Machine learning for beam alignment in mmWave networks

Irched Chafaa

Thèse de doctorat

NNT : 2021UPASG044

université PARIS-SACLAY

Apprentissage automatique pour l'alignement des faisceaux dans les réseaux à onde millimétrique

Machine learning for beam alignment in mmWave networks

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 580, Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat: Réseaux, information et communications
Unité de recherche : Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.
Référent : CentraleSupélec

Thèse présentée et soutenue à Paris-Saclay, le 07/09/2021, par

Irched CHAFAA

Composition du Jury

Marceau COUPECHOUX Professeur, Télécom Paris, France	Président
Mehdi BENNIS Associate Professor, University of Oulu, Finland	Rapporteur & Examinateur
Laura COTTATELLUCCI Professeure, Friedrich-Alexander University of Erlangen-Nuremberg, Germany	Rapporteure & Examinatrice
Walid SAAD Professeur, Virginia Tech University, USA	Examinateur
Ghaya REKAYA-BEN OTHMAN Professeure, Télécom Paris, France	Examinatrice
Marwa CHAFII Maîtresse de conférences, ENSEA, France	Examinatrice
Romain NEGREL Maître de conférences, ESIEE Paris, France	Invité

Direction de la thèse

Mérouane DEBBAH Professeur, Centre de recherche Lagrange en mathématiques et calculs, France	Directeur de thèse
E.Veronica BELMEGA Maîtresse de conférences, ENSEA, France	Co-encadrante

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Title: Machine learning for beam alignment in mmWave networks

Keywords: mmWave, beam alignment, multi-armed bandits, unsupervised deep learning, federated learning

Abstract: To cope with the ever increasing mobile data traffic, an envisioned solution for future wireless networks is to exploit the large available spectrum in the millimeter wave (mmWave) band. However, communicating at these high frequencies is very challenging as the transmitted signal suffers from strong attenuation, which leads to a limited propagation range and few multipath components (sparse mmWave channels). Hence, highly-directional beams have to be employed to focus the signal energy towards the intended user and compensate all those losses. Such beams need to be steered appropriately to guarantee a reliable communication link. This represents the so called beam alignment problem where the beams of the transmitter and the receiver need to be constantly aligned. Moreover, beam alignment policies need to support devices mobility and the unpredicted dynamics of the network, which result in significant signaling and training overhead affecting the overall performance.

In the first part of the thesis, we formulate the beam alignment problem via the adversarial multi-armed bandit framework, which copes with arbitrary network dynamics including non-stationary or adversarial components. We propose online and adaptive beam alignment policies relying only on one-bit feedback to steer the beams of both nodes of the communication link in a distributed manner. Building on the well known exponential weights algorithm (EXP3) and by exploiting the sparse nature of mmWave channels, we propose a modified policy (MEXP3), which comes with optimal theoretical guarantees in terms of asymptotic regret. Moreover, for finite horizons, our regret upper-bound is tighter than that of the original EXP3 suggesting better performance in practice. We then introduce an additional modification that accounts for the temporal correlation between successive beams and propose another beam alignment policy (NBT-MEXP3).

In the second part of the thesis, deep learning tools are investigated to select mmWave beams in an access point – user link. We leverage unsupervised deep learning to exploit the channel knowledge at sub-6 GHz and predict beamforming vectors in the mmWave band; this complex channel-beam mapping is learned via data issued from the DeepMIMO dataset and lacking the ground truth. We also show how to choose an optimal size of our neural network depending on the number of transmit and receive antennas at the access point. Furthermore, we investigate the impact of training data availability and introduce a federated learning (FL) approach to predict the beams of multiple links by sharing only the parameters of the locally trained neural networks (and not the local data). We investigate both synchronous and asynchronous FL methods. Our numerical simulations show the high potential of our approach, especially when the local available data is scarce or imperfect (noisy). At last, we compare our proposed deep learning methods with reinforcement learning methods derived in the first part. Simulations show that choosing an appropriate beam steering method depends on the target application and is a tradeoff between rate performance and computational complexity.

Titre: Apprentissage automatique pour l'alignement des faisceaux dans les réseaux à onde millimétrique

Mots clés: ondes millimétriques, alignement des faisceaux, bandits manchots, apprentissage profond non supervisé, apprentissage fédéré

Résumé: Pour faire face à la croissance exponentielle du trafic des données mobiles, une solution possible est d'exploiter les larges bandes spectrales disponibles dans la partie millimétrique du spectre électromagnétique. Cependant, le signal transmis est fortement atténué, impliquant une portée de propagation limitée et un faible nombre des trajets de propagation (canal parcimonieux). Par conséquent, des faisceaux directifs doivent être utilisés pour focaliser l'énergie du signal transmis vers son utilisateur et compenser les pertes de propagation. Ces faisceaux ont besoin d'être dirigés convenablement pour garantir la fiabilité du lien de communication. Ceci représente le problème d'alignement des faisceaux pour les systèmes de communication à onde millimétrique. En effet, les faisceaux de l'émetteur et du récepteur doivent être constamment ajustés et alignés pour combattre les conditions de propagation difficiles de la bande millimétrique. De plus, les techniques d'alignement des faisceaux doivent prendre en compte la mobilité des utilisateurs et la dynamique imprévisible du réseau. Ceci mène à un fort coût de signalisation et d'entraînement qui impacte les performances des réseaux.

Dans la première partie de cette thèse, nous reformulons le problème d'alignement des faisceaux en utilisant les bandits manchots (ou multi-armed bandits), pertinents dans le cas d'une dynamique du réseau imprévisibles et arbitraire (non-stationnaire ou même antagoniste). Nous proposons des méthodes en ligne et adaptatives pour aligner indépendamment les faisceaux des deux nœuds du lien de communication en utilisant seulement un seul bit de feedback. En se basant sur l'algorithme des poids exponentiels (EXP3) et le caractère parcimonieux du canal à onde millimétrique, nous proposons une version modifiée de l'algorithme original (MEXP3) avec des garanties théoriques en fonction du regret asymptotique. En outre, pour un horizon du temps fini, notre borne supérieure du regret est plus serrée que celle de l'algorithme EXP3, indiquant une meilleure performance en pratique. Nous introduisons également une deuxième modification qui utilise les corrélations temporelles entre des choix successifs des faisceaux dans une nouvelle technique d'alignement des faisceaux (NBT-MEXP3).

Dans la deuxième partie de cette thèse, des outils de l'apprentissage profond sont examinés pour choisir des faisceaux dans un lien point d'accès – utilisateur. Nous exploitons l'apprentissage profond non supervisé pour utiliser l'information des canaux au-dessous de 6 GHz afin de prédire des faisceaux dans la bande millimétrique; cette fonction canal-faisceau complexe est apprise en utilisant des données non-annotés du dataset DeepMIMO. Nous discutons aussi le choix d'une taille optimale pour le réseau de neurones en fonction du nombre des antennes de transmission et de réception au point d'accès. De plus, nous étudions l'impact de la disponibilité des données d'entraînement et introduisons une approche basée sur l'apprentissage fédéré pour prédire des faisceaux dans un réseau à plusieurs liens en partageant uniquement les paramètres des réseaux de neurones entrainés localement (et non pas les données locales). Nous envisageons les méthodes synchrones et asynchrones de l'approche par apprentissage fédéré. Nos résultats numériques montrent le potentiel de notre approche particulièrement au cas où les données d'entrainement sont peu abondantes ou imparfaites (bruitées). Enfin, nous comparons nos méthodes basées sur l'apprentissage profond avec celles de la première partie. Les simulations montrent que le choix d'une méthode convenable pour aligner les faisceaux dépend de la nature de l'application et présente un compromis entre le débit obtenu et la complexité du calcul.

Maison du doctorat de l'Université Paris-Saclay
2ème étage aile ouest,Ecole normale supérieure Paris-Saclay
4 avenue des sciences,
91190 Gif sur Yvette, France

Acknowledgement

This work was carried out at the ETIS/ENSEA UMR 8051 Laboratory in CY Cergy Paris University. For this, I would like to thank the laboratory for welcoming me to the ICI team. More specifically, I would like to express my sincere gratitude and heartfelt thanks to my advisor Associate professor/HDR **E.V Belmega** for her outstanding advising approach, reliable time availability and also her continuous support, via her kindness and encouraging words, throughout the different milestones of my doctoral thesis. Your work ethics and methodology will serve as a model for my future career. My sincere thanks also go to my thesis director Pr. **Mérouane Debbah** for his valuable advice, insightful directions and making the time for me despite his busy schedule.

My respectful gratitude goes to all members of the jury for kindly accepting to participate in my PhD defense.

I would like to thank my lab-mates, **Anastacia**, **Habiba**, **Tarek** and **Amine** for the stimulating work environment and the fun moments, which helped to ease the stressful periods of my thesis. Special thanks go to **Tarek** for his technical assistance during times of need.

Special words of gratitude and love go to my family for their unconditional support and to my friends from outside of the academic world.

Finally, I want to thank every person who has contributed to enrich my doctoral journey.

