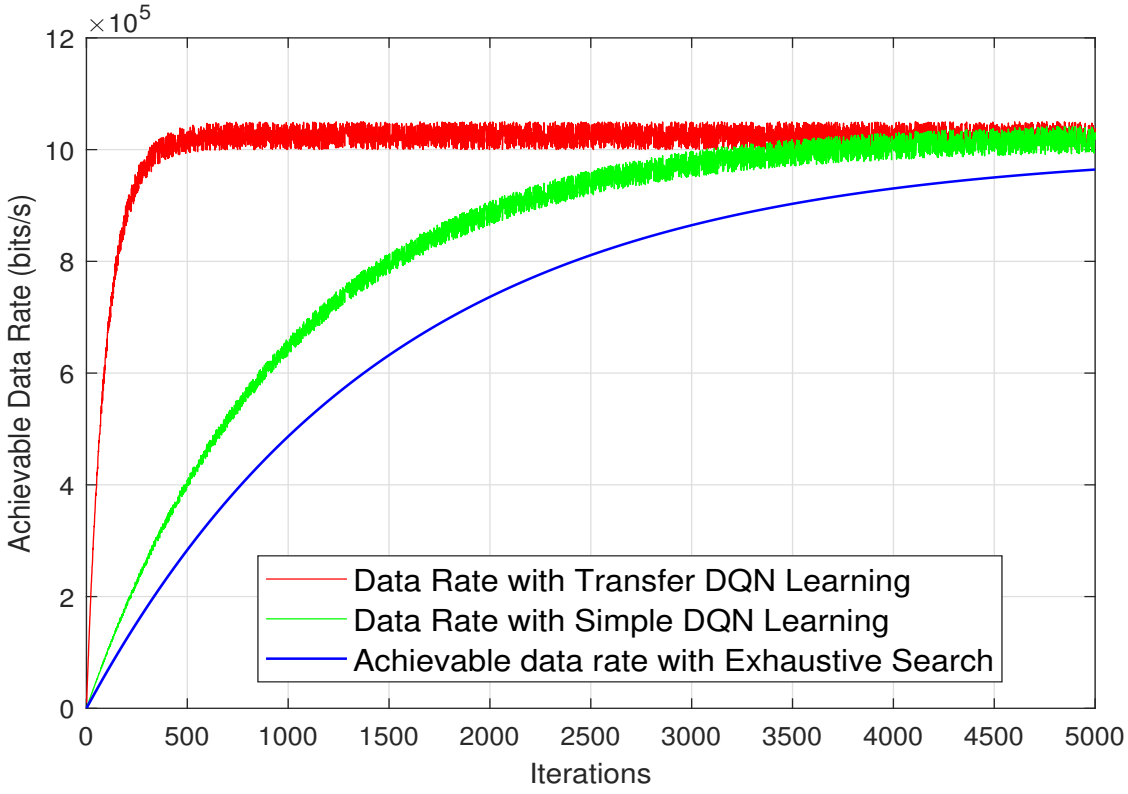
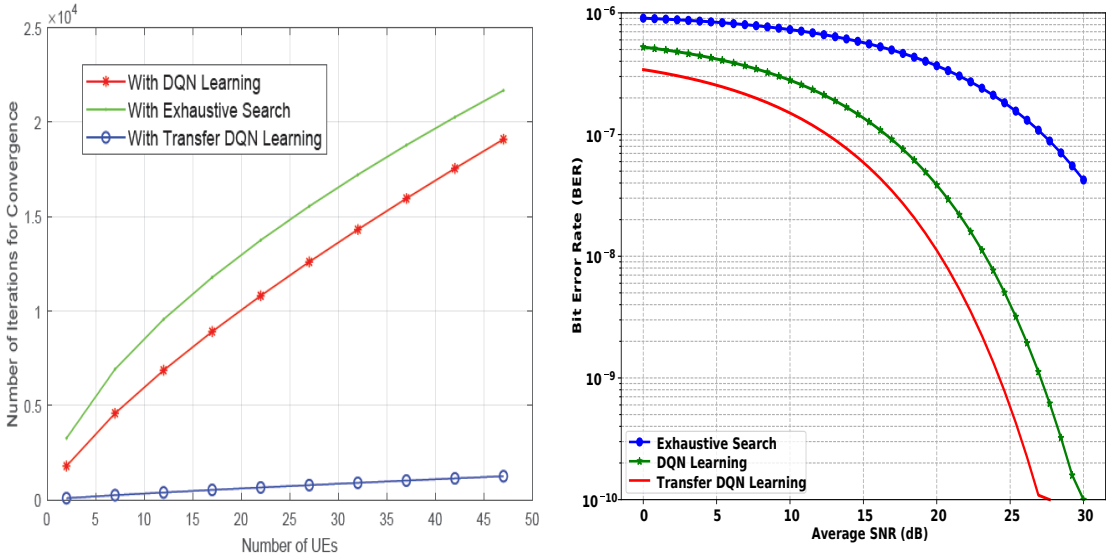


Figure 3.3: Graph showing the behavior for achievable sum-rate with the number of iterations when a new UE enters the room

Fig. 3.3 shows the behavior of achievable sum-rate with the number of iterations when a new UE enters the experimental room. The sum-rate achieved with transfer DQN learning is compared with the sum-rate achieved with DQN learning and with exhaustive search (ES) algorithm. It can be seen that both the DQN learning algorithms outperform the ES algorithm. ES algorithm reaches its maximum performance in around 5000 iterations.



(a) (b)

Figure 3.4: (a) Graph depicting the behavior for the number of iterations vs the number of UEs for the new incoming UE in the room (b) Comparison of BER vs SNR

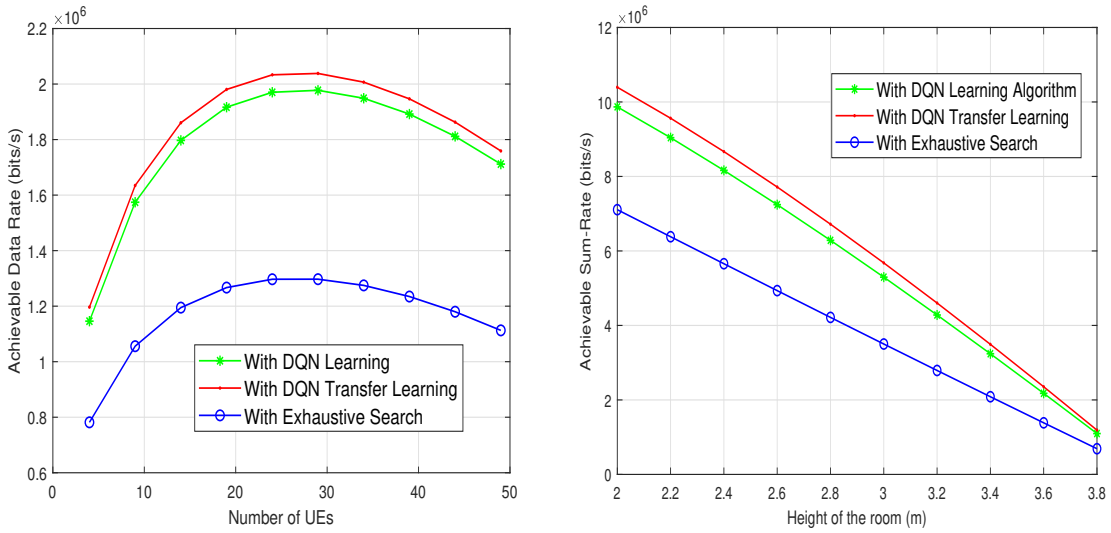
The maximum value it is able to reach is 9 Mbits/s (or Mbps). A normal DQN learning algorithm takes around 3500 iterations to reach its maximum possible sum-rate value of 10 Mbps. However, the value achieved by DQN learning algorithm is achieved by transfer DQN learning in less than 500 iterations.

Fig. 3.4a shows the comparison of number of iterations with the varying number of UEs present inside the room. The number of UEs present inside the room are varied from 2 to 48. With 4 UEs, the outcomes for different schemes investigated in Fig.3.4a can be matched with Fig. 3.3. Further, as the number of UEs increase, a sharp increase is observed in the number of iterations for convergence of both DQN learning and ES. However, DQN transfer learning shows comparatively a very slow increase. For 22 UEs inside the room, ES converges in nearly 13748 iterations, DQN learning converges in 10000 iterations, while transfer DQN learning converges in just 650 iterations. The data can be seen in Table 3.2. Fig. 3.4b shows the bit error rate (BER) vs SNR curve. As expected ES performs worst, with BER remaining near 10^{-7} even at 30 dB. Transfer DQN learning shows clear gains and significantly outperforms both the DQN learning and ES achieving the best result of 10^{-10} at 25–27 dB. By using prior knowledge, transfer DQN learning ensures faster convergence and significantly improved BER in hybrid WiFi/LiFi systems.

Table 3.2: No. of UEs vs No. of Iterations for different schemes

UEs Nos./ Iterations	DQN Learning	ES	Transfer DQN Learning
2	1788	3261	83
7	4577	6916	243
12	6857	9556	387
22	10804	13748	650
32	14310	17214	897
47	19093	21680	1247

As the number of UEs are increasing, it increases the possibility of presence of a UE near to the newly entering UE, which further increases the possibility of reliable transfer of information to it.

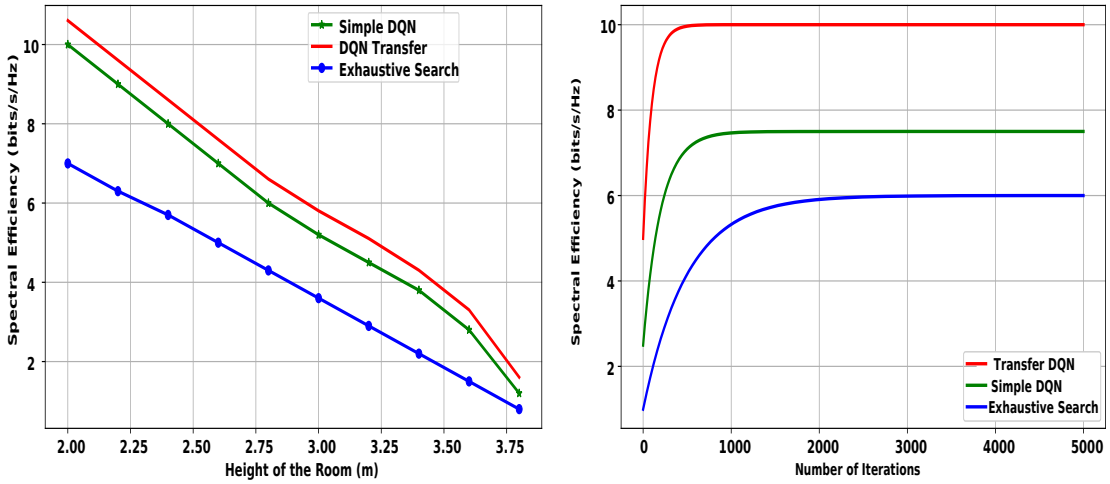


(a) (b)

Figure 3.5: (a) Graph depicting the behavior for achievable sum-rate with the number of UEs when a new UE enters the room (b) Graph depicting the behavior for the achievable sum-rate vs the height of the room for a new UE entering the room

Fig.3.5a shows the effect of number of static UEs on the achievable data rate (bits/s) of the proposed transfer learning scheme, the DQN learning scheme and the ES mechanism. The number of UEs vary from 4 to 49, while a new UE is entering into the room. The DQN transfer learning scheme outperforms both the ES and DQN learning. Note that DQN learning achieves an optimal performance after more number of iterations. Thus, the final achievable data rate it achieves is comparable to that of transfer learning. However, it pays the cost of $7-10$ times more iterations to obtain the same achievable data

rate as obtained by transfer learning. Thus, it is actually significantly inferior as compared to transfer learning. Note that the achievable data rate first increases with the increase in the number of UEs. This happens due to the increased number of AP-UE links. However, as the number of UEs increases beyond 28, a decrease in the achievable sum-rate is observed. This is because more UEs form more AP-UE links which also increases interference to every UE. Therefore, after a certain number of UEs in the room, a decrease in the achievable sum-rate is observed.



(a) (b)

Figure 3.6: (a) Comparison of spectral efficiency vs height of the room (b) Comparison of spectral efficiency vs number of iterations

Fig.3.5b shows the variation of the achievable sum-rate with the height of the room. The DQN transfer learning algorithm outperforms the DQN learning algorithm and the ES algorithm. From (3.1), it can be seen that the channel gain reduces as the distance between the transmitter and the receiver increases. A decrease in the channel gain magnitude means more distortion in the received signal. An increase in room height increases transmitter-receiver separation. Thus, increasing room height leads to a sharp decline in the achievable sum-rate value for all the algorithms under investigation.

Fig.3.6a shows spectral efficiency (bits/s/Hz) versus room height. As height increases, path loss and dispersion reduce the SNR, leading to lower spectral efficiency. Transfer learning consistently outperforms DQN and ES, demonstrating superior robustness.

Fig.3.6b shows spectral efficiency (bits/s/Hz) versus number of iterations. Transfer DQN learning converges rapidly to nearly 10 bits/s/Hz, outperforming both DQN and ES. DQN learning improves over ES, however, remains less efficient without transfer learning. ES is computationally intensive, converges slowly, and yields lower spectral efficiency. Overall, transfer DQN learning achieves faster convergence and higher efficiency, making it the optimal choice in hybrid WiFi/LiFi systems.

3.6 Conclusion

In this chapter, the resource allocation problem for hybrid WiFi/LiFi system has been addressed. The problem of joint optimization of bandwidth, association parameter, and transmission power has been addressed by the proposed algorithm. The proposed algorithm successfully overcomes the non-concavity issue in this joint optimization problem, particularly for the case of a newly entering UE into the set-up with the help of transfer learning. The proposed algorithm uses the already gathered data for the UE nearest to the new UE. Simulations verify that proposed algorithm shows 10% more achievable sum-rate with 14.28 lesser percentage of iteration to converge. For new UEs the maximum achievable sum-rate is achieved with 54% lesser number of iterations.

3.7 Appendix A

3.7.0.1 Elaboration on the integer optimization and non-concavity issues in $\mathcal{P}$

Note that our objective function is the sum-rate r_d, which is the cumulative of all the individual AP to UE link data rates r_{de}. The data rate r_{de} is given by Shannon's capacity and is a function of SINR as

$$r_{de} = B_{de} \log_2(1 + SINR_{de})$$

From (3.38) it can be seen that SINR is a function of association parameter a_{de} which takes integer values 1 and 0 for UE e being associated and not being associated to AP d

respectively. Thus, the objective function defined in this work is a non-concave integer optimization problem. For the sake of completeness, the non-concavity of r_d in a_{de}, B_{de} and P_d can be proven as in our problem $\mathcal{P}$ for $d \in \mathcal{D}\backslash\{0\}$ the objective function is given as

$$r_d = \sum_{e\in\mathcal{E}} \frac{1}{2} a_{de} B_{de} \times \log_2 \left(1 + w \frac{C G_{de}^2 P_d^2}{N_0^{\text{LiFi}} B_{de} + \sum_{k\in\mathcal{D}\backslash\{d\}} \rho_e C G_{ke}^2 P_k^2 \left(1 - \prod_{e'\in\mathcal{E}\backslash\{e\}} (1-\alpha_{ke'})\right)^2} \right) . \tag{3.38}$$

Due to the presence of indicator function a_{de} the system is non-concave. Since we know that the sum of logarithmic function is strictly concave. However, due to a_{de}, r_d will be neither concave nor convex. To prove that, let us take a system as $d = 1, 2$ and $e = 1, 2$. Let us assume $a_{11} = 1, a_{12} = 0, a_{21} = 0$ and $a_{22} = 1$ and define vector $\mathbf{x} = \{x_1, x_2, x_3, x_4\}$ where $x_1 = B_{11}, x_2 = B_{22}, x_3 = P_1$, and $x_4 = P_2$ and $\sqrt{w}CG_{11} = a$, $\sqrt{\rho}CG_{12} = b$, $\sqrt{\rho}CG_{21} = c$, $\sqrt{w}CG_22 = d$ and $N_0 = g$. Then,

$$r_d(\mathbf{x}) = \frac{1}{2} x_1 \log_2 \left(1 + \frac{a^2 x_3^2}{g x_1 + b^2 x_4^2}\right) + \frac{1}{2} x_2 \log_2 \left(1 + \frac{d^2 x_4^2}{g x_2 + c^2 x_3^2}\right) . \tag{3.39}$$

To check concavity, we find the Hessian matrix of r_d wrt $\mathbf{x}$, i.e., $\nabla^2_{\mathbf{x}} r_d(\mathbf{x})$. The elements of $\nabla^2_{\mathbf{x}} r_d$ are obtained as follows

$$\frac{d^2 r_d}{dx_1^2} = -\frac{a^2 g x_3^2 ((a^2 g x_3^2 + 2c^2 g x_4^2) x_1 + 2a^2 c^2 x_4^2 x_3^2 + 2c^4 x_4^4}{ln(2)(g x_1 + c^2 x_4^2)^2 (g x_1 + a^2 x_3^2 + c^2 x_4^2)^2}, \tag{3.40}$$

$$\frac{d^2 r_d}{dx_2^2} = -\frac{d^2 g x_4^2 ((d^2 g x_4^2 + 2b^2 g x_3^2) x_2 + 2d^2 b^2 x_3^2 x_4^2 + 2b^4 x_3^4}{ln(2)(g x_2 + c^2 x_3^2)^2 (g x_2 + d^2 x_4^2 + b^2 x_3^2)^2}, \tag{3.41}$$

$$\begin{aligned} \frac{d^2 r_d}{dx_3^2} = &-\frac{-2d^2 b^2 x_4^2 x_2}{(b^2 x_3^2 + g x_2)^2 (\frac{d^2 x_4^2}{b^2 x_3^2 + g x_2} + 1)} + \frac{8 d^2 b^4 x_4^2 x_2 x_3^2}{(b^2 x_3^2 + g x_2)^3 (\frac{d^2 x_4^2}{b^2 x_3^2 + g x_2} + 1)} - \\ &\frac{4 d^4 b^4 x_4^4 x_2 x_3^2}{(b^2 x_3^2 + g x_2)^4 (\frac{d^2 x_4^2}{b^2 x_3^2 + g x_2} + 1)} + \frac{2a^2 x_1}{(g x_1 + c^2 x_4^2)(\frac{a^2 x_3^2}{g x_1 + c^2 x_4^2} + 1)} - \\ &\frac{4 a^2 x_1 x_3^2}{(g x_1 + c^2 x_4^2)^2 (\frac{a^2 x_3^2}{g x_1 + c^2 x_4^2} + 1)}, \end{aligned} \tag{3.42}$$

and

$$\frac{d^2r_d}{dx_4^2} = -\frac{-2a^2c^2x_3^2x_1}{(c^2x_4^2+gx_1)^2(\frac{a^2x_3^2}{c^2x_4^2+gx_1}+1)} + \frac{8a^2c^4x_3^2x_1x_4^2}{(c^2x_4^2+gx_1)^3(\frac{a^2x_3^2}{c^2x_4^2+gx_1}+1)} - \frac{4a^4c^4x_3^4x_1x_4^2}{(c^2x_4^2+gx_1)^4(\frac{a^2x_3^2}{c^2x_4^2+gx_1}+1)} + \frac{2d^2x_2}{(gx_2+b^2x_3^2)(\frac{d^2x_4^2}{gx_2+b^2x_3^2}+1)} - \frac{4d^2x_2x_4^2}{(gx_2+b^2x_3^2)^2(\frac{d^2x_4^2}{gx_2+b^2x_3^2}+1)}. \tag{3.43}$$

It is evident that $\frac{d^2r_d}{dx_1^2}$ and $\frac{d^2r_d}{dx_2^2}$ are always negative. However, the nature of $\frac{d^2r_d}{dx_3^2}$ and $\frac{d^2r_d}{dx_4^2}$ is not constant. It can become positive or negative for varying values of B_{11}, B_{22}, P_1 and P_2. Thus, for the LiFi network, r_d will neither be concave nor convex in B_{11}, B_{22}, P_1, and P_2. Next, we investigate the behavior of r_d in the WiFi network. The system model consists of only one WiFi AP indexed as $d = 0$. The achievable data rate for the WiFi network is given as

$$r_d = \sum_{e\in\mathcal{E}} a_{0e}B_{0e}\log_2\left(1+\frac{P_0G_{0e}}{N_0^{\text{WiFi}}B_{0e}}\right). \tag{3.44}$$

For simplicity in calculations, let us consider $a_{0e} = 1$, $|\mathcal{E}| = 1$, $N_0^{\text{WiFi}} = a$ and a vector $\mathbf{x} = \{x_1x_2\}$ where $x_1 = B_{01}$ and $x_2 = P_0G_{0e}$. The achievable rate can be written as

$$r_d(\mathbf{x}) = x_1\log_2\left(1+\frac{x_2}{ax_1}\right). \tag{3.45}$$

The Hessian matrix $\nabla_{\mathbf{x}}^2 r_d(\mathbf{x})$ elements are obtained as

$$\frac{d^2r_d(\mathbf{x})}{dx_1^2} = \frac{-x_2^2}{ln(2)x_1(ax_1+x_2)^2}, \tag{3.46}$$

and

$$\frac{d^2r_d(\mathbf{x})}{dx_2^2} = \frac{-a^2x_1}{ln(2)(ax_1+x_2)^2}. \tag{3.47}$$

It is evident that both the elements of $\nabla_{\mathbf{x}}^2 r_d(\mathbf{x})$ are negative. The higher order of $|\mathcal{E}|$ will result to sum of such similar functions. We know that r_d will be jointly concave in B_{0e} and P_0. However, it is neither concave nor convex in the indicator function a_{0e}. Thus, jointly it will be neither concave nor convex in a_{0e}, B_{0e} and P_0.

Chapter 4

Actor-critic DDPG in Hybrid RF/LiFi systems

As discussed in the previous chapter, a hybrid RF and LiFi network combines the strengths of RF and LiFi technologies. RF offers broad coverage, while LiFi provides high data rates. As these technologies operate on non-interfering spectra, they can co-exist without interfering with each other. This setup not only enhances the data rate but also makes the network more reliable, especially when physical obstacles might block signals. However, resource management in hybrid RF/LiFi networks is challenging due to the dynamic environment and the differing characteristics of the two technologies. Effective resource allocation maximizes data rate in these networks. DQN based mechanisms can overcome the issue of non-concavity. However, they struggle in dynamic systems with large dimensions and continuous action spaces.

In this chapter, we introduce a model-free DRL approach to address the resource allocation problem in hybrid RF/LiFi networks. Our DRL model is designed to handle real-world conditions, considering factors like blockages and user mobility. Unlike traditional methods that need extensive modeling and assumptions, our approach learns directly from interacting with the environment, making it highly adaptable and robust. Simulation results demonstrate that our method enhances resource utilization and overall network per-

formance, achieving a 62.8% increase in sum-rate and a 42.8% improvement in optimal transmit power compared to conventional methods.

4.1 Overview

In the recent years, wireless internet data traffic has seen a phenomenal growth across the RF spectrum [125]. The growth is estimated to fully saturate the RF spectrum by 2035 [2], highlighting significant future challenges. OWC, such as LiFi, present a promising possibility to alleviate the burden on RF communications. LiFi uses the unlicensed light spectrum, offering high-data rate and bidirectional, and multi-user connectivity without interfering with RF signals [11, 126, 127]. However, challenges such as inefficiency in NLOS scenarios, random orientations, and UE mobility introduce significant channel quality variations, underscoring the need to integrate RF and LiFi communications for robust and mobile broadband solutions [128], [129]. The integrated system, known as a hybrid RF/LiFi system, has emerged as a valuable addition to HetNets, alleviating the strain on overloaded only-RF communication systems [130].

