February 2023
Master's Degree Thesis

Deep Reinforcement Learning-Based Coordinated Beamforming for mmWave Massive MIMO Vehicular Networks

Graduate School of Chosun University
Department of Computer Engineering
Pulok Tarafder

Deep Reinforcement Learning-Based Coordinated Beamforming for mmWave Massive MIMO Vehicular Networks

mmWave 대규모 MIMO 차량 네트워크를 위한
심층강화학습 기반 빔형성 기술 연구

February 24, 2023

Graduate School of Chosun University
Department of Computer Engineering
Pulok Tarafder

Deep Reinforcement Learning-Based Coordinated Beamforming for mmWave Massive MIMO Vehicular Networks

Advisor: Prof. Wooyeol Choi, Ph.D.

A thesis submitted in partial fulfillment of the requirements for a master's degree

October 2022

Graduate School of Chosun University
Department of Computer Engineering
Pulok Tarafder

This is to certify that the master's thesis of

Pulok Tarafder

has been approved by examining committee for
the thesis requirement for the master's degree.

타라프더 풀록의
석사학위논문을 인준함

위원장	조선대학교 교수	신석주	(인)
위원	조선대학교 교수	강문수	(인)
위원	조선대학교 교수	최우열	(인)

2022년 12월

조선대학교 대학원

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Deep Reinforcement Learning-Based Coordinated Beamforming for mmWave Massive MIMO Vehicular Networks

Pulok Tarafder
Advisor: Prof. Wooyeol Choi, Ph.D.
Department of Computer Engineering
Graduate School of Chosun University

With the increase in the number of connected devices, to facilitate more users with high-speed transfer rates and enormous bandwidth, millimeter-wave (mmWave) technology has become one of the promising research sectors in both industry and academia. As a critical enabler for beyond fifth-generation (B5G) technology, mmWave beamforming for mmWave has been studied for many years. Multi-input multi-output (MIMO) system, which is the baseline for beamforming operation, rely heavily on multiple antennas to stream data in mmWave wireless communication systems. Moreover, high-speed mmWave applications face challenges such as blockage and latency overhead. Furthermore, the efficiency of the mobile systems is severely impacted by the high training overhead required to discover the best beamforming vectors in large antenna array mmWave systems. In order to mitigate the stated challenges, in this thesis, we propose a novel deep reinforcement learning (DRL) based coordinated beamforming scheme where multiple base stations (BSs) serve one mobile station (MS) jointly. The constructed solution then uses a proposed DRL model

and predict the suboptimal beamforming vectors at the BSs out of possible beamforming codebook candidates. This solution enables a complete system that facilitates highly mobile mmWave applications with dependable coverage, minimal training overhead, and low latency. Numerical results demonstrate that our proposed algorithm remarkably increase the achievable sum rate capacity for the highly mobile mmWave massive MIMO scenario while ensuring low training and latency overhead.

한 글 요 약

mmWave 대규모 MIMO 차량 네트워크를 위한 심층강화학습 기반 빔형성 기술 연구

타라프더 플록
지도교수: 최우열
컴퓨터공학과
조선대학교 대학원

연결 장치의 수가 증가함에 따라 더 많은 사용자가 고속 전송 속도와 엄청난 대역폭을 사용할 수 있도록 밀리미터파(mmWave) 기술은 산업계와 학계 모두에서 유망한 연구 분야 중 하나가 되었습니다. 5세대 이상(B5G) 기술을 위한 중요한 인에이블러로서 mmWave용 mmWave 빔포밍은 수년 동안 연구되어 왔습니다. 빔포밍 동작의 기준이 되는 MIMO(Multi-Input Multi-Output) 시스템은 mmWave 무선 통신 시스템에서 데이터를 스트리밍하기 위해 다중 안테나에 크게 의존합니다. 또한 고속 mmWave 애플리케이션은 막힘 및 대기 시간 오버헤드와 같은 문제에 직면해 있습니다. 또한 모바일 시스템의 효율성은 대형 안테나 어레이 mmWave 시스템에서 최상의 빔포밍 벡터를 발견하는 데 필요한 높은 교육 오버헤드에 의해 심각한 영향을 받습니다. 이러한 문제를 해결하기 위해 이 논문에서는 여러 기지국(BS)이 하나의 이동국(MS)에 공동으로 서비스하는 새로운 DRL(Deep Reinforcement Learning) 기반 조정 빔포밍 방식을 제안합니다. 그런 다음 구성된 솔루션은 제안된 DRL 모델을 사용하여 가능한 빔포밍 코드북 후보 중에서 BS에서 차선책 빔포밍 벡터를 예측합니다. 이 솔루션은 신뢰할 수 있는 적용 범위, 최소한의 교육 오버헤드 및 짧은 대기 시간으로 이동성이 높은 mmWave 애플리케이션을 용이하게 하는 완전한 시스템을

가능하게 합니다. 수치 결과는 제안된 알고리즘이 낮은 교육 및 대기 시간 오버헤드를 보장하면서 이동성이 높은 mmWave 대규모 MIMO 시나리오에 대해 달성 가능한 합계 속도를 크게 높일 수 있음을 보여줍니다.

I. INTRODUCTION

With the recent advancements in 5G, it is not ambitious to expect that 5G will enable 1000× more data traffic than the wideley established current 4G standards [1], [2]. Foreseeing the rise is users in increased traffic demands, facilitating these massive users and serving great quality cellular networks require high frequency waves. Recently, millimeter wave (mmWave) communication has attracted significant interest in designing 5G wireless communication systems owing to its advantages in reducing spectrum scarcity and enabling high data speeds [3]. The ranges of mmWave frequency band lies between 30 GHz to 300 GHz. This higher frequencies however travels very short distance due to their physical limitations in the spectrum and demonstrates high path loss [4]. Consequently, higher frequencies require smaller cellular cells to overcome the challenges such as path loss and blockage [5]. The massive multiple-input multiple-output (mMIMO) can use hundreds of antennas simultaneously to propagate signal in the same time-frequency resource and serve tens of users at the same time [6]. The mMIMO techniques can be utilized to perform highly directional transmissions thanks to the short wavelength of mmWave, which makes it physically feasible to equip a lot of antennas at the transceiver in a cellular network and can significantly improve network capacity [7]–[10].

Vehicles are getting more sensors as driving gets more automated, resulting in increasingly higher data rates. Beamforming in mMIMO makes it possible to serve distance users with mmWaves, even users that are not stationary. Therefore, the only practical method for large bandwidth connected automobiles is mmWave mMIMO communication [11]. As a result, mmWave mMIMO systems can serve mobile vehicles effectively considering the proper beam is selected. Due

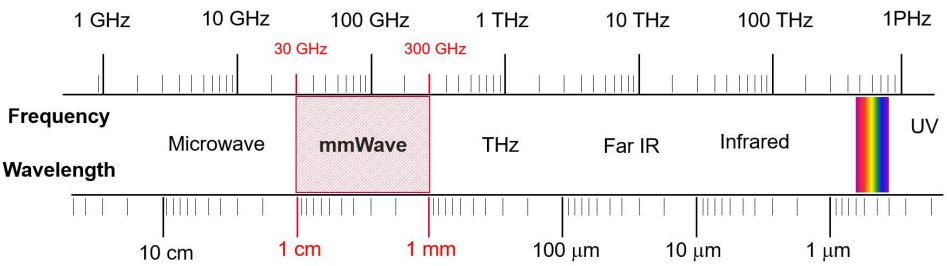

Figure 1: The mmWave frequency band.

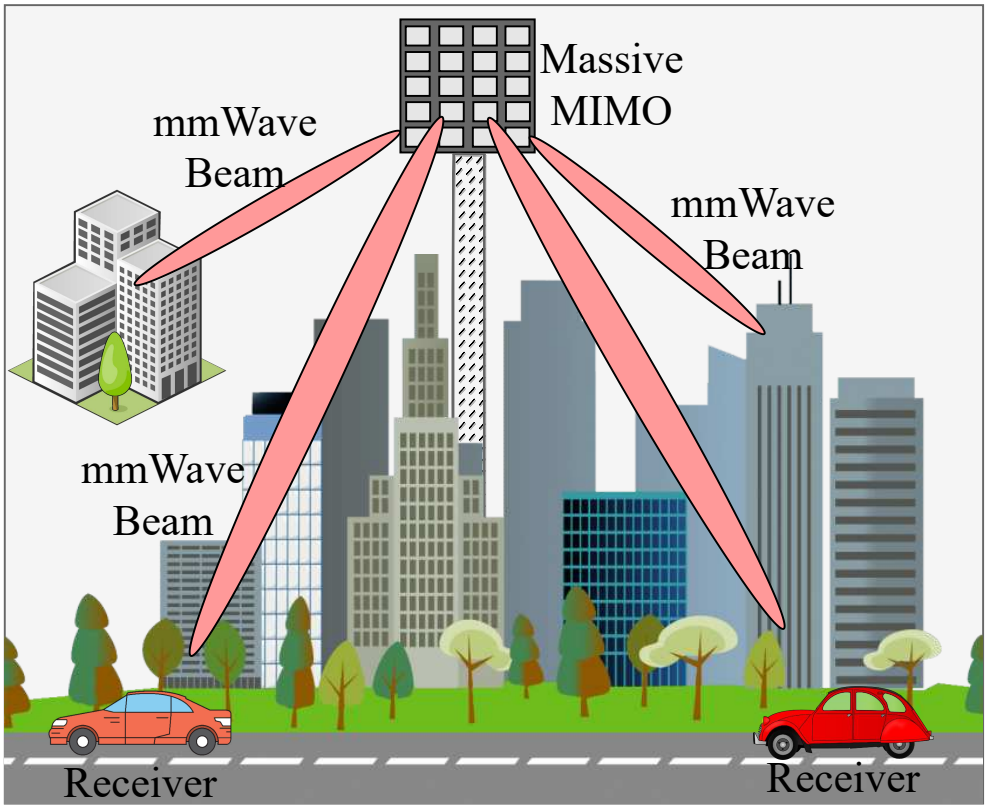

Figure 2: Illustration of mMIMO beamforming.