Contents

List of Figures

List of Tables

Abbreviations

2D Two Dimensional

3D Three Dimensional

3GPP 3rd Generation Partnership Project

5G Fifth Generation

ABF Analog Beamforming

ACK Acknowledgement

AFL Asynchronous Federated Learning

AoA Angle of Arrival

AoD Angle of Departure

AP Access Point

BA Beam Alignment

CDF Cumulative Distribution Function

CL Centralized Learning

CSI Channel State Information

DAC Digital to Analog Converter

DBF Digital Beamforming

DL Deep Learning

DNN Deep Neural Network

DRL Deep Reinforcement Learning

EXP3 Exponential Weights for Exploration and Exploitation

FD Full Duplex

FL Federated Learning

HD Half Duplex

IEEE Institute of Electrical and Electronics Engineers

IL Individual Learning

ITU International Telecommunication Union

LOS Line of Sight

MAB Multi Armed Bandit

MEXP3 Modified Exponential Weights for Exploration and Exploitation

MIMO Multiple Input Multiple Output

ML Machine Learning

mmWave millimeter wave

NACK Negative Acknowledgment

NBT-MEXP3 Nearest neighbour-aided Beam Tracking Modified Exponential Weights for Exploration and Exploitation

NLOS Non Line of Sight

NOMA Non Orthogonal Multiple Access

OFDM Orthogonal Frequency Division Multiplexing

RF Radio Frequency

RL Reinforcement Learning

Rx Receiver

SFL Synchronous Federated Learning

SINR Signal to Interference plus Noise Ratio

SI Self Interference

SNR Signal to Noise Ratio

Tx Transmitter

UCB Upper Confidence Bound

ULA Uniform Linear Array

1

Introduction

Because of the congestion in the sub-6 GHz microwave spectrum, the millimeter wave (mmWave) band, which typically refers to frequencies higher than 20 GHz, has been considered as a promising solution for future wireless networks [8, 9] to provide more available bandwidth, and thus achieve the high data rates required by data-hungry applications such as transmitting ultra high definition videos.

Propagation at mmWave frequencies is mainly characterized by a significant attenuation of the transmitted signal due to the high free-space path loss [10] and additional losses when the signal penetrates other objects or gets absorbed by particles in the wireless environment (the atmosphere) [11, 12]. This suggests the use of highly directional beams by using large antenna arrays jointly with beamforming techniques [13, 14] to compensate for the high propagation loss. Luckily, the small wavelength at the mmWave band allows to place a large number of antenna elements in relatively small size arrays which yields a large beamforming gain [15] by focusing the signal's power toward the intended user's equipment. The specific characteristics of mmWave systems lead to a new set of challenges facing their practical deployment as we describe in the next chapter. In this thesis, we focus on one fundamental challenge, which is the steering of the directional beams in dynamic mmWave networks.

Indeed, this type of directional communications lead to the so called beam alignment problem where the beams of the transmitter and the receiver need to be constantly aligned and well steered, before data transmission, to guarantee a reliable communication link and achieve the desired communication performance. The beam alignment problem is also subject to a large training overhead that grows fast with the resolution of the beamforming codebook and the large number of antennas.

Moreover, the beam alignment policies need to support user mobility and to cope with unpredictable and possibly non-stochastic variations of the wireless network (e.g., caused by the users' behavior and intermittent connectivity). Under such time-varying conditions, adjusting the beam-directions and identifying optimal beamforming vectors implies significant additional signaling and training overhead, which affects the overall performance. Hence, online beam alignment policies capable of adapting on-the-fly to such changes become necessary to enable future mobile mmWave applications such as virtual reality headsets, autonomous vehicles, etc.

1.1 State of the art

Given its importance, the beam alignment problem has been addressed extensively in the literature. Early approaches can be divided into two main categories: beam training and compressed sensing [16]. The first approach consists of training from a set of candidate beamforming vectors through exhaustive search [17] or adaptive hierarchical search [18, 19] to identify the best beam-direction in terms of a given metric (e.g., signal-to-noise ratio (SNR)). For instance, in the IEEE 802.11ad standard [19] the beam search is done with wide beams whose widths are reduced progressively following a multi-level hierarchical scheme. The main limitation of such approaches resides in the large training feedback and coordination overhead making it unsuitable for mobile mmWave applications.

The compressed sensing methods [20, 21] estimate the channel parameters, such as the propagation path gains and angles of arrival/departure, and use them to construct beamforming vectors to steer the beams. These methods exploit the mmWave channel's sparsity, which reduces the training delay compared to the first category. However, they scale poorly with the number of antennas and require a precise prior knowledge of the channel structure and sparsity [16]. Moreover, these methods rely on strong assumptions regarding the temporal variation of the channel (either static or stochastic) during the estimation phase, which poses several issues when the channel is highly dynamic and possibly non-stochastic.

Another common disadvantage of classical approaches is that they do not exploit the results of the past beam alignment experiences, which are related to the wireless environment (receiver's location, environment geometry,...). The two approaches are unable to exploit the mapping function relating the wireless setup to the beam alignment results because it involves many environment parameters and changes from one setup to another which makes it non-trivial to find a closed-form expression for this mapping function [16].

More recently, machine learning (ML) tools have been considered because of their ability to learn unknown models and solve hard online optimization problems. Machine learning allows to learn the implicit mapping function relating the beam alignment results to the environment parameters. This allows to design beam alignment strategies in dynamic wireless networks that rely on less stringent assumptions and are more data-oriented. In this context, two main ML frameworks have been exploited to steer the beams in mmWave networks: reinforcement learning and deep learning.

Reinforcement learning (RL) [22] relies on online interactions between the learning agent and its wireless environment, which results in a feedback used to determine the beam direction on-the-fly. The authors in [23] used Q-learning to select beams that meet quality of service requirements at the user end. Another beam alignment policy based on Q-learning is proposed in [24] to optimize the network capacity. In [25], deep reinforcement learning is leveraged to perform beam alignment for multi-user mmWave systems. In this thesis, we focus on the application of a particular case of reinforcement learning: multi-armed bandits (MAB) for the beam alignment, which has received a significant interest in the recent literature.

In the MAB framework, the beam alignment is cast into a sequential decision-making problem, in which the devices (e.g., a central node or the transmitter/receiver) choose at each step an arm (i.e a beam-direction), out of a finite set of choices, and then observe a reward (e.g., the resulting SNR at the receiver). The devices then learn the best beam-direction by jointly *exploiting* the observed past rewards and *exploring* new beam-directions. The authors in [26] proposed the unimodal beam alignment algorithm that restricts the search set of the best directions by using the correlation between consecutive beams and the unimodality of the power of the received signal. An online beam alignment algorithm for mmWave vehicular communications was introduced in [27], which uses the vehicle's direction of arrival as a contextual information. In [28], another beam alignment policy was investigated, which requires the perfect knowledge of the data rates for all chosen beam-directions. The policy in [29] incorporates the receiver's location as an out-of-band additional information to improve the beam alignment. In [30], the proposed policy aims at reducing the beam alignment delay by exploiting previously acquired knowledge about the channel. Another algorithm, based on MABs, is also proposed in [31] for beam alignment and tracking.

All the existing MAB-based policies cited above depend on a central authority, which first chooses jointly the best pair of beam-directions of the transmitter and the receiver, and then feedbacks the result to both devices, resulting in a high signaling overhead. Moreover, these approaches are deterministic and exploit the so-called upper confidence bound (UCB) from the stochastic bandits framework [32]. This implies that they are relevant only in stochastic and stationary wireless environments and can not account for other possibly non-stationary components such as the unpredictable behaviour and connectivity patterns of interfering devices.

In general, RL policies don not require an offline training phase and have a relatively low computational complexity. However, they require a certain exploration time to identify good beam-directions (at the beginning of the communication or after significant channel variations), which may affect their performance in latency-sensitive applications.

Building on the wide success of neural networks, data-driven approaches have recently found their way in wireless communications with supervised and unsupervised deep learning (DL) [33]. Several existing works exploit neural networks as universal approximators to learn the relation between the wireless environment and optimal beam-directions. In [16], a set of access points coordinate to serve one user by learning an appropriate beam via supervised DL, which exploits an omnidirectional mmWave uplink signal. The centralized nature of this approach implies heavy signaling between the different access points and a central entity, which increases the training overhead. Moreover, the omnidirectional mmWave uplink signal transmitted by a single antenna can be very limited in range and power. In [6, 34], a neural network is trained to map sub-6 GHz channel information to beamforming vectors at mmWave for a single transmitter-receiver link. The proposed approach is based on classification, in which the beamforming vectors are selected from a predefined discrete codebook, yielding a sub-optimal solution. The authors in [35] used the federated learning (FL) framework to map the mmWave channels into analog beamformers in a multi-user downlink network. Therefore, the proposed policy requires the knowledge of the mmWave channel

matrices, which are more difficult to estimate and require larger training overhead compared to sub-6 GHz channels. A deep neural network is trained in [36] following the unsupervised learning to choose beamforming vectors for mmWave communications. The work in [37] proposes a beamforming design method using unsupervised learning to maximize the network's rate. A recent work [38] employs a deep neural network to select jointly a mmWave base station and beam providing the best communication performance in heterogeneous cellular networks.

Although promising, the DL policies rely on the availability of sufficient amount of relevant training data and an optimal design of the neural network architecture. Acquiring training data is currently not trivial, being both expensive and time-consuming; not to mention the privacy and security issues it may raise.