In hybrid RF/LiFi systems, optimal resource allocation remains a crucial inherently complex integer non-concave problem. Conventional convex optimization based resource allocation algorithms often fail in obtaining global optimum solutions to such problems. These solutions are based on presumption of values for at least one parameter. However, presuming a value does not make the system robust for optimal resource allocation. DQN based mechanisms can overcome this issue. However, it's limitation lies in its capability to manage only discrete and low-dimensional action spaces and fails in dynamic systems with large dimension and continuous action spaces [32, 31, 17]. Dynamic systems such as hybrid RF/LiFi are often continuous rather than discrete. DQN does not support continuous action spaces. It requires discretization of the action space, hence struggles and leads to suboptimal performance. In our proposed dynamic hybrid RF/LiFi system, the action space grows exponentially with the large size of the rooms and number of UEs. DQN is forced to discretize continuous actions which makes it slow and computationally expen-

sive. Therefore, we are motivated on exploring a machine learning technique designed to optimize future rewards in a large dynamic environment. Our objective is to devise an intelligent system that maximizes the overall average throughput for all UEs. To address these objectives, we investigate the application of off-policy DRL, which is a fusion of reinforcement learning and deep learning techniques [131].

In this Chapter, we propose to apply a model-free off-policy actor-critic algorithm to learn policies in high dimensional continuous action spaces. A model-free approach means that the agent (CU) does not require prior knowledge of the environment; instead, it learns optimal policies by directly interacting with the environment and collecting experiences over time. Our proposed DDPG-based algorithm is based on the DRL mechanism. DDPG is an off-policy-based DRL technique, where an agent can pick a policy not solely from its current policy (observational policy) but also from its past experiences (behavioral policy). Off-policy learning means that the agent learns from past experiences stored in a replay buffer rather than relying only on the current policy. It can learn from past experiences. It follows the actor-critic architecture, which enables it to learns both the optimal policy and the value function efficiently. It supports continuous action and state spaces, which leads to faster convergence and improved stability. A key feature is that it is simple and easy to implement.

4.2 Chapter Contributions

Our main contribution in this chapter are as follows

- *Novelty:* We have developed a virtual environment for a hybrid RF/LiFi system that closely emulates real-world conditions, including the dynamicity. To the best of our knowledge, this is the first instance of such a system being created with the application of DDPG.
- *Dynamic set-up:* Our model accounts for the dynamic nature of the environment by considering the introduction of new UEs into the room and incorporating load balancing constraints, thus reflecting a realistic and dynamic system configuration.

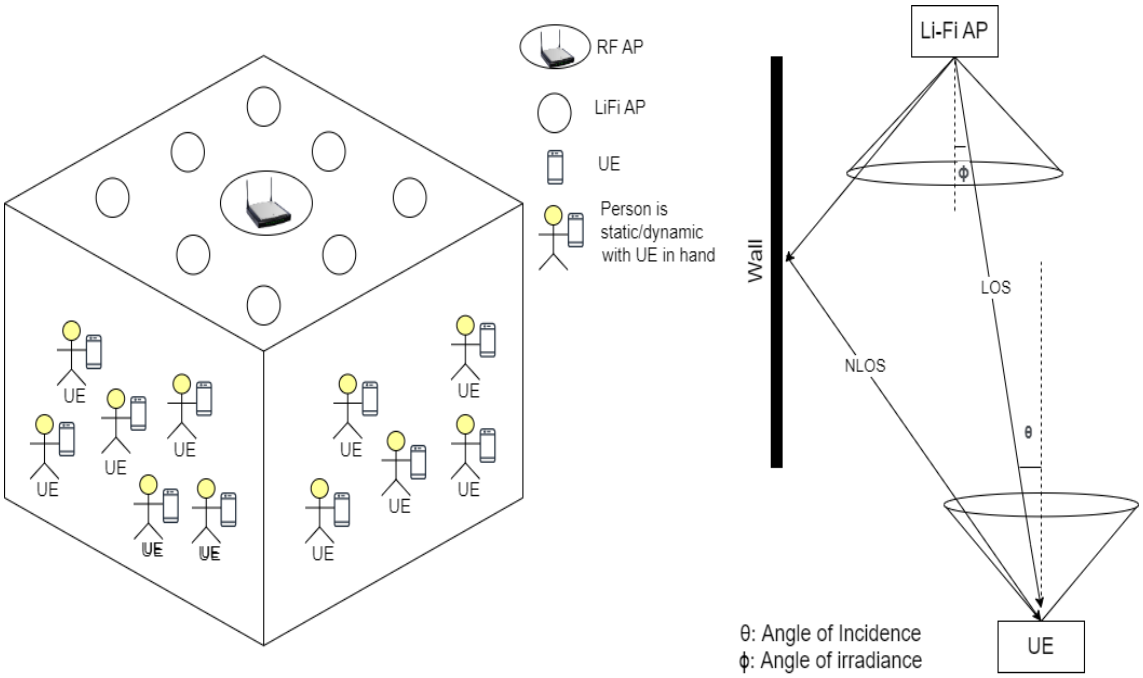


Figure 4.1: Hybrid RF/LiFi Environment, and signal propagation from LiFi AP-UE

- *Continuous action space:* The proposed state-of-the-art algorithm utilizes a continuous state and action space, resulting in a more optimal resource utilization, higher achievable data rates, and reduced transmission power.

- *DRL algorithm:* Using the DRL-based DDPG algorithm which can handle continuous state and action space, a comprehensive framework has been made to implement the optimization goal of maximizing the achievable sum-rate.

The rest of the chapter is organized as follows: Section 4.3 explains the system model under consideration. Section 4.4 explains the problem considered. Section 4.6 describes the illustrated framework. In Section 4.7, simulations have been explained, and Section 4.8 concludes the chapter.

4.3 System Model

Fig. 4.1 illustrates an indoor scenario, typically representing a hall where users may be static or moving at variable speeds, i.e., they are dynamic. Each user holds a UE. Multiple LED APs are deployed on the ceiling, known as LiFi APs. Additionally, an RF AP is

Table 4.1: Important notations and their meanings

Notation	**Meaning**
e	Index of APs
f	Index of UEs
k	The interferer AP index
r_{ef}	The achievable data rate between the eth AP and the fth UE
r_e	Downlink data rate of the eth AP
$B_{\text{max}}^{\text{LiFi}}$	Maximum BW allotted to LiFi AP
$B_{\text{max}}^{\text{RF}}$	Maximum BW allotted to RF AP
$\mathcal{E}$	Set of APs
$\mathcal{F}$	Set of UEs
m	Lambertian coefficient
θ_{ef}	Angle of incidence
ϕ_{ef}	Angle of irradiance
A_{pd}	Area of PD
CC	Channel capacity
B_{ef}	BW allotted to the fth UE by the eth AP
G_{ef}	Channel gain between eth AP and fth UE
ρ_f	Responsivity of the receiver PD
α_{ef}	Indicator function showing association of AP e-UE f
n_f^{rf}	RF noise power
n_f^{v}	LiFi noise power
P_e	Power transmitted by LiFi AP e
$P_{\text{max}}^{\text{RF}}$	Maximum transmit power of the RF AP
$P_{\text{max}}^{\text{LiFi}}$	Maximum transmit power of the LiFi AP

installed on the ceiling, creating a hybrid RF/LiFi setup. Together, the RF and LiFi APs provide uninterrupted connectivity to the UEs. In this indoor scenario, UEs connect to the LiFi AP if they are within LOS. When LOS components are absent from the LiFi AP, UEs connect to the RF AP to maintain ubiquitous connectivity.

Fig. 4.1 also shows an enlarged version of the light communication from the LiFi AP, providing significant details about the LiFi communication model. This link diagram illustrates the relationship between the angle of irradiance from the LiFi AP coverage area and the angle of incidence at the UE's FOV.

The link diagram demonstrates communication between the LiFi AP and a UE, illustrating both the direct LOS and indirect paths (NLOS) through which light signals can travel from the LiFi AP to the UE in a hybrid RF/LiFi system. The LOS path provides the strongest signal. In contrast, the NLOS path, involving reflections, may result in weaker and more

distorted signals due to delays and scattering.

The light emitted by the LiFi AP forms an angle ϕ, which represents the angle of irradiance. This it the angle between the vertical axis (normal to the LiFi AP surface) and the direction in which light is emitted. When received at the UE, the light forms an angle θ, representing the angle of incidence. This is the angle between the vertical axis (normal to the UE surface) and the direction from which the light is received. Both angles are crucial as they affect the coverage area, signal strength, and overall system performance.

4.4 Problem Formulation

Let $\mathcal{E}$ denote the AP ensemble, with individual APs indexed as $e = 0, 1, 2, \ldots, |\mathcal{E}|$ in the set up shown in Fig. 4.1. The RF AP situated at the center of the ceiling is indexed as $e = 0$. The LiFi APs, serving as LED light sources, are indexed as $e = 1, 2, \ldots, |\mathcal{E}| - 1$. The set $\mathcal{F}$ comprises UEs with indices $f = 0, 1, 2, \ldots, |\mathcal{F}|$. Each UE is assumed to maintain a constant height *t* above the floor. UEs have the option to establish connections with either LiFi or RF APs to receive data, contingent upon the AP offering superior data rates. Additionally, the model accommodates the inclusion of a new UE entering the scenario.

4.4.1 Propagation Channel Modeling

In hybrid RF/LiFi, the signal is transmitted over optical and RF channels. We first explain in detail the channel gain parameters for both the channels.

4.4.1.1 LiFi Channel Gain

Let a UE f connect to a LiFi AP in LOS. The DC channel gain G^v_{ef} is governed by the Lambertian law model as [44]

$$G^v_{ef} = \frac{(m+1)A_{\text{pd}}\cos^m\theta_{ef}\cos\phi_{ef}T_{\text{opt}}(\phi_{ef})g(\phi_{ef})}{2\pi d^2_{ef}}, \tag{4.1}$$

where $T_{\text{opt}}(\phi_{ef})$ is the receiving optical filter's gain, which is either 1 or a constant value within the receiver's FOV, θ_{ef} denotes the incidence angle at UE f from the eth AP, ϕ_{ef} is the angle of irradiance at AP e from fth UE, d_{ef} represents the distance between the fth UE and the eth AP, and m represents the exponent of the Lambertian radiation pattern, expressed as $m = -\frac{\ln 2}{\ln \cos \phi_{1/2}}$, with $\phi_{1/2}$ corresponding to the semi-angle at half illuminance. The concentrator gain, denoted as $g(\phi_{ef})$, is given as

$$g(\phi_{ef}) = \begin{cases} \frac{n^2}{\sin^2 \phi_{\text{FOV}}} \text{ if } 0 \leq \phi_{ef} \leq \varphi_{\text{FOV}} \\ 0 \qquad\quad \text{if } \phi_{ef} > \varphi_{\text{FOV}}, \end{cases} \tag{4.2}$$

where φ_{FOV} denotes the receiver's FOV, while n is the refractive index expressed as the ratio of the speed of light in vacuum to that in the optical material.

The mathematical modeling of channel gain G_{NLOS} in NLOS conditions, following the steps performed in (3.5)-(3.8), is obtained as [128]

$$G_{\text{NLOS}} = \sum_{l=0}^{\infty} G^{(l)}, \tag{4.3}$$

where l denotes the reflection index, and channel gain $G^{(l)}$ after the l reflection. The non-ideal channel gain between the eth LiFi AP and the fth UE is expressed as

$$G_{ef} = G_{\text{NLOS}} + G^{v}_{ef} \text{ for } e \in \mathcal{E} \backslash \{0\}. \tag{4.4}$$

4.4.1.2 RF Channel Gain

The signal transmitted by RF AP and received by UE f follows the RF signal propagation model. The signal received accounts for both fading and path loss. The received RF signal power channel gain is modeled using the WINNER-II [118] channel model. The channel gain between the UE and RF AP is modeled as

$$G_{0f} = L d_{t_{0f}}^{-a_{0f}} \chi_{0f}, \tag{4.5}$$

where χ_{0f} represents the Nakagami fading channel, $d_{t_{0f}}$ is the distance between UE and RF AP, and a_{0f} indicates the path-loss exponent. Here, the value of L is determined by $L = 10^{X/10}$, where $X = M + N \log_{10}\left(\frac{f_c}{5}\right)$ describes a propagation model, where f_c represents the carrier frequency in GHz, and M and N are constants specific to the propagation environment. In LOS scenario, the values are $M = 46.8$ and $N = 20$, whereas in NLOS scenario, the values are $M = 43.8$ and $N = 20$.

The channel parameter χ_{0f} models the amplitude variations of a wireless signal due to multipath propagation. The PDF of χ_{0f} is characterized by the Nakagami-κ distribution as

$$PDF(r) = \frac{2\kappa^{\kappa}}{\Gamma(\kappa)\Omega^{\kappa}} r^{2\kappa-1} \exp\left(-\frac{\kappa}{\Omega} r^2\right), \quad r \geq 0, \tag{4.6}$$

where $\kappa \geq 0.5$ is the fading parameter, indicating the severity of fading and $r = \sqrt{\|v_{0f}\|^2}$ is amplitude of received signal, Ω is the average power of the received signal, and $\Gamma(\kappa)$ is the Gamma function. This versatile Gamma distribution encompassing various fading conditions. For $\kappa = 1$, it approximates to the Rayleigh distribution, representing severe fading with no LOS component. For $\kappa > 1$ to ∞, it approximates to the Rician fading distribution.

4.4.1.3 Received RF and LiFi Signals

As mentioned earlier, our ultimate goal is the maximization of data rate for users, characterized by achievable sum-rate maximization. Achieving sum-rate maximization involves a step-by-step analysis of data transmission and reception at the transmitter and the receiver respectively. Thus, we explain the mathematical model of the received data, which is the transmitted data multiplied by the channel gain and added with the noise at the receiver and interference signals. Once the data is received at the receiver, the achievable sum-rate is formulated with the help of Shannon capacity formula, which tells the data rate on an individual transmitter - receiver link. Let $X = [x_0, x_1, \ldots, x_{|\mathcal{E}|}]$ be the signal vector transmitted by the LiFi APs and RF AP. When the UE f communication is established

from an RF AP $e = 0$, the signal received by a UE f is given as[17]

$$v_{0f} = \sqrt{G_{0e}P_0} \times x_0 + n_f^{rf}, \tag{4.7}$$

where n_f^{rf} represents AWGN.

When UE f communicates with LiFi APs, the signal received at UE f is given as

$$v_{ef} = \rho_f G_{ef} P_e x_e + \sum_{k \in \mathcal{E} \setminus \{e\}} \rho_f G_{kf} P_k x_k D_k(\alpha_{kf'}) + n_f^v, \tag{4.8}$$

where ρ_f represents responsivity, and n_f^v is thermal and shot noise. Additionally, for idle APs, the term $D_k(\alpha_{kf'})$ is expressed as $D_k(\alpha_{kf'}) = \left(1 - \prod_{f' \in \mathcal{F} \setminus \{f\}} (1 - \alpha_{kf'})\right)$. Here, $\alpha_{kf'}$ is an indicator function indicating the association of UE *f'* with AP *k*, given as

$$\alpha_{kf'} = \begin{cases} 1 \text{ if UE } f' \text{ is associated to AP } k \\ 0 \text{ otherwise.} \end{cases} \tag{4.9}$$

Based on the parameters formulated in this section, we next formulate the Shannon capacity based achievable sum-rate expression.

4.4.2 Shannon Capacity and Achievable Sum-Rate

The CC of an RF network is obtained with the Shannon capacity formula. However, LiFi communication uses IM/DD of light signals. The capacity of LiFi AP-UE link is formulated with modified Shannon's capacity formula [119]. The Shannon capacity formula is a function of the bandwidth and SINR values. SINR is formulated as the ratio of product of signal power and channel gain to product of noise and bandwidth. Since we are using multiple LiFi APs, consideration of interfering AP signal power is done in the denominator of the SINR expression (4.13). Finally the throughput is formulated in terms of Shannon capacity and it is maximized with constraints imposed on bandwidth and power.

The authors in [119] have approximated the CC of IM/DD (CC_{LiFi}) using Shannon's

capacity formula for lower bound is given as

$$CC_{LiFi} = \frac{1}{2} B \log_2 \left(1 + w \frac{\rho^2 P_t^2}{\sigma^2}\right), \tag{4.10}$$

where w is a constant defined as $e/2\pi$ (with e being Euler's number), B denotes modulation bandwidth, P_t signifies optical power received, and σ^2 stands for Gaussian noise power. The factor of 1/2 appears due to different constraints. On the other hand, for an RF network the CC can be obtained with the direct application of Shannon capacity formula as

$$CC_{RF} = B \log_2 \left(1 + \frac{P_t^2}{\sigma^2}\right), \tag{4.11}$$

Hence, achievable data rate of a hybrid RF/LiFi system is given as

$$r_{ef} = \begin{cases} B_{0f} \log_2 \left(1 + SINR_{0f}\right), & \text{for } e = 0 \text{ and} \\ \frac{1}{2} B_{ef} \log_2 \left(1 + wSINR_e f\right), & \text{for } e \in \mathcal{E} \backslash \{0\}, \end{cases} \tag{4.12}$$

where $SINR_{0f}$ and $SINR_{ef}$ are the SINRs on the RF and LiFi links, respectively, and are provided as follows:

$$\begin{aligned} SINR_{0f} &= \frac{P_0 G_{0f}}{n_f^{\mathrm{rf}} B_{0f}}, \text{ and} \\ SINR_{ef} &= \frac{G_{ef}^2 P_e^2}{n_f^{\mathrm{v}} B_{ef} + \sum_{k \in \mathcal{E} \backslash \{f\}} \rho_e G_{kf}^2 P_k^2 \left(1 - \prod_{f' \in \mathcal{F} \backslash \{f\}} (1 - \alpha_{kf'})\right)^2}, \end{aligned} \tag{4.13}$$

where the bandwidths of the RF AP-UE pair and LiFi AP-UE pair are denoted as B_{0f} and B_{ef}, respectively. With these parameters, the throughput of the AP can be determined as

$$r_e = \sum_{f \in \mathcal{F}} \alpha_{ef} r_{ef}. \tag{4.14}$$

The achievable sum-rate is formulated as

$$r = \sum_{e \in \mathcal{E}} r_e. \tag{4.15}$$

4.4.3 Problem Addressed

The maximization of achievable sum-rate is subjected to constraints on bandwidth allocation, transmission power, SINR, and load balancing. As the allocation strategy involves an indicator parameter, this problem becomes a non-convex integer optimization problem. To solve this problem, conventional optimization algorithms presumes value of one of the optimization parameters and subsequently optimizing other parameters. However, as all the optimization parameters impact each other, such presumptions can lead to suboptimal results.

The issue related to presumption of values was earlier solved with a DQN learning algorithm based approach [32, 31]. However, this approach faces limitations as DQN is unable to handle continuous and high-dimensional action spaces, restricting its application in dynamic systems. To address this issue, we propose the application of DDPG algorithm to large and dynamic hybrid RF/LiFi systems. The optimization goal is explained in detail in the next section.

4.5 Achievable Sum-rate Maximization

As mentioned earlier, our aim is to design a DDPG based algorithm to achieve sum-rate maximization while adhering to all the constraints and load balancing. The DDPG algorithm associates the UE to an AP which satisfies the constraints and assigns the optimal transmit power and bandwidth in the system. The action vector represents the value of the transmission power, bandwidth, and association parameters. The state vector contains the observation state and constraints that must be fulfilled to make the reliable connection between a UE and an AP. After observing the state of the environment (if the constraints are satisfied), DRL agent (CU) chooses an action that maximizes the reward by satisfying all the constraints.

The fundamental objective of hybrid RF/LiFi communication system is to maximize the total achievable sum-rate. However, this goal is subject to constraints such as bandwidth, transmission power, and SINR. We assume that LiFi APs have maximum available band-

width $B_{\text{max}}^{\text{LiFi}}$ and maximum allowable transmission power $P_{\text{max}}^{\text{LiFi}}$, while RF APs have similar parameters $B_{\text{max}}^{\text{RF}}$ and $P_{\text{max}}^{\text{RF}}$, and P_e is the power transmitted by LiFi AP e. Additionally, let y_{ef} represent the minimum SINR threshold for a $e-f$ pair. Denoting z as $[B_{ef}, P_e, \alpha_{ef}]$, the resource allocation problem is formulated as follows:

$$\mathcal{P}: \max_{B_{ef},P_e,\alpha_{ef}} r_e, \qquad \text{for } e \in |\mathcal{E}|, f \in |\mathcal{F}|, \tag{4.16}$$

subject to

$$\mathcal{C}_1: \sum_{f\in\mathcal{F}} \alpha_{ef} B_{ef} \leq B_{\text{max}}^{\text{LiFi}}, \quad \text{for } e \in \mathcal{E}\backslash\{0\}, \tag{4.17}$$

$$\mathcal{C}_2: \sum_{f\in\mathcal{F}} \alpha_{ef} B_{ef} \leq B_{\text{max}}^{\text{RF}}, \quad \text{for } e = 0, \tag{4.18}$$

$$\mathcal{C}_3: 0 \leq P_e \leq P_{\text{max}}^{\text{LiFi}}, \qquad \text{for } e \in \mathcal{E}\backslash\{0\}, \tag{4.19}$$

$$\mathcal{C}_4: 0 \leq P_0 \leq P_{\text{max}}^{\text{RF}}, \quad \text{for } e = 0, \text{and} \tag{4.20}$$

$$\mathcal{C}_5: SINR_{ef} \;\; \geq \;\; y_{ef}, \text{ for } e \in \mathcal{E}, f \in \mathcal{F}. \tag{4.21}$$

When equality is attained in constraint (4.21), the following criteria are obtained to prevent SINR violation

$$\begin{aligned} &1 - \textstyle\sum_{e\in\mathcal{E}} \sum_{f\in\mathcal{F}} \zeta_{ef} > 0, \text{ and} \\ &\textstyle\sum_{e\in\mathcal{E}} \sum_{f\in\mathcal{F}} \kappa_e \zeta_e \leq 1, \end{aligned} \tag{4.22}$$

where,

$$\zeta_{ef} = \left(1 + \frac{1}{y_{ef}}\right)^{-1}, \text{ and} \tag{4.23}$$

$$\kappa_{ef} = \frac{n_0 B_{ef}}{(G_{ef} P_e / y_{ef}) - n_0 B_{ef}} + 1. \tag{4.24}$$

Possibilities exist that a particular AP offers higher SINR than other APs, resulting into a huge number of UEs associating with it. The sum-rate in this case remains unaffected, however, the individual UE data rates may see a drastic reduction. Such a condition,

known as load imbalance, must be avoided. To ensure load balancing, we impose the final constraint on this system as

$$\mathcal{C}_6 : b_e \leq b, \tag{4.25}$$

where b_e is the number of UEs associated with an AP and b is the maximum number of UEs that can be associated to an AP.

The problem in (4.16) is a non-concave integer optimization problem. Conventional optimization algorithms assume one of these values. Therefore, we apply model-free DRL algorithms to solve the problem.

4.6 Framework

We formulate an action space composed of transmit power, bandwidth, and association parameter. The state space represents the two parts. The former has information about the location and positioning of UE and AP. The latter ensures that all the constraints mentioned in (4.17) to (4.25) are satisfied. If the constraints are not satisfied, the association will be aborted and a new iteration will be initiated. When all the constraints are satisfied, the reward, i.e, achievable sum-rate is maximized for an AP - UE pair.