to the fundamental differences between mmWave communications and current microwave-based communication technologies (e.g., 2.4 GHz and 5 GHz), the

mmWave systems present difficulties, such as a high sensitivity to shadowing and a significant signal attenuation[12]. In this thesis, in order to overcome these issues and allow mMIMO environments where a highly non-stationary active user is present, we introduce a coordinated beamforming scheme utilizing deep reinforcement learning (DRL) to select the optimal beam for a vehicular communication system. First, a deep Q-network (DQN) algorithm is created to handle the beam selection problem as a Markov decision process (MDP). Then, by ensuring that the limitations of the beam selection matrix are met, our goal is to choose the best beams to maximize the sum rate for the user served.

A. Related Works

There have been few standard traditional approached for beamforming or beam selection. In [13], Gao *et al.* followed an exhausting search approach for beamforming which demonstrates very high complexities in the system. Pal *et al.* [14] on the other hand followed a minor different approach where iterate through the users and beams for determining the best possible beamforming matrices. This approach is also executed with high complexity algorithm.

On the other hand, DL-based approaches shows promising results in terms of application complexity and viability. Alkhateeb *et al.* [15] derived a high mobility supported mmWave massive MIMO based DL enabled coordinated beamforming scheme for an outdoor scenario. To formulate their design, they utilized distributed BSs simultaneously to serve a mobile user. They predicted the optimal beams using traditional DL approach, and compared the achievable rate performance of their DL method with optimal achievable rate of beamforming. Zhang *et al.* [16] proposed a multi-user massive MIMO coordinated beamforming scheme for heterogeneous networks (HetNets)

focusing on energy efficiency (EE) based on convolutional neural network (CNN) approach. They designed and used a multi-user huge MIMO HetNets optimization challenge to maximize EE with less complexity and compute delay. In order to accomplish end-to-end autonomous beamforming, [17] introduced a constrained deep neural network (constrained-DNN) based beamforming technique. This method uses a NN in place of the beamforming matrices used in conventional beamforming. In [18], in-depth experiments for coordinated multipoint transmission at 73 GHz were carried out in a downtown Brooklyn urban open square setting. The results of the analysis showed that serving a user jointly at the same time by many BSs can achieve a considerable coverage improvement. Moreover, another work on BS coordination, where a user is concurrently given access by many BSs, may be used to generate a significant coverage increase and is demonstrated by Maamari *et al.* in an analysis of the performance of heterogeneous mmWave cellular networks in [19]. Gupta *et al.* in [20] investigated the scope of a minimum of one LOS case when the users are served with line of sight (LOS) connections. The results showed that the density of coordinating BSs should scale with the square of the blockage density in order to maintain the same LOS connection. Although[18]–[20] established how BS coordination significantly increased coverage, they lack the analysis of producing coordinated beamforming vectors.

Beamforming challenges and training overhead management gets more complicated when the high mobile system is incorporated into the mmWave massive MIMO systems.

In order to enable high-speed, long-range, and reliable transmission in millimeter-wave 60 GHz wireless personal area networks (60 GHz WPANs), Wang *et al.* [21] introduced a beamforming approach applied in the media

access control (MAC) layer on top of various physical layer (PHY) designs. [11] suggested a new strategy to lower the overhead for beam alignment by utilizing DSRC and/or sensor information as side information. Afterwards, they provided detailed examples of how to leverage location data from dedicated short-range communication (DSRC) to lessen the overhead of beam alignment and tracking in mmWave vehicle-to-everything (V2X) applications.

Va *et al.* on the other hand proposed a multipath fingerprint database using the vehicle's position (for example, as determined by GPS) to gain information of probable pointing directions for accurate beam alignment. The power loss probability is a parameter used in the method to measure misalignment precision and is used to enhance candidate beam selection. Moreover, two candidate beam selection techniques are created, one of which uses a heuristic, and the other aims to reduce the likelihood of misalignment. In addition, Zhou *et al.* [22] proposed a DQN-based algorithm to train and determine the optimal receiver beam direction with the purpose of maximizing average received signal power (RSP).

However, there are various drawbacks to designing beamforming vectors solely based on location data and RSP. First, narrow-beam systems may not function effectively with position-acquisition sensors like GPS because of their poor precision, which is typically in the range of meters. Second, these technologies are unable to handle indoor applications, since GPS sensors perform poorly inside of structures. Additionally, the beamforming vectors depend on the environment's shape, obstructions, etc. in addition to the transmitter and receiver's locations. Also, RSP can experience severe penetration power loss because of the vehicle's metal body.

B. Contributions

In this thesis, for highly mobile mmWave applications, we provide a novel DRL approach for highly mobile mmWave communication architecture. As part of our suggested method, a coordinated beamforming system is used, in which a number of BSs concurrently provide access to a single non-stationary user. In this approach, a deep learning (DL) network exclusively utilizes beam patterns and learns how to anticipate the BSs beamforming vectors from the signals obtained at the scattered BSs. Here, the idea behind this is that the propagated waves collectively acquired at the scattered BSs indicate a distinctive multi-path signature of both the user position and its surroundings. There are several benefits to the suggested approach. First, the suggested technique can accommodate not only LOS but non LOS (NLOS) framework without the need for specialized position-acquiring devices because beamforming prediction is based on the uplink received signals rather than position data. Second, only omni received pilots, which may be retrieved with minimal overhead training, are needed for the determination of the best beams. Furthermore, because the DL model trains and responds to any environment, it does not need any training before deployment in the suggested system. The proposed deep learning model also inherits the coverage and reliability improvements of coordination, since it is coupled with the coordinated beamforming mechanism.

The contributions of the proposed beamforming scheme are summarized as follows:

- We develop a simple coordinated beamforming scheme where several BSs that employ RF beamforming and are connected to a central cloud processing unit that uses baseband processing, which serves a mobile user

at once. To increase the platform's effective achievable rate, we define a training and design issue for the central baseband processing and for BSs RF beamforming vectors. The trade-off between the beamforming training overhead and the achievable sum rate using the proposed beamforming vectors is taken into account when determining the effective achievable rate for highly mobile mmWave systems.

- For the selected system, we construct a fundamental coordinated beamforming technique that relies on uplink training for creating the RF and baseband beamforming vectors. The BSs choose their RF beamforming vectors from a predetermined codebook as part of this baseline approach. The baseband beamforming is then designed by a central processor to guarantee consistent incorporating at the user. We demonstrate that the standard beamforming technique achieves the best attainable rates in a few unique but crucial situations. However, this technique has a significant training cost, which encourages the use of machine learning models.

- We introduce a system operation of machine learning modeling of a unique combined DRL and coordinated beamforming solution. The main concept of the suggested technique is to anticipate the RF beamforming vectors of the coordinating BSs using just beam patterns, i.e., with very little training overhead. The proposed approach also enables minimal coordination overhead harvesting of coordinated beamforming improvements with wide coverage and low latency, making the method a viable solution for highly mobile mmWave applications.

C. Thesis Layout

The thesis is organized as follows. In Chapter II, we present the system and channel model of our communication system. Then in chapter III, we describe the problem statement, fundamentals of RL and DRL. Next, in Chapter III, we present our proposed solution and simulation analysis of priority-based joint resource allocation with DQL. Then in Chapter IV, we describe the problem statement, proposed solution and simulation analysis of DRL-Based coordinated beamforming approach. And finally, we conclude the thesis in Chapter V.

II. System and Channel Model

In this section, we elaborate our coordinated mmWave system and channel model. Additionally, each model's main assumptions are highlighted.

In this section, we discuss the chosen frequency-selective coordinated mmWave system and channel models. Additionally, each model's main assumptions are highlighted.

A. System model

We analyze a mmWave enabled vehicular communication architecture shown in Fig. 3, where N BSs are concurrently providing service to one mobile station (MS). Each BS is equipped with M number of antennas, and each BS is linked to a central processing unit in the cloud. In the interests of simplicity, we assume that each BS utilizes analog-only beamforming with networks of phase shifters and has a single RF chain [23]. In this thesis, we use the assumption that the MS is equipped with only one antenna.