1.2 Contributions and manuscript organization

The contributions of this thesis can be split into two main parts. In the first part of the thesis, we focus squarely on *distributed* beam alignment policies that do not require the existence of a central node nor rely on any assumptions regarding the network dynamics, as opposed to existing work [26, 27, 28, 29, 30]. For this, we build on the exponential weights algorithm for exploration and exploitation (EXP3) [39] to define novel beam alignment policies capable to adapt to such *arbitrary and unpredictable environments*. To the best of our knowledge, our work is the first to exploit exponential weights for beam alignment in mmWave networks. Compared with traditional schemes, our policies aim at learning the best beam-directions in an adaptive, online manner without relying on pre-deployed training every time the channel changes. Indeed, it is possible to simultaneously transmit data while tracking good beams from the beginning of the transmission. Of course, this comes at a cost in terms of high outage levels in the early stages of the learning process. The main advantage of our adaptive policies is that this cost happens only, in the beginning of the transmission or when the channel undergoes significant changes, and that they do not require dedicated training every channel coherence time (nor to optimize the training phase duration, which has a crucial impact on the data transmission efficiency). Finally, our online beam alignment policies do not require the perfect knowledge of the channel and relies solely on one-bit of feedback (ACK/NACK type) that basically captures whether the target SNR has been reached at the receiver.

In the second part of the manuscript, we propose an unsupervised deep learning method to map uplink sub-6 GHz channels into downlink mmWave beamforming vectors. Our proposed neural network takes the sub-6 GHz uplink channels as input and outputs directly the corresponding mmWave beamforming vectors, by exploiting the patterns and features in the input channels (channel statistics and environment characteristics). Unlike supervised DL methods, our approach does not require ground-truth data and, thus, does not require the computation of the optimal downlink beam for each training sample. As opposed to existing work [6, 34], we formulate the channel-beam mapping problem as a regression and not a classification, which implies that the predicted beams

have a continuous (and, hence higher) angular resolution and can overcome the sub-optimality caused by a quantized beamforming codebook. Compared to [6], we design a simpler neural network architecture, with less training parameters, optimized based on a communication-tailored loss function such that the predicted beamforming vectors maximize the communication rate directly. Hence, our approach combines both model (via the communication rate function) and data-oriented ingredients (via the neural network) and takes advantage of both worlds. As opposed to [35], our method requires only the available channel state information (CSI) at sub-6 GHz to predict the mmWave beams, which is much easier to acquire than the mmWave CSI.

Moreover, we study the more general mmWave network composed of multiple access point – user links. We propose a federated learning scheme to predict the mmWave beamforming vectors locally, at each access point, to preserve their data privacy and avoid the heavy signaling of centralized learning. We investigate both synchronous and asynchronous uploading of the local models. The asynchronous approach allows to reduce the uploading communication cost during the training phase and contributes to power savings as only one access point trains its neural network at each iteration. These advantages come at the cost of a certain performance degradation in terms of rate, which illustrates the tradeoff between *training cost vs. rate performance*. Finally, we compare classical RL methods with DL ones in terms of rate performance and computational complexity.

This thesis manuscript is organized as follows: In chapter 2, we introduce the mmWave channels and describe their propagation characteristics and benefits for wireless communications. Then, the main challenges of such high frequencies are explained. In this chapter, we also present briefly the three main paradigms of machine learning with a particular focus on their advantages and drawbacks for resource management problems. Finally, a comparison between classic reinforcement learning and deep learning is provided.

In chapter 3, we present the first part of the contributions of the thesis. We model the beam alignment in arbitrarily dynamic mmWave networks as an adversarial MAB problem, in which the transmitter and the receiver select their own beam-directions individually while relying only on a 1-bit of feedback. We investigate the optimal size of the beamforming codebook to reduce the exploration cost. Then, we propose a novel modified exponential weights (MEXP3) algorithm that exploits the sparse nature of mmWave channels. We then prove that the new MEXP3 has the no-regret property and that the average regret decays to zero. We introduce a further reward modification and propose the nearest neighbor-aided beam tracking modified exponential-weight algorithm (NBT-MEXP3), which exploits the temporal correlation between consecutively aligned beams to restrict the beam search to the neighborhood of a previously found good beam. At last, numerical results are provided to show that the proposed policies offer better practical performance compared to other existing policies especially in terms of outage and throughput for both single and multi-path channels.

In chapter 4, the second part of the contributions is presented. We propose to leverage unsupervised deep learning to steer the beams of the access points towards their users in the mmWave band (downlink) using the sub-6 GHz (uplink) channel estimations. We show via numerical simulations how to tune the optimal size of our neural network in

function of the number of transmit and receive antennas. After this, a distributed beam steering scheme based on federated learning is introduced and evaluated in a multi-link mmWave network. At the end of this chapter, we provide a detailed comparison between the proposed deep learning method and the multi-armed bandit algorithms from the previous chapter.

Finally, chapter 5 provides conclusions of the research conducted during this thesis and different perspectives for future work.

1.3 Publications and posters

This thesis has lead to the following publications and poster presentations:

Journals

[J2sub] **I. Chafaa**, R. Negrel, E.V. Belmega, and M. Debbah, "Unsupervised deep learning for mmWave beam steering exploiting sub-6 GHz channels", submitted to IEEE Trans. on Wireless Commun, Apr. 2021.

[J1] **I. Chafaa**, E.V. Belmega, and M. Debbah, "One-bit Feedback exponential learning for beam alignment in mobile mmWave", IEEE Access, pp.194575-194589, Oct. 2020.

International conferences

[C3] **I. Chafaa**, R. Negrel, E.V. Belmega, and M. Debbah, "Federated channel-beam mapping: from sub-6GHz to mmWave", IEEE WCNC, Workshop on Distributed Machine Learning for Future Communications and Networking, Mar. 2021.

[C2] **I. Chafaa**, E.V. Belmega, and M. Debbah, "Exploiting Channel Sparsity for Beam Alignment in mmWave Systems via Exponential Learning", IEEE ICC, Open Workshop on Machine Learning for Communications (ML4COM), Dublin, Ireland, Jun. 2020.

[C1] **I. Chafaa**, E. V. Belmega, and M. Debbah, "Adversarial Multi-armed Bandit for mmWave Beam Alignment with One-bit Feedback", ACM ValueTools 2019, Palma de Mallorca, Spain, Mar. 2019.

Posters

[P3] **I. Chafaa**, "Adversarial Multi-armed Bandits for mmWave Beam Alignment with One-Bit Feedback", IEEE Training School: Machine learning for communications, Paris, France, 2019.

[P2] **I. Chafaa**, "Online Exponential Learning for Beam-Alignment in dynamic millimeter wave Systems", Meet-up Doctorants & Industrie, Paris, France 2019.

[P1] **I. Chafaa**, "Adversarial Multi-armed Bandits for mmWave Beam Alignment with One-Bit Feedback", Journée des doctorants, ETIS, Cergy-Pontoise, France 2019.

2

mmWave communications and machine learning

2.1 Introduction

Under the pressure of the exponential growth of mobile data traffic, the wireless communication infrastructure needs to migrate to higher frequencies with larger available bandwidth. The mmWave band is a strong candidate to increase the spectral efficiency of the current systems. In this chapter, we provide an overview about this band, its propagation characteristics and advantages for wireless networks. Then, we identify the main challenges facing the deployment of fully wireless networks at the mmWave band. Building on the recent success of machine learning in wireless networks, we present the main paradigms of machine learning and discuss their advantages, drawbacks and applications for the resource allocation problem in mmWave networks.

2.2 Characteristics and challenges of mmWave communications

In this section, we describe the propagation properties of the mmWave signal. We also describe the main challenges of mmWave systems including the beam alignment problem, which represents the focus of this thesis.

2.2.1 The mmWave band

The current communication industry has focused its wireless applications on the electromagnetic spectrum which is below 6 GHz. However, there is an under-utilized spectrum beyond the traditional sub-6 GHz band, including the mmWave frequency range, with promising large available bandwidths. The mmWave band refers to frequencies higher than 20 GHz and up to 300 GHz.

The International Telecommunication Union (ITU) and the 3rd Generation Partnership Project (3GPP) have conducted measurements campaigns to identify the best candidate frequencies for mmWave wireless systems and formulate their corresponding standards. The first step was focused on mmWave frequencies under 40 GHz. Since 2018, the second step involved frequencies up to 100 GHz. As a result of the

research work of different standard bodies, a set of three mmWave frequency bands surfaced as strong candidates for use in wireless networks: 28 GHz, 39 GHz and 72 GHz [40]. These bands suffer less from atmospheric absorption, and thus become more suitable for long distance communications. Moreover, measurement campaigns [7, 41, 42, 43, 44, 45, 4, 46] demonstrated their feasibility (in terms of range, path loss exponent, penetration loss, etc) in multi-path urban environments for line-of-sight (LOS) and non-line-of-sight (NLOS) communications. The 60 GHz band is not considered particularly for outdoor wireless networks given the severe atmospheric absorption and the fact that it has already been commercialized for an indoor application, which is the WiGig (wireless gigabit): a 60 GHz WiFi (wireless fidelity) [47].

The 28 GHz band This band [26.5GHz − 29.5GHz] [40] has attracted the interest of early field trials from several service providers since it represents the lowest available mmWave frequencies, yielding more favorable propagation conditions and lower device complexity. In fact, it remains the most tested mmWave band so far. It can offer up to 850 MHz of contiguous bandwidth. Extensive measurement campaigns in New York city [7, 41] confirmed the possible exploitation of this band for wireless networks in both LOS and NLOS scenarios. Test results brought more understanding to the propagation characteristics of the 28 GHz band in urban environments mainly regarding the path loss models, reflection properties and penetration loss for different construction materials. In 2015, Samsung measurements also demonstrated that a communication link can be established for over 200 meters of distance at this band [40] in urban environments.