4.6.1 Action Space $\mathcal{A}$

The multiple agents in the system take decisions based on the present state of the environment. Since we are using DDPG therefore, we are considering continuous multi-dimensional action space. Each variable in the action space involves three parameters, transmission power (P_{e_i}), AP - UE link bandwidth (B_{ef_i}), and association parameter (α_{ef_i}), $1 \leq i \leq l$ where l is the size of action space for each eth LiFi AP - fth UE link. Allocating downlink bandwidth and transmission power to the APs has a substantial impact on the system's possible sum-rate. Additionally, we are also determining the affiliation of UEs with APs for downlink data reception. DDPG's continuous action space provides greater precision compared to DQN's discrete action spaces. Each agent's action (RF and LiFi APs) represents the transmit power, bandwidth, and association parameter

vector for each user in the system. If the user is not in LOS with the LiFi AP, their transmit power and bandwidth is 0, and its gets associated to RF AP. The RF AP uses the actor-network's output to perform actions. We describe the action space parameters one-by-one:

1. **The allocation strategy association parameter (α_{ef_i})**: We first include the allocation strategy into the action space. In this regard, the first parameter included in the action space is the association parameter α_{ef_i}. The association parameter is an indicator variable which tells the association of an AP to a UE. The association is based on the level of SINR received. DRL agent works such that the UE gets associated to an AP which provides the higher SINR in it FOV with load balancing.

2. **Bandwidth (B_{ef_i})**: Once the association parameter has been formulated, the next issue is to find strategies for adjusting their bandwidths. Thus, the second parameter included in the action space is the bandwidth (B_{ef_i}). It is known that CC is directly proportional to the bandwidth. The bandwidth allocated to an RF AP is given as (B_{0f_i}) where as, for the LiFi AP it is given as (B_{ef_i}) such that $e \in \mathcal{E}\backslash\{0\}$. As the available spectrum is limited, we have applied the constraint on bandwidth usage. Similarly, like the allocation of transmit power, the CU allocates the bandwidth to the AP by keeping the constraint satisfied. The AP provides the association to UEs while balancing the load. Throughout this process the CU monitors the action so that the value chosen does not exceed the maximum limit in order to achieve the higher data rate.

3. **Transmission Power (P_{e_i})**: The third parameter included in the action space is the transmission power (P_{e_i}). The process of maximizing the achievable sum-rate must also consider eye safety and power saving. As we have put up a constraint on maximum power that a CU can allocate to an AP, hence among the values of action vector, highest value is chosen from the action vector to maximize the achievable sum-rate.

The action space $\mathcal{A}$ is thus formulated as

$$\begin{aligned}\mathcal{A} =& \{(P_{e_1}, B_{ef_1}, \alpha_{ef_1}), (P_{e_2}, B_{ef_2}, \alpha_{ef_2}), \ldots, (P_{e_l}, B_{ef_l}, \alpha_{ef_l})\} \\ =& \{a_{ef_1}, a_{ef_2}, \ldots, a_{ef_l}\},\end{aligned} \tag{4.26}$$

where P_e and B_{ef} are power and bandwidth allocations, respectively, for the e^{th} AP and f^{th} UE link. α_{ef} denotes the association parameter of an UE e and AP f link. These values are scaled to match operational parameters within the environment, with P_{e_l} constrained within $[P_{\text{min}}, P_{\text{max}}]$ and B_{ef_l} within $[B_{\text{min}}, B_{\text{max}}]$, where $P_{\text{min}}, B_{\text{min}}$ and $P_{\text{max}}, B_{\text{max}}$ are the minimum and maximum permissible levels for power and bandwidth, respectively.

4.6.2 State Space

In the proposed hybrid RF/LiFi system, the state space s_f provides the observation of the agent (CU) in the environment, which includes

1. The locations of UEs and APs at an instantiation, represented by $x_f, y_f, \text{AP}_{x_f}$ and AP_{y_f}.

2. The outcome of the action of the agent (CU) observed on the environment. The action of CU is choosing a bandwidth, power, and association parameter. The outcome of this action is observed on the constraints imposed on the system, represented by $C_1, C_2, C_3, C_4, C_5, C_6$.

In [17, 31], authors have used state space as a constraint. For the environment's state vector, we employ the constraints along with UEs-APs location at any given moment. The state s_f for UE f can be described as $[x_f, y_f, \text{AP}_{x_f}, \text{AP}_{y_f}, C_1, C_2, C_3, C_4, C_5, C_6]$, where x_f, y_f are the normalized coordinates of UEs f's location, and $\text{AP}_{x_f}, \text{AP}_{y_f}$ are the normalized location of the AP e serving UE f. The latter part C_1 to C_6 are the constraints mentioned in (4.17) to (4.25). The rationale for choosing constraints in the state space is that it directly affects the optimal resource allocation and data rate. Hence, the state vector

for f users is defined as $[s = s_1, s_2, ..., s_f]$, where state variable s_f for user f is given as

$$s_f = [x_f, y_f, \text{AP}_{x_f}, \text{AP}_{y_f}, C_1, C_2, C_3, C_4, C_5, C_6]. \tag{4.27}$$

Note that C_m takes values in the set $\{0, 1\}$ for all $m = 1, 2, \ldots, 6$, with a value closer to 1 indicating the degree to which it effectively satisfies the corresponding constraint. C_m becomes zero when any of the constraints is not satisfied.

4.6.3 Reward $R(s, a)$

The aim of the reward function is to optimize network performance and maximize overall achievable sum-rate. It is mathematically expressed as

$$R(s, a) = \sum_{e \in \mathcal{E}} r_e = \sum_{e \in E} \sum_{f \in \mathcal{F}} \alpha_{ef} r_{ef}, \tag{4.28}$$

where α_{ef} and r_{ef} are the parameters obtained after one complete inference cycle of DDPG algorithm is completed. To use DDPG, we must define the problem in continuous state space, action space and rewards as mentioned. In DDPG, neural networks are trained using the set ($\text{s}_t, a_t, r_t, s_{t+1}$).

4.7 Simulation Results

We consider the system model shown in Fig. 4.1. To assess its practicality, we utilize a Gymnasium environment developed in Python. Simulation data are taken from [50]. The setup includes 8 LiFi APs surrounding the RF AP arranged on the room ceiling. The area of the room considered is $10m \times 10m$ with specific parameters mentioned in Table 4.2.

4.7.1 Performance Analysis

We compare our proposed DDPG algorithm with PPO, twin-delayed deep deterministic policy gradient (TD3), DQN, and double deep Q network (DDQN) on several performance

Algorithm 2 DDPG for resource allocation in hybrid RF/LiFi System

Initialize critic network $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer D
for episode = 1, M **do**
    Initialize a random process N for action exploration
    Receive initial observation state s_1
    for t = 1, T **do**
        Select action $a_t = \mu(s_t|\theta^\mu) + N_t$ according to the current policy and exploration noise
        Execute action a_t and observe reward r_t and new state s_{t+1}
        Store transition (s_t, a_t, r_t, s_{t+1}) in D
        Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from D
        Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
        Update critic by minimizing the loss:

$$L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$$

        $\theta^Q \leftarrow \theta^Q - \alpha \nabla_{\theta^Q} L$
        Update actor policy using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s_i}$$

        Update the target networks:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$

    end for
end for

metrics. We initially opted for DQN as a benchmark because DDPG efficiently handles continuous action spaces, unlike the DQN-based learning algorithm which is suited for discrete actions. The evaluation is based on detailed simulations running over 2000 episodes within a hybrid RF/LiFi system with learning rates 0.0003. It outperforms the DQN in all the scenarios. To mark the robustness of our proposed algorithm, we compared it with several other mature algorithms like DDQN, TD3 and PPO. Remarkably, DDPG outperforms all of them, i.e., PPO, TD3, DQN, and DDQN in all the metrics.

Fig. 4.2 compares the performance of DDPG, PPO, TD3, DQN, and DDQN in terms of the achievable sum-rate (in Mbps) over the number of episodes.

Table 4.2: Simulation Parameters

Li-Fi Parameter	**Values**
Number of RF AP	1
Number of LiFi APs	8
Height of Room	5m
Number of UEs	8
Area of PD	1 cm^2
LiFi AP average optical power	9.2 W
Noise at LiFi AP	10^{-21} A^2/Hz
Responsivity	0.28 A/W
FOV at UE	$60°$
Users Speed	0 to 0.5 m/s randomly
Area Size	10 $\times 10m^2$
Bandwidth	20 MHz
Semi-Angle of LiFi APs	$70°$
Hyper-parameter	**Values**
Number of episode	2000
Learning rate	0.0003
Discount factor	0.9
DDPG architecture	**Activation**
Fully connected actor layer	Sigmoid
Fully connected critic layer	ReLU

DDPG shows consistent improvement over the episodes, with the sum-rate eventually stabilizing at approximately 1750 Mbps. This indicates that DDPG successfully learns to optimize the sum-rate efficiently over time. PPO demonstrates similar steady growth as DDPG, eventually reaching a slightly lower sum-rate, stabilizing close to 1700 Mbps. This indicates strong performance in optimizing the sum-rate, though it is slightly less efficient compared to DDPG. This is likely because under low-noise conditions and slow movements of UEs, DDPG achieves better performance. TD3, a variant of DDPG, performs satisfactorily but remains inferior to DDPG. Although TD3 is also an actor-critic method, it employs twin critics and target policy smoothing to reduce overestimation bias. However, in this set-up, the state spaces and action spaces are relatively low-dimensional, with comparatively low complexity and minimal noise. Due to this, DDPG's simplicity makes it perform better in this set-up. DQN exhibits high volatility, with frequent fluctuations, showing instability and inconsistent improvement. Similarly, DDQN also shows significant volatility and fails to consistently improve the sum-rate. DDPG and PPO show

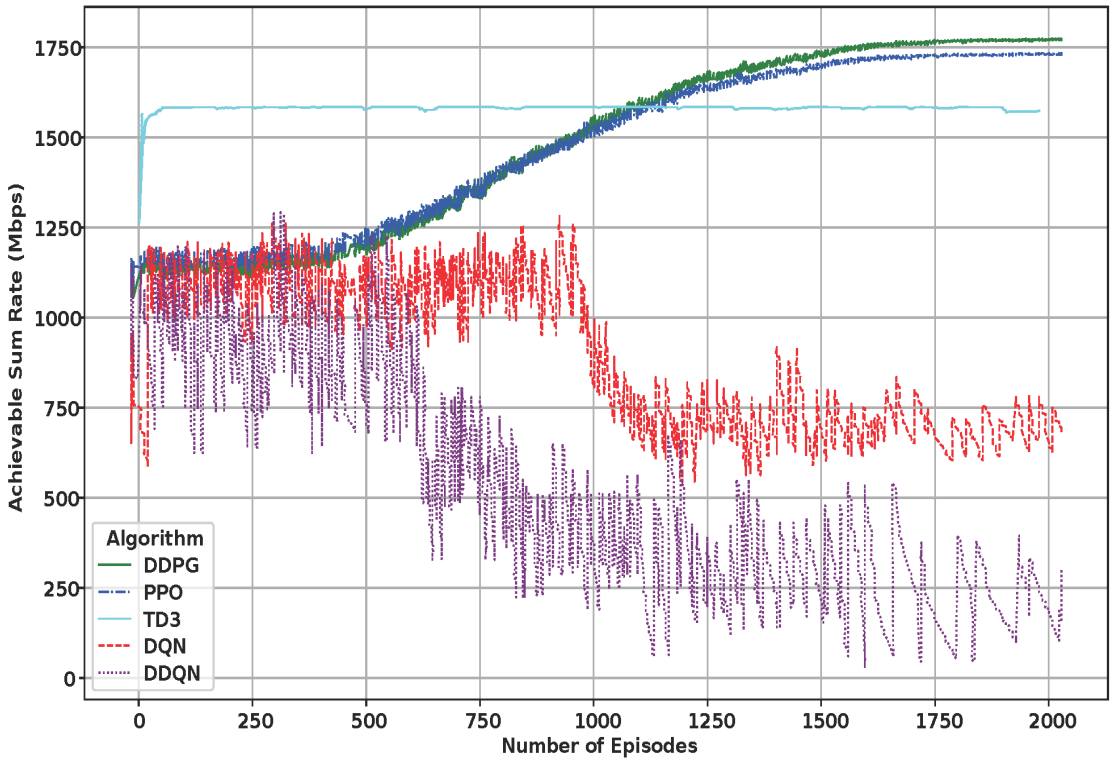
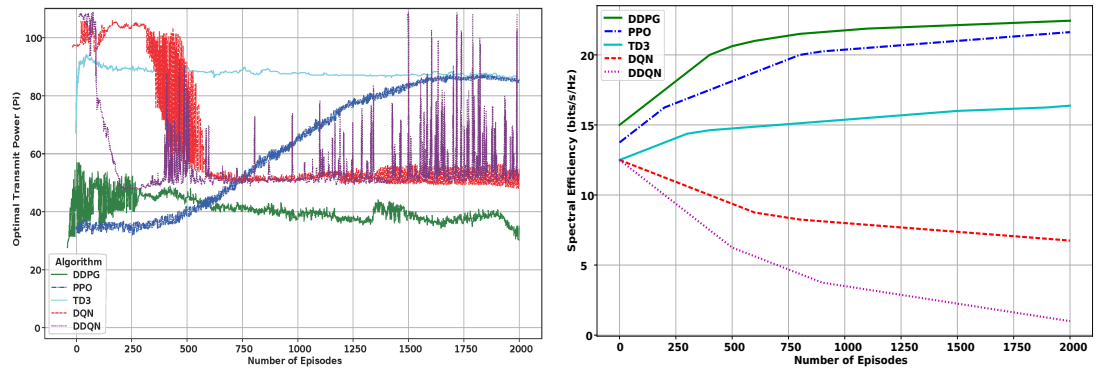


Figure 4.2: Comparison of DDPG with DQN, DDQN, PPO, and TD3 based learning algorithms at learning rate 0.0003.

stable learning and convergence to higher sum-rates compared to DQN and DDQN. This is likely due to their ability to handle continuous action spaces more effectively. Among the algorithms, DDPG and PPO achieve the best results, indicating their suitability for the considered large and dynamic set-up. TD3 also performs satisfactorily, however, remains slightly behind DDPG and PPO. DQN and DDQN, being better suited for discrete action spaces, struggle in this environment.

Fig. 4.3a compares the performance in terms of the optimal transmit power over the number of episodes. For DDPG, the transmit power fluctuates initially but stabilizes around $35 - 40$ mW after approximately 500 episodes. This suggests that DDPG is relatively stable for optimal transmit power, optimizing it in the best possible manner in this set-up. For PPO, the transmit power initially starts at around 35 mW but reaches higher values as the learning progresses compared to DDPG. TD3 quickly stabilizes at a transmit power of around $70-80$ mW. While TD3 shows stable performance, it requires higher transmission power. DQN displays significant volatility in transmit power, starting at around $90 - 100$



(a) (b)

Figure 4.3: (a) Convergence of DRL algorithms in terms of mean power consumption at learning rate 0.0003.(b) Comparison of spectral efficiency vs number of episodes

mW and reducing to around 50 mW. DQN and DDQN struggle to find a stable policy for transmission power in this set-up. This instability indicates difficulties in optimizing transmit power, likely due to their inability to handle continuous action and state spaces. DDPG, PPO, and TD3 shows more stable convergence. This indicates their superior ability to handle continuous action spaces. It indicates that algorithms supporting continuous action spaces (like DDPG, PPO and TD3) are better suited for the set-up considered in this chapter. Fig.4.3b shows spectral efficiency versus number of episodes. DDPG achieves the highest efficiency, with PPO close behind, while TD3 converges slower with lower efficiency. DQN and DDQN perform poorly due to their limitation to discrete state-action spaces. Overall, actor–critic methods (DDPG, PPO, TD3) clearly outperform Q-learning-based methods, underscoring the need for continuous state-action frameworks in dynamic hybrid RF/LiFi systems.

Fig. 4.4 and 4.7 illustrate the variation in the achievable sum-rate as a function of the ceiling height in an indoor environment. They compare the impact of different learning rates on system performance. As expected, in both cases, the achievable sum-rate declines as the ceiling height increases. This is likely as the channel gain is inversely proportional to the square of the distance between an UE and an AP, causing it to decrease with increasing ceiling height. Additionally, variations in learning rates influence the convergence behavior, thereby affecting the achievable sum-rate. The results clearly show that the

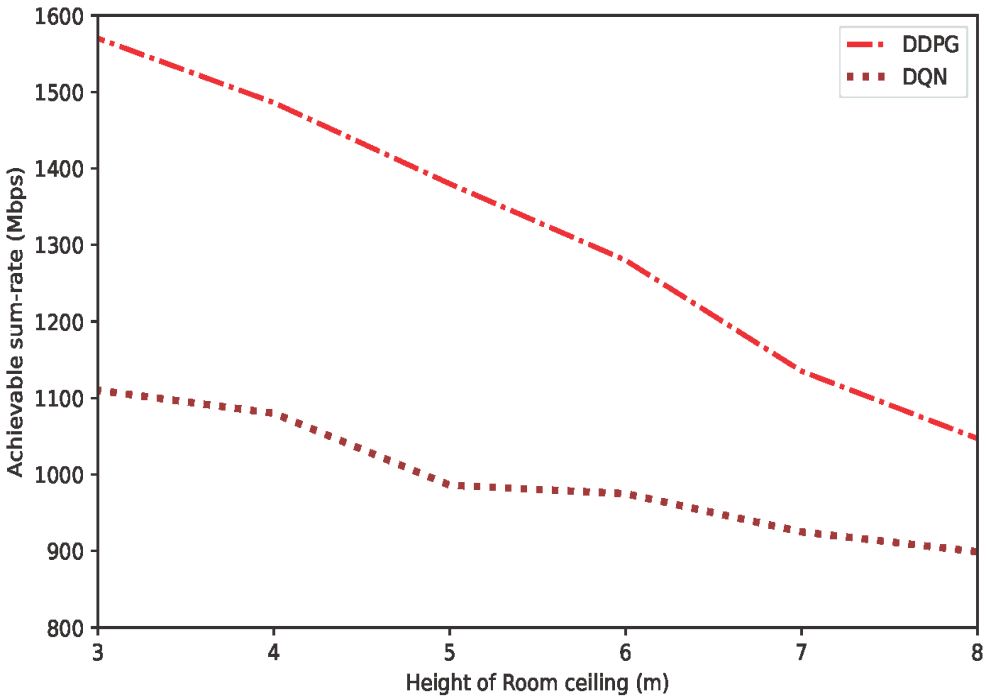


Figure 4.4: Convergence of proposed algorithm with ceiling height at learning rate 0.0003

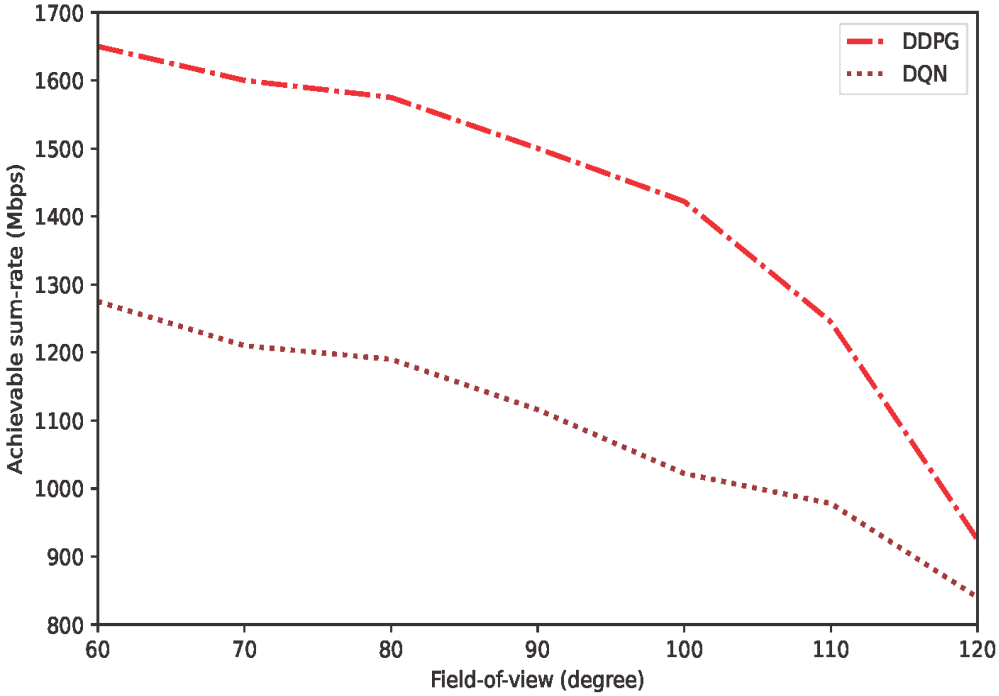


Figure 4.5: Convergence with field of view at learning rate 0.0003.

higher learning rate makes the convergence smoother with achieving higher sum-rate.

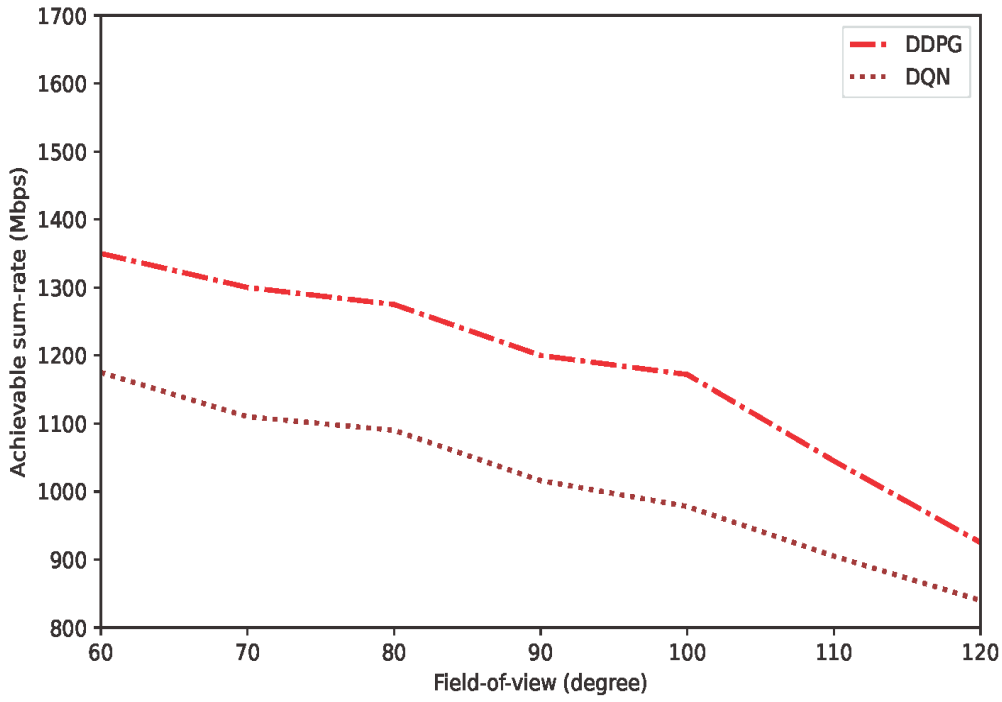


Figure 4.6: Convergence with field of view at learning rate 0.03

Fig. 4.5 and 4.6 show the achievable sum-rate for different values of the FOV. As expected, increasing the FOV angle results in a reduction of the overall achievable sum-rate. This is likely because a larger FOV allows the reception of more undesired multipath components and background interference degrades the signal quality received at UEs. As a result, the achievable sum-rate for a UE decreases. A narrower FOV helps in focusing more on the desired received signal with lesser interference signals, leading to achieve higher sum-rate and vice-versa.