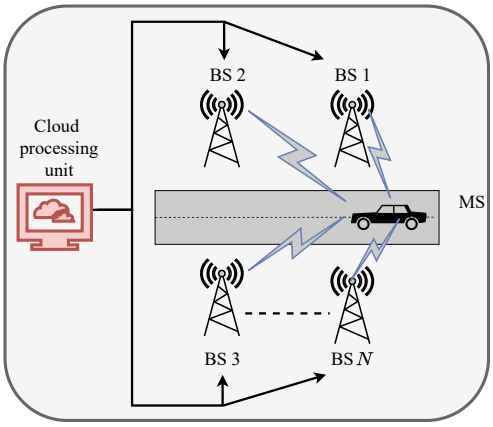

Figure 3: A downlink mmWave massive MIMO vehicular beamforming system.

The signals are precoded using a $N \times 1$ digital precoder $\mathbf{f_k} \in \mathbb{C}^{N \times 1}$. The

frequency domain signals are then converted into the time domain using N K-point inverse fast Fourier Transforms (IFFTs). Afterwards, each BS n performs a time-domain analog beamforming and then transmits the resulting signal. At the receiver end, the received signal is converted to the frequency domain using a K-point FFT, presuming perfect synchronization of frequency and carrier offset. The received signal at k^{th} subcarrier at n^{th} BS is denoted by

$$\mathbf{y}_k = \sum_{n=1}^{N} \mathbf{h}_{k,n}^T \mathbf{x}_{k,n} + \mathbf{n}_k, \tag{1}$$

where $\mathbf{x}_{k,n}$ is the transmitted complex baseband signal, $\mathbf{h}_{k,n}$ is the $M \times 1$ channel vector between the MS and BS, $\mathbf{n}_k \in \mathbb{C}^{M \times 1}$ is the received noise at the BS with independent and identically complex (i.i.c.) additive white Gaussian noise (AWGN) distribution with zero mean and variance σ^2.

B. Channel model

We consider a L clustered geometric wideband model for our mmWave cellular channel [24]–[26]. For each cluster l, it is assumed that $l = 1, \cdots, L$ contributes one ray with a temporal delay $\tau_l \in \mathbb{R}$, and azimuth/elevation angles of arrival (AoA) is θ_l, ϕ_l. Let $p_{rc}(\tau)$ be a pulse shaping function for T_S-spaced signaling assessed at τ seconds, and let ρ_n signify the path-loss between the user and the n^{th} BS [27]. The delay-d channel vector in this model $\mathbf{h}_{d,n}$ between the user and the n^{th} BS, is as follows

$$\mathbf{h}_{d,n} = \sqrt{\frac{M}{\rho_n}} \sum_{l=1}^{L} \beta_l p(dT_s - \tau_l) \mathbf{a}_n(\theta_l, \phi_l), \tag{2}$$

where $\mathbf{a}_n$ denotes the array response vector of the n^{th} BS. Considering the delay-d channel in (2), for subcarrier k, our frequency domain channel vector $\mathbf{h}_{kn}$ can be formulated as

$$\mathbf{h}_{k,n} = \sum_{d=0}^{D-1} \mathbf{h}_{d,n} \exp\left(-j\frac{2\pi k}{K}d\right). \tag{3}$$

Our adopted block-fading channel model $\{\mathbf{h}_{k,n}\}_{k=1}^{K}$ is considered to remain constant throughout the channel coherence time, abbreviated T_C, and it is dependent on user the mobility and the channel multi-path components [28].

III. Fundamentals of Reinforcement Learning

To understand the proposed solution, we have to understand the fundamentals of reinforcement learning (RL) and deep reinforcement learning (DRL). Thus, we briefly discuss the internal structure, decision-making process, and the convergence process of RL, DRL in this chapter.

A. Reinforcement learning

RL is different from traditional ML (supervised or unsupervised) models. RL consists of a decision maker or agent that interacts with an environment that is placed in.The agent will receive some representation of the environment's condition at each time step. The agent chooses an action based on this depiction. The environment is then changed to a new state after that. As a result of its prior action, the agent receives a reward during the process.

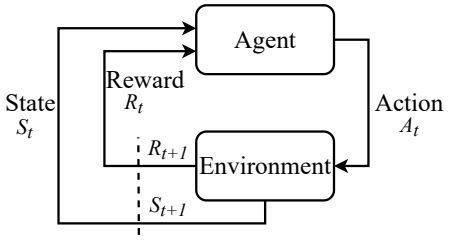

Figure 4: MDP Process for RL.

The decision-making process in RL is formulated with Markov decision process (MDP). MDP provides sequential stochastic control for the agent in the decision-making process [29]. In general, MDP has four components, (S, A, R, P) where S is the state, A is the action, R is the immediate reward for an action taken, and P is the state transition probability. As illustrated in Fig. 4, in any time step t, the agent interacts with the environment, and observes a current state

S_t and performs an action A_t. After that, the agent is awarded a reward R_t. At the same time, the agent experiences a new transition of the state S_t to S_{t+1} and so on. Meanwhile, the agent tries to meet its primary aim, which is finding a policy π that returns the possible maximum accumulated reward. The agent eventually aims in maximizing the anticipated discounted total reward indicated by $max[\sum_{t=0}^{T} \delta R_t(S_t, \pi(S_t))]$, where the discount factor is $\delta \in [0,1]$. This is the discounted reward which forms the Bellman equation otherwise known as Q-function as follows

$$Q(S_t, A_t) = (1-\alpha) \times Q(St, At) + [R + \delta(maxQ(S_{t+1}, A_t))], \tag{4}$$

where α denotes the learning rate.

Algorithm 1 RL Algorithm with Q-learning

1: $Q(S,A) = 0$
2: Init $\alpha, \delta, \varepsilon$
3: **for** $t = 1, 2, ..., T$ **do** Select A_t for S_t, according to ε
4: Get immediate R_t
5: Get S_{t+1}
6: Update $Q(S,A)$ via MDP
7: $S_t \leftarrow S_{T+1}$
8: $\pi(s)$ = arg max $Q(S,A)$

Q-learning is another name for RL with a Q-function. The agent first explores each state of the environment while performing numerous actions, and then uses the Q-function to create a Q-table for each state-action pair. Afterwards, the agent starts executing actions for the highest Q-value possible from the Q-table to exploit the environment. Subsequently, the agent begins exploring or exploiting

the environment based on the likelihood, and this strategy is referred to as the ε-greedy policy. A representation of an example is presented in Algorithm 1.

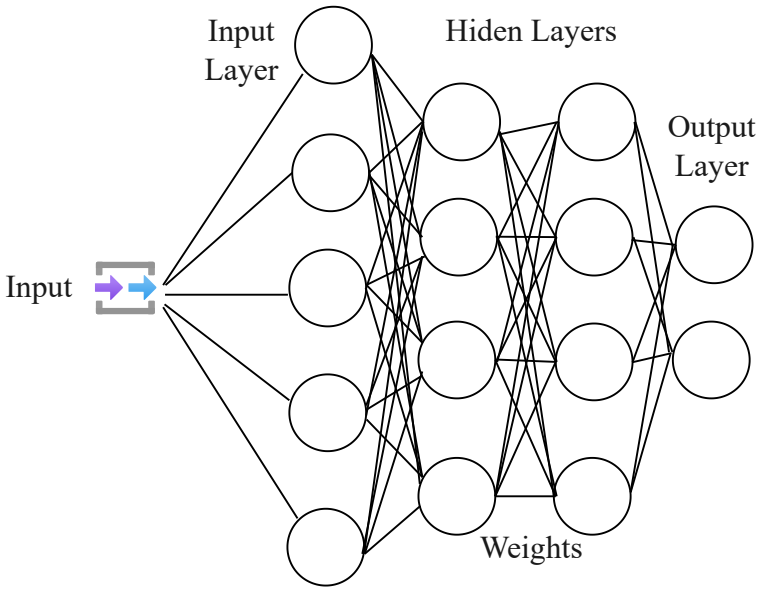

Figure 5: Neural Network.

B. Deep reinforcement learning

The regular Q-learning algorithm is mostly suitable in an environment where the action space and state space is comparatively small. As a result, the Q-learning system starts to become complicated and starts to perform poorly as the state and action space becomes larger. Consequently, even though this algorithm became very widely accepted and implemented in recent years as a result of its effectiveness in solving complicated sequential decision-making issues Q-learning still has some limitations such as tackling very large state space. To handle difficult sequential decision-making issues for such systems, DRL

combines DL techniques with RL. DL is especially helpful for solving issues with high-dimensional state spaces. As a result of its capacity to learn many levels of abstraction from data, DL enables RL to handle more challenging problems with less prior information. For applications in real works scenarios, tackling iteration complexities and large state space does not bring sufficient performance when applying RL alone. To overcome the challenges of RL, researchers implemented a deep neural network (DNN) in place of Q-table and called it deep Q-leaning (DQN) [30]. We illustrate a standard DQN algorithm in Fig. 7. By learning from the data, the DNN's primary objective is to eliminate manual computation. Technically, any DNN is a non-linear structure which mimics the structure of the human brain and can train to carry out tasks including classification, prediction, visualization, and decision-making [31]. It is made up of neurons stacked at different levels. A layer of input, two layers of hidden and one layer of output are normally included, all of which are coupled as depicted in Figure. 5 [32]. The initial layers or known as input layers takes input and pass it to the deep inside layers or hidden layers with the assistance of the input neurons. Consequently, the output layer receives the data once it has been conveyed from the hidden layer. During this process, all the neuron accumulate a weighted input, an activation function, and an output. Depending on the neuron's input, the activation function dictates the output [33]. According to this definition, the activation function acts as a trigger that is dependent on the weighted input.