The 39 GHz band This band ranges from 37 GHz to 40 GHz in the mmWave spectrum [40]. It can provide up to 1.6 GHz of contiguous bandwidth to support high capacity services. Its limited propagation allows for dense frequency reuse in urban areas and indoor environments as reported in [42, 43, 44, 45]. The 39 GHz band is also used for fixed point-to-point backhaul links to meet the increasing demand for backhaul capacity [48].

The 73 GHz band Also known as the E-band [48], it occupies the range [71GHz − 76GHz] of the mmWave spectrum. One particularity of this band, compared to the previous ones, is its larger 2 GHz contiguous bandwidth available for mobile communications, which gives it an advantage regarding the achievable data throughput. Despite the high frequency (higher attenuation), several measurement campaigns [4, 46] showed the viability of this band for high data rates and short range links (less than 200 m). One practical example is the prototype developed by Huawei and Deutsche Telekom, at the 2016 mobile world congress, of a multi-user multiple-input and multiple-output (MIMO) system operating at 73 GHz, providing potentially more than 20 Gbit/s throughput for individual users [40].

Even though different mmWave bands may differ in some propagation characteristics or in the amount of available bandwidth, they all share the following common key advantages for future wireless networks:

- A large available spectral bandwidth, as opposed to the congested sub-6 GHz spectrum, to support a wide range of wireless applications and services and meet the continuously growing demand of mobile traffic. Indeed, mmWave technologies represent a solid driver to achieve the ambitious requirements of future wireless networks, which can not be possible with the sub-6 GHz band, such as guaranteeing a uniform minimum data rate of 1 Gbit/s for every user, supporting rigorous latency requirements (at most 1 ms), supporting high mobility applications, etc.
- The high loss encountered by the transmitted signal, in this band, can be of advantage to mmWave networks. First, the transmitted power can be confined within both a limited angular domain and propagation range, when using narrow beams, which can contribute to limiting the interference between different mmWave links operating in the same frequency band. Moreover, indoor mmWave networks can incur limited interference due to the high penetration loss of mmWave signals by different construction materials. Also, the limited propagation range allows more frequency reuse for short distance communications and even enhance the security aspect of the transmitted signal by limiting its propagation distance.
- The small wavelength in the mmWave band allows to incorporate a large number of radiating elements in small size antenna arrays, which are essential for beamforming techniques to provide high power gains and compensate for the significant path loss at this band.

2.2.2 Wireless propagation characteristics

A fundamental issue is to understand the propagation characteristics of millimeter waves to be able to deploy new mmWave wireless networks. Understanding the mmWave channel would allow to conceive appropriate air interface and multiple access techniques, new system architectures and signal processing methods dedicated to this particular band. The results of the extensive measurement campaigns conducted during these last years in different environments (urban and rural) [7, 41, 42, 43, 44, 45, 4, 46] helped to understand better the behaviour of a transmitted mmWave signal and demonstrated the feasibility of its deployment in wireless systems. In the following, we regroup the different propagation characteristics into two main properties and emphasize on the viability of the mmWave band for future wireless networks despite its generally disadvantageous propagation conditions compared to the traditional sub-6 GHz band.

Limited range

The propagation of millimeter waves is highly influenced by different meteorological conditions which imposes limitations on the range of mmWave systems. The transmitted signal is mainly affected by free space path loss, some gases in the atmosphere such as oxygen, penetration loss and precipitation (rain, snow, fog) by reducing its power through absorption and dispersion phenomena [49].

Free space path loss One of the major early concerns of transmitting at extremely high frequencies was the signal's power attenuation due to propagation in free space. According to Friis equation [10], the received power of a transmitted signal is inversely proportional to the square of the carrier frequency, which increases the free space path loss at the mmWave band.

Gas absorption Millimeter waves get absorbed by atmospheric gases particularly oxygen and water vapor [1]. This phenomenon depends highly on the signal's carrier frequency, which means that the transmitted power attenuation is significant only for some frequencies as shown in figure 2.1. For instance, one high absorption by oxygen occurs at the $57 - 64$ GHz band (15 dB/km). These absorption peaks lead to an even more limited range, but outside of these bands the propagation is not highly affected. For example, there is an attenuation of only $0,02$ dB/km and $0,09$ dB/km by oxygen and water vapor respectively for the 28 GHz band, which motivates its popularity for future outdoor mmWave networks.

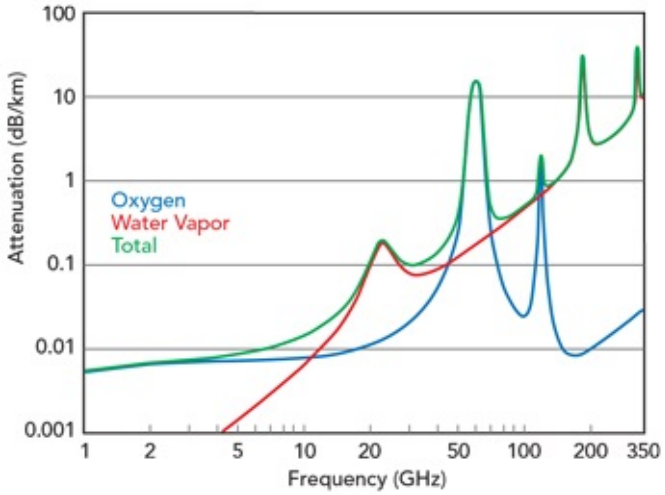

Figure 2.1: Atmospheric absorption of electromagnetic waves [1]

Penetration loss Another reason for the limited range of millimeter waves is the high loss incurred by the signal when penetrating other objects and materials. To understand the mmWave propagation in urban environments, penetration loss measures have been conducted in New York city for the 28 GHz band [7]. The main results are summarized in table 2.1.

The typical outdoor surfaces materials present high penetration loss of $40,1$ dB and $28,3$ dB for tinted glass and brick respectively, which shows how difficult it can be for millimeter waves to penetrate buildings. This needs to be taken in consideration when elaborating the link power budget of a mmWave system to compensate for the power loss of the signal. For indoor materials, the penetration losses are relatively less significant.

environment	material	thickness (cm)	penetration loss (dB)
outdoor	tinted glass	3.8	40.1
	brick	185.4	28.3
indoor	clear glass	<1.3	3.6
	wall	38.1	6.8

Table 2.1: Penetration loss at 28 GHz [7]

Precipitation loss The rain represents another concern for mmWave communications since the raindrops are practically of the same size as the wavelength, at this band, which causes the dispersion of the signal. The rain attenuation depends on the raindrops characteristics such as their size and fall rate. Fig 2.2 [2] shows this attenuation for different rainfall rates. For a high rainfall rate of 25 mm/h, the signal is attenuated by 7 dB/Km at 28 GHz and 10 dB/Km at 73 GHz. Given the fact that a mmWave coverage area is expected to be limited (100 m to 200 m), these losses become less significant. For example, the rain attenuation at the 28 GHz band can reach at most $1,4$ dB. Snow and fog also contribute to the attenuation of the mmWave signal.

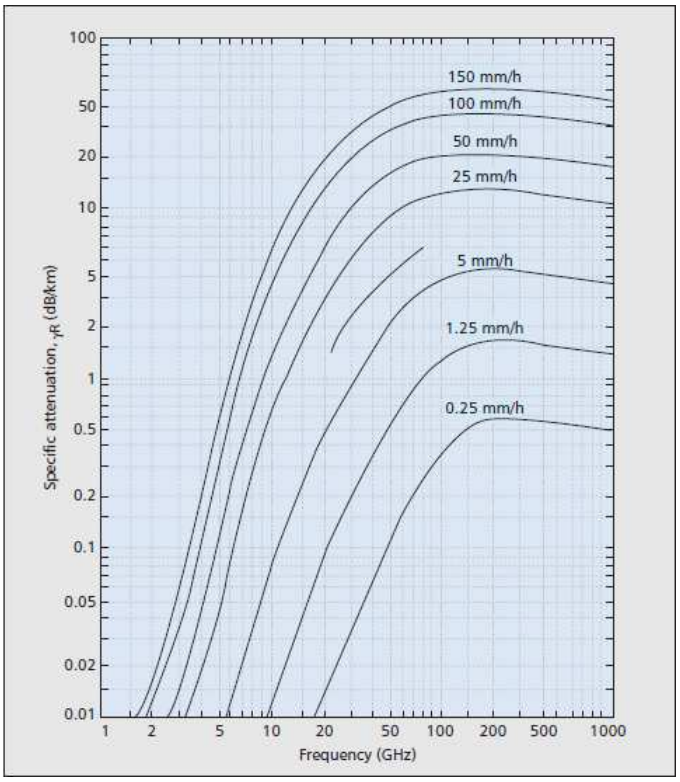

Figure 2.2: Rain attenuation for different rainfall rates [2]

Experimental data show that the effect of snow depend on its consistency [50]. For humid snow, measures show higher attenuation than the one incurred because of the rain. However, it becomes smaller for less humid snow. Regarding the fog effect, it is less significant than the rain and will not affect the mmWave signal notably [51].