Furthermore, it can be observed that the proposed DDPG algorithm consistently outperforms the DQN in terms of achievable sum-rate. This superior performance of DDPG is due to its ability to handle continuous action spaces more effectively. In contrast, DQN suffers because it relies on discrete action spaces.

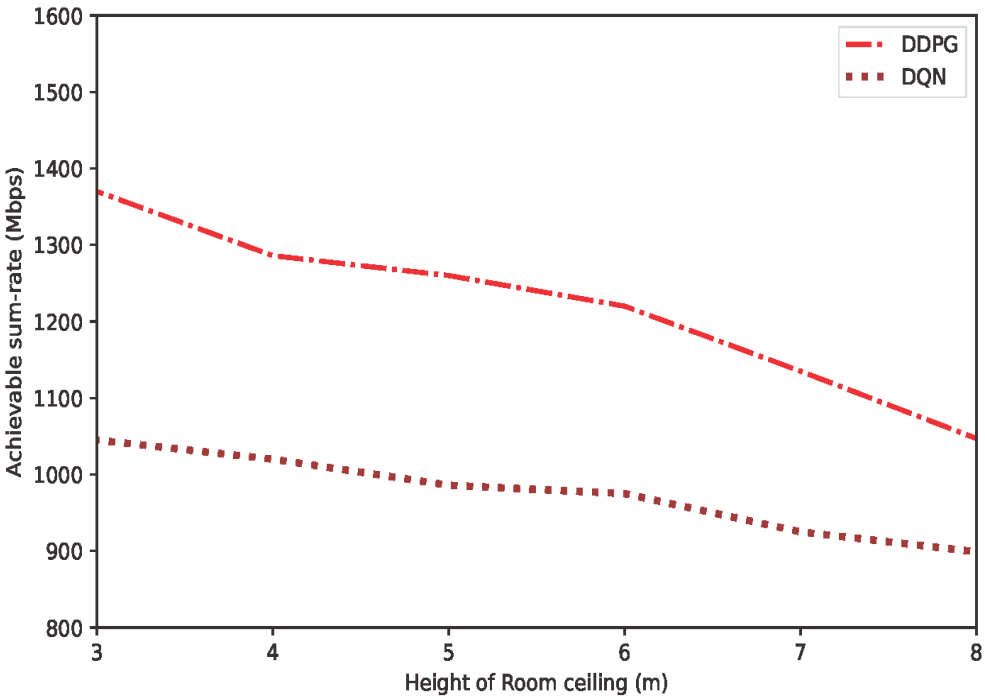


Figure 4.7: Convergence of proposed algorithm with ceiling height at learning rate 0.03

4.8 Conclusion

In this chapter, we investigate our proposed DDPG method for resource allocation in dynamic and hybrid RF/LiFi systems. Remarkably, we created a near-realistic scenario environment using Gymnasium. The proposed algorithm was compared with several mature DRL algorithms that support both the discrete and continuous action and state spaces. Simulation results show that DDPG outperforms all of them. Also, with it's capability to handle continuous action space, DDPG shows 62.8% better performance than DQN in terms of achievable sum-rate and 42.8% in terms of optimal transmit power.

Chapter 5

A2C and PPO with Random orientation in Hybrid RF/VLC

As mentioned earlier, VLC has emerged as a promising technology, delivering high-speed data transmission for 5G and beyond communication. Nevertheless, its susceptibility to blockages demands a co-deployment with traditional RF systems to ensure uninterrupted connectivity. This co-deployment, known as a hybrid RF/VLC system, is a subset of HetNets and offers interoperability, energy efficiency, and optimal resource utilization. In hybrid RF/VLC, efficient resource allocation and load balancing are crucial. In previous chapters, it was discussed that existing DQN learning-based methods designed to address these issues, fail in large and dynamic environments. In this chapter, we further investigate alternative approaches for optimal resource allocation and load balancing in dynamic and large hybrid RF/VLC systems, to achieve maximum data rates for users. Additionally, we take random orientation of UEs into account. We propose two model-free on-policy DRL based schemes, namely A2C and PPO, for efficient resource allocation in this set-up. Simulation results show that the A2C and PPO based schemes outperform the DQN learning scheme by 31.3% and 32.5%, respectively, in terms of data rates. The proposed schemes also outperform DDPG in data rate maximization by up to 8.1% and 9.7%, respectively.

5.1 Overview

The recent CISCO report indicates 29.3 billion internet users and network devices worldwide [3]. Due to RF spectrum limitations, alternative wireless communication using unlicensed optical spectrum, like VLC, has gained prominence, especially for indoor use [21]. VLC is energy-efficient, offers illumination, data transmission, positioning, and is immune to RF interference, making it suitable for sensitive environments [132]. Standalone VLC deployment is impractical due to its susceptibility to blockages. Its co-deployment with RF creates a hybrid RF/VLC system. Hybrid RF/VLC is a part of HetNets, that can enhance capacity, mobility, and energy efficiency [32]. Optimal resource allocation in these systems remains a key research focus [17, 20]. Particularly in downlink resource allocation, non-concavity involved in the joint optimization problem of the downlink bandwidth, transmission power of the APs, and the integer association parameter is a crucial issue [21].

Several classical optimization algorithms have been proposed in the existing literature [22, 20, 23, 24, 25, 133, 26, 134] to address the dual issues of non-concavity and integer optimization. However, conventional optimization mechanisms often rely on assuming values for at least one of the optimization parameters and finding the best values for the remaining parameters [27]. Notably, as the optimization parameters jointly impact the data rate, their joint optimization incorporating the interplay between them is necessary without making any presumptions on their values. The assignment of downlink power and bandwidth has a direct impact on SINR, and vice versa. Presuming a value for the downlink bandwidth, AP transmit power, or association parameter results to suboptimal outcome.

Deep learning, a subset of machine learning, has been found efficient in solving problems involving integer optimization or combined optimization of multiple variables [17] and for optimal resource allocation in hybrid RF/VLC [17, 29, 31]. In [17], the authors have used DQN learning to solve the joint optimization problem of association parameter, bandwidth and transmission power. Wang et al. in [29] use deep learning for seamless

Table 5.1: Related Works

Parameter	Proposed Approach	Pros	Cons
Optimization of transmit power [30], transmit power and user scheduling [32]	Distributed DRL - DDPG [30], actor-critic model-free [32]	Better performance, adapt changes [30], renewable energy harvesting [32]	Higher implementation complexity [30], no load balancing [32]
Achievable sum-rate [31], [17], [135]	DQN-transfer learning [31], DQN [17], limited-content and limited-frequency feedback [135]	High data rate and fewer number of iterations [31], moment to moment update [17], both downlink and uplink [135]	No load balancing, off-policy [31] [17], fixed bandwidth and average power [135]
EE [33], [22], [23], [50], [34], EE and spectral efficiency (SE) [25],	DRL[33], Dinkelbach's algorithm, successive convex approximation [22], ϵ-constraint method [23], Dinkelbach-type procedure [50], iterative joint user association and power control [25], DRL [34]	Enhanced QoS [33], improved cell edge user experience [22], LOS blockages and intercell interference [23], power and bandwidth efficient allocation [50], high EE and SE [25], both uplink and downlink [34]	No load balancing and random orientation [33], interference and no load balancing [22], convexity, sum-rate and power trade-off [23], no association parameter and load balancing[50], no load balancing [25] [34]
Maximization of proportional fairness[20], outage probability [35]	Dual decomposition method-Karush-Kuhn-Tucker [20], deep learning [35]	Increased fairness [20], dynamicity and human blockers consideration [35]	Time and bandwidth equal allocation [20], no random orientation and no load balancing [35]
Seamless handover protocol and data-rate [29], resource allocation and load balancing with handover [136]	DRL [29], game theory and OFDM Access resource allocation [136]	Increased data-rate [29], random orientation [136]	Fixed movement path, no load balancing [29], no bandwidth and power constraints [136]
Secrecy capacity [36] [37],	DQN [36], DDPG [37]	Enhanced secrecy capacity reliable data rate [36] [37],	No random orientation and load balancing [36], [37]

handover and increased downlink data rate in an ultradense deployment of VLC APs. DRL based solutions have been used against the heuristic methods to improve the stability and optimize the transmit power [30]. In [31], transfer learning has been proposed for performing optimal resource allocation to maximize the data rate.

Although, the primary goal of these works is sum-rate maximization, the DQN algorithms [17, 31] struggle with efficiency in large, dynamic environments due to the expansive action space and off-policy nature [32]. The increasing number of UEs and their dynamism further expand the action space. Since DQN operates as an off-policy algorithm, it uses the experience replay to learn and has no agent to pick the actions. Thus, it learns from recorded sessions and performs suboptimally. Further, off-policy DRL [29, 30], which works with discrete action spaces, faces challenges in large areas as continuous action spaces need to be converted to discrete ones, introducing high quantization noise. In contrast, on-policy algorithms, where agents select their own actions, offer better stability and performance by using multiple agents learning in parallel [137]. On-policy algorithms also handle practical considerations like random orientation of UEs in a more efficient manner.

In this chapter, we investigate the use of on-policy DRL algorithms, specifically A2C and PPO. These algorithms handle both discrete and continuous action spaces, making them suitable for optimization in large, dynamic setups. A2C and PPO use a model-free DRL approach to solve non-concave optimization problems in hybrid RF/VLC systems, efficiently managing large, dynamic conditions and expanding action spaces. These policy-based methods allow multiple agents to learn in parallel by interacting with the environment, optimizing all variables simultaneously. Additionally, they offer better sample complexity and ease of implementation [138].

5.2 Chapter Contributions

Our main contributions in this chapter are as follows:

1. *Practicality of the set-up and the problem*: We consider optimization in a dynamic 20×20 meter hall with variable speed users and random device orientations.

2. *Deep Learning model for a large hybrid RF/VLC setup*: We develop a model-free DRL communication model for optimizing UE-AP connectivity.

3. *Development of deep learning framework*: A multi-objective optimization framework for maximizing the sum-rate by optimizing association parameters, transmission power, and downlink bandwidth subject to various constraints has been introduced for a large and dynamic system.

4. *Incorporating load balancing*: We formulate a comprehensive set of constraints that incorporates load balancing along with minimum SINR requirements, transmit power limitations, bandwidth considerations, and association parameters.

   The rest of the chapter is organised as follows: Section 5.3 illustrates system model, the proposed mechanism to solve the optimization problem is discussed in Section 5.4. Section 5.5 discusses the performance of proposed schemes and Section 5.6 concludes the chapter.

5.3 System Model

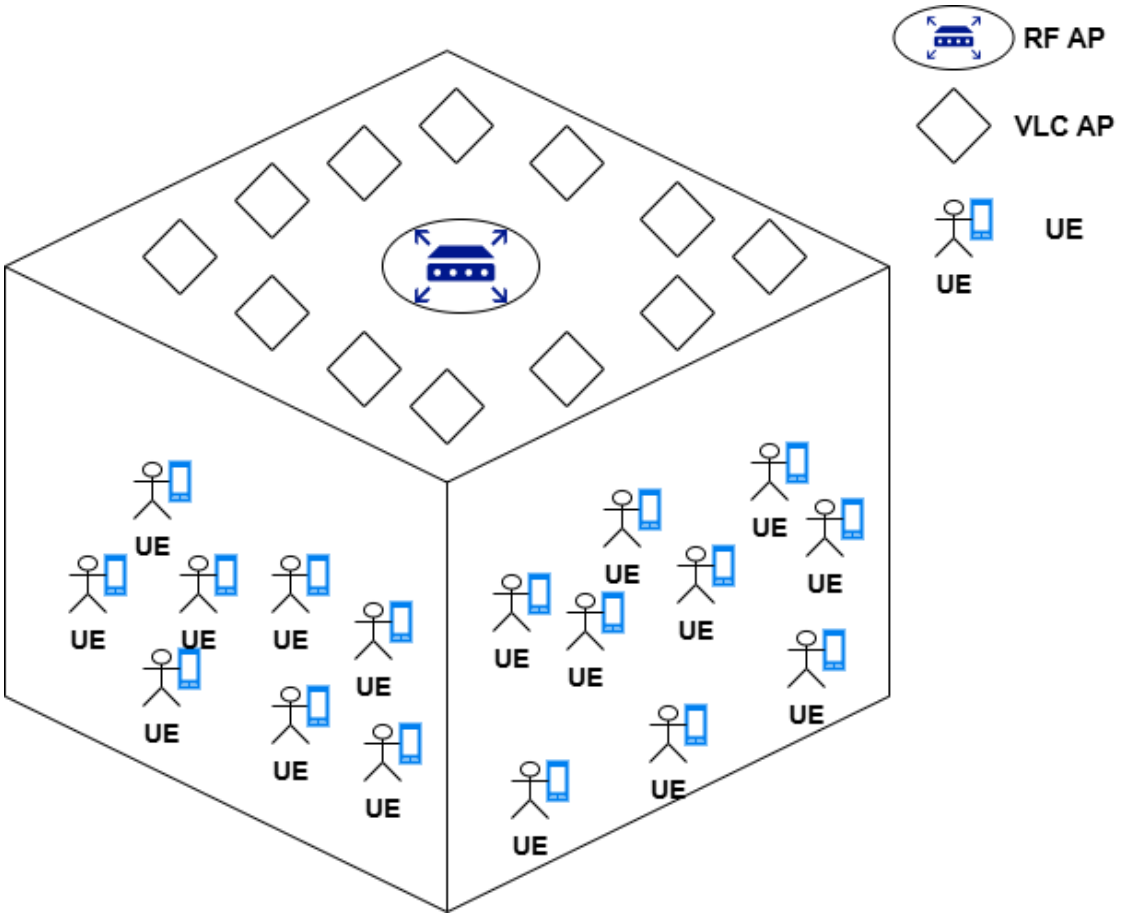


Figure 5.1: Hybrid RF/VLC Environment

Let $\mathcal{N}$ denote the set of VLC and RF APs in a 20×20 meter hall. The system model includes $|\mathcal{N}| - 1$ VLC APs and 1 RF AP, all on the ceiling. A CU near an RF AP dynamically manages the network. If a UE only receives NLOS components from a VLC AP, it connects to another VLC AP or the RF AP with the maximum SINR. The CU handles bandwidth allocation, transmission power control, and UE-AP association. APs are indexed as $i = 1, 2, \ldots, |\mathcal{N}|$, with VLC APs as $i = 1, 2, \ldots, |\mathcal{N}| - 1$ and the RF AP as $i = 0$. APs are at height h from the UEs, which are indexed as $j = 0, 1, 2, \ldots, |\mathcal{U}|$.

5.3.1 VLC System Modeling

According to [50] the optimal channel gain CG_{ij}^{v}, in LOS signal for VLC is given as

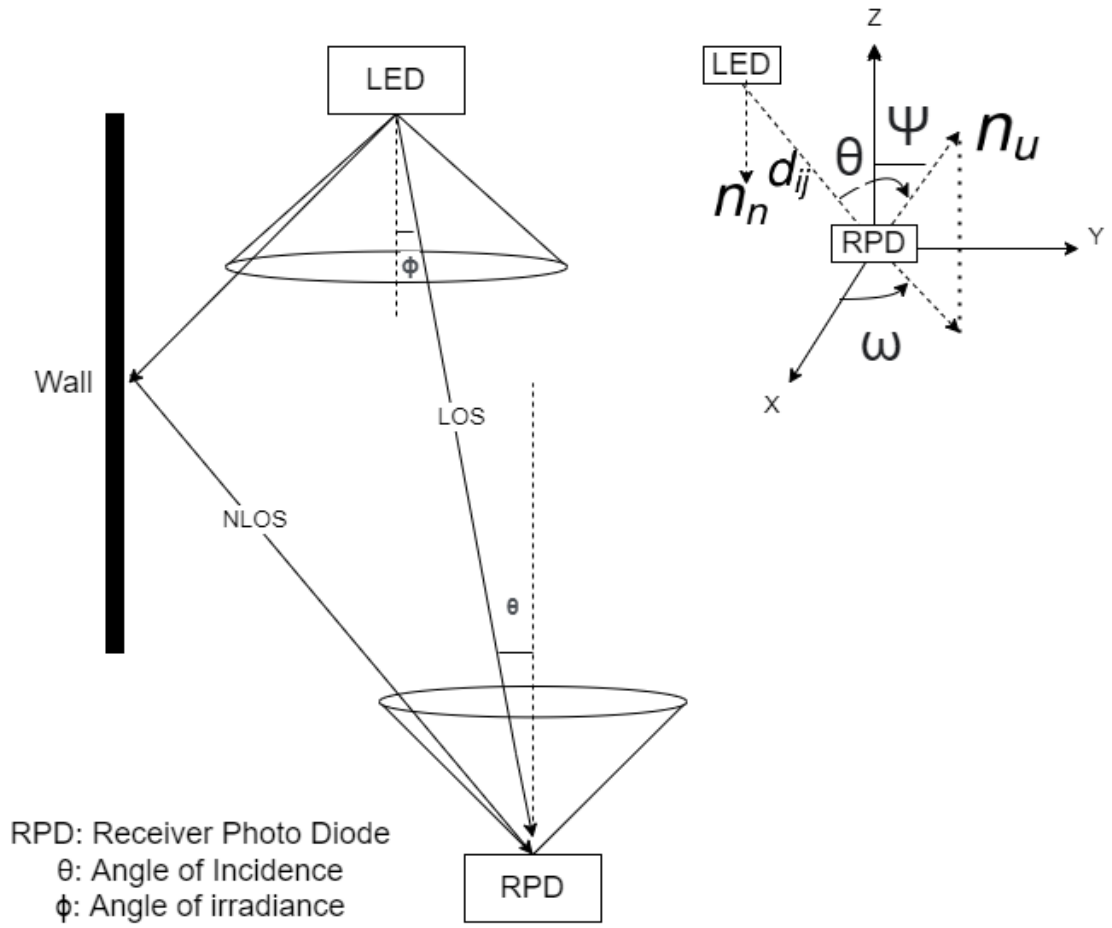


Figure 5.2: Downlink geometry in LOS-NLOS scenario with polar and azimuth random orientation angle of UE in VLC Environment

$$CG_{ij}^{v} = \frac{(m+1)A_{\text{pd}}\cos^{m}\theta_{ij}\cos\phi_{ij}T_{\text{opt}}(\phi_{ij})g(\phi_{ij})}{2\pi d_{ij}^{2}}, \tag{5.1}$$

where $T_{\text{opt}}(\phi_{ij})$ is the optical filter gain (constant or unity within the receiver's FOV), θ_{ij} is the incidence angle at UE j from AP i, ϕ_{ij} is the irradiance angle at AP i, and d_{ij} is the distance between UE j and AP i.

The concentrator gain $g(\phi_{ij})$ in (5.1) is written as

$$g(\phi_{ij}) = \begin{cases} \frac{n^2}{\sin^2\phi_{\text{FOV}}} & \text{if } 0 \le \phi_{ij} \le \varphi_{\text{FOV}} \\ 0 & \text{if } \phi_{ij} > \varphi_{\text{FOV}}, \end{cases} \tag{5.2}$$

where φ_{FOV} is the FOV angle of the UE, n is the refractive index, and the order of the Lambertian radiation profile m is given as

Table 5.2: Important notations and their meanings

Notation	**Meaning**
i	Index of APs
j	Index of UEs
k	The interferer AP index
r_{ij}	The achievable data rate between the ith AP and the jth UE
r_i	Downlink data rate of the ith AP
$B_{\text{max}}^{\text{VLC}}$	Maximum BW allotted to VLC AP
$B_{\text{max}}^{\text{RF}}$	Maximum BW allotted to RF AP
P_i	Transmit power of the ith AP
P_t	Total optical power received from an AP
$\mathcal{N}$	Set of APs
$\mathcal{U}$	Set of UEs
m	Lambertian coefficient
θ	Angle of incidence
ϕ	Angle of irradiance
$CG^{(l)}$	DC Channel gain after lth reflection
A_{pd}	Area of PD
CG_{ij}	Channel gain between ith AP and $j$yh UE
CC	Channel capacity
B_{max}^{RF}	Maximum bandwidth allotted to the jth UE by the ith AP
ρ_j	Responsivity of the receiver PD
$\mho_{ij}$	Indicator function showing association of AP i-UE j
N_0^r	RF noise power
N_0^v	VLC noise power
$P_{\text{max}}^{\text{RF}}$	Maximum transmit power of the RF AP
$P_{\text{max}}^{\text{VLC}}$	Maximum transmit power of the VLC AP

$$m = -\frac{\ln 2}{\ln \cos \phi_{1/2}}, \tag{5.3}$$

where $\phi_{1/2}$ is semi-angle at half illuminance. In a typical indoor scenario, some transmitted light components get reflected from the walls and are received by the photo diode. For such NLOS light components, the channel gain is given as [139]

$$CG_{\text{NLOS}} = \sum_{l=0}^{\infty} CG^{(l)}, \tag{5.4}$$

where *l* represents the reflection index, and the source LED exhibits channel gain $CG^{(l)}$

after *l*th reflection which is further expressed as

$$CG^{(l)} = \int_S CG_1 CG_2 \dots CG_{l+1} OP_q^{(l)} dA_s, \tag{5.5}$$

where dA_s represents the tiny area of reflection surface, while $OP_q^{(l)}$ is the optical power of the light ray component after undergoing l reflections emitted from the qth transmitting VLC AP. $CG_1, CG_2, \dots, CG_{l+1}$ are the DC channel gains of each traced path for the reflected component expressed as [139], [116]

$$\begin{aligned} &CG_1 = \tfrac{(m+1)A_s}{2\pi d_1^2} \cos^m(\theta_1) \cos(\phi_1), CG_2 = \tfrac{A_s}{\pi d_2^2} \cos^m(\theta_2) \cos(\phi_2), \\ &\dots, CG_{l+1} = \tfrac{A_s}{\pi d_{p+1}^2} \cos^m(\theta_{l+1}) \cos(\phi_{p+1}) T_{\mathrm{s}}(\phi_{l+1}) g(\phi_{l+1}), \end{aligned} \tag{5.6}$$

where A_s is the area of the wall where incident light hits, θ_i and ϕ_i for $i = 1, 2, \dots, l+1$ are the angles of irradiance and incidence at the l reflections.