The agent uses backpropagation during the training phase to adjust the weighted values of the input of the neurons depending on the outputs of the output layer. The agent compares the policy DNN model's output to the target DNN model and calculates the error [34]. The agent then uses backpropagation to update the policy DNN. The usual term for this procedure is optimization with

gradient descent. The agent uses policy DNN to update the target DNN after a certain amount of time. ERM is added to the DQL framework to help the optimal policy converge more steadily [35], [36]. The agent performs various actions and records the current states, rewards received, upcoming states, and ERM actions [35], [36]. Afterwards, the agent trains the policy DNN using a small batch of data from the device [37]. As a result, the agent uses the learned DNN to carry out its decision-making activity efficiently and promptly. We illustrated the mechanism of DQN simply in Fig. 7 and Algorithm 2 to have a better understanding.

Algorithm 2 The Deep Q-learning Algorithm with Q-learning

Init policy, target DQN with random w, w'
Init experience replay memory (ERM)
Init ε
for $t = 1, 2, ..., T$ **do** Select A_t for S_t, according to ε
    Get immediate R_t
    Get S_{t+1}
    Put $(S_t, A_t, R_t, S_{t+1}) \rightarrow$ ERM
    Form random sample mini batch of (S_t, A_t, R_t, S_{t+1}) from ERM
    Optimize w of DNN policy using MDP with gradient descent
    $w' \leftarrow w$ after T

IV. DRL-Based Coordinated Beamforming

In this chapter, we introduce a baseline DRL coordinated beamforming approach for a highly mobile vehicular mmWave communication system. To present the proposed solution, we first describe the problem formulation, then derive the novel DRL based approach for beamforming. In this chapter, we also present the environment setup, dataset generation, simulation parameters, and performance analysis for our proposed scheme.

A. Problem Statement

1. Problem Formulation

For a vehicular mmWave based 5G network, serving any user or MS is challenging because of the dynamic and varying environment characteristics. When signal interference, fading effect, and network congestion are considered, that we subsequently describe as the environment dynamics [38], it becomes much more complicated to serve the receiver end by maintaining eMBB, mMTC, and URLLC standards. As a result, traditional static schemes have become obsolete when performing large-scale beamforming operations. To achieve the highest level of sum rate, reduce the overhead, and tackle the large RF beamforming vector arrays, adaptive beam selection approaches are best suited for this specific task. With this motivation, in this paper, we exploit the DRL's capability of tackling varying environments to maximize the achievable data rate by selecting the optimal beam for mmWave vehicular networks in a coordinated approach.

In this paper, considering a set of beamforming vectors $\{\mathbf{f}_n^{BF}\}_{n=1}^{N}$, our focus is to formulate a beam selection matrix to optimize the downlink achievable rate

of the mmWave vehicular beamforming system. The user maximum achievable rate can be derived as

$$\mathbf{R}_a = \frac{1}{K}\sum_{k=1}^{K}\log_2\left(1+SNR\left|\sum_{n=1}^{N}\mathbf{h}_{k,n}^{T}\mathbf{f}_n^{BF}\right|^2\right), \tag{5}$$

B. Proposed DRL-based Beam Selection

In a multicell mmWave mMIMO downlink scenario, a large uniform planar array (UPA) are installed on a BS. In this thesis, we select 4 BS with 32×8 UPA resulting in M = 256 antenna arrays for each BS. We used publicly available DeepMIMO [39] dataset to generate the channel matrices between the BSs and the MS. For our model, we adopt the 'O1_60' dataset, which consists of an outdoor setting with two streets and one intersection and the system operates at 60 GHz mmWave band. The scenario holds 3 user grids (UGs): UG1, UG2, and UG3 in the Cartesian coordinate system. Combining three UGs, this scenario can employ 1,184,923 users where each user represents a possible position of our MS. For our case, we generate the scenario for 54,481 user coordinate positions. Moreover, each BS generates 64 beams, resulting V = 256 beams generation altogether from 4 BSs. Our MS is mobile in nature and our goal is to serve the MS with the best beam, coordinating with 4 BSs.

We propose a DRL framework which utilizes DQN to train and optimize the beam selection assignment. Typically, the DQN technique consists of an environment and an agent using a deep neural network (DNN). The agent engages with the environment before performing any action. Here, the BSs acts as an agent. At the beginning, the agent starts exploring the environment, moving from one state to another, at that point it has very less information about the

environment. As the agent explores the environment, it gathers information and starts to take action by exploiting the environment with the help of reward function. In any timestep t, if the current state is S_t, the agent will receive an immediate reward R_t assessing the performed action A_t using the DNN. The agent also get to take the next state S_{t+1} as input from the environment in the same timestep. Depending upon the performed A_t, the agent receives a reward R_t, if the taken action can achieve good sum rate, then the agent will also receive a good R_t. The agent gains knowledge of its surroundings and develops an ideal beam selection assignment strategy by foreseeing future events. The DNN algorithm learns this policy π at each timestep as it continues to move forward with the next timesteps.

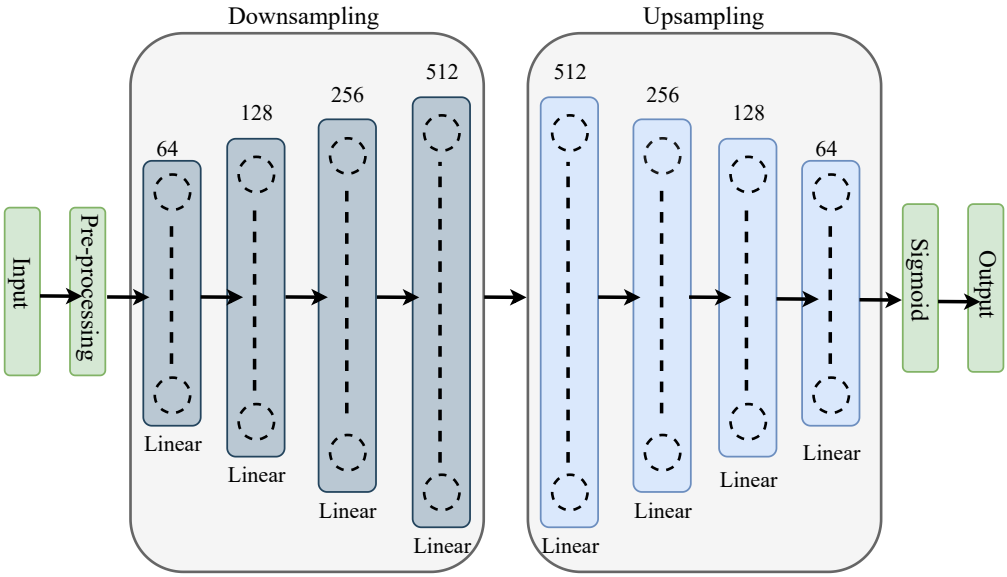

Figure 6: Proposed DNN Architecture

Next, we formulate our state, action, and reward functions as follows:

- **State:** We utilize the channel matrices for all the BSs as the state of our

environment. The complex channel matrices are constructed incorporating the bandwidth, user position, noise figure, and noise power. If the environment has Z states each having V beams, then, the state space with $Z \times V$ can be represented as $S = \tilde{S_1}, \tilde{S_2}, \tilde{S_3}, \cdots, \tilde{S_Z}$.

- **Action:** The goal of the agent is to assign beam for serving from the action space A. At each episode for a set of S, the agent has to take $Z \in A$ actions while maintaining one action per V elements from the S. Out of the $Z \times V$, the target of the agent is choosing a beam which will maximize the data rate.