Although these different losses contribute to limit the range of mmWave communica-

tions, they can be compensated by high gain antenna arrays. This has been experimentally verified in [3] by conducting path loss measures in an anechoic room using a patch antenna (3 GHz) and an antenna array (30 GHz) of the same physical size as shown in Fig. 2.3. The results show that the same path loss, across different distances, occurs when establishing a communication link using 3 GHz path antenna and 30 GHz antenna array. Moreover, the path loss decreases by 20 dB when the antenna array is used at both the transmitting and the receiving ends. The previously mentioned measurement campaigns also showed that the path loss exponent, in the mmWave band, is comparable to the one in traditional sub-6 GHz systems when antenna arrays are employed. For a 200 m to 300 m link, the path loss exponent lies between 3.2 and 4, 58 for NLOS scenarios depending on the scattering environment and between 1.68 and 2.3 for LOS scenarios. All these findings confirm the possible exploitation of mmWave frequencies for a wireless link within a limited range compared to the traditional wireless systems.

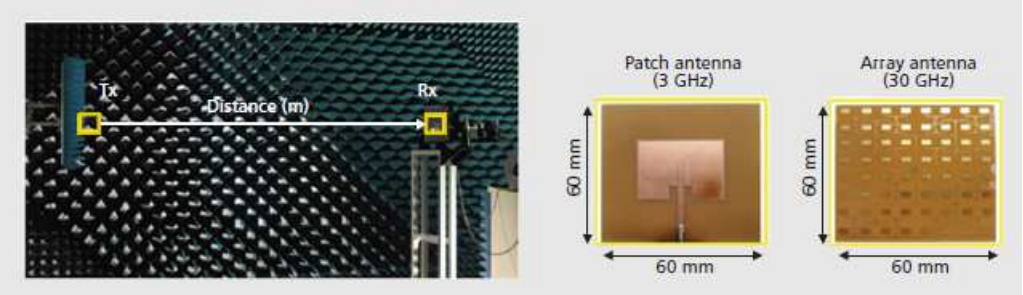

Figure 2.3: Experimental setup for path loss measurements [3]

Channel sparsity

When it comes to multi-path channel components, mmWave channels are sparse. The number of dominant scatters for frequencies below 3 GHz varies between 4 and 9 whereas it is mainly less than 4 or 5 for the mmWave band [52]. In fact, an important dispersion occurs during the signal's trajectory, which results in more power loss and less significant paths of propagation. Moreover, the angle spread after reflection is found to be very small compared to the microwave band. This increases the antenna sensitivity and any small perturbation of the antenna's orientation could impact the reliability of the transmission link.

To illustrate the channel sparsity, we refer to Fig. 2.4, which shows different received energy lobes at 73 GHz in the azimuth plane for a NLOS environment [4]. It can be noted that there are only three significant multi-path components. The power delay profile, in Fig. 2.5 , for the same setting, confirms the sparsity of the channel. There are only three multi-path components with a received power higher than the threshold (fixed according to the receiver sensitivity).

In conclusion, the high path loss is in fact manageable and does not exclude the mmWave band from being used in wireless networks provided the use of antenna arrays with beamforming techniques. However, it is required to use new models for the

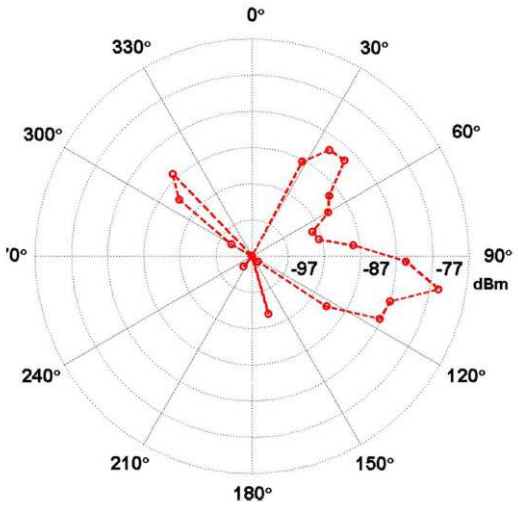

Figure 2.4: Received power lobes at 73 GHz [4]

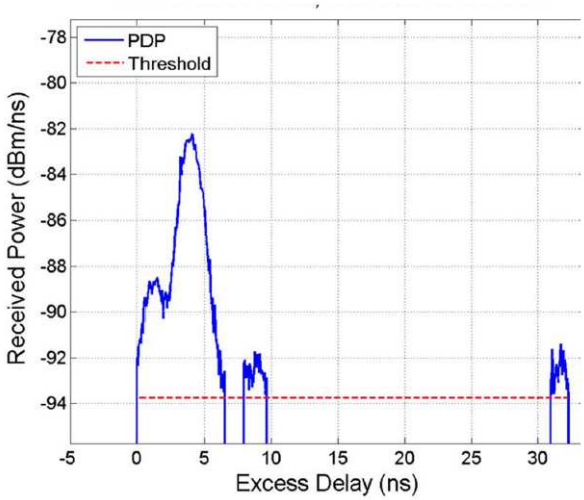

Figure 2.5: Received power delay profile at 73 GHz [4]

mmWave channel since the widely used Rayleigh model [53] can not be applied because of the sparsity of the channel. Extensive work on new mmWave channel models have been conducted by different fifth generation (5G) research groups during the last years [11]. Moreover, new algorithms that are adapted to the low rank channel matrix, are needed for different operations such as hybrid precoding and channel estimation as demonstrated in [54] and [18] respectively. In fact, the channel sparsity can even be leveraged to propose efficient algorithms as we will see in our proposed MEXP3 algorithm, which will be presented in chapter 3 of this manuscript.

2.2.3 Wireless communication challenges

The mmWave band represents a strong candidate for future wireless networks to face the ever-growing data traffic, and yet there is still no mmWave commercial cellular network. As a matter of fact, there are still challenges facing the deployment of mmWave

networks, that essentially originate from the specific characteristics of this large band of the radio spectrum compared to traditional systems. In the following, we present four major challenges for mmWave communications. However, our work of this thesis is focused on one particular challenge: the beam alignment problem.

Beamforming architecture for mmWave systems

Beamforming is a classic signal processing technique, where antenna arrays are used to focus the signal in a desired direction [55]. This could be done at the transmitting side alone or at both ends of the communication link to guarantee more power gains and improve the link power budget. Future mmWave systems will depend essentially on beamforming techniques to overcome the high propagation losses. Additionally, beamforming could be used to reduce multi-user interference thanks to the spatial selectivity of highly directional beams. Nevertheless, its practical implementation introduces some new challenges.

We can distinguish three beamforming schemes. The digital beamforming (DBF) is performed in the baseband, where each radio frequency (RF) chain's signal is multiplied by a certain weight as shown in Fig. 2.6. DBF offers flexibility and good performance at the expense of the system's complexity, cost and power consumption since each antenna requires a complete dedicated RF chain (amplifier, local oscillator, analog/digital converter,...) [54]. For instance, a digital beamformer working at Gbit/s rates and using 64 antennas consumes up to 200 W [56]. Consequently, DBF alone is impractical for mmWave systems.

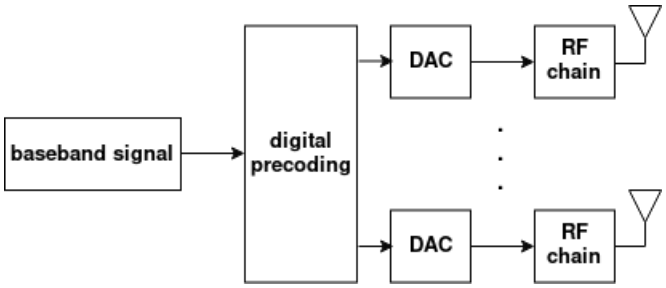

Figure 2.6: Digital beamforming architecture at the transmitter

The analog beamforming (ABF) (Fig. 2.7) consists of manipulating each antenna's signal with complex coefficients using controlled phase shifters and variable gain amplifiers [57], [58]. The ABF architecture has low cost and low complexity and consumes less power compared to DBF, since it is not required to connect each antenna to a whole dedicated RF chain. The ABF has been adopted for the indoor mmWave communication standard IEEE 802.11ad [59].

Even though the ABF is economically more attractive, it is less flexible and does not support multi-stream transmissions [60] and introduces some additional hardware related constraints (quantized phase values) which limits the processing that can be done in an analog domain. Therefore, there is a complexity vs. performance tradeoff that advocates for a hybrid beamforming (Fig. 2.8) which combines ABF and DBF schemes. This hybrid beamforming architecture was first presented in [61], [62] and

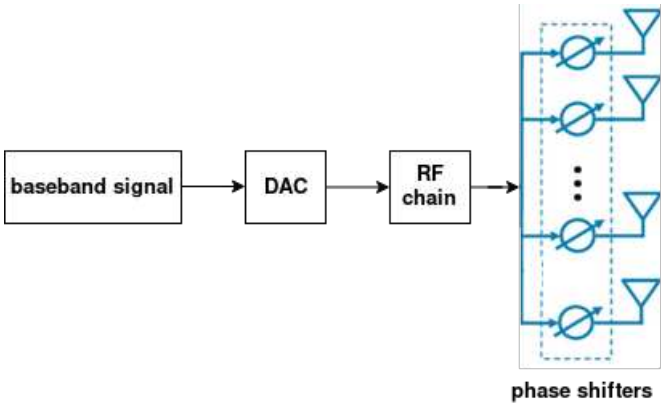

Figure 2.7: Analog beamforming architecture at the transmitter

then received a lot of interest and lead to several variant architectures [63]. Such hybrid schemes benefit from having less RF chains than antennas, which reduces the cost and power consumption while maintaining the flexibility provided by the digital domain to perform different signal processing algorithms.