The random orientation of UEs due to users' angular hand movements impacts θ_{ij}, making it a random variable [135, 136]. This effect can be characterized using the polar angle ψ and azimuth angle ω, as shown in Fig. 5.2. The statistics of the LOS depend on UE orientation. In (5.1), ϕ_{ij} remains unaffected by random orientation and is expressed as

$$\cos \phi_{ij} = -\frac{\mathbf{n_u} \cdot \mathbf{d_{ij}}}{|\mathbf{d_{ij}}|}, \tag{5.7}$$

where $\mathbf{n_u}$ is the UE's normal vector. Considering random orientation, the UE's normal vector n_u in the Cartesian coordinate system is defined as [135]

$$\cos \theta_{ij} = \frac{\mathbf{n_u}.\mathbf{d_{ij}}}{d_{ij}} = \left(\frac{x_n - x_u}{d}\right) \sin\psi \cos\omega + \left(\frac{y_n - y_u}{d}\right) \sin\psi \sin\omega + \left(\frac{z_n - z_u}{d}\right) \cos\psi, \tag{5.8}$$

where $[x_n, y_n, z_n]$ and $[x_u, y_u, z_u]$ denote the position vector of AP and UE, respectively, $\mathbf{d_{ij}}$ is the distance vector between VLC AP and UE. For simplicity of notation d_{ij} is written as d. For simplification, we can rewrite (5.8) as

$$g(\psi) = \cos \theta_{ij} = a \sin\psi + b \cos\psi, \tag{5.9}$$

where

$$a = -\left(\frac{x_n - x_u}{d}\right)\cos\omega - \left(\frac{y_n - y_u}{d}\right)\sin\omega, \tag{5.10}$$

and

$$b = \left(\frac{z_n - z_u}{d}\right). \tag{5.11}$$

The probability density function of $\cos\theta_{ij}$, approximated with truncated Laplace distribution, is given as

$$\widetilde{f}_{cos\theta_{ij}} = \frac{1}{\Delta\left(\widehat{\mu}_\psi, \widehat{b}_\psi, \widehat{\tau}_{max}\right)} exp\left(-\frac{|\widehat{\tau} - \mu_\psi|}{\widehat{b}_\psi}\right), \tag{5.12}$$

where τ represents the realization of the random variable $\cos\theta_{ij}$, and

$$\Delta(\hat{\mu}_\psi, \hat{b}_\psi, \hat{\tau}_{\max}) = 2\hat{b}_\psi\left(1 - \frac{1}{2}\exp\left(\frac{\hat{\mu}_\psi - \hat{\tau}_{\max}}{\hat{b}_\psi}\right) - \frac{1}{2}\exp\left(\frac{-1 - \hat{\mu}_\psi}{\hat{b}_\psi}\right)\right) \tag{5.13}$$

with $\tilde{f}_{\cos\theta_{ij}}$ in the support range of $-1 \leq \hat{\tau} \leq \hat{\tau}_{\max}$, $\widehat{\mu}_\psi = a\sin\mu_\psi + b\cos\mu_\psi$, and $\widehat{b}_\psi = b_\psi\,|a\cos\mu_\psi - b\sin\mu_\psi|$.

The total optical power received P_t from a single LED includes the LOS and NLOS components, and is expressed as

$$P_t = (CG_{\text{NLOS}} + CG_{ij}^v)P_i = CG_{ij}P_i \text{ for } i \in \mathcal{N}\backslash\{0\}, \tag{5.14}$$

where P_i is the power transmitted by the VLC AP i and $CG_{ij} = CG_{\text{NLOS}} + CG_{ij}^v$ is the effective channel gain between AP i and UE j for $i \in \mathcal{N}\backslash\{0\}$. The experimental evaluations use mean P_i, which is P_i averaged over $\widetilde{f}_{\cos\theta_{ij}}$. Also, LEDs emit light with wavelength λ and their spectral power distribution is given as $P_i(\lambda)$. Thus, P_i and P_t are given as

$$P_i = \int_\lambda P_i(\lambda)d\lambda, \tag{5.15}$$

and

$$P_t = \int_{\theta_{ij}} P_t(\theta_{ij})\widetilde{f}_{\cos\theta_{ij}}d\theta_{ij}. \tag{5.16}$$

5.3.2 Radio Frequency Channel Model

We consider RF signal transmission in the 60 GHz wideband region having an orthogonal frequency division multiplexing (OFDM) bandwidth B_{RF} [140]. The channel between the UE and RF AP is modelled as

$$CG^k = \sqrt{10^{-L(d_r)/10}} CG_w^n, \tag{5.17}$$

where CG^k is the complex channel transfer function between the kth sub channel and serving RF AP, and the separation distance is d_r the corresponding large-scale fading loss $L(d_r)$ in dB is given by [141]

$$L(d_r) = L(do) + 10\gamma \log_{10}(d/do) + X, \tag{5.18}$$

where at a distance of $d_o = 1$ m, the reference path loss $L(d_o)$ is 68 dB. The path loss exponent is $\gamma = 1.6$, and the shadowing component X follows a Gaussian distribution with a mean of zero and standard deviation $\sigma = 1.8$ dB [142].

Note that in LOS propagation, the channel transfer function is given as

$$CG_{0j} = \sqrt{10^{-L(d)/10}} \left(\sqrt{\frac{K}{1+K}} CG_d + \sqrt{\frac{1}{1+K}} CG_s \right), \tag{5.19}$$

$CG_d = \sqrt{1/2}(1+j)$ represents the direct path fading channel, while the scattered path fading channel is modeled as a complex Gaussian random variable, $CG_s \sim \mathcal{CN}(0,1)$. Here, j denotes the imaginary unit. The Rician factor K is set to 10 dB [141].

5.3.3 Communication Model

Let $X = [x_0, x_1, \ldots, x_{|\mathcal{N}|}]$ be the signal vector transmitted by the VLC APs and RF AP. When UE j is associated to the RF AP $i = 0$, it will receive the signal y_{0j} represented as

$$y_{0j} = \sqrt{CG_{0j} P_{RF}} \times x_0 + {N_0}^r, \tag{5.20}$$

where ${N_0}^r$ is the AWGN.

Conversely, if UE j is associated with a VLC AP i such that $i \in \mathcal{N}\backslash\{0\}$, y_{ij} will be represented as

$$y_{ij} = \rho_j C G_{ij} P_i x_i + \sum_{k \in \mathcal{N}\backslash\{0\}} \rho_j C G_{kj} P_k x_k D_k(\mho_{kj'}) + {N_0}^v, \tag{5.21}$$

where ρ_j is the responsivity of the receiving PD at the UE j, ${N_0}^v$ accounts for both the shot noise and thermal noise, and $D_k(\mho_{kj'}) = \left(1 - \prod_{j' \in \mathcal{N}\backslash\{0\}} (1 - \mho_{kj'})\right)$, where $\mho_{kj'}$ represents an indicator function indicating the association of the AP k with UE j' such that

$$\mho_{kj'} = \begin{cases} 1 \text{ if AP } k \text{ is associated to UE } j' \\ 0 \text{ otherwise,} \end{cases}. \tag{5.22}$$

In (5.21), we assume $\mho_{ij} = 1$ to indicate that UE j is connected to AP i. The desired combination is the AP i - UE j pair, while AP k interferes at UE j. The parameter $D_k(\mho_{kj'})$ ensures that AP k transmits only if it is associated with at least one UE j', where $j' \neq j$. If AP k is idle and not transmitting to any UE, $D_k(\mho_{kj'}) = 0$, indicating it is switched off.

5.3.4 Achievable Data Rate

When a UE connects to an RF AP, CC follows the Shannon's formula. For association with a VLC AP using IM/DD, the CC is lower-bounded [119] as

$$CC = \frac{1}{2} B \log_2 \left(1 + w \frac{\rho^2 P_t^2}{\sigma^2}\right), \tag{5.23}$$

where $w = e/2\pi$, ρ is the responsivity, B denotes the modulation bandwidth, and σ^2 the Gaussian distribution noise power.

Following (5.20), (5.21), and (5.23), we express the instantaneous achievable data rate at

the UE j as

$$r_{ij} = \begin{cases} B_{0j} \log_2 (1 + SINR_r), & \text{for } i = 0 \\ \frac{1}{2} B_{ij} \log_2 (1 + wSINR_v), & \text{for } i \in \mathcal{N} \backslash \{0\}, \end{cases} \tag{5.24}$$

where $SINR_r$ and $SINR_v$ are lower bounds for RF and VLC respectively and are given as

$$\begin{aligned} SINR_r &= \frac{P_{RF} C G_{0j}}{N_0 B_{0j}}, \text{ and} \\ SINR_v &= \frac{C G_{ij}^2 P_i^2}{N_0^v B_{ij} + \sum_{k \in \mathcal{N} \backslash \{0\}} \rho_j C G_{kj}^2 P_k^2 \left(1 - \prod_{j' \in \mathcal{N} \backslash \{0\}} (1 - \mho_{kj'})\right)^2}, \end{aligned} \tag{5.25}$$

where B_{ij} is the bandwidth of the AP i - UE j link. Based on the (5.24) the expression of throughput of AP i can be expressed as

$$r_i = \sum_{j \in \mathcal{U}} \mho_{ij} r_{ij}. \tag{5.26}$$

5.3.5 Problem Statement

Our objective is to find the optimal $\mho_{ij}$, B_{ij} and $P_i \; \forall \; i, j$ for achieving maximum r_i. The formulation of resource allocation problem is as follows

$$\mathcal{P} : \max_{B_{ij}, P_i, \mho_{ij}} r_i, \qquad \text{for } i \in |\mathcal{N}|, j \in |\mathcal{U}|, \tag{5.27}$$

subject to constraints:

$$\mathcal{C}_1 : \sum_{j \in \mathcal{U}} \mho_{ij} B_{ij} \leq B_{\text{max}}^{\text{VLC}}, \quad \text{for } i \in \mathcal{N} \backslash \{0\}, \tag{5.28}$$

where $B_{\text{max}}^{\text{VLC}}$ denotes the maximum bandwidth allocated to the VLC AP. Similarly,

$$\mathcal{C}_2 : \sum_{j \in \mathcal{U}} \mho_{ij} B_{ij} \leq B_{\text{max}}^{\text{RF}}, \quad \text{for } i = 0, \tag{5.29}$$

as it is known that available RF bandwidth is very low and expensive as compare to VLC. We have to exploit the available resources optimally. Therefore, this constraint ensures

that the maximum bandwidth allocated to the RF AP should not exceed the $B_{\text{max}}^{\text{RF}}$. Similarly, constraint is imposed on the transmission power considering eye safety and power budget, the constraint on VLC AP for maximum power is given as

$$\mathcal{C}_3 : 0 \leq P_i \leq P_{\text{max}}^{\text{VLC}}, \qquad \text{for } i \in \mathcal{N} \backslash \{0\}, \tag{5.30}$$

a similar constraint on maximum transmission power of RF AP is formulated as follows

$$\mathcal{C}_4 : 0 \leq P_0 \leq P_{\text{max}}^{\text{RF}}, \quad \text{for } i = 0, \tag{5.31}$$

and a constraint has been considered on SINR to achieve reliable communication as

$$\mathcal{C}_5 : SINR_{ij} \;\; \geq \;\; \gamma_{ij}, \text{ for } i \in \mathcal{N}, j \in \mathcal{U}. \tag{5.32}$$

When equality is achieved in (5.32), then the subsequent conditions are derived to avoid SINR constraint violation [17]

$$\begin{aligned} &1 - \textstyle\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{U}} \zeta_{ij} > 0, \text{ and} \\ &\textstyle\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{U}} \kappa_i \zeta_i \leq 1, \end{aligned} \tag{5.33}$$

where

$$\zeta_{ij} = \left(1 + \frac{1}{\gamma_{ij}}\right)^{-1}, \text{ and} \tag{5.34}$$

$$\kappa_{ij} = \frac{N_0 B_{ij}}{(C G_{ij} P_i / \gamma_{ij}) - N_0 B_{ij}} + 1. \tag{5.35}$$

If one AP offers high SINR, many UEs may associate with it, lowering individual data rates. To balance the load, a constraint limits the number of UEs per AP to u as

$$\mathcal{C}_6 : u_i \leq u, \tag{5.36}$$

where u_i is the number of UEs associated to AP$_i$.

The constraint in (5.32) ensures the minimum SINR for reliable AP-UE communication,

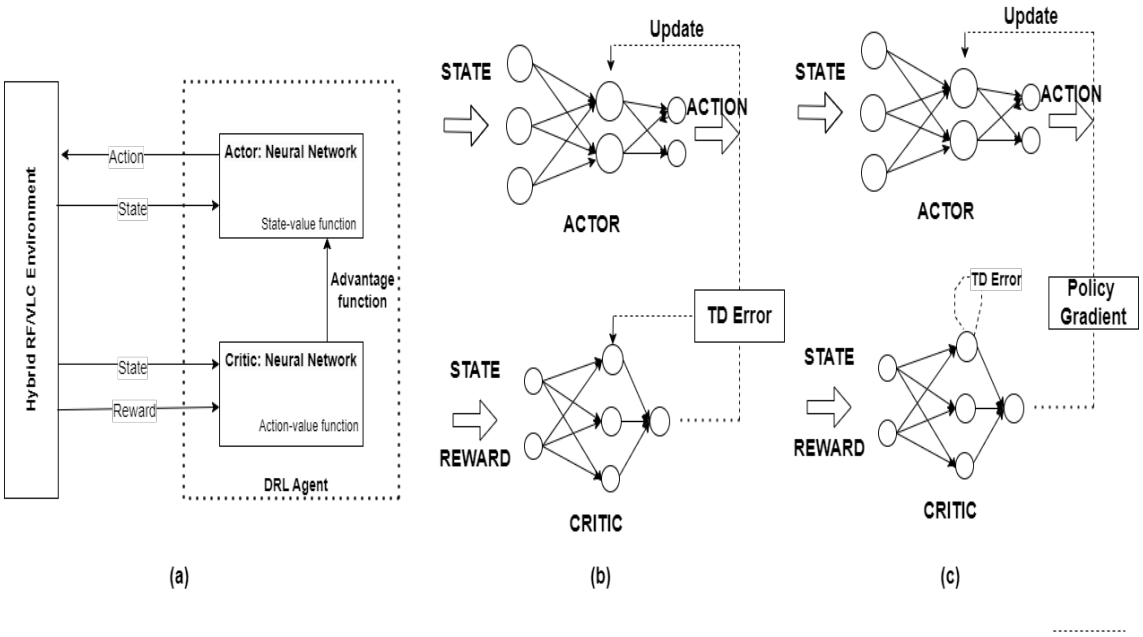


Figure 5.3: a) Flowchart of the proposed DRL network (b) Actor and critic stages of PS1 algorithm (c) Actor and critic stages of PS2 algorithm.

preventing interference from other APs. If interference violates this constraint, the DRL mechanism adjusts AP transmission power to comply. Constraints (5.33), (5.34), and (5.35) are designed to satisfy (5.32). Failure to meet these constraints results in AP interference.

The problem in (5.27) is a jointly non-concave integer optimization problem in $\mho_{ij}$, B_{ij}, and P_i, which conventional algorithms can't solve without presuming values. We apply A2C and PPO DRL algorithms to solve (5.27) subject to constraints (5.28)-(5.36).

5.4 Proposed Mechanism to solve $\mathcal{P}$

To solve the problem $\mathcal{P}$ in (5.27), we propose two on-policy model-free DRL techniques namely A2C and PPO.

5.4.1 Framework

We propose using A2C and PPO for resource allocation in a dynamic hybrid RF/VLC system. These policy gradient methods update policy parameters step-by-step to optimize the

problem in a continuous action space. Given the stochastic dynamic environment in (5.27), policy-based approaches perform better by avoiding quantization noise, which causes oscillation and non-optimal convergence in discrete spaces. Both mechanisms operate on continuous state and action spaces, $\mathcal{S}$ and $\mathcal{A}$.

5.4.1.1 Action Space $\mathcal{A}$

In this study, the action space is a continuous multi-dimensional space where each dimension corresponds to the control parameters for a single VLC AP. For a network with n VLC APs, the action a is represented as:

$$\begin{aligned}\mathcal{A} =& \{(P_{i_1}, B_{ij_1}, \mho_{ij_1}), (P_{i_2}, B_{ij_2}, \mho_{ij_2}), \ldots, (P_{i_n}, B_{ij_n}, \mho_{ij_n})\} \\ =& \{a_{ij_1}, a_{ij_2}, \ldots, a_{ij_n}\},\end{aligned} \tag{5.37}$$

where $P_{ij_l} \in [P_{\text{min}}, P_{\text{max}}]$ and $B_{ij_l} \in [B_{\text{min}}, B_{\text{max}}]$ represent the power and bandwidth allocation for each AP-UE link, respectively. Values for P_i and B_{ij} are scaled to match operational parameters, with $P_{\text{min}}, P_{\text{max}}$ and $B_{\text{min}}, B_{\text{max}}$ being the minimum and maximum permissible levels.

5.4.1.2 State Space $\mathcal{S}$

State space primarily consists of two parts. The former part consists of the dynamics of the user, namely its location and speed of motion. The latter part shows the satisfaction of the constraints (5.28)-(5.36) resulting from the actions in (5.37). For u users in the network, the state s can be described as $[s = \{s_1, s_2, \ldots, s_u\}]$ with each user state s_j given as

$$s_j = \left[x_j, y_j,\ v_{x_j}, v_{y_j},\ \text{AP}_{x_j},\ \text{AP}_{y_j}, \mathbf{C}\right], \tag{5.38}$$

where x_j, y_j are normalized coordinates of user j, v_{x_j}, v_{y_j} are its normalized velocity components, APx_j, APy_j are normalized AP locations serving UE j, and $\mathbf{C} = [C_1, C_2, ..., C_6]$, where C_1 to C_6 represent constraints (5.28)-(5.32) and (5.36). Each $C_m \in 0, 1$ indicates the degree to which its respective constraint is satisfied; C_m approaches 1 when the con-

straint is well-satisfied. If any constraint is not met, the corresponding C_m is set to 0. Note that **C** is common for all UEs.

5.4.1.3 Reward Function

The reward function $R(s, a)$ aims to optimize network performance by satisfying all constraints and maximizing the achievable sum-rate. If every constraint variable in state space is non-zero ($C_m \neq 0$), the reward is received in terms of the achievable sum-rate as

$$R(s, a) = r_i, \tag{5.39}$$

further if any $C_m = 0$, a value lower than the current $R(s, a)$ is returned to the CU.

As it can be seen in (5.39), the data rate of each individual ith AP (r_i) is first maximized. During this maximization, the constraint C_6 ensures that the number of UEs associated to an AP do not exceed a particular value. Thus, one AP is not crowded with UE connections and load balancing is ensured across all the APs. Since there is an interference term in the Shannon's capacity formula, the maximization of an individual ith AP takes into account the other APs also. Simultaneously there is a constraint on SINR. Thus, UEs are not associated to an AP only for the sake of reaching the upper bound allowed for that AP. Once the data rates of all the APs is maximized, the sum of all the data rates, $R = \sum_{i=1}^{N} r_i$ is considered as sum-rate. The sum-rate is calculated in the last step of the algorithm. Therefore, for all the evaluations in the simulation section, the sum-rate has been considered.

5.4.2 A2C based Proposed Scheme 1 (PS1)

PS1 combines policy and value-based methods using two neural networks: the actor and the critic. These networks operate on state space $\mathcal{S}$ and action space $\mathcal{A}$. The actor observes the environment and selects actions $a \in \mathcal{A}$ to maximize rewards, while the critic evaluates the policy and computes rewards/penalties using temporal difference (TD) error and the advantage function based on received rewards.

Algorithm 3 Implementation of PS1

Initialize Actor network with weights θ.
Initialize Critic network with weights ϕ.
Initialize Controlling Unit (CU).
Initialize discount factor γ.
Initialize learning rates $\alpha_\theta, \alpha_\phi$.
for each episode **do**
  Initialize system state $s = \{s_1, s_2, \ldots, s_u\}$ from CU
  Initialize episode reward $R = 0$
  for each step in episode **do**
    Choose an action $a \in \mathcal{A}$ based on $\pi_\theta(a|s)$
    Take action a, update system state to s' and receive reward R
    Update s' in CU
    Compute TD error $A = r + \gamma V_\phi(s') - V_\phi(s)$ and advantage A(s, a) = Q(s, a) - V(s)
    Update θ using A to optimize the policy:

$$\theta \leftarrow \theta + \alpha_\theta \nabla_\theta \log \pi_\theta(a|s) A$$

    Update ϕ to minimize mean square error (MSE) $= (r + \gamma V_\phi(s') - V_\phi(s))^2$:

$$\phi \leftarrow \phi - \alpha_\phi \nabla_\phi \text{MSE}$$

    $s = s'$
    Aggregate rewards R
  end for
  Store R for performance evaluation.
  if convergence criteria met **then**
    Break
  end if
end for

For any given state $s \in \mathcal{S}$, the actor network generates an action $a \in \mathcal{A}$, to maximize the expected return $J(\pi)$, defined as $J(\pi) = \mathbb{E}_{\tau\sim\pi}\left[\sum_{t=0}^{\infty}\gamma^t R(s_t, a_t)\right]$, where, τ is a trajectory, γ is the discount factor, and $R(s_t, a_t)$ is the reward at time t. For each state s the critic estimates the value function $V(s)$ as $V(s) = \mathbb{E}_{a\sim\pi}\left[\sum_{t=0}^{\infty}\gamma^t R(s_t, a_t)|s_0 = s\right]$. To address high variance in policy gradient methods, the critic computes the advantage function $A(s, a) = Q(s, a) - V(s)$, where $Q(s, a)$ is the action-value function, defined as

$$Q(s, a) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^t R(s_t, a_t)|s_0 = s, a_0 = a\right]. \tag{5.40}$$

Both the actor and critic networks are updated iteratively. The critic with stochastic gradient descent and the actor to maximize $J(\pi) + \lambda A(s, a)$. The critic aims to minimize the TD error for better estimates of value-function. This leads the policy towards better advantage function estimate and better convergence. The next state value function is denoted as $V(s')$ and the TD error is given as $\delta = R(s, a) + \gamma V(s') - V(s)$.

The CU uses the PS1 algorithm to optimize resource allocation for each VLC AP, aiming efficient serving and maximize data rates. The actor neural network selects optimal actions for power, bandwidth, and association parameters based on the system's current state vector $\mathcal{S}$.The critic network updates value function estimates based on these actions, refining future decisions. Guided by a multi-objective reward function R, this iterative process ensures effective resource utilization and data rate maximization, while ensuring the hybrid RF/VLC system approaches an optimal state, maximizing the expected return $J(\pi)$ over time.

5.4.3 PPO based Proposed Scheme 2 (PS2)

The PS1 algorithm can cause drastic policy updates that may hinder learning. To avoid this, we propose PS2, known for its efficiency and stability in complex environments.

Like PS1, PS2 uses (5.40) to calculate the action-value function. However, PS2 improves stability by using a clipping approach to constrain policy updates, ensuring consistency with the old policy. This method creates a surrogate objective function and takes small

Algorithm 4 Implementation of PS2

Initialize Actor network with weights θ and old weights $\theta_{\text{old}} = \theta$.
Initialize Critic network with weights ϕ.
Initialize Controlling Unit (CU).
Initialize discount factor γ.
Initialize learning rates $\alpha_\theta, \alpha_\phi$.
Initialize clipping parameter ϵ.
for each episode **do**
    Initialize system state $s = \{s_1, s_2, \ldots, s_u\}$ from CU
    Initialize episode reward $R = 0$
    for each step in episode **do**
        Choose an action $a = [a_1, a_2, \ldots, a_n]$ based on $\pi_\theta(a|s)$
        Take action a, update system state to s' and receive reward R
        Update s' in CU
        Compute Advantage $A = r + \gamma V_\phi(s') - V_\phi(s)$
        Calculate policy ratio $r_t(\theta) = \frac{\pi_\theta(a|s)}{\pi_{\theta_{\text{old}}}(a|s)}$
        Update θ using clipped objective:

$$\theta \leftarrow \theta + \alpha_\theta \mathbb{E}_t \left[\min\left(r_t(\theta)A, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)A\right)\right]$$

        Update ϕ to minimize MSE $= (r + \gamma V_\phi(s') - V_\phi(s))^2$:

$$\phi \leftarrow \phi - \alpha_\phi \nabla_\phi \text{MSE}$$

        $s = s'$
        Aggregate rewards R
    end for
    $\theta_{\text{old}} = \theta$
    Store R for performance evaluation.
    if convergence criteria met **then**
        Break
    end if
end for

steps for optimal convergence. The optimization of the PS2 objective function is given as

$$\mathcal{L}^{CLIP}(\theta) = \mathbb{E}_t \left[\min \left(R_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t \right) \right], \tag{5.41}$$

where $r_t(\theta)$ represents the ratio of the probability of an action under the new and old policies. $\hat{A}_t$ is the advantage function estimator at time t and ϵ is the hyperparameter limiting policy change in a single update.