- **Reward:** In our reward function, we first derive the data rate for each channel as follows:

$$\mathbf{R}_r = \log_2 \left(1 + SNR \left| \sum_{n=1}^{N} \mathbf{h}_{k,n}^T \mathbf{f}_n^{BF} \right|^2 \right), \tag{6}$$

  For every action the agent takes, we calculate the data rate of the chosen action and feed it as the reward value. Our aim is to acquire the highest possible cumulative reward as it obtains reward for each action, according to

$$\mathbf{R}_{rmax} = \arg\max \sum_{k=1}^{K} \log_2 \left(1 + SNR \left| \sum_{n=1}^{N} \mathbf{h}_{k,n}^T \mathbf{f}_n^{BF} \right|^2 \right), \tag{7}$$

With this state, action, and reward function, we propose the DNN architecture as shown in Fig. 6 as the policy controller for the beam selection. The DNN takes the place of the Q-table and calculates the Q-values for each environment state-action pair. Deriving probabilities for each beam selection for each state space is the primary objective of the DNN, and this probability can be defined by $Q(S,A)$

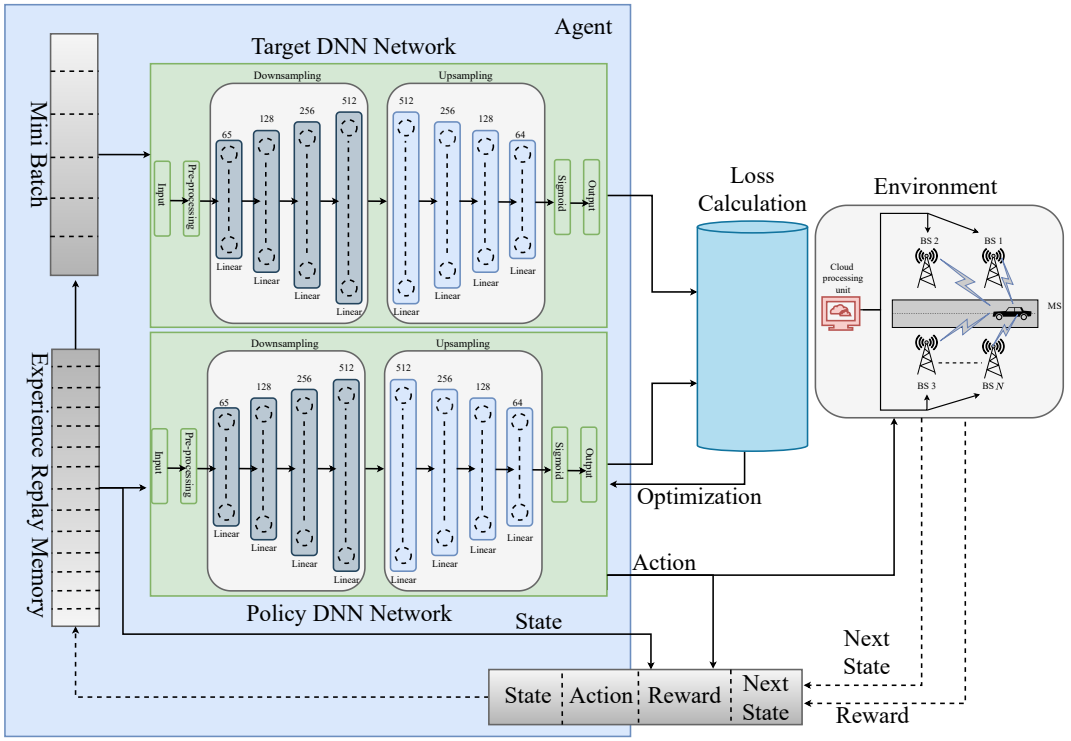

Figure 7: DQN Framework.

of the DQN algorithm. We select the best beam out of $V = 64$ candidate beams, coordinately with 4 BSs.

1. Training

For our training phase of our model, we used Adam optimizer [40] with a learning rate of 0.0005. Our DRL model minimizes the error of our training in the DNN using the Smooth$_{L1}$ loss function [41]. If we have a batch of size N, the unreduced loss can be described as following equations 8 and 9 [42].

$$\ell(x,y) = L = \{l_1, \ldots, l_N\}^T \tag{8}$$

Table 1: Adopted DeepMIMO dataset parameters

Parameters	**Values**
Scenario	'O1_60'
Active BS	3,4,5,6
Receivers	R1000 - R1300
Frequency band	60 GHz
Bandwidth	500 MHz
Number of OFDM subcarriers	1024
Subcarrier limit	64
Number of paths	5
BS antenna shape	$1 \times 32 \times 8$
Receiver antenna shape	$1 \times 1 \times 1$

where

$$l_n = \begin{cases} 0.5\,(x_n - y_n)^2 \,/\, \text{beta}, , & \text{if } |x_n - y_n| < \text{ beta} \\ |x_n - y_n| - 0.5 * \text{ beta}, & \text{otherwise} \end{cases} \tag{9}$$

The value of beta set to default 1. We have implemented the training and the testing of our DRL scheme with PyTorch as a [43] backend. The initial data generation of the outdoor scenario 'O1_60' was executed using MATLAB. We implemented our system in an Intel Core i9-9900x CPU workstation powered with 128 GB of ram and 2 NVIDIA Titan V GPUs.

C. Simulation Analysis

In this section, we evaluate the proposed DRL based coordinated beamforming approach

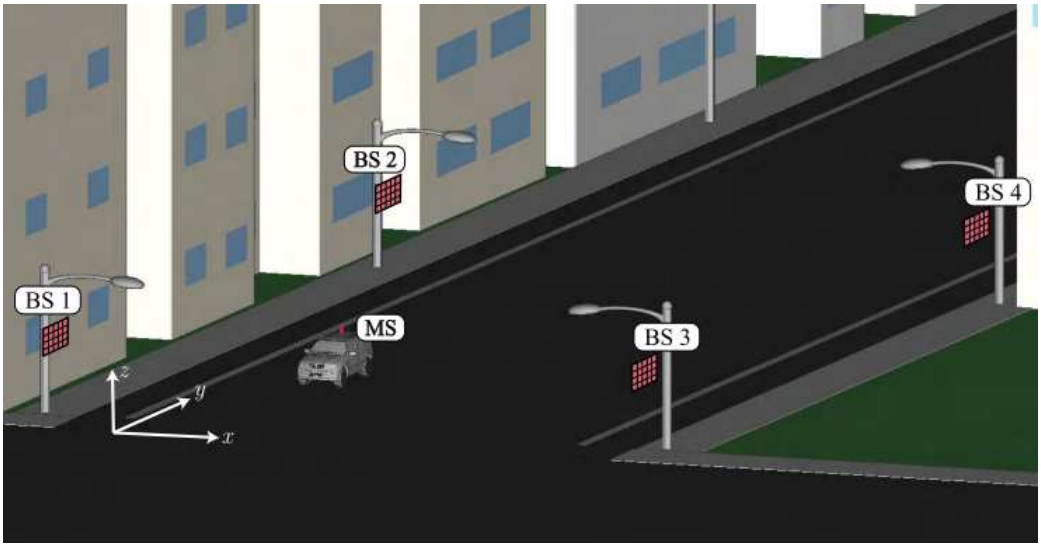

Figure 8: A figure containing 4 BSs serving one MS.

in different case studies by comparing it with traditional DL architecture [15]. The following subsections describe the simulation environment, preparation of the dataset, and the performance evaluation.

1. Simulation Environment

Wireless InSite [44] is an industry grade ray tracing tool which is commonly used in mmWave massive MIMO research. For our methodology, we used the popular publicly available DeepMIMO [39] dataset generated by the Wireless InSite. We used the 'O1_60' outdoor scenario of two streets and one intersection, which is a mmWave communication scenario operating at 60 GHz as illustrated in Fig. 8.

On the other hand, Fig. 9 demonstrates a top view of the user grid arrangements of the scenario. For this adopted scenario, 4 BSs are equipped on the top of 4 lamp posts to concurrently provide beam coverage for one MS coordinately. The lamps are located 60m away, side by side. For the ray

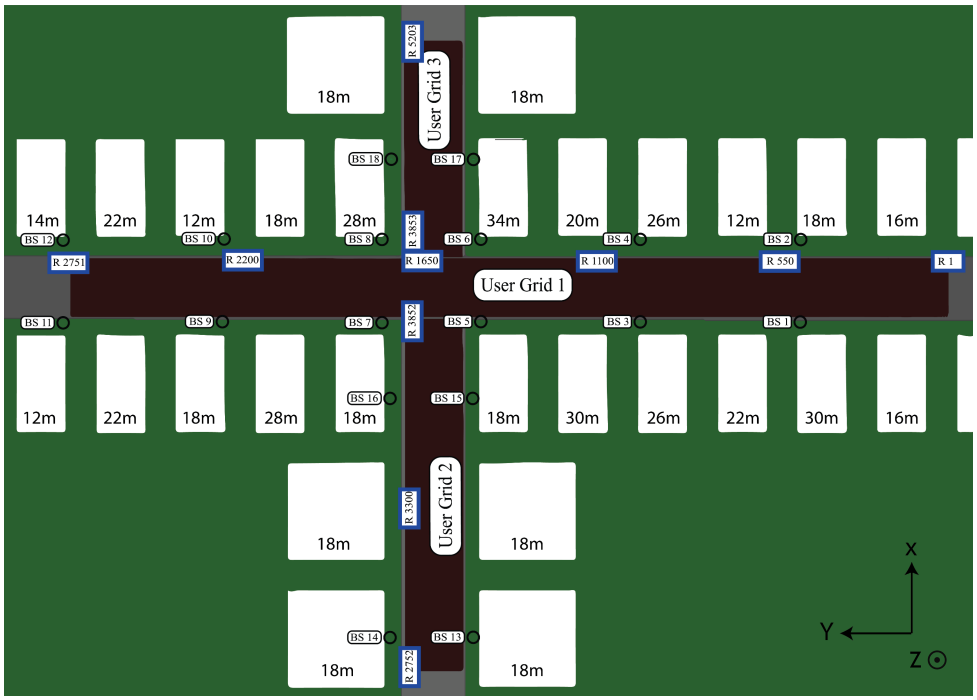

Figure 9: The top view of the ’O1’ scenario.