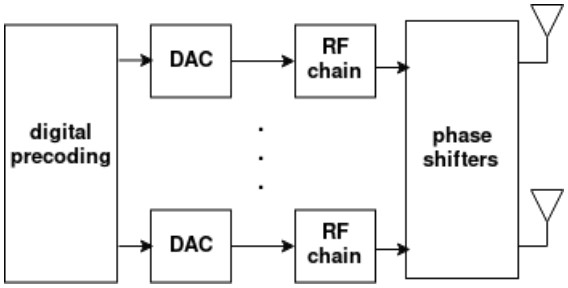

Figure 2.8: Hybrid beamforming architecture at the transmitter

To sum up, the complexity vs. performance tradeoff between DBF and ABF encourages the hybrid configuration. There is still no known hybrid architecture that provides the same performance as the fully digital beamforming. Instead, they provide asymptotic performance that best tradeoff complexity and performance to be chosen according to the application. In practice, the hardware characteristics still pose some constraints such as RF impairments (quantization noise, non linearity of power amplifiers, ...). The improvement of hybrid beamforming remains an interesting future research axis.

Beam alignment problem

Due to the high path loss encountered in mmWave frequencies, the transmitter communicates with its users via directional beams to focus the signal's power toward the intended user. As a result, it becomes crucial to steer the beam toward the right direction and not to interfere with other users. Similarly, the user's beam needs to be adjusted as well to establish a reliable communication link with the transmitter. This problem is known as the beam alignment problem (also known as beam steering or beam training). The results reported in [64] show the importance of beam alignment;

a misalignment of 18 degrees in a 7 degree beam-width system yields to 17 dB loss, which translates into 6 Gbit/s reduction of maximum throughput.

An immediate method to solve the beam alignment problem is to estimate the mmWave channel parameters (angles of departure/arrival and propagation paths gains) to construct the appropriate beams accordingly. However, estimating the mmWave channel is challenging and not practical especially for dynamic channels. Highly mobile users require constant beam alignment to maintain a reliable link despite the channel's variations, which impose a large training and feedback overhead on the system. Moreover, multiple antennas are often connected to a single RF chain (hybrid beamforming), which means that the system can not acquire a direct observation of the channel for each receive antenna. Furthermore, the use of large antenna arrays and the high path loss complicate the channel estimation for each receiving antenna in the mmWave band. Hence, the need for new beam alignment policies, requiring partial or no channel state information, to avoid estimating explicitly the mmWave channel and relieve the system from the large training and feedback overhead, is timely and important.

To sum up, beam alignment is a fundamental inevitable step in all mmWave systems (cellular networks, device to device or vehicular communications). This is why beam alignment represents a main focus for this thesis. The major limiting factors of beam alignment policies can be grouped into:

- Time constraints: mmWave systems need to align their beams relatively quickly, within the channel's coherence time (or the beam coherence time) before major changes in the wireless environment;
- Mobility conditions: more training and feedback overhead are introduced because of the mobility of users that require a constant update of the alignment, which poses more performance demands on the beam alignment policies.
- Limited feedback: performing beamforming with limited feedback is important for mmWave systems to reduce the training overhead and achieve low latency systems. With limited feedback policies, the system can encounter some performance degradation and a latency improvement. This highlights a tradeoff performance vs. low feedback, which requires further investigation.

Full duplex mmWave communication

Full duplex (FD) technology consists of transmitting and receiving simultaneously in the same frequency band. Given the large antenna arrays employed in mmWave systems, a partition of the array can be used to form beams for data transmission whereas the other partition can be used as a receiver. Since the base stations are expected to have larger number of antennas compared to the users, FD can be used only at the base station with half duplex (HD) users. The base station can transmit a signal to a user while receiving another signal from other users.

In theory, FD would allow to double the system capacity when compared to an HD system since it allows the antenna array to transmit and receive at the same time. The FD presents other advantages as it reduces the end-to-end feedback delays and increases

the spectrum utilization efficiency and the overall network efficiency. However, FD suffers from practical challenges and limitations because of the Self-Interference (SI) [65], where a part of the transmitted beam can interfere with the adjacent receiving array. Other practical limitations include traffic constraints and increased inter-cell interference [65]. Even though the transmitted signal is known in digital baseband, it remains challenging, in practice, to remove it completely because of SI channel estimation errors, interference from other users, noise added by the RF impairments and also the notable power difference between transmitted and received signals. Moreover, performing SI cancellation increases the system's complexity.

It is important to note that the SI channel is different from the usual mmWave end-to-end channel. The SI channel represents the channel between the transmitting and the receiving array at the same end. The main difference is that it incorporates near-field propagation model LOS components with far-field propagation model NLOS components [66]. This implies that more channel modeling and estimation work is needed to eliminate this kind of interference.

The potential use of FD technology for mmWave communication is studied in [66] and two antenna configurations are proposed. One configuration uses the same antenna array for transmitting and receiving while the second one uses separate antenna arrays. The second configuration offers better SI mitigation than the first one since we can separate the two arrays and position them far enough from each other given their relatively small dimensions (especially for a base station). The transmitter/receiver isolation of the first configuration depends mainly on the isolation capabilities of the used hardware components. Classical SI cancellation methods, which are used for traditional sub-6 GHz systems revealed unsuitable to efficiently mitigate the SI at the mmWave band [65, 66], which necessitates new approaches adapted to the mmWave band such as the ones proposed in [67, 66, 3, 68].

Lastly, the FD technology promises an important spectral efficiency improvement for mmWave communications provided that certain issues are resolved. First, the SI cancellation remains a primal challenge for practical FD mmWave systems and requires more research investigation to attain good isolation between the transmitting and receiving arrays. The SI management is also still limited by the impact of hardware impairments. Second, practical implementation issues, including FD systems that support bandwidths of GHz order and directional transmission/reception mode, need to be addressed. The actual FD systems support up to 40 MHz bandwidths [69]. An FD architecture with directional transmission and omnidirectional reception is proposed in [70]. Advanced SI management policies and hardware components represent open research axes.

Network densification

To enhance the mmWave wireless capacity, networks are expected to have a large scale cell densification. Network densification can be achieved via densification over space (spatial densification) and/or frequency (spectral aggregation). Spatial densification consists of increasing the number of deployed base stations in a geographic area besides increasing the number of antennas at both ends of the communication link (user device

and base station). Spectral aggregation designates combining larger parts of the electromagnetic spectrum from different bands (from sub-6 GHz to mmWave frequencies) [71].

Spatial densification challenges Since macro-cell deployment, in the mmWave band, is costly and requires further site planning, spatial densification implies low power indoor or outdoor small-cells (100 m to 200 m radius [72]), which will be operating next to each other. Theoretically, this can increase the mutual interference. However, directional antenna arrays would help to reduce this interference (angular isolation effect). According to [73], measurements at the 60 GHz band, show that the interference between cells is negligible for outdoor networks. The significant spatial densification yields different cellular architectures with overlapping coverage of dense base stations, but with low inter-cell interference. This requires new protocols for fast cell switching to handle the intermittent mmWave link. On the other hand, the interference in indoor environment requires more management such as power control and transmission coordination [9] due to the large number of access points present in crowded places (such as shopping malls or large business buildings, which can contain a large number of data-hungry devices). Several interference management protocols have been proposed to deal with the large network densification [9].

Spectral aggregation challenges Utilizing large parts of the radio spectrum from different bands (and possibly non adjacent) leads to heterogeneous networks functioning simultaneously at different bands (microwave and mmWave). This creates a new hardware challenge regarding the design of RF transceivers, which need to be cost and power efficient to support the spectral aggregation. The heterogeneous topology also creates a need for more standardization work of site planning and network resources management [74], [75].

Finally, we emphasize that the main challenge considered in this thesis is the beam alignment problem. Full duplex and network densification are emerging technologies for mmWave systems, which introduce several challenges that are not treated in this thesis and left for future investigation.

2.3 Machine learning tools for mmWave communications

In this section, we summarize some machine learning tools, which are most relevant to this thesis. We present a general overview of the three main machine learning paradigms with their benefits and drawbacks, and their potential use for wireless communications and more specifically for beam alignment in mmWave networks. More technical details and specific algorithms, employed in this thesis, will be detailed throughout the rest of the manuscript.

2.3.1 Supervised learning

The supervised learning approach consists of using labeled data to train a learning model [1] to classify them into a given number of categories (classification) or predict continuous outcome (regression). Each data sample is composed of an input feature paired with a ground-truth label or variable to be predicted, which represents the correct output. The model is trained to identify patterns and relationships between the input and the desired output to provide accurate labels for new unseen data. Supervised learning problems can be solved by various machine learning algorithms, depending on the data and the problem at hand such as support vector machines, decision trees, linear regression, etc [76].