In the present system, the CU employs PS2 to determine optimal actions based on the state of the VLC APs. Each action, similar to the PS1 approach, signifies specific configurations like power, bandwidth, and association parameter for every VLC AP.

5.4.4 Network Time and Training Complexity Discussion

In this section, we perform an analysis of the complexity of the proposed schemes.

5.4.4.1 Network Time Complexity

To assess the complexity of value or policy functions, we categorize them as either critic-only (DQN) or actor–critic (DDPG, PS1, PS2). Let k represent the feature size of the traffic state and a the action space size.

Each network consists of $L+2$ hidden layers with m neurons. An empirical study [143] suggests optimal neuron counts of 128 or 256 with hidden layer depths of two or three. The time complexity for DQN is given by $\mathcal{O}(km + Lm^2 + ma) = \mathcal{O}(Lm^2 + (k+a)m)$. Both DQN and policy gradient methods share this complexity. For DDPG, PS1, and PS2, the time complexity of separate critic and actor networks is $\mathcal{O}(Lm^2 + (k+a)m)$. Since m is much larger than a, k, and L, and comparable to N, the time complexity for DQN and DDPG simplifies to $\mathcal{O}(Lm^2)$, while for PS1 and PS2, it is $\mathcal{O}(2Lm^2)$.

5.4.4.2 Training Complexity

DRL algorithms use on-policy (PS1, PS2) or off-policy (DQN, DDPG) strategies. Let d_1 and d_2 be update steps for off-policy and on-policy methods, respectively, with b epochs

Table 5.3: Complexity for different DRL algorithms

Algorithm	Network time Complexity	Training Complexity (forward)	Training Complexity (backward)
DQN	$O(Lm^2)$	$O(d+2bf_1)$	$O(bf_1)$
DDPG	$O(Lm^2)$	$O(d+2bf_1)$	$O(2bf_1)$
PS1	$O(2Lm^2)$	$O(d+bf_2)$	$O(2bf_2)$
PS2	$O(2Lm^2)$	$O(d+2bf_2)$	$O(4bf_2)$

per model update. The simulation steps per episode are $d = \frac{T}{\Delta t}$, with update frequencies $f_1 = \frac{T}{d_1 \Delta t}$ (off-policy) and $f_2 = \frac{T}{d_2 \Delta t}$ (on-policy). The replay buffer capacity is C.

Typically, $d_1 < d_2$ as on-policy methods require longer sampling. For on-policy, buffer capacity matches d_2, while off-policy methods sample until convergence.

Training involves forward and backward passes. For DQN and DDPG, the forward pass is $O(d + 2bf_1)$, with backward complexity of $O(bf_1)$ and $O(2bf_1)$. For PS1, forward and backward passes are $O(d + 2bf_2)$ and $O(2bf_2)$, respectively. PS2 has a backward complexity of $O(4bf_2)$. Generally, $d \gg b$, f_2 and is comparable to $10bf_1$.

5.4.5 Floating Point Operations (FLOPs)

For a DRL agent, the time complexity for one forward and one backward pass can be describe in metric of FLOPs [144]. The number of FLOPs for forward pass N_{for} is given as

$$N_{\text{for}} \approx (\alpha_{\text{m}} + 1) \sum_{l=1}^{L} n^l n^{l-1}. \tag{5.42}$$

The number of FLOPs for backward pass N_{back} is given as

$$N_{\text{back}} \approx n_{\text{trn}} (\alpha_{\text{m}} + 1) \sum_{l=1}^{L} n^l n^{l-1}, \tag{5.43}$$

where α_m denotes the number of FLOPs of one multiplication operation, n^l and n^{l-1} are the input and output neurons in lth layer. L is number of layers with training samples n_{trn} and number of epochs n_e.

5.4.6 Dynamic Bandwidth Allocation

The number of UEs associated with each VLC AP will change dynamically. Note that quasi-static period is the time span for which a specific number of UEs are associated with an AP. As the UEs move around dynamically, this number of UEs likely to change. Thus, the change in this number depends on the motion of the UEs. The user speed produces a Doppler shift. This Doppler shift gives us a coherence time T_c. T_c is the time span for which a specific number of UEs are associated to an AP. Thus, considering the dynamics of the system under consideration, T_c comes out nearly equal to 373 ms for sitting and 134 ms for walking [135].

The fundamental requirement for our proposed schemes to be successful is that T_c must be more than the execution time for one iteration of the neural network, T_{nn}. When $T_c > T_{nn}$, the execution of the DRL algorithm will be completed within the time span in which a specific number of UEs is associated with an AP. Thus, the bandwidth allocation will be done within T_c time span. We calculate T_{nn} with FLOPs. When we put the values in above expression (5.42) and (5.43), when L=4, α_m=1, epochs or episodes = 5000, $n_{trn} = 100$. Some of the values are taken from [144], we get T_{nn} around 57 ms which is much less than T_c. Hence, during the quasi-static period, the channel state information (CSI) is constant [145]. In indoor environments, both VLC and RF channels maintain constant CSI for a short duration called T_c. So atleast for that time the UE will be associated to an AP.

5.5 Performance Evaluation

The experimental simulation set-up discussed in the system model has been created using python language in the gymnasium environment. For practicality, the simulation environment is assumed to be dynamic with UEs randomly orientated. PS1 and PS2 have been compared with the DDPG and DQN based algorithms as benchmarks. Notably, DDPG and DQN are state-of-the-art techniques, that have been used in the latest available works aligning with our objective in this work [30, 17, 37, 33, 36]. Additionally, the proposed algorithms are also compared with optimization algorithms such as epsilon-greedy, random

and ES.

The details of simulation parameters, some of which are taken from [50], have been mentioned in Table 5.3 and the set-up is shown in Fig. 5.3.

5.5.1 Baseline Strategies

The DDPG (DRL based approach) algorithm, DQN algorithms, epsilon-greedy and ES are explained as follows:

5.5.1.1 DDPG and DQN based algorithms

DQN learning is suitable for discrete action space. However, PS1 and PS2 can efficiently handle continuous state and action space. Thus, we use continuous action space DRL technique DDPG as a benchmark. Both DDPG and DQN uses experience replay and hence are more sample efficient. However, due to this they can diverge from current policy and are less stable. DQN may handles complex scenarios but it depends on quantization to support discrete action spaces. Quantization hinders its ability to find an optimal policy. DDPG operates in a continuous action space, effectively avoiding the limitations of quantization noise and leading to an optimal policy. DDPG combines Q-learning and policy gradient methods. However, DDPG is highly susceptible to hyperparameters tuning. Similar to PS1 and PS2 architecture, it also has actor-critic approach and supports continuous state and action spaces. DQN and DDPG uses explicit exploration and less exploitation as both of them uses experience replay. PS1 and PS2 strategies have much more exploration than exploitation, and in PS2 the clipping ensures both stable exploitation and controlled exploration.

5.5.1.2 Exhaustive Search

ES, or brute force search, is a simple algorithm that finds solutions by checking all possible options within given constraints. It guarantees the best solution. Thus, it has been considered as a benchmark. However, it may be impractical for large action spaces due to its exponential time complexity.

Table 5.4: Simulation Parameters

Hybrid RF/VLC environment parameter	**Values**
Number of RF APs	1
Number of VLC APs	12
Number of UEs	10
Area of Photodiode	1 cm^2
VLC AP Average Optical Power	9.2 W
RF Transmit Power	100.0 mW
VLC Transmit Power	10.0 mW (for each VLC AP)
Noise at VLC AP	10^{-21} A^2/Hz
Responsivity	0.28 A/W
FOV at UE	$60°$
Order of Lambertian	1.2
Users Speed	0 to 1.5 m/s randomly
Area Size	20×20 m^2
Frequency Range	2400 to 2500 MHz
Semi-Angle of VLC APs	$70°$
Hyper-parameters	**Values**
Number of epochs	5000
Learning rate	0.01, 0.03, 0.0003
Discount factor	0.99
Number of steps	200
Mini batch size	64
Clipping range for PS2	0.2

5.5.1.3 Epsilon - greedy

Epsilon - greedy is a simple reinforcement learning method. It balances the exploitation and exploration trade off between the random action taken with probability ϵ, and the action with the highest known reward (greedy choice) with probability $1 - \epsilon$. However, in large action spaces it might yield suboptimal results and converge slowly due to constant ϵ value and randomness that may not adapt well as it spends a significant amount of time on random exploration. Note that in results epsilon-greedy is written as Epsilon-G.

5.5.2 Hyperparameters

The hyperparameters for training include epochs, learning rate, discount factor, steps, mini-batch size, and PS2 clipping rate. An epoch is a full pass through the dataset updating model parameters. The learning rate determines the step size towards minimizing the loss function, with higher rates enabling faster updates but risking overshooting, and

lower rates offering smoother but slower learning. The discount factor balances immediate and future rewards. Mini-batch size is the number of samples processed before updating parameters. The PS2 clipping range limits policy changes to stabilize training by clipping probability ratios within 1 ± 0.2.

5.5.3 Results and Discussion

Fig. 5.4 shows the convergence of the proposed algorithm with a learning rate of 0.03. It can be observed that the high learning rate causes significant policy fluctuations, highlighting DDPG's limited learning effectiveness. In contrast, the proposed schemes outperforms DDPG, with the PS2 achieving about 27% better sum-rate than the PS1. They are also compared with DQN, ES, epsilon-greedy and random. DQN is inconsistent and stabilizes at a lower sum-rate with minimal improvement. ES is highly variable and unreliable for larger, dynamic scenarios. Random selection of actions show consistent low data rates, indicating no learning. Epsilon-greedy also stabilize at lower data rates. Note that DQN and ES curve has been smoothed for better visualization.

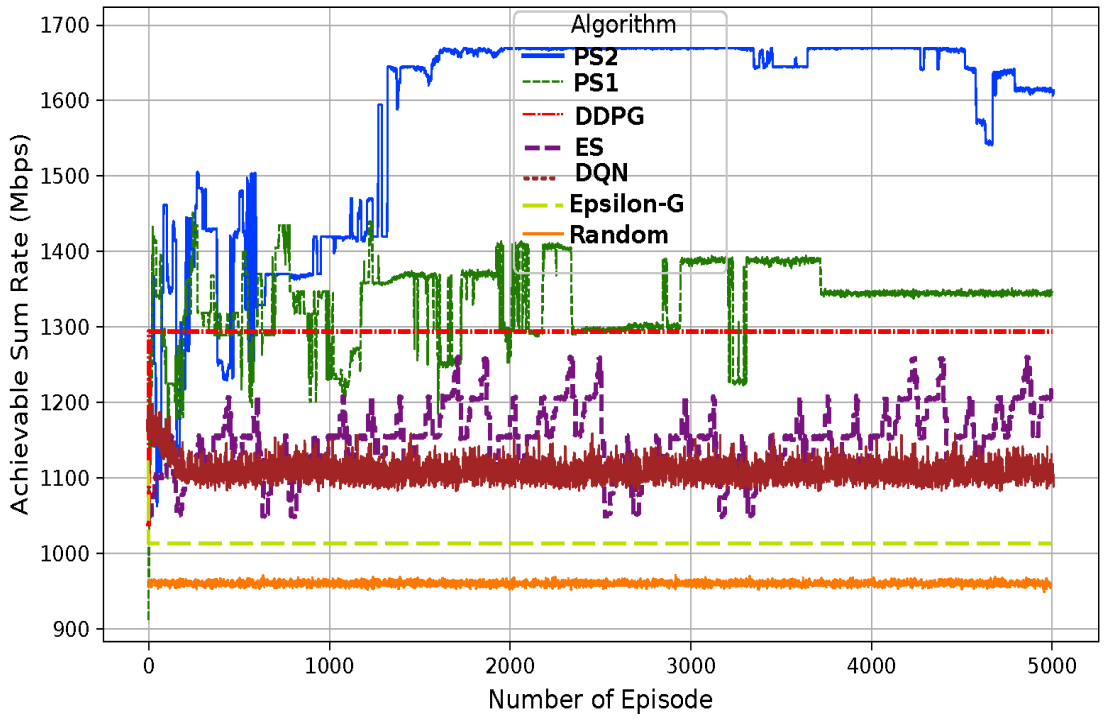


Figure 5.4: Convergence of DRL algorithms with learning rate 0.03

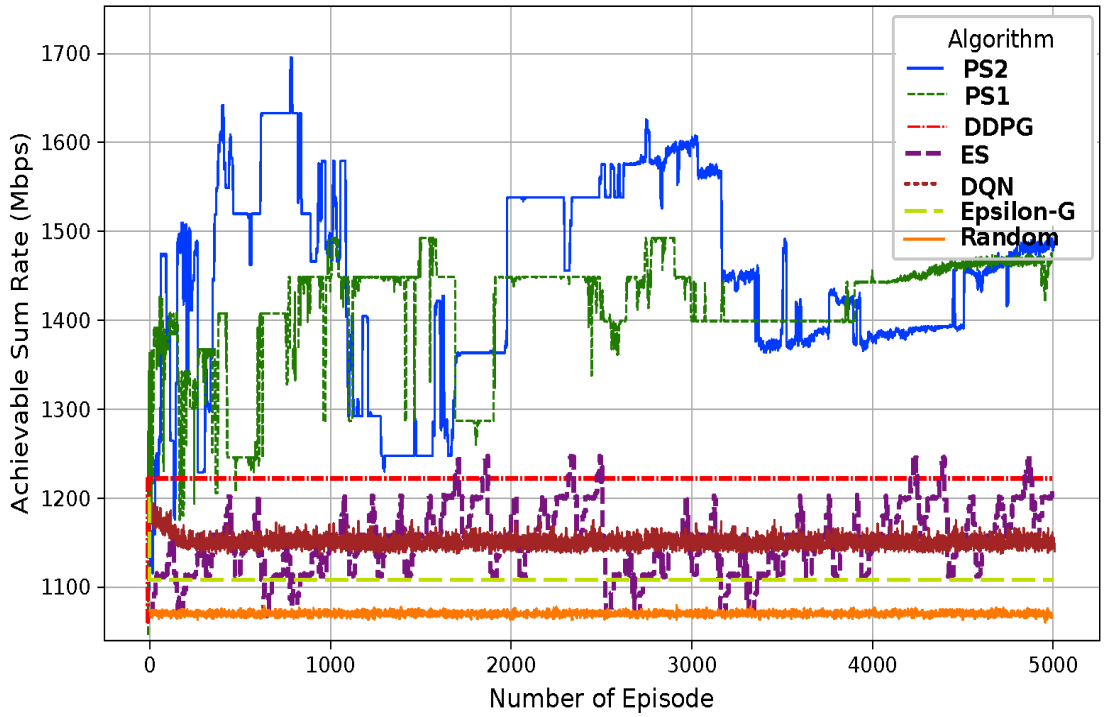


Figure 5.5: Convergence of DRL algorithms with learning rate 0.01.

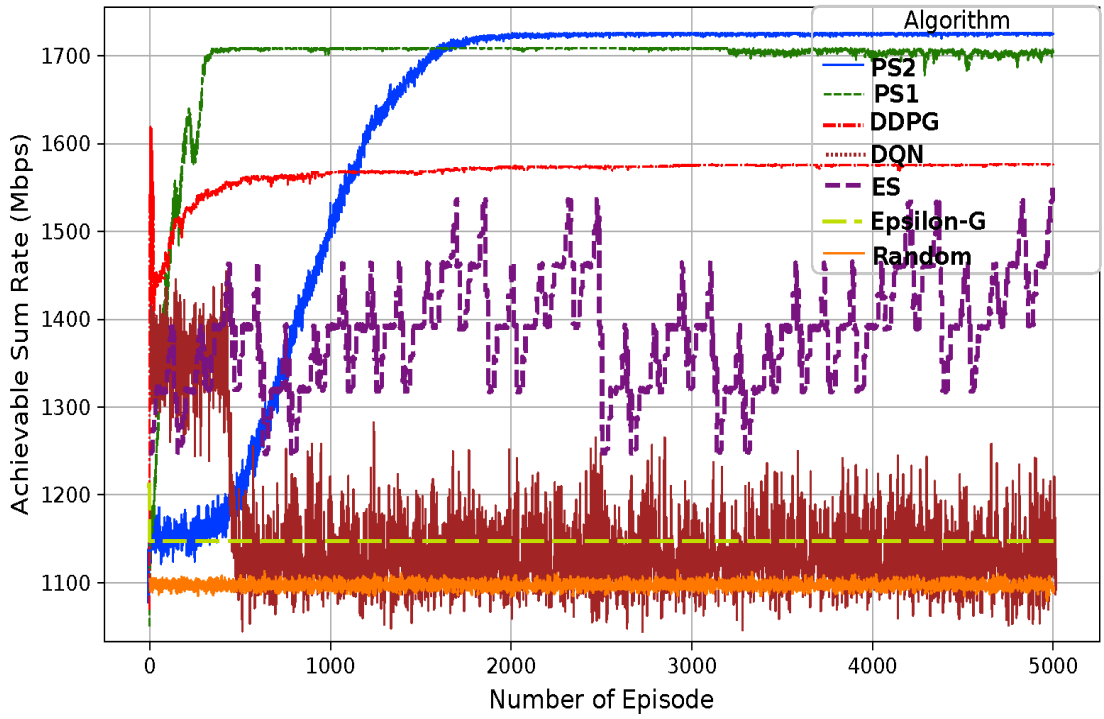


Figure 5.6: Convergence of DRL algorithms with learning rate 0.0003

We then reduced the learning rate to 0.01 to assess policy learning. Fig. 5.5 shows the convergence of DRL algorithms at this rate. DDPG performs poorly, while PS2 outperforms PS1 by about 20%. DQN's performance declines with more episodes, ES is highly variable, random and epsilon-greedy shows constant data rates.

Fig. 5.6 and 5.7 demonstrates the convergence of the proposed algorithm's achievable sum-rate and mean power consumption as the learning rate decreases to 0.0003. Lowering the learning rate facilitates smoother policy convergence, with PS1 exhibiting faster learning compared to PS2 due to the latter's clipping objective function.

In Fig. 5.6, PS1 and PS2 demonstrate the highest and most stable performance, with PS2 outperforming PS1. DDPG performs moderately. DQN is inconsistent and stabilizes at a lower sum-rate with minimal improvement. ES showing higher variability and unreliability for larger, dynamic scenarios. Thus, PS1 and PS2 are the most effective for maximizing sum-rate, while DQN and ES remains impractical due to high variability. Random and epsilon-greedy shows stagnancy and hence impractical for use.

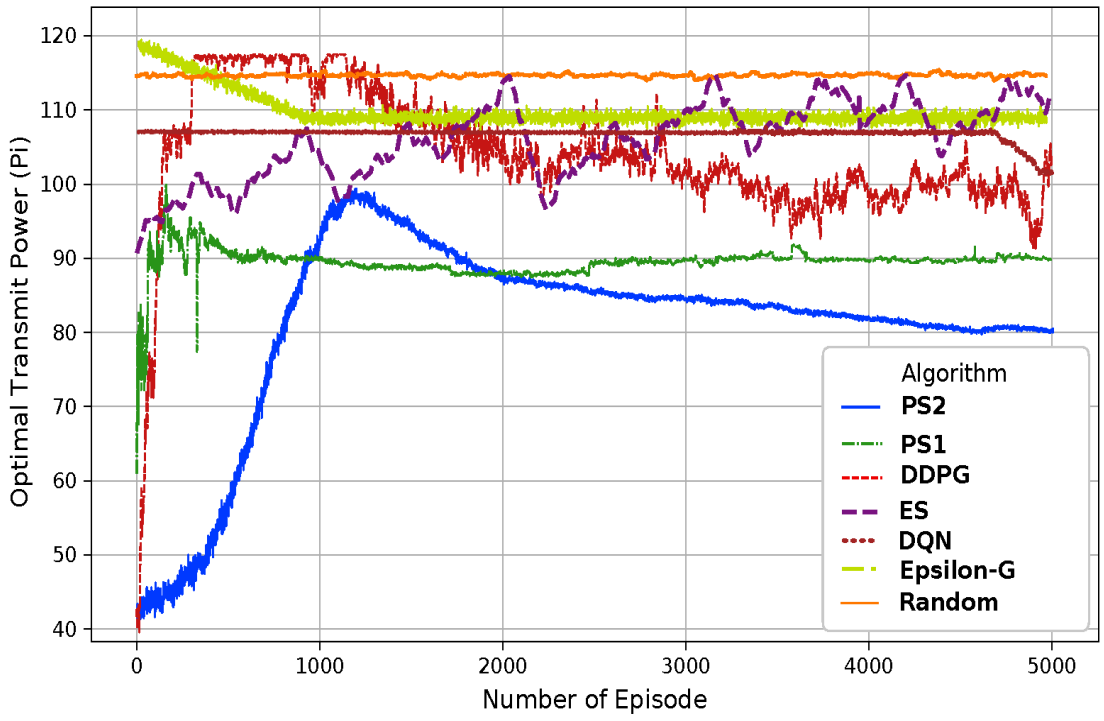


Figure 5.7: Comparison of optimal transmit power utilization.

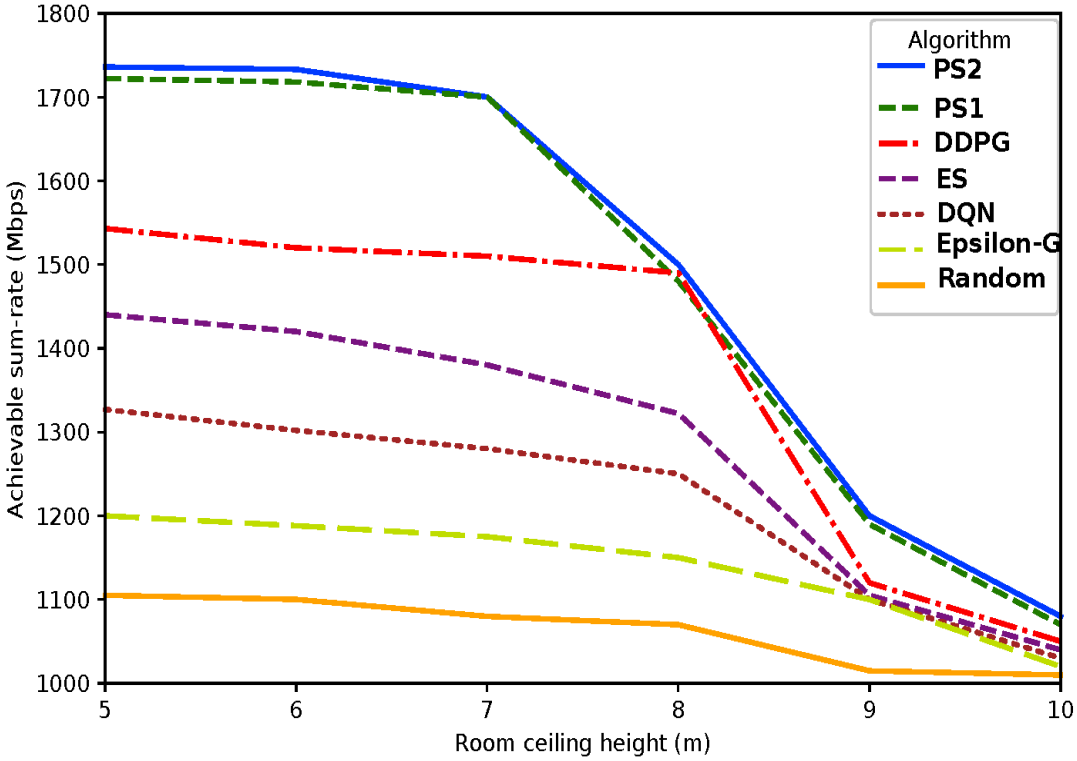


Figure 5.8: Achievable sum-rate vs. room ceiling height

Fig. 5.7 compares the optimal transmit power of PS1, PS2, DDPG, DQN, ES, random, and epsilon-greedy algorithms. DDPG shows high variability, while PS1 and PS2 are more stable. PS1 stabilizes quickly around 90mW, and PS2 reaches 95mW, maintaining stability around 82mW. Both PS1 and PS2 achieve stable performance with lower optimal transmit power, about 30% and 32% better than DDPG, respectively. For applications requiring low transmission power and stability, PS1 and PS2 outperform DDPG. DQN exhibits higher data rated and ES exhibits higher variance, hence less consistent performance. Random selection of actions show constant transmit power, indicating no learning. Epsilon-greedy methods show improvement over time but stabilize at higher power consumption. Thus, random and epsilon-greedy are less effective, with the former showing no learning and the latter consuming more power.