Table 2: Simulation parameters for the DRL model

Parameters	**Values**
Beams per BS distribution	16
Total beams	64
Transmit power	30 dBm
Learning rate (LR)	0.0005
Discount factor (γ)	0.999
Epsilon (ε)	[1, 0.1, 0.001]
Batch size	96
Number of episodes	250
Data instances	200

tracing, DeepMIMO used 60 GHz international telecommunication union (ITU) standards for the materials used in the environment, in this case its buildings. Every BS is installed on the 6m elevation having 32×8 antenna elements. The MS is incorporated with a single antenna on top of the vehicle. During the uplink training, we assumed a transmit power of 30 dBm for the MS. The adopted DeepMIMO parameters for dataset generation and the simulation parameters used in this work are summarized in the Table 1 and Table 2 respectively.

2. Performance Analysis

In this subsection, we will evaluate our achieved performance in terms of sum rate and will compare our rate with the traditional ML approach.

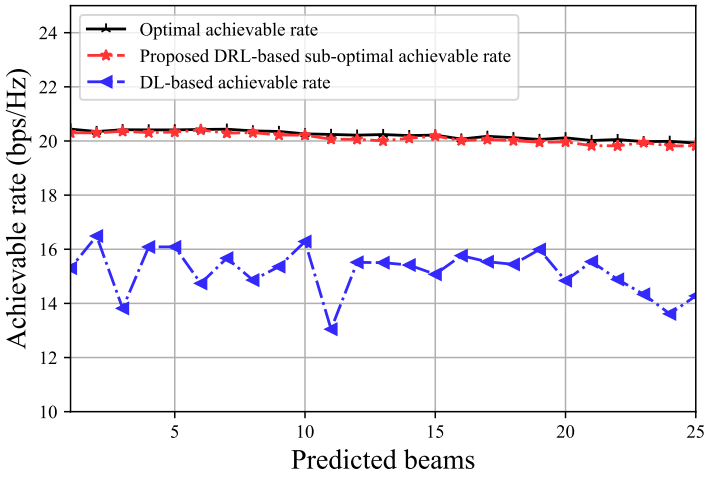

Figure 10: A comparison of effective achievable rate without overhead consideration.

Fig. 10 represents the performance analysis of our proposed model having 3 performance matrices. We plot an effective achievable rate based on our DRL, conventional DL, and optimal data rate. It is clear that our proposed DRL outperforms by a large margin and demonstrates suboptimal performance. In Fig. 10, we did not consider any beam training or latency overhead.

Communication system requires overhead. There are instances when overhead is essential for interoperability and successful communications, yet there are other occasions when overhead is unnecessary. For vehicular mmWave communication, when the user is mobile, one of the most viable communication overheads is velocity because the connectivity between the BS and the user gets affected by the velocity. For fast-moving users, it needs fast beam switching from

the BS, otherwise, because of the delay, the user might not get service on time from the BS as it moves away from its current position.

In Fig. 11 we compare DRL and DL-based beamforming performance with the optimal beamforming performance by incorporating overhead. In this stage, we consider the 64 beam training overhead, with coherence time at 40 kmph speed. It is visible that, even though our suboptimal performance experienced a slight decrease, the DRL beamforming achievable rate is still significantly higher than the DL approach.

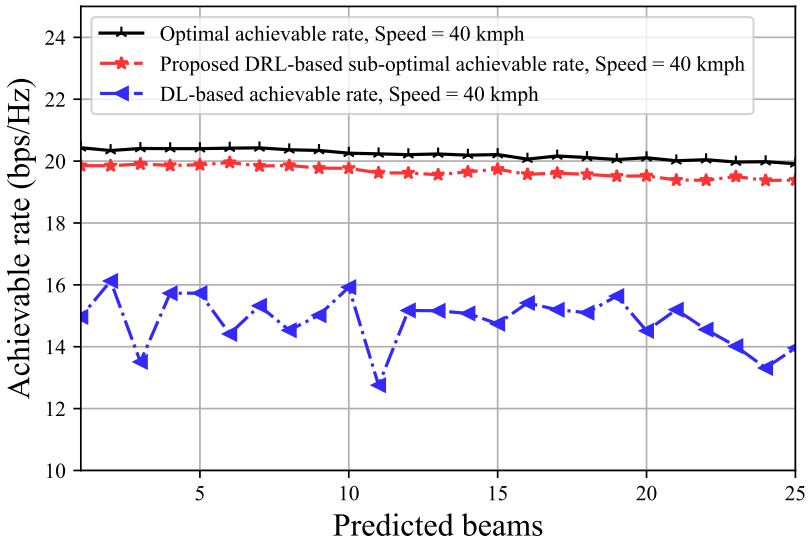

Figure 11: A comparison of effective achievable rate including overhead consideration (40 kmph).

We also compared the achievable rate versus different user position at 80 kmph and 120 kmph speed in Fig. 12 and Fig. 13 respectively. The results followed similar trends. Our DRL-based approach outperformed the DL approach by a large margin and demonstrated suboptimal performance. As the user

position moved, the achievable rate saw a slight but steady decrease over the period. However, for the traditional DL-based approach, the performance was inconsistent.

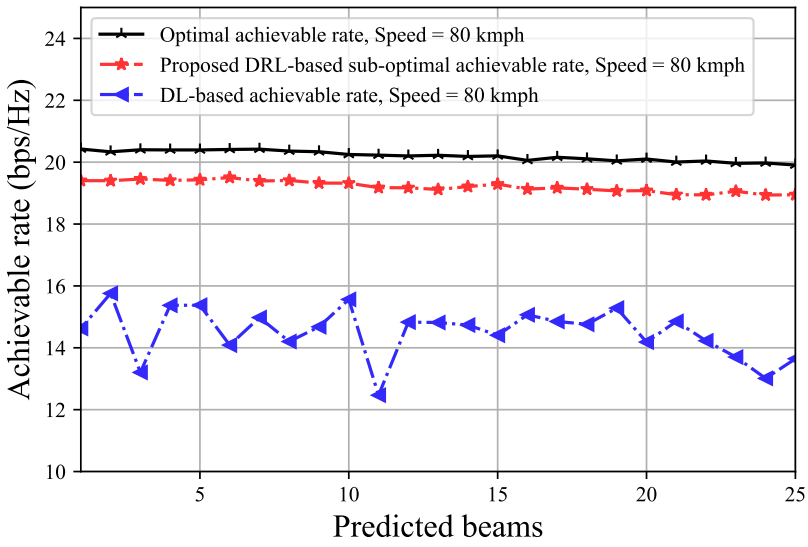

Figure 12: A comparison of effective achievable rate including overhead consideration (80 kmph).

Moreover, we compared our DRL-based average achievable sum rate for all three overhead speed side by side in Fig. 14. The performance was similarly very consistent throughout the plot, and the achievable rate of declination due to the increased overhead was negligible.

We also compared the performance of our proposed DRL scheme by varying SNR. In Fig. 15 it is demonstrated how the performance of our model varies at two different SNR levels which are low SNR at 10 dB, and high SNR at 30 dB. Previous results containing 38.65 dB SNR portrayed higher results. In the figure, it is portrayed that after the SNR was reduced to 30 dB, the initial performance

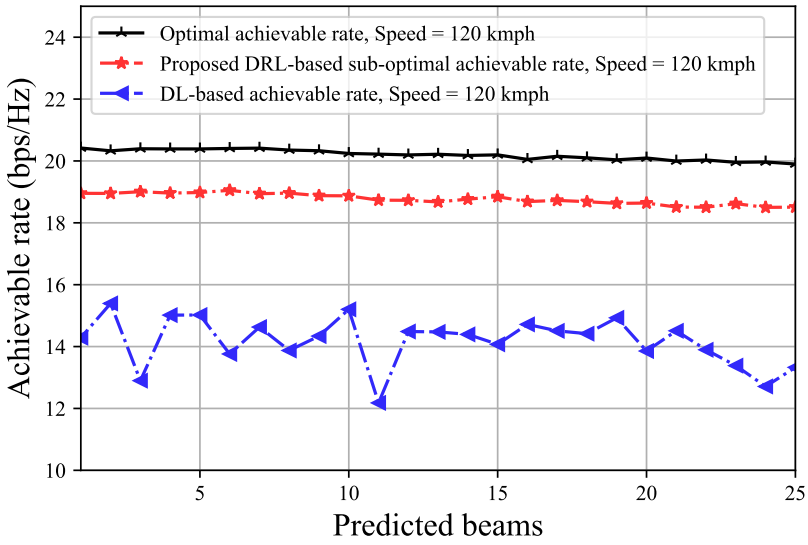

Figure 13: A comparison of effective achievable rate including overhead consideration (120 kmph).

dropped by 14.27% in terms of the average sum rate for our DRL method. Also, in Fig. 15, we have illustrated the performance of our DRL model at SNR of 10 dB. It is noticeable that, for another 20 dB of SNR drop, the performance declined by another 38.51%.