Another recent approach is exploiting deep learning tools. Deep learning refers to a machine learning field, which exploits deep neural networks (DNNs) containing multiple hidden layers to perform successive processing operations and extract high level features from the available training data [77]. In other words, deep learning leverages the approximation capabilities of DNNs to predict an output from relevant training data. As opposed to traditional machine learning methods (e.g., support vector machines), deep learning does not require the dedicated pre-processing step called feature extraction, which provides a representation of raw data that can be used by classic algorithms. Indeed, the layers of a DNN can learn directly an implicit representation of the raw data. In other words, the feature extraction is already included in the processing performed by the DNN. Supervised deep learning is the sub-field of deep learning exploiting supervised learning techniques via DNNs.

Supervised deep learning has been used to solve real world problems in many fields. Resource allocation in wireless communications is no exception [33] given many relevant resource allocation problems are NP-hard such as sum-rate maximization, beamforming problems and energy-efficiency maximization. Moreover, the optimal resource allocation policy depends on time-varying system parameters (e.g user positions, channels, number of user, etc), which means that the optimization problem needs to be solved frequently leading to a significant complexity overhead. Supervised deep learning (and supervised learning in general) allows to collect and interpret data to learn an unknown mapping between the system parameters and the corresponding optimal resource allocation policies. In [78, 79, 33], a detailed overview of several applications of supervised deep learning for resource management in future wireless networks is provided. As an example, the authors in [80] proposed a supervised deep learning approach for power control in massive MIMO systems. Supervised deep learning has also been used for problems related to the physical layer such as channel estimation, localization, data decoding, etc [33].

Although supervised deep learning has been used for various problems of wireless communications, practical implementation still faces some challenges. The proposed resource management solutions require offline training with sufficient amount of correctly labeled data. First, obtaining such data can be time-consuming and costly in wireless

[1] A learning model refers to the outcome of a ML algorithm that was learned using training data. For instance, for neural networks it represents a specific structure with weights' values.

networks and raises privacy issues. Second, labeling the available data introduces more cost and needs to be accurate to avoid prediction errors.

2.3.2 Unsupervised learning

In unsupervised learning, the model is trained using raw data without the knowledge of the ground-truth output corresponding to each input feature. Using only input features, the learning algorithm identifies patterns, similarities and differences within the data to carry out the desired task. Unsupervised learning algorithms are mainly categorized into clustering (identify groups within data) or association (predict rules describing the data) [81]. Unsupervised deep learning refers to the sub-field of deep learning using DNNs with the unsupervised learning approach.

Unlike supervised learning, unsupervised learning models can cluster or classify data on their own without the need for labeled data. This means that they can even provide unexpected findings that are not considered during the labeling in supervised learning. Also, it frees the learning model from the burden of labeling all the training data (unlabeled data is cheap and easy to collect and store). Moreover, unsupervised learning yields generative models capable of dealing with new information. For instance, if a classifier is trained to recognise two categories and is fed with an input of a third category, it would classify it incorrectly with one of the first two categories whereas an unsupervised learning model would be able to identify it as a different category. However, the resulting output can be less accurate since the model is learning from raw data without prior knowledge of the ground-truth as in supervised learning. Also, the model's complexity can be impacted by the increase of the input dimensions because the learning process will take more time to analyse all possibilities within larger input features without the aid of a ground-truth output.

Despite the popularity of supervised deep learning in wireless networks, there are several applications of unsupervised deep learning too as discussed in [78, 79, 33, 82]. In [36], an unsupervised machine learning scheme based on auto encoders is proposed to identify mmWave beamformers. An unsupervised DNN is trained in [83] to conceive a multi-user precoding scheme and improve the spectral efficiency of non-orthogonal multiple access (NOMA) mmWave systems. Another example is [84] where the authors investigate unsupervised learning tools for the user clustering and power allocation problem in NOMA mmWave networks.

Semi-supervised learning When the the available data is mainly unlabeled, but still contains a small fraction of labeled samples, semi-supervised learning can be used as a combination of both supervised and unsupervised learning. It determines structures and correlations between all the samples (as in unsupervised learning), exploits the labeled data to best predict a ground-truth for the unlabeled ones and then uses the entire data to train a supervised learning model to make predictions for unseen data. This approach is pertinent when supervised learning algorithms are needed but the data is not fully labeled (a common issue in real-world ML problems). It allows to avoid the time-consuming and expensive process of labeling all the training data.

Some of the practical applications of the semi-supervised learning include webpage classification, text document classifiers, facial recognition, etc [85]. Recent work also includes applications for resource management. In [86], semi-supervised learning is used to detect anomalies in mobile wireless networks and adapt their resource allocation policies. The authors in [87] exploit semi-supervised deep learning to recognize different radio technologies accessing the spectrum to define spectrum management policies and mitigate interference. The work in [88] considers environmental (indoor or outdoor) recognition of a mobile device using semi-supervised learning to adapt their resources for greater efficiency. In [89], a semi-supervised beam selection method is also proposed for mmWave networks. More applications of semi-supervised learning are summarized in [82].

2.3.3 Reinforcement learning

The reinforcement learning paradigm covers ML techniques that enable an agent to learn in an interactive environment using the feedback obtained as a result of its own actions and experiences. As illustrated in Fig. 2.9, the agent takes an action in its environment. Then, the action gives rise to a reward and a state (situation of the agent in the environment), which are fed back to the learning agent. Generally, RL algorithms are either model-based or model-free.

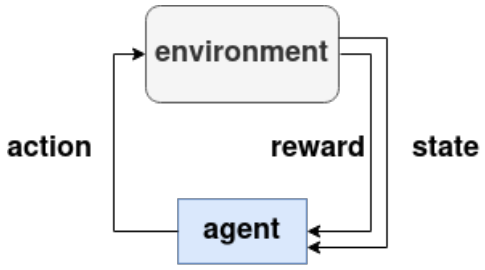

Figure 2.9: Typical reinforcement learning illustration.

Model-based learning [90] seeks to model the environment then chooses the optimal policy using the learned model. The main drawback of this approach lies in the fact that it relies on more assumptions and approximations for a given task, which limits it to that specific type of tasks. In model-free learning, the agent establishes its optimal policy over time by systematic trial and error. Its optimal policy needs to choose actions that tend to increase the overall sum of rewards while still exploring its environment (exploitation vs. exploration tradeoff). Recent development of DNNs also gave birth to a new approach: deep reinforcement learning (DRL) [91] combining the RL paradigm with DNNs by allowing a feedback loop between a neural network based algorithm and the environment. In other words, the agent learns over a time-varying dataset due to its interaction with the environment. One particular case of model-free RL, which is most relevant to the thesis, is multi-armed bandits (MAB) [32].

Multi-armed bandits

MAB represents the simplest possible RL problem with a single state and only self transitions [22]. Following the typical framing of a RL problem as in Fig. 2.9, the learning agent chooses an action from a finite set of actions, and thus observes a reward. For the next stage of the repeated decision process, the agent chooses again an action based on the information provided by the previously received reward. The name of this approach comes from the analogy in the original problem formulation [92] where a gambler (agent) faces slot machines (each called a one-armed bandit). When an arm (action) is pulled, the gambler receives a reward. The gambler aims to maximize its expected gain without knowing the optimal arm, which illustrates the need for data exploitation and exploration. Many algorithms have been proposed in the literature. The most popular ones are the upper confidence bound (UCB) and the exponential weight algorithm for exploration and exploitation (EXP3) [32].

One main difference between MABs and other RL approaches (e.g., Q-learning [93]) lies in its simple formulation and implementation. In [94], the authors refer to MABs as one-state or stateless particular case of RL. In other words, a MAB agent picks an action, gets a reward and starts again without a state transition (the one state being having access to all arms with unknown reward distributions). This means that the MABs ignores the state and learns to balance the exploitation vs. exploration tradeoff simply via the reward mechanism. Contextual MABs [94] extends this model by keeping the one-state formulation and making decisions based on information about the state of the environment (context) besides previous reward observations (feedback). Note that the general RL problem can be seen as an extension of contextual bandits.

Do recent RL approaches render MABs obsolete? The MAB-based algorithms require low computation capabilities and less memory compared to other RL algorithms. Moreover, MAB-based algorithms come with theoretical performance guarantees in terms of regret. The fundamental exploration vs. exploitation tradeoff makes MAB problems challenging and still applicable to practical problems. The exploration aspect in MABs is well understood especially on the theory side as opposed to advanced DRL approaches where the exploration is still limited to simple epsilon-greedy strategies, or sampling from stochastic policies [91]. Therefore, MAB algorithms are relevant in many practical problems such as website advertising optimization, clinical trials, financial portfolio design, etc [95]. Resource allocation problems for wireless networks (mmWave included) are no exception. The nature of the MAB formulation makes it a suitable approach to tackle the problems of beam selection in mmWave systems [96], transmission power allocation [97] and channel (frequency subcarrier) selection [98], which are among the main radio resource management problems. All these recent works [96, 97, 98] proposed MAB-based resource management policies, which are crucial for future wireless networks serving a wide variety of users and requiring efficient radio resource management.

On the other hand, the simple one-state formulation of MABs makes them unsuitable, compared to general RL methods, for problems with different states where an action may cause a state transition. Furthermore, RL problems with high-dimensional state

space and action space are handled better with recent DRL approaches, which are based on scaling up prior RL work to high-dimensional problems such as an agent learning from large visual inputs and data from different sensors. In face of this dimensionality issue, the powerful approximation properties of DNNs are leveraged to predict optimal actions, based on a training over samples from the state or action space, instead of tabular (Q-learning) and other classic non-parametric methods (MABs), which become impractical with large problems [91].