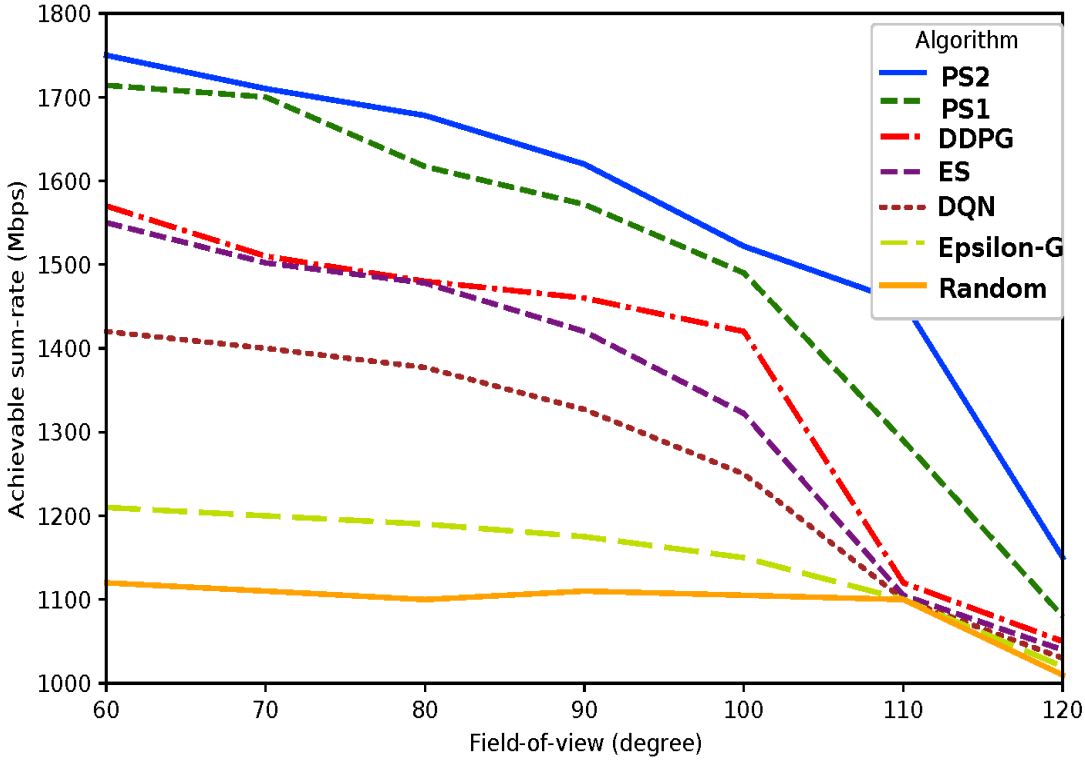


Figure 5.9: Achievable sum-rate vs. FOV of receiver.

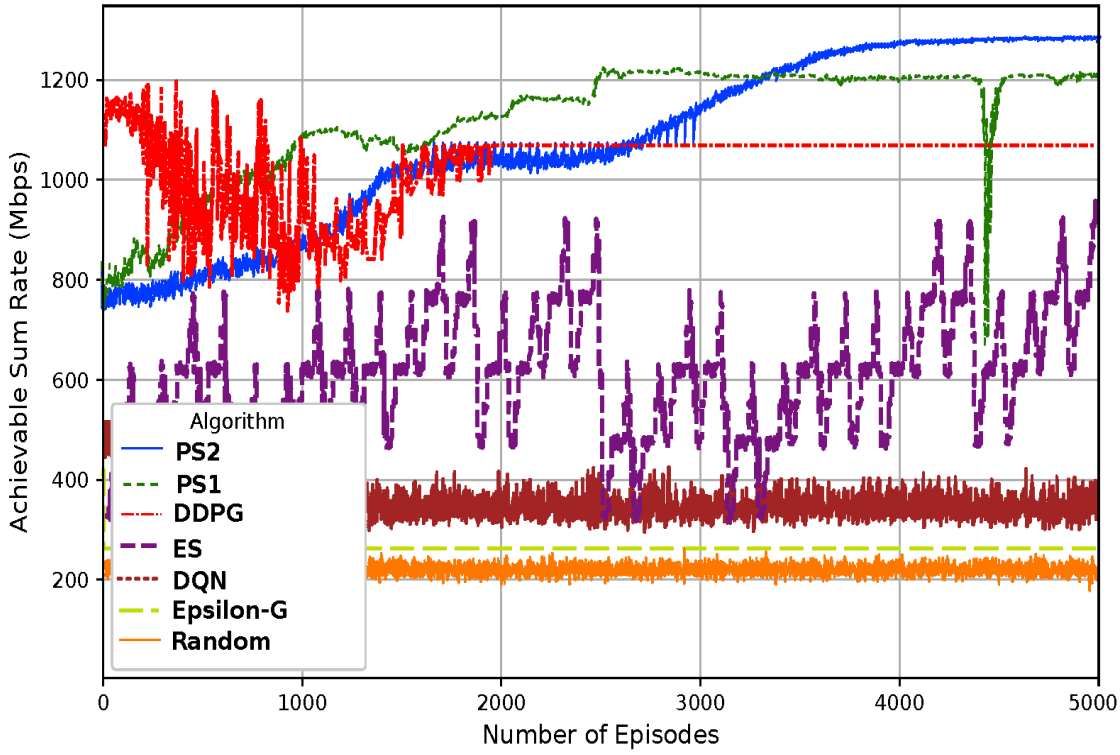


Figure 5.10: Convergence of DRL algorithms for scalability

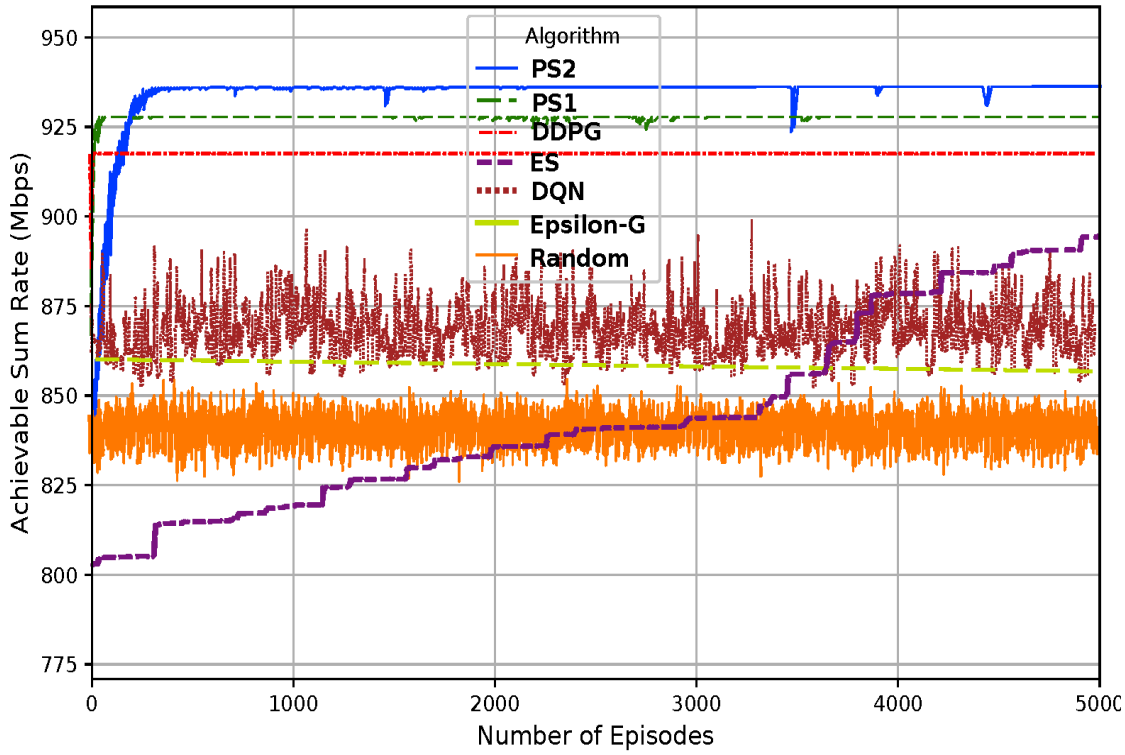


Figure 5.11: Convergence of DRL algorithms for ith AP.

Fig. 5.8 depicts the impact of ceiling height variation on achievable sum-rate, showing a decrease as ceiling height increases due to larger receiver FOV resulting in increased interference. Fig. 5.9 illustrates the decrease in achievable sum-rate with increased FOV. Simulation results shows that DDPG, DQN, ES, Random and epsilon greedy are struggling to learn while PS1 and PS2 perform well with large continuous action spaces. Also, higher learning rates enable faster learning but reduce convergence stability.

Fig. 5.10 illustrates the performance in a 50m $\times$ 50m set-up to show the scalability of proposed schemes. PS1 and PS2 achieves the highest sum-rate consistently. Both of them outperform the baseline algorithms. DDPG struggles initially but shows stable convergence later. ES and DQN shows high variability in performance throughout, with large oscillations. Both epsilon-greedy and random performing poor and not learning effectively.

Fig. 5.11 shows performance for ith AP. PS1 and PS2 perform consistently well, achieving high sum-rates relatively early and maintaining them. DDPG also shows stable per-

formance. ES achieves a stable, high data rate but it is computationally intensive. DQN fluctuates due to discretization and shows negligible improvement. Epsilon-greedy and random fluctuates more and maintains a low, relatively constant sum-rate, as it does not learn from the environment.

5.5.4 Discussion on Optimality and Performance Order

Both PS1 and PS2 are on-policy DRL techniques, while DDPG and DQN are off-policy. On-policy algorithms learn the value of the executed policy, including exploration, whereas off-policy algorithms learn from experiences in a replay buffer. Given the dynamic environment, on-policy methods are favored for their stability and adaptability. They utilize the advantage function to reduce variance and enhance learning precision. PS2 effectively minimizes large policy changes with its clipped surrogate objective, making it suitable for environments with large action spaces, like our hybrid RF/VLC system. Additionally, PS2 is less sensitive to hyperparameter tuning, likely explaining the performance order of PS2, PS1, and DDPG, as confirmed by our simulation results.

5.6 Conclusion

This study presents an on-policy DRL solution for the non-concave joint optimization of dynamic resource allocation in hybrid RF/VLC networks, including load balancing. The continuous action space involves association parameters, bandwidth, and transmission power. We applied PS1 and PS2, efficient model-free on-policy DRL algorithms, to handle continuous action spaces. These algorithms demonstrated superior performance in optimizing resource allocation within the dynamic environment. Simulation results show that PS1 and PS2 achieve faster convergence and better data rates, improving up to 8.1% and 9.7% compared to existing DDPG-based algorithms.

Chapter 6

Conclusion and Future scope

This chapter concludes the research work presented in the thesis, highlighting the key discoveries and new contributions. Several theories and algorithms were explored and proposed during this study, which may expand the scope of future investigations. Additionally, we explored opportunities for further research based on the findings of this work.

In recent decades, interest in exploring machine learning and deep learning techniques in resource allocation of hybrid RF/VLC systems and the HetNets area has grown widely. This thesis evaluates the application of deep learning and DRL technologies for resource allocation in hybrid RF/VLC systems. The resource allocation consists of parameters like bandwidth, transmission power, and the association parameter. Our goal is to maximize the achievable sum-rate of the hybrid RF/VLC systems. To achieve this, near-realistic channel models, including blockage and random orientation, are proposed. It also explores possible DRL techniques for efficient resource allocation. The contributions of this thesis include creating near-realistic channel models, efficiently allocating resources, achieving sum-rate maximization and exploring latest DRL algorithms. The thesis is organized into chapters, each focusing on one of these contributions.

In Chapter 2, an overview of VLC, hybrid RF/VLC, deep learning and DRL techniques is mentioned. The chapter explains the importance of resource allocation with focusing on the issue of non-concavity in the joint optimization problem. It also presents a detailed

review of machine learning and deep learning frameworks applied in the domain. Finally, the chapter identifies gaps in existing research and suggests new algorithms and areas in hybrid RF/VLC systems.

In Chapter 3, the joint optimization problem of resource allocation in hybrid WiFi/LiFi systems is formed and solved with DQN. The proposed algorithm effectively optimizes bandwidth, association parameters, and transmission power simultaneously. By using DQN, the approach successfully resolves the non-concavity issue in this joint optimization problem. The proposed DQN transfer learning algorithm is particularly beneficial when a new UE enters the system. For the new UE, the algorithm utilizes existing data from the nearest UE to improve efficiency. Simulations show that the proposed algorithm achieves a higher sum-rate while requiring fewer iterations to converge. Also, when new UEs are introduced, the algorithm reaches the maximum possible sum-rate with reduction in the number of iterations.

In Chapter 4, we proposed a DRL algorithm that works in a continuous action space. The proposed DDPG algorithm works efficiently with continuous action space for the joint optimization problem. In addition to resource allocation, we also considered load balancing in hybrid RF/LiFi systems. Remarkably, we tested this strategy in an environment that closely resembled a real-world scenario, which was created in a gymnasium using Python. We compared our proposed DDPG algorithm with several well-established DRL algorithms that are designed to handle both discrete and continuous action and state spaces. The simulation results revealed that the proposed DDPG algorithm outperformed the baseline strategies. DDPG not only efficiently handles continuous action spaces, but also shows significantly better performance.

In Chapter 5, the study aims to provide an on-policy DRL based solution for optimizing dynamic resource allocation in hybrid RF/VLC networks with consideration of load balancing. The study also considers the random orientation of UEs in the set-up. The joint optimization considers variables such as association parameters, bandwidth, and transmission power, all represented in a continuous action space. The continuous action space

provides more accuracy while considering the random orientation of UEs. To handle this, two efficient model-free on-policy DRL algorithms, A2C and PPO, were applied. These algorithms improved resource distribution in the dynamic environment and achieved better performance. Simulation results show that A2C and PPO offer faster convergence and higher data rates than the baseline strategies.

6.1 Future Scope

- *Multi-tier expansion:* In this thesis, the proposed work on hybrid RF/VLC is an example of two-tier HetNets. However, some future generation networks can have three tier or multi-tier hybrid RF/VLC networks. Thus, an extension of the current study in three or multi-tier HetNets could be a promising future direction for the current research. Since HetNets consist of various types of cells, including macro, micro, pico, and femtocells, which work together to enhance network coverage and capacity. In 5G, HetNets leverage multiple radio access technologies and frequency bands to meet the increasing demand for data rates and connectivity in large venues.

- *Real Time Implementation:* The proposed mechanism can be implemented in real time test bed involving a combination of useful hardware and software. The three main components for test bed implementation are transmitter, receiver, and signal processing. The VLC component will use LED arrays as transmitters, positioned strategically on the ceiling to provide coverage across the room. The RF AP, such as a Wi-Fi router, can be installed to provide overlapping wireless coverage. Deep learning algorithms proposed can be implemented on a server or a high-performance computing unit that connects to both the VLC and RF networks. The deep learning model receives real-time data on user positions, orientations, CSI, and QoS requirements, gathered through a combination of sensors, such as PDs for VLC. Use of universal software radio peripheral (USRP) software defined radios (SDR) is cost efficient and versatile to handle signal processing and connectivity. The USRP interfaces with a host computer where DRL algorithms are implemented. The exe-

cution of the DRL algorithms will allow the USRP to adapt its transmission power and bandwidth dynamically, enabling efficient data distribution between the RF and VLC components.

- *Aerial expansion with multiple RF APs:* The aerial expansion of the system involves using multiple RF APs. This thesis work can be extended by increasing the number of RF and VLC APs in larger areas. Adding more APs will help create a more reliable and robust system setup.
- *Theft Protection and Secrecy:* The work in this thesis does not consider secrecy capacity. Theft protection and secrecy refer to measures designed to safeguard data from unauthorized access or theft. Future work may include secrecy capacity, which ensures that sensitive information remains secure. By using this approach, the system prevents illegitimate users from accessing or stealing the data.
- *Emergency Situations:* The current work, particularly in chapter 5, assumes normal conditions in public places. Future extensions may consider disaster conditions, such as when fire alarms are raised and users start rushing to the emergency exit doors. In such indoor conditions, the user speeds may exceed 1.5 m/s. The number of users in a particular area may change abruptly. Such a study can be significantly helpful for aiding communication solutions to disaster management units.

References

[1] FCC Staff Technical Paper. “Mobile Broadband: The Benefits of Additional Spectrum”. In: *Federal Communications Commission (FCC)* Washington, DC (October 2010).

[2] Tezcan Cogalan and Harald Haas. “Why would 5G need optical wireless communications?” In: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE. 2017, pp. 1–6.

[3] U Cisco. “Cisco annual internet report (2018–2023) white paper”. In: *Cisco: San Jose, CA, USA* 10.1 (2020), pp. 1–35.

[4] Ericsson. “Ericsson Mobility Report”. In: November (2024), pp. 1–40.

[5] R.W. Burns and Institution of Electrical Engineers. *Communications: An International History of the Formative Years*. Communications: An International History of the Formative Years. Institution of Engineering and Technology, 2004. ISBN: 9780863413278. URL: `https://books.google.co.in/books?id=7eUUy8-VvwoC`.

[6] Alexander Graham Bell. “Upon the production and reproduction of sound by light”. In: *Journal of the society of Telegraph Engineers* 9.34 (1880), pp. 404–426.

[7] Toshihiko Komine and Masao Nakagawa. “Fundamental analysis for visible-light communication system using LED lights”. In: *IEEE transactions on Consumer Electronics* 50.1 (2004), pp. 100–107.

[8] P. H. Pathak et al. “Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges”. In: *IEEE Communications Surveys Tutorials* 17.4 (2015), pp. 2047–2077.

[9] Sridhar Rajagopal, Richard D Roberts, and Sang-Kyu Lim. “IEEE 802.15. 7 visible light communication: modulation schemes and dimming support”. In: *IEEE Communications Magazine* 50.3 (2012), pp. 72–82.

[10] Eldad Perahia and Robert Stacey. *Next generation wireless LANs: 802.11 n and 802.11 ac*. Cambridge university press, 2013.

[11] Harald Haas et al. “Introduction to indoor networking concepts and challenges in LiFi”. In: *Journal of Optical Communications and Networking* 12.2 (2020), A190–A203.

[12] Aleksandar Damnjanovic et al. “A survey on 3GPP heterogeneous networks”. In: *IEEE Wireless communications* 18.3 (2011), pp. 10–21.

[13] Ahmed R Elsherif et al. “Resource allocation and inter-cell interference management for dual-access small cells”. In: *IEEE Journal on Selected Areas in Communications* 33.6 (2015), pp. 1082–1096.

[14] Siavash Bayat et al. “Distributed user association and femtocell allocation in heterogeneous wireless networks”. In: *IEEE Transactions on Communications* 62.8 (2014), pp. 3027–3043.

[15] Hisham Abuella et al. “Hybrid RF/VLC systems: A comprehensive survey on network topologies, performance analyses, applications, and future directions”. In: *IEEE Access* 9 (2021), pp. 160402–160436.

[16] Dushyantha A Basnayaka and Harald Haas. “Hybrid RF and VLC systems: Improving user data rate performance of VLC systems”. In: *2015 IEEE 81st vehicular technology conference (VTC Spring)*. IEEE. 2015, pp. 1–5.

[17] Shivanshu Shrivastava et al. “Deep Q-network learning based downlink resource allocation for hybrid RF/VLC systems”. In: *IEEE Access* 8 (2020), pp. 149412–149434.

[18] Sylvester Aboagye et al. “Joint access point assignment and power allocation in multi-tier hybrid RF/VLC HetNets”. In: *IEEE Transactions on Wireless Communications* 20.10 (2021), pp. 6329–6342.

[19] Shayan Zargari et al. “Resource allocation of hybrid VLC/RF systems with light energy harvesting”. In: *IEEE Transactions on Green Communications and Networking* 6.1 (2021), pp. 600–612.

[20] Vasilis K Papanikolaou et al. “On optimal resource allocation for hybrid VLC/RF networks with common backhaul”. In: *IEEE Transactions on Cognitive Communications and Networking* 6.1 (2020), pp. 352–365.

[21] Rong Zhang et al. “Anticipatory association for indoor visible light communications: Light, follow me!” In: *IEEE Transactions on Wireless Communications* 17.4 (2018), pp. 2499–2510.

[22] Konstantinos G. Rallis et al. “Energy Efficient Cooperative Communications in Aggregated VLC/RF Networks With NOMA”. In: *IEEE Transactions on Communications* 71.9 (2023), pp. 5408–5419. DOI: 10.1109/TCOMM.2023.3292486.

[23] Sylvester Aboagye et al. “Energy-Efficient Resource Allocation for Aggregated RF/VLC Systems”. In: *IEEE Transactions on Wireless Communications* 22.10 (2023), pp. 6624–6640. DOI: 10.1109/TWC.2023.3244871.

[24] Shiyuan Sun et al. “Joint resource management for intelligent reflecting surface–aided visible light communications”. In: *IEEE Transactions on Wireless Communications* 21.8 (2022), pp. 6508–6522.

[25] Sylvester Aboagye et al. “VLC in Future Heterogeneous Networks: Energy– and Spectral–Efficiency Optimization”. In: *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*. 2020, pp. 1–7. DOI: 10.1109/ICC40277.2020.9148909.

[26] Haotong Cao et al. “Softwarized Resource Allocation in Digital Twins-Empowered Networks for Future Quantum-Enabled Consumer Applications”. In: *IEEE Transactions on Consumer Electronics* 70.1 (2024), pp. 800–810. DOI: 10.1109/TCE.2024.3370052.

[27] Rui Jiang et al. “Joint user association and power allocation for cell-free visible light communication networks”. In: *IEEE Journal on Selected Areas in Communications* 36.1 (2017), pp. 136–148.

[28] Hongji Huang et al. “Deep Learning for Physical-Layer 5G Wireless Techniques: Opportunities, Challenges and Solutions”. In: *IEEE Wireless Communications* 27.1 (2020), pp. 214–222. DOI: 10.1109/MWC.2019.1900027.

[29] Liqiang Wang et al. “Deep reinforcement learning-based adaptive handover mechanism for VLC in a hybrid 6G network architecture”. In: *IEEE Access* 9 (2021), pp. 87241–87250.