Furthermore, in Fig. 16, we have illustrated the convergence of our proposed algorithm. It can be seen that the achievable sum rate converges with a time step t in terms of loss. It is observable from the loss plot that, after approximately 3.2×10^6 iterations, our model converged successfully.

Overall, we can state that the performance of our model significantly rises as the SNR increases. Our proposed DRL architecture is robust and flexible in various conditions such as different SNRs and different velocities.

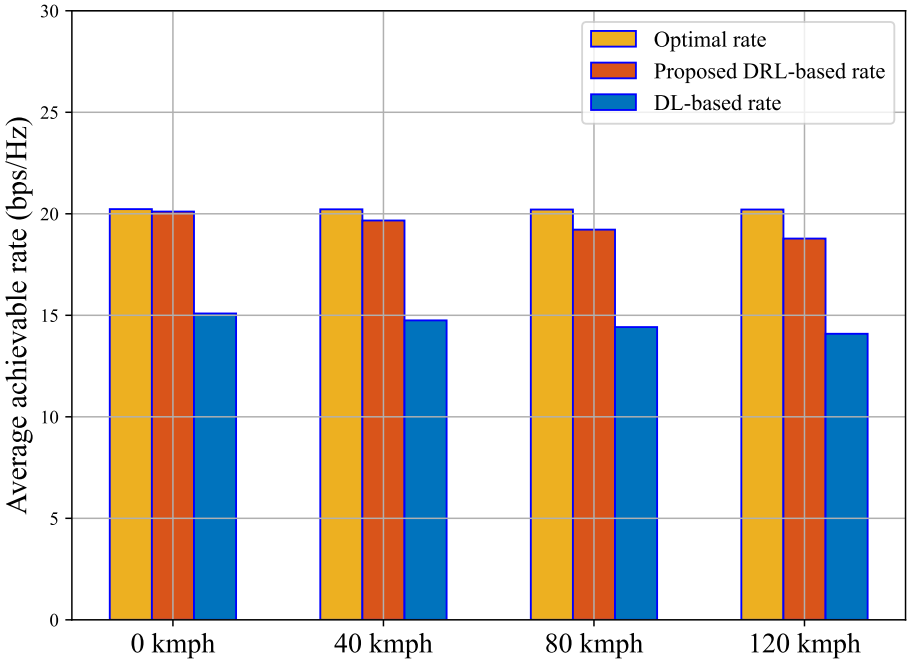

Figure 14: Average effective achievable rate at different speed.

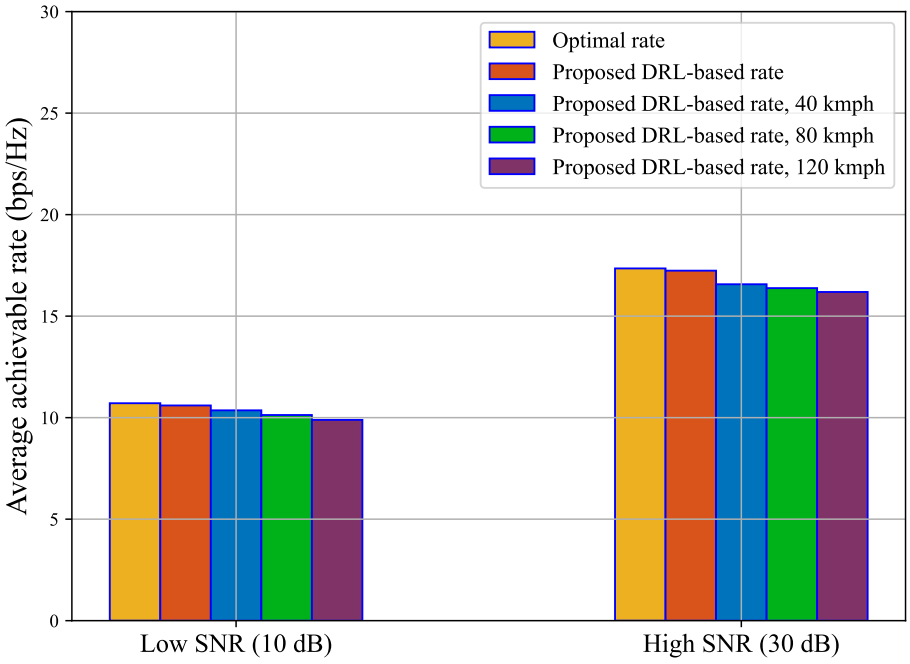

Figure 15: Effective achievable rate comparison at high and low SNR.

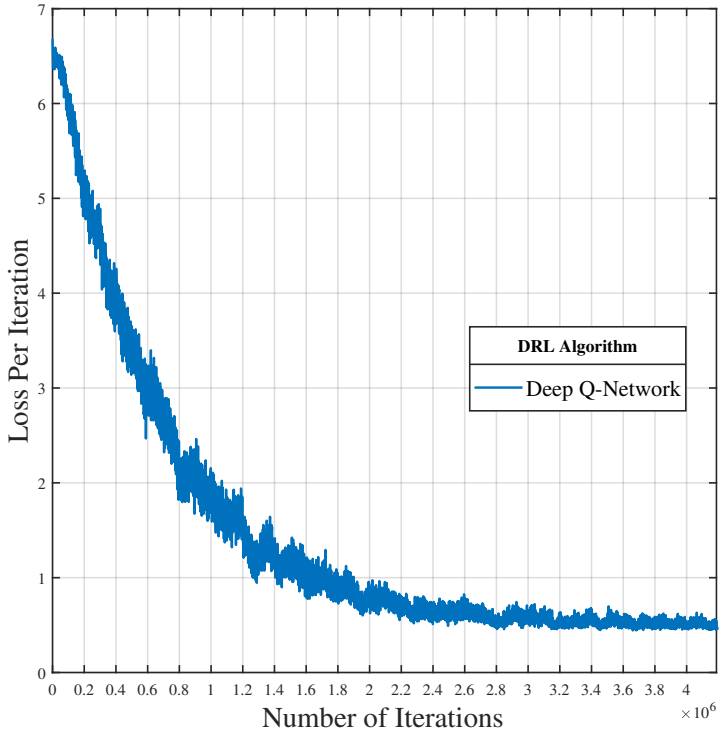

Figure 16: Loss convergence plot for the proposed DQN-based coordinated beamforming.

V. CONCLUSION

In this thesis, we propose a sub-optimal beam selection scheme with deep *Q*-learning which enables high mobile applications in mmWave massive MIMO systems. The key idea is to utilize the powerful exploration-exploitation strategy of DRL to derive the optimal beam selection policy which learns the mapping of omni-received uplink pilot and learn sub-optimal beam mapping. The presented solution requires small training overhead and beam overhead while ensuring very close achievable sum rate standards close to optimal performance. The proposed solution also ensures reliable coverage and shorter latency while serving beam towards the highly MS mmWave user end.

PUBLICATIONS

A. Journals

1. H. Islam, P. Tarafder **and** W. Choi, "LSTM-GRU model-based channel prediction for one-bit massive MIMO system," *Under Review in IEEE Transactions on Vehicular Technology*, 2022.

2. P. Tarafder **and** W. Choi, "MAC protocols for mmWave communication: A comparative survey," *Sensors*, **jourvol** 22, **number** 10, **page** 3853, 2022.

B. Conferences

1. P. Tarafder, M. Kang **and** W. Choi, "A comparative study on centralized MAC protocols for 60 GHz mmWave communications," **in** *2021 International Conference on Information and Communication Technology Convergence (ICTC)* IEEE, 2021, **pages** 888–892.

REFERENCES

[1] F. B. Saghezchi, J. Rodriguez, S. Mumtaz *et al.*, "Drivers for 5g: The 'pervasive connected world'," *Fundamentals of 5G Mobile Networks*, **pages** 1–27, 2015.

[2] T. Chen, M. Matinmikko, X. Chen, X. Zhou **and** P. Ahokangas, "Software defined mobile networks: Concept, survey, and research directions," *IEEE Communications Magazine*, **jourvol** 53, **number** 11, **pages** 126–133, 2015.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta **and** P. Popovski, "Five disruptive technology directions for 5g," *IEEE Communications Magazine*, **jourvol** 52, **number** 2, **pages** 74–80, 2014. DOI: 10.1109/MCOM.2014.6736746.

[4] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai **and** J. Rodriguez, "Millimeter-wave massive mimo communication for future wireless systems: A survey," *IEEE Communications Surveys Tutorials*, **jourvol** 20, **number** 2, **pages** 836–869, 2018. DOI: 10.1109/COMST.2017.2787460.

[5] Q. C. Li, H. Niu, A. T. Papathanassiou **and** G. Wu, "5g network capacity: Key elements and technologies," *IEEE Vehicular Technology Magazine*, **jourvol** 9, **number** 1, **pages** 71–78, 2014. DOI: 10.1109/MVT.2013.2295070.

[6] E. G. Larsson, O. Edfors, F. Tufvesson **and** T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, **jourvol** 52, **number** 2, **pages** 186–195, 2014. DOI: 10.1109/MCOM.2014.6736761.