[30] Bekir Sait Ciftler, Abdulmalik Alwarafy, and Mohamed Abdallah. “Distributed DRL-based downlink power allocation for hybrid RF/VLC networks”. In: *IEEE Photonics Journal* 14.3 (2021), pp. 1–10.

[31] Tanya Verma et al. “Transfer learning for resource allotment in dynamic hybrid WiFi/LiFi communication systems”. In: *Optics Communications* 546 (2023), p. 129761.

[32] Yifei Wei et al. “User scheduling and resource allocation in HetNets with hybrid energy supply: An actor-critic reinforcement learning approach”. In: *IEEE Transactions on Wireless Communications* 17.1 (2017), pp. 680–692.

[33] Mohamad Azizi et al. “Efficient AoI-Aware Resource Management in VLC-V2X Networks via Multi-Agent RL Mechanism”. In: *IEEE Transactions on Vehicular Technology* 73.9 (2024), pp. 14009–14014. DOI: 10.1109/TVT.2024.3392738.

[34] Helin Yang et al. “Learning-based energy-efficient resource management by heterogeneous RF/VLC for ultra-reliable low-latency industrial IoT networks”. In: *IEEE Transactions on Industrial Informatics* 16.8 (2019), pp. 5565–5576.

[35] Parvez Shaik et al. “Hybrid RF-VLC Communication System: Performance Analysis and Deep Learning Detection in the Presence of Blockages”. In: *IEEE Transactions on Vehicular Technology* (2024), pp. 1–15. DOI: 10.1109/TVT.2024.3425729.

[36] Duc M. T. Hoang et al. “Joint Design of Adaptive Modulation and Precoding for Physical Layer Security in Visible Light Communications Using Reinforcement Learning”. In: *IEEE Access* 12 (2024), pp. 82318–82332. DOI: 10.1109/ACCESS.2024.3412055.

[37] Danya A. Saifaldeen et al. “DRL-Based IRS-Assisted Secure Hybrid Visible Light and mmWave Communications”. In: *IEEE Open Journal of the Communications Society* 5 (2024), pp. 3007–3020. DOI: 10.1109/OJCOMS.2024.3395425.

[38] Alin-Mihai Căilean and Mihai Dimian. “Current challenges for visible light communications usage in vehicle applications: A survey”. In: *IEEE Communications Surveys & Tutorials* 19.4 (2017), pp. 2681–2703.

[39] Rong Zhang et al. “Visible light communications in heterogeneous networks: Paving the way for user-centric design”. In: *IEEE wireless communications* 22.2 (2015), pp. 8–16.

[40] Sridhar Rajagopal, Richard D Roberts, and Sang-Kyu Lim. “IEEE 802.15. 7 visible light communication: modulation schemes and dimming support”. In: *IEEE Communications Magazine* 50.3 (2012), pp. 72–82.

[41] Samuel M Berman et al. “Human electroretinogram responses to video displays, fluorescent lighting, and other high frequency sources”. In: *Optometry and vision science* 68.8 (1991), pp. 645–662.

[42] Ram Sharma et al. “Optimal LED deployment for mobile indoor visible light communication system: Performance analysis”. In: *AEU-International Journal of Electronics and Communications* 83 (2018), pp. 427–432.

[43] Zhiyong Du et al. “Context-aware indoor VLC/RF heterogeneous network selection: Reinforcement learning with knowledge transfer”. In: *IEEE Access* 6 (2018), pp. 33275–33284.

[44] Thilo Fath and Harald Haas. “Performance comparison of MIMO techniques for optical wireless communications in indoor environments”. In: *IEEE Transactions on Communications* 61.2 (2012), pp. 733–742.

[45] Jia Wang et al. “A general channel model for visible light communications in underground mines”. In: *China Communications* 15.9 (2018), pp. 95–105.

[46] Bugra Turan et al. “Vehicular VLC frequency domain channel sounding and characterization”. In: *2018 IEEE Vehicular Networking Conference (VNC)*. IEEE. 2018, pp. 1–8.

[47] Dominic C O’Brien et al. “Home access networks using optical wireless transmission”. In: *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE. 2008, pp. 1–5.

[48] Chia-Lung Tsai and Zhong-Fan Xu. “Line-of-sight visible light communications with InGaN-based resonant cavity LEDs”. In: *IEEE photonics technology letters* 25.18 (2013), pp. 1793–1796.

[49] Yunlu Wang, Dushyantha A Basnayaka, and Harald Haas. “Dynamic load balancing for hybrid Li-Fi and RF indoor networks”. In: *2015 IEEE international conference on communication workshop (ICCW)*. IEEE. 2015, pp. 1422–1427.

[50] Mohamed Kashef et al. “Energy efficient resource allocation for mixed RF/VLC heterogeneous wireless networks”. In: *IEEE Journal on Selected Areas in Communications* 34.4 (2016), pp. 883–893.

[51] Abdallah Khreishah et al. “A hybrid RF-VLC system for energy efficient wireless access”. In: *IEEE Transactions on Green Communications and Networking* 2.4 (2018), pp. 932–944.

[52] Hany Elgala, Raed Mesleh, and Harald Haas. “Indoor optical wireless communication: potential and state-of-the-art”. In: *IEEE Communications magazine* 49.9 (2011), pp. 56–62.

[53] Harald Haas et al. “What is lifi?” In: *Journal of lightwave technology* 34.6 (2015), pp. 1533–1544.

[54] Rose Qingyang Hu and Yi Qian. *Heterogeneous cellular networks*. John Wiley & Sons, 2013.

[55] Irina Stefan and Harald Haas. “Hybrid visible light and radio frequency communication systems”. In: *2014 IEEE 80th vehicular technology conference (VTC2014-Fall)*. IEEE. 2014, pp. 1–5.

[56] Dushyantha A Basnayaka and Harald Haas. “Design and analysis of a hybrid radio frequency and visible light communication system”. In: *IEEE Transactions on Communications* 65.10 (2017), pp. 4334–4347.

[57] Sylvester Aboagye et al. “Energy-efficient resource allocation for aggregated RF/VLC systems”. In: *IEEE Transactions on Wireless Communications* 22.10 (2023), pp. 6624–6640.

[58] Minghua Chen et al. “Markov approximation for combinatorial network optimization”. In: *IEEE transactions on information theory* 59.10 (2013), pp. 6301–6327.

[59] Qiaoyang Ye et al. “User association for load balancing in heterogeneous cellular networks”. In: *IEEE Transactions on Wireless Communications* 12.6 (2013), pp. 2706–2716.

[60] Mohanad Obeed et al. “Joint optimization of power allocation and load balancing for hybrid VLC/RF networks”. In: *Journal of Optical Communications and Networking* 10.5 (2018), pp. 553–562.

[61] Mai Kafafy et al. “Power efficient downlink resource allocation for hybrid RF/VLC wireless networks”. In: *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2017, pp. 1–6.

[62] Shivanshu Shrivastava et al. “Deep Q-network learning based downlink resource allocation for hybrid RF/VLC systems”. In: *IEEE Access* 8 (2020), pp. 149412–149434.

[63] Irina Stefan, Harald Burchardt, and Harald Haas. “Area spectral efficiency performance comparison between VLC and RF femtocell networks”. In: *2013 IEEE international conference on communications (ICC)*. IEEE. 2013, pp. 3825–3829.

[64] Mohamed Kashef, Mohamed Abdallah, and Naofal Al-Dhahir. “Transmit power optimization for a hybrid PLC/VLC/RF communication system”. In: *IEEE Transactions on Green Communications and Networking* 2.1 (2017), pp. 234–245.

[65] Abdallah Khreishah et al. “A hybrid RF-VLC system for energy efficient wireless access”. In: *IEEE Transactions on Green Communications and Networking* 2.4 (2018), pp. 932–944.

[66] Sylvester Aboagye et al. “VLC in future heterogeneous networks: Energy–and spectral–efficiency optimization”. In: *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE. 2020, pp. 1–7.

[67] Jiaxuan Chen, Zhaocheng Wang, and Tianqi Mao. “Resource management for hybrid RF/VLC V2I wireless communication system”. In: *IEEE Communications Letters* 24.4 (2020), pp. 868–871.

[68] Abhishek Gupta et al. “On association and bandwidth allocation for hybrid RF/VLC systems”. In: *2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. IEEE. 2018, pp. 1–6.

[69] Bugra Turan and Sinem Coleri. “Machine learning based channel modeling for vehicular visible light communication”. In: *IEEE Transactions on Vehicular Technology* 70.10 (2021), pp. 9659–9672.

[70] Huy Q Tran and Cheolkeun Ha. “Machine learning in indoor visible light positioning systems: A review”. In: *Neurocomputing* 491 (2022), pp. 117–131.

[71] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[72] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*. Vol. 1. MIT press Cambridge, MA, USA, 2017.

[73] Nguyen Cong Luong et al. “Applications of deep reinforcement learning in communications and networking: A survey”. In: *IEEE communications surveys & tutorials* 21.4 (2019), pp. 3133–3174.

[74] Kai Arulkumaran et al. “A brief survey of deep reinforcement learning”. In: *arXiv preprint arXiv:1708.05866* (2017).

[75] Christopher John Cornish Hellaby Watkins. “Learning from delayed rewards”. In: (1989).

[76] Christopher JCH Watkins and Peter Dayan. “Q-learning”. In: *Machine learning* 8 (1992), pp. 279–292.

[77] Alia Asheralieva and Yoshikazu Miyanaga. “An autonomous learning-based algorithm for joint channel and power level selection by D2D pairs in heterogeneous cellular networks”. In: *IEEE transactions on communications* 64.9 (2016), pp. 3996–4012.

[78] Zhong Li, Cheng Wang, and Chang-Jun Jiang. “User association for load balancing in vehicular networks: An online reinforcement learning approach”. In: *IEEE Transactions on Intelligent Transportation Systems* 18.8 (2017), pp. 2217–2228.

[79] E. Ghadimi et al. “A reinforcement learning approach to power control and rate adaptation in cellular networks”. In: *2017 IEEE International Conference on Communications (ICC)*. 2017, pp. 1–7.

[80] Volodymyr Mnih et al. “Human-level control through deep reinforcement learning”. In: *Nature* 518.7540 (Feb. 2015), pp. 529–533. ISSN: 00280836. URL: `http://dx.doi.org/10.1038/nature14236`.

[81] S. Wang et al. “Deep reinforcement learning for dynamic multichannel access”. In: *International Conference on Computing, Networking and Communications (ICNC)*. 2017, pp. 1–7.

[82] Y. Wei et al. “User Scheduling and Resource Allocation in HetNets With Hybrid Energy Supply: An Actor-Critic Reinforcement Learning Approach”. In: *IEEE Transactions on Wireless Communications* 17.1 (2018), pp. 680–692.

[83] X. Wan et al. “Reinforcement Learning Based Mobile Offloading for Cloud-Based Malware Detection”. In: *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. 2017, pp. 1–6.

[84] Z. Xu et al. “A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs”. In: *2017 IEEE International Conference on Communications (ICC)*. 2017, pp. 1–6.

[85] Ursula Challita, Walid Saad, and Christian Bettstetter. “Cellular-Connected UAVs over 5G: Deep Reinforcement Learning for Interference Management”. In: *CoRR*

abs/1801.05500 (2018). arXiv: 1801.05500. URL: http://arxiv.org/abs/1801.05500.

[86] Y. He, N. Zhao, and H. Yin. "Integrated Networking, Caching, and Computing for Connected Vehicles: A Deep Reinforcement Learning Approach". In: *IEEE Transactions on Vehicular Technology* 67.1 (2018), pp. 44–55.

[87] K. Zheng et al. "Big data-driven optimization for mobile networks toward 5G". In: *IEEE Network* 30.1 (2016), pp. 44–51.

[88] C. Jiang et al. "Machine Learning Paradigms for Next-Generation Wireless Networks". In: *IEEE Wireless Communications* 24.2 (2017), pp. 98–105.

[89] H. Zhu et al. "Exploring Deep Learning for Efficient and Reliable Mobile Sensing". In: *IEEE Network* 32.4 (2018), pp. 6–7.

[90] M. Wang et al. "Machine Learning for Networking: Workflow, Advances and Opportunities". In: *IEEE Network* 32.2 (2018), pp. 92–99.

[91] Y. Xin et al. "Machine Learning and Deep Learning Methods for Cybersecurity". In: *IEEE Access* 6 (2018), pp. 35365–35381.

[92] Zubair Md Fadlullah et al. "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems". In: *IEEE Communications Surveys & Tutorials* 19.4 (2017), pp. 2432–2455.

[93] Q. Mao, F. Hu, and Q. Hao. "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey". In: *IEEE Communications Surveys Tutorials* 20.4 (2018), pp. 2595–2621.

[94] M. Chen et al. "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial". In: *IEEE Communications Surveys Tutorials* 21.4 (2019), pp. 3039–3071.

[95] Xiping Wu et al. "Hybrid LiFi and WiFi Networks: A Survey". In: *IEEE Communications Surveys Tutorials* 23.2 (2021), pp. 1398–1420. DOI: 10.1109/COMST.2021.3058296.

[96] M. Obeed et al. “On Optimizing VLC Networks for Downlink Multi-User Transmission: A Survey”. In: *IEEE Communications Surveys Tutorials* 21.3 (2019), pp. 2947–2976.

[97] Sungwook Kim. “Hybrid RF/VLC Network Spectrum Allocation Scheme Using Bargaining Solutions”. In: *IEEE Access* 10 (2022), pp. 20019–20028. DOI: 10.1109/ACCESS.2022.3153327.

[98] Sylvester Aboagye et al. “Matching Theory-Based Joint Access Point Assignment and Power Allocation in Hybrid RF/VLC HetNet”. In: *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*. 2020, pp. 1–7. DOI: 10.1109/GLOBECOM42002.2020.9322597.

[99] Weihua Wu, Fen Zhou, and Qinghai Yang. “Adaptive Network Resource Optimization for Heterogeneous VLC/RF Wireless Networks”. In: *IEEE Transactions on Communications* 66.11 (2018), pp. 5568–5581. DOI: 10.1109/TCOMM.2018.2831207.

[100] Tanuja Dogra and Manoranjan Rai Bharti. “User pairing and power allocation strategies for downlink NOMA-based VLC systems: An overview”. In: *AEU - International Journal of Electronics and Communications* 149 (2022), p. 154184. ISSN: 1434-8411. DOI: https://doi.org/10.1016/j.aeue.2022.154184. URL: https://www.sciencedirect.com/science/article/pii/S1434841122000814.

[101] S. C. Chen, N. Bambos, and G. J. Pottie. “Admission control schemes for wireless communication networks with adjustable transmitter powers”. In: *Proceedings of INFOCOM ’94 Conference on Computer Communications*. 1994, 21–28 vol.1.

[102] Xiaoxin Qiu and K. Chawla. “On the Performance of adaptive modulation in cellular systems”. In: *IEEE Transactions on Communications* 47.6 (1999), pp. 884–895.

[103] Jianqing Fan et al. “A theoretical analysis of deep Q-learning”. In: *Learning for dynamics and control*. PMLR. 2020, pp. 486–489.

[104] Stevo Bozinovski. “Reminder of the first paper on transfer learning in neural networks, 1976”. In: *Informatica* 44.3 (2020).

[105] A. Fulgosi S. Bozinovski A. Santic. “Normal teaching strategy in pair-association in the case teacher:human learner:machine. (original in Croatian: Normalna strategija obicavanja u obucanju asocojacije parova u slucaju ucitelj:covjek-ucenik:masina)”. In: *Proc. Conf. ETAN, Banja Luka* 21.IV (1977), pp. 341–346.

[106] S. Bozinovski. “Experiments with nonbiological systems teaching. (original in Macedonian: Eksperimenti na obucuvanje na nebioloski sistemi)”. In: *Proc. Conf. ETAN, Zadar* 22.IV (1978), pp. 371–379.

[107] S. Bozinovski. “Teaching space: A representation concept for adaptive pattern classification. COINS Technical Report”. In: *University of Massachusetts, Amherst* No (1981), pp. 81–28.

[108] Stevo Božinovski. “A representation theorem for linear pattern classifier training”. In: *IEEE transactions on systems, man, and cybernetics* 1 (1985), pp. 159–161.

[109] David E Rumelhart, James L McClelland, PDP Research Group, et al. *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*. The MIT press, 1986.

[110] Lorien Y Pratt, Jack Mostow, and Candace A Kamm. “Direct transfer of learned information among neural networks”. In: *Proceedings of the ninth National conference on Artificial intelligence-Volume 2*. 1991, pp. 584–589.

[111] Chuanqi Tan et al. “A survey on deep transfer learning”. In: *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*. Springer. 2018, pp. 270–279.

[112] Qiyang Zhao, David Grace, and Tim Clarke. “Transfer learning and cooperation management: balancing the quality of service and information exchange overhead in cognitive radio networks”. In: *Transactions on Emerging Telecommunications Technologies* 26.2 (2015), pp. 290–301.

[113] Meiyu Wang et al. “Transfer learning promotes 6G wireless communications: Recent advances and future challenges”. In: *IEEE Transactions on Reliability* 70.2 (2021), pp. 790–807.

[114] Helin Yang et al. “Learning-based energy-efficient resource management by heterogeneous RF/VLC for ultra-reliable low-latency industrial IoT networks”. In: *IEEE Transactions on Industrial Informatics* 16.8 (2019), pp. 5565–5576.

[115] John R Barry. *Wireless infrared communications*. Vol. 280. Springer Science & Business Media, 1994.

[116] John R. Barry and David G. Messerschmitt. *Wireless Infrared Communications*. USA: Kluwer Academic Publishers, 1994. ISBN: 0792394763.

[117] Zabih Ghassemlooy et al. *Visible light communications: theory and applications*. CRC press, 2017.

[118] “IST-4-027756 WINNER II D1.1.2 V1.2. (2008, Feb.). WINNER II Channel Models. Available: http://www.ist-winner.org”. In: pp. 241–245.

[119] A. Lapidoth, S. M. Moser, and M. A. Wigger. “On the capacity of free-space optical intensity channels”. In: *2008 IEEE International Symposium on Information Theory*. 2008, pp. 2419–2423.

[120] S. Hranilovic and F. R. Kschischang. “Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise”. In: *IEEE Transactions on Information Theory* 50.5 (2004), pp. 784–795.

[121] A. A. Farid and S. Hranilovic. “Capacity Bounds for Wireless Optical Intensity Channels With Gaussian Noise”. In: *IEEE Transactions on Information Theory* 56.12 (2010), pp. 6066–6077.

[122] A. Chaaban, J. Morvan, and M. Alouini. “Free-Space Optical Communications: Capacity Bounds, Approximations, and a New Sphere-Packing Perspective”. In: *IEEE Transactions on Communications* 64.3 (2016), pp. 1176–1191.

[123] S. Pietrzyk and G. J. M. Janssen. “Radio resource allocation for cellular networks based on OFDMA with QoS guarantees”. In: *IEEE Global Telecommunications Conference, 2004. GLOBECOM ’04.* Vol. 4. 2004, 2694–2699 Vol.4.

[124] S. J. Pan and Q. Yang. “A Survey on Transfer Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (2010), pp. 1345–1359.

[125] Helena Serpi and Christina Tanya Politi. “Radio environment maps for indoor visible light communications aided by machine learning”. In: *AEU-International Journal of Electronics and Communications* 170 (2023), p. 154866.

[126] MSM Gismalla et al. “Design of an optical attocells configuration for an indoor visible light communications system”. In: *AEU-International Journal of Electronics and Communications* 112 (2019), p. 152946.

[127] Lei Qian, Xuefen Chi, and Linlin Zhao. “Analysis of effective capacity for visible light communication systems with mobility support”. In: *AEU-International Journal of Electronics and Communications* 88 (2018), pp. 38–43.

[128] Cheng Chen, Dushyantha Basnayaka, and Harald Haas. “Non-line-of-sight channel impulse response characterisation in visible light communications”. In: *2016 IEEE International Conference on Communications (ICC)*. May 2016, pp. 1–6. DOI: 10.1109/ICC.2016.7511382.

[129] Zhengquan Zhang et al. “6G wireless networks: Vision, requirements, architecture, and key technologies”. In: *IEEE vehicular technology magazine* 14.3 (2019), pp. 28–41.

[130] Rong Zhang et al. “Visible light communications in heterogeneous networks: Paving the way for user-centric design”. In: *IEEE wireless communications* 22.2 (2015), pp. 8–16.

[131] Xianfu Chen et al. “Age of information aware radio resource management in vehicular networks: A proactive deep reinforcement learning perspective”. In: *IEEE Transactions on wireless communications* 19.4 (2020), pp. 2268–2281.

[132] Borja Genoves Guzman, Alejandro Lancho Serrano, and Víctor P Gil Jimenez. “Cooperative optical wireless transmission for improving performance in indoor scenarios for visible light communications”. In: *IEEE Transactions on Consumer Electronics* 61.4 (2015), pp. 393–401.

[133] Keshav Singh et al. “Joint active and passive beamforming design for RIS-aided IBFD IoT communications: QoS and power efficiency considerations”. In: *IEEE Transactions on Consumer Electronics* 69.2 (2022), pp. 170–182.

[134] Chung Kit Wu et al. “State-of-the-Art and Research Opportunities for Next-Generation Consumer Electronics”. In: *IEEE Transactions on Consumer Electronics* 69.4 (2023), pp. 937–948. DOI: 10.1109/TCE.2022.3232478.

[135] Mohammad Dehghani Soltani et al. “Modeling the random orientation of mobile devices: Measurement, analysis and LiFi use case”. In: *IEEE Transactions on Communications* 67.3 (2018), pp. 2157–2172.

[136] Zhihong Zeng et al. “Realistic indoor hybrid WiFi and OFDMA-based LiFi networks”. In: *IEEE Transactions on Communications* 68.5 (2020), pp. 2978–2991.

[137] Volodymyr Mnih et al. “Asynchronous methods for deep reinforcement learning”. In: *International Conference on Machine Learning*. PMLR. 2016, pp. 1928–1937.

[138] John Schulman et al. “Proximal policy optimization algorithms”. In: *arXiv preprint arXiv:1707.06347* (2017).

[139] J.M. Kahn and J.R. Barry. “Wireless infrared communications”. In: *Proceedings of the IEEE* 85.2 (1997), pp. 265–298. DOI: 10.1109/5.554222.

[140] Dajana Cassioli, Luca Alfredo Annoni, and Stefano Piersanti. “Characterization of path loss and delay spread of 60-GHz UWB channels vs. frequency”. In: *2013 IEEE International Conference on Communications (ICC)*. IEEE. 2013, pp. 5153–5157.

[141] Dushyantha A Basnayaka and Harald Haas. “Design and analysis of a hybrid radio frequency and visible light communication system”. In: *IEEE Transactions on Communications* 65.10 (2017), pp. 4334–4347.

[142] Dajana Cassioli and Nikola Rendevski. “A statistical model for the shadowing induced by human bodies in the proximity of a mmWaves radio link”. In: *2014 IEEE International Conference on Communications Workshops (ICC)*. IEEE. 2014, pp. 14–19.

[143] John Schulman. “The nuts and bolts of deep RL research”. In: *NIPS Deep RL Workshop*. 2016.

[144] Pengyu Cong and Chenyang Yang. “Number of FLOPs of Training DNNs for Learning Precoding”. In: *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. 2023, pp. 1–6. DOI: 10.1109/VTC2023-Spring57618.2023.10200945.

[145] Mohammad Dehghani Soltani et al. “Bidirectional user throughput maximization based on feedback reduction in LiFi networks”. In: *IEEE Transactions on Communications* 66.7 (2018), pp. 3172–3186.