[7] J. Hoydis, S. ten Brink **and** M. Debbah, "Massive mimo in the ul/dl of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, **journvol** 31, **number** 2, **pages** 160–171, 2013. DOI: 10.1109/JSAC.2013.130205.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, **journvol** 9, **number** 11, **pages** 3590–3600, 2010. DOI: 10.1109/TWC.2010.092810.091092.

[9] A. Ghosh, T. A. Thomas, M. C. Cudak *et al.*, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, **journvol** 32, **number** 6, **pages** 1152–1163, 2014. DOI: 10.1109/JSAC.2014.2328111.

[10] F. Rusek, D. Persson, B. K. Lau *et al.*, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, **journvol** 30, **number** 1, **pages** 40–60, 2013. DOI: 10.1109/MSP.2011.2178495.

[11] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat **and** R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Communications Magazine*, **journvol** 54, **number** 12, **pages** 160–167, 2016. DOI: 10.1109/MCOM.2016.1600071CM.

[12] P. Tarafder **and** W. Choi, "MAC protocols for mmWave communication: A comparative survey," *Sensors*, **journvol** 22, **number** 10, **page** 3853, 2022.

[13] X. Gao, L. Dai, Z. Chen, Z. Wang **and** Z. Zhang, "Near-optimal beam selection for beamspace mmwave massive mimo systems," *IEEE Communications Letters*, **jourvol** 20, **number** 5, **pages** 1054–1057, 2016. DOI: 10.1109/LCOMM.2016.2544937.

[14] R. Pal, K. V. Srinivas **and** A. K. Chaitanya, "A beam selection algorithm for millimeter-wave multi-user mimo systems," *IEEE Communications Letters*, **jourvol** 22, **number** 4, **pages** 852–855, 2018. DOI: 10.1109/LCOMM.2018.2803805.

[15] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu **and** D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, **jourvol** 6, **pages** 37 328–37 348, 2018. DOI: 10.1109/ACCESS.2018.2850226.

[16] Y. Zhang, B. Zhang, H. Wang, T. Zhang **and** Y. Qian, "Deep learning-based coordinated beamforming for massive mimo-enabled heterogeneous networks,"
in*2021 IEEE Global Communications Conference (GLOBECOM)* 2021, **pages** 1–6. DOI: 10.1109/GLOBECOM46510.2021.9685628.

[17] J. Tao, Q. Wang, S. Luo **and** J. Chen, "Constrained deep neural network based hybrid beamforming for millimeter wave massive mimo systems," **in***ICC 2019 - 2019 IEEE International Conference on Communications (ICC)* 2019, **pages** 1–6. DOI: 10.1109/ICC.2019.8761742.

[18] G. R. MacCartney, T. S. Rappaport **and** A. Ghosh, "Base station diversity propagation measurements at 73 ghz millimeter-wave for 5g coordinated multipoint (comp) analysis," **in***2017 IEEE Globecom Workshops (GC Wkshps)* 2017, **pages** 1–7. DOI: 10.1109/GLOCOMW.2017.8269045.

[19] D. Maamari, N. Devroye **and** D. Tuninetti, "Coverage in mmwave cellular networks with base station co-operation," *IEEE Transactions on Wireless Communications*, **jourvol** 15, **number** 4, **pages** 2981–2994, 2016. DOI: 10.1109/TWC.2016.2514347.

[20] A. K. Gupta, J. G. Andrews **and** R. W. Heath, "Macrodiversity in cellular networks with random blockages," *IEEE Transactions on Wireless Communications*, **jourvol** 17, **number** 2, **pages** 996–1010, 2017.

[21] J. Wang, Z. Lan, C. woo Pyo *et al.*, "Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems," *IEEE Journal on Selected Areas in Communications*, **jourvol** 27, **number** 8, **pages** 1390–1399, 2009. DOI: 10.1109/JSAC.2009.091009.

[22] X. Zhou, X. Zhang, C. Chen *et al.*, "Deep reinforcement learning coordinated receiver beamforming for millimeter-wave train-ground communications," *IEEE Transactions on Vehicular Technology*, **jourvol** 71, **number** 5, **pages** 5156–5171, 2022. DOI: 10.1109/TVT.2022.3153928.

[23] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh **and** A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE Journal of Selected Topics in Signal Processing*, **jourvol** 10, **number** 3, **pages** 436–453, 2016. DOI: 10.1109/JSTSP.2016.2523924.

[24] T. S. Rappaport, S. Sun, R. Mayzus *et al.*, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE Access*, **jourvol** 1, **pages** 335–349, 2013. DOI: 10.1109/ACCESS.2013.2260813.

[25] M. R. Akdeniz, Y. Liu, M. K. Samimi *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, **jourvol** 32, **number** 6, **pages** 1164–1179, 2014. DOI: 10.1109/JSAC.2014.2328154.

[26] M. K. Samimi **and** T. S. Rappaport, "Ultra-wideband statistical channel model for non line of sight millimeter-wave urban channels," **in***2014 IEEE Global Communications Conference* 2014, **pages** 3483–3489. DOI: 10.1109/GLOCOM.2014.7037347.

[27] P. Schniter **and** A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," **in***2014 48th Asilomar Conference on Signals, Systems and Computers* 2014, **pages** 273–277. DOI: 10.1109/ACSSC.2014.7094443.

[28] V. Va, J. Choi **and** R. W. Heath, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Transactions on Vehicular Technology*, **jourvol** 66, **number** 6, **pages** 5014–5029, 2017. DOI: 10.1109/TVT.2016.2622164.

[29] M. Ponsen, M. E. Taylor **and** K. Tuyls, "Abstraction and generalization in reinforcement learning: A summary and framework," **in***International Workshop on Adaptive and Learning Agents* Springer, 2009, **pages** 1–32.

[30] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang **and** L.-C. Wang, "Deep reinforcement learning for mobile 5g and beyond: Fundamentals, applications, and challenges," *IEEE Vehicular Technology Magazine*, **jourvol** 14, **number** 2, **pages** 44–52, 2019. DOI: 10.1109/MVT.2019.2903655.

[31] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare **and** J. Pineau. 2018.

[32] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu **and** F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, **jourvol** 234, **pages** 11–26, 2017.

[33] D. Bau, J.-Y. Zhu, H. Strobelt, A. Lapedriza, B. Zhou **and** A. Torralba, "Understanding the role of individual units in a deep neural network," *Proceedings of the National Academy of Sciences*, **jourvol** 117, **number** 48, **pages** 30 071–30 078, 2020.

[34] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

[35] V. Mnih, A. P. Badia, M. Mirza *et al.*, "Asynchronous methods for deep reinforcement learning," **in***International conference on machine learning* PMLR, 2016, **pages** 1928–1937.

[36] K. Arulkumaran, M. P. Deisenroth, M. Brundage **and** A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, **jourvol** 34, **number** 6, **pages** 26–38, 2017.

[37] T. P. Lillicrap, J. J. Hunt, A. Pritzel *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[38] M. Sana, A. De Domenico, W. Yu, Y. Lostanlen **and** E. Calvanese Strinati, "Multi-agent reinforcement learning for adaptive user association in dynamic mmwave networks," *IEEE Transactions on Wireless Communications*, **jourvol** 19, **number** 10, **pages** 6520–6534, 2020. DOI: 10.1109/TWC.2020.3003719.

[39] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," **in***Proc. of Information Theory and Applications Workshop (ITA)* San Diego, CA, 2019, **pages** 1–8.

[40] D. P. Kingma **and** J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[41] R. Girshick, "Fast r-cnn," **in***Proceedings of the IEEE international conference on computer vision* 2015, **pages** 1440–1448.

[42] *Smoothl1loss*. [Online]. Available: `https://pytorch.org/docs/stable/generated/torch.nn.SmoothL1Loss.html`.

[43] A. Paszke, S. Gross, F. Massa *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, **jourvol** 32, 2019.

[44] Remcom, *Wireless InSite*, `http://www.remcom.com/wireless-insite`.

ACKNOWLEDGEMENTS

I want to express my gratefulness to all the individuals who have supported me in the process of completing my Master's degree and research. Firstly, I really would like to take this opportunity to express my gratitude to Prof. Wooyeol Choi, my supervisor, for allowing me to pursue my Master's degree at Chosun University. His constant inspiration, support, and insightful recommendations have led and pushed me throughout my studies and research. His continuous supervision and direction have aided me in producing high-quality research. I will be eternally grateful to him for instilling in me the values of professionalism, organizational skills, and concentration. I also would like to convey my heartfelt gratitude to Prof. Seok Joo Shin and Prof. Moon Soo Kang, members of the thesis committee, for their constructive remarks and helpful ideas. Furthermore, I am glad for the opportunity to work in the Department of Computer Engineering at Chosun University with such a diversified batch of students, teachers, and staff. I want to thank Smart Networking Lab for giving me such an excellent opportunity and an environment to develop academically. My lab colleagues have been a source of moral and intellectual support for me. In addition, I want to express my appreciation to all of my Bangladeshi seniors and friends at Chosun University for their compassion and cooperation in making my life in South Korea easy and joyful. Lastly, I want to express my gratitude to my parents, relatives, and friends for their constant and unwavering support throughout my difficult times. It would have been difficult for me to do anything without their motivation and direction.