To sum up, the multi-armed bandit algorithms represent a particular case of the RL paradigm and remain useful for practical problems where decisions need to be made sequentially and under uncertainty. When the problem is of low dimensionality w.r.t the action apace and can be formulated as a one-state (or stateless) RL problem, the MAB framework provide efficient simple algorithms with theoretical guarantees to choose optimal actions as we propose in the next chapter for the beam alignment problem in mmWave networks. For larger problems, other RL and DRL approaches become more appropriate.

2.3.4 Classic reinforcement learning vs deep learning

The deep learning terminology comprises machine learning techniques involving deep neural networks. Deep learning applications for wireless networks have gained large interest in the recent years [33], which rises the question of the viability of classic RL approaches (such as MABs) for future wireless networks.

Both reinforcement and deep supervised learning aim to find a mapping between an input and an output. The difference lies in the fact that the first framework uses feedback signal (rewards) resulting from online interactions with the environment whereas the latter employs labeled data during an offline training phase. Compared to unsupervised deep learning (which also requires offline training), reinforcement learning seeks to find an optimal action to maximize the total cumulative rewards of the agent where in unsupervised learning the objective is to identify similarities and differences within the training data. Furthermore, classic RL algorithms are usually less complex when it comes to computation. The complexity of deep learning methods is higher and increases with the dimensionality of the problem and the DNN architecture.

To settle the score, both approaches have potential in wireless networks according to the nature of the problem at hand and the objective of the learning. The recent explosion of neural network applications does not exclude classic RL from research investigation and practical applications. Instead, technical requirements and application characteristics, such as latency, computation power, energy consumption and data availability for training, have to be taken into account to leverage one approach over the other. For instance, it is less efficient to use complex neural network solutions for simple problems that can be solved using classic RL. In chapter 4, we illustrate the comparison between MABs and deep learning methods for the particular beam alignment problem in mmWave systems.

2.4 Conclusion

In this chapter, we have presented the mmWave channel with a detail description of its propagation characteristics and challenges. The difficult propagation conditions, at this band, can be overcome by large antenna arrays and beamforming techniques. Moreover, we have presented the main machine learning paradigms and compared them in term of advantages and drawbacks for resource allocation problems. In the following chapters, we exploit two different machine learning frameworks (MAB and DL) to solve the beam alignment problem in mobile mmWave networks.

3

Multi-armed bandits for mmWave beam alignment

3.1 Introduction

This chapter represents the first part of the thesis contributions. It is dedicated to the exploitation of multi-armed bandits for the mmWave beam alignment problem. First, we formulate the beam alignment as an adversarial MAB problem. We investigate the optimal size for the beamforming codebook. Then, we propose online beam alignment policies based on the exponential weights algorithm (EXP3) and the sparse nature of mmWave channels. Also, we discuss the no-regret property for the proposed policies. Finally, we illustrate the performance of the beam alignment policies via numerical results in practical simulations settings.

3.2 System model

We consider a point-to-point mmWave MIMO system, as depicted in Fig. 3.1, consisting of a fixed transmitter (Tx), equipped with M_T antennas and $N_T \leq M_T$ radio frequency (RF) chains, and a mobile receiver (Rx) equipped with M_R antennas and $N_R \leq M_R$ RF chains. Both nodes communicate via directional beams which point towards certain spatial directions determined by the hybrid (analog and digital) beamforming vectors $\mathbf{f}_i \in \mathbb{C}^{M_T}, i \in \{1, \ldots, A\}$ and $\mathbf{w}_j \in \mathbb{C}^{M_R}, j \in \{1, \ldots, A\}$ used at the transmitter and the receiver respectively.

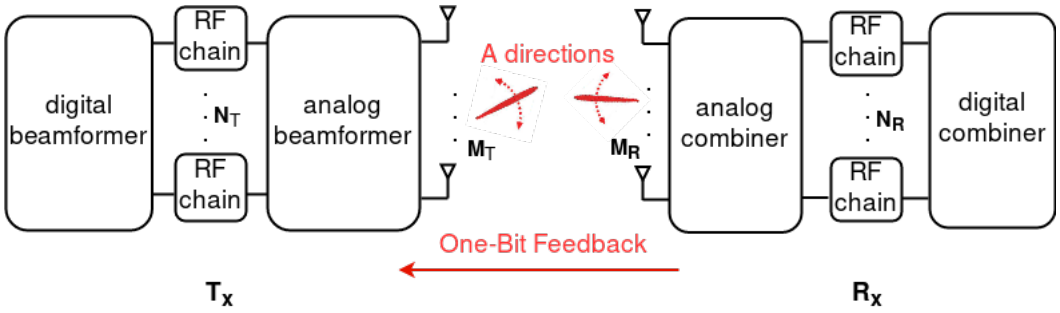

Figure 3.1: Beam alignment in a point-to-point mmWave MIMO system.

3.2.1 Beamforming codebook

The codebook consists of a set of A hybrid beamforming vectors designed offline using the procedure in [99, Algorithm 1], which offers high beamforming gains compared to other existing codebooks in the literature [99]. The codebook size, $A = 2^n$, $n \in \mathbb{N}^\star$, represents the total number of all possible beam-directions.

The analog beamformers designed to steer the transmitted signal into a particular spatial direction are implemented using phase shifters, which cover uniformly the angular domain between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Since each beamforming vector corresponds to a unique direction, our proposed online policies will select a suitable beamforming vector (or equivalently the beam-direction) to meet the SNR requirements.

The main role of the digital weights is to optimize the beamforming gain of the different beams (with respect to an ideal beam pattern). They are fixed and tuned as in [99]. Also, the digital part of the codebook enables transmission via multiple data streams, which could be exploited in multi-user systems. Such a transmission mode is not possible when using only an analog beamformer with a single RF chain.

3.2.2 mmWave channel model

The transmitted signal in the mmWave band experiences limited scattering. Therefore, we use the well-known narrowband geometric model [18, 13, 28] with L propagation paths

$$\mathbf{H}(t) = \sqrt{\frac{M_T M_R}{\rho}} \sum_{l=1}^{L} \alpha_\ell \, \mathbf{a}_R(\theta_\ell)\mathbf{a}_T(\phi_\ell)^\dagger \, \mathrm{e}^{j2\pi\nu_\ell t}, \tag{3.1}$$

where ρ represents the average path loss [100]; $\alpha_\ell \sim \mathcal{N}(0, \sigma^2_{\alpha_\ell})$, $\ell \in \{1, 2, .., L\}$ is the complex path gain assumed to follow a Gaussian distribution; σ_{α_ℓ} is the average power gain; ϕ_ℓ and θ_ℓ are the angles of departure (AoD) and arrival (AoA) respectively; ν_ℓ is the Doppler shift of the ℓ^{th} path; $\mathbf{a}_T(\theta_\ell)$ and $\mathbf{a}_R(\phi_\ell)$ are the array steering vectors for the transmitter and the receiver. Assuming a uniform linear array (ULA), $\mathbf{a}_T(\theta_\ell)$ and $\mathbf{a}_R(\phi_\ell)$ can be expressed as:

$$\mathbf{a}_R(\theta_\ell) = \frac{1}{\sqrt{M_R}}[1, \mathrm{e}^{j\frac{2\pi}{\lambda}d\sin(\theta_\ell)}, ..., \mathrm{e}^{j(M_R-1)\frac{2\pi}{\lambda}d\sin(\theta_\ell)}]^T, \tag{3.2}$$

$$\mathbf{a}_T(\phi_\ell) = \frac{1}{\sqrt{M_T}}[1, \mathrm{e}^{j\frac{2\pi}{\lambda}d\sin(\phi_\ell)}, ..., \mathrm{e}^{j(M_T-1)\frac{2\pi}{\lambda}d\sin(\phi_\ell)}]^T, \tag{3.3}$$

where d is the distance between the antenna elements within the array and λ represents the wavelength of the transmitted signal.

While most of the results in this chapter hold irrespective from the channel model and network dynamics, we use the narrowband model in (3.1) to illustrate the performance of the proposed algorithms. Note that the geometric model is widely used for the mmWave band due to its ability to capture the sparse nature of the mmWave channel as a combination of a limited number of propagation paths, which are described by their main parameters such as gain and angles of departure/arrival. It encloses

physical properties of the signal propagation including the dependence on environment geometry, frequency band, etc., which are important for a machine learning method to identify good beams.

3.2.3 User mobility

To model the mobility of the receiver, we exploit the boundless mobility model adopted in [5] assuming a bounded two-dimensional movement area. This model is memory-based and incorporates temporal correlations in the update of the user's speed and direction, which leads to realistic settings and time-varying channel matrices $\mathbf{H}(t)$. It also allows to impose limitations on the linear speed, acceleration and rotation speed, and thus offers a good tradeoff between accuracy and flexibility [5]. In this model, the speed $v(t)$ and direction of movement $\Theta(t)$ are updated every channel coherence time T_c as follows:

$$\begin{cases} v(t+T_c) = \min\{\max\{v(t)+\Delta v,\ 0\},\ v_{\max}\}, \\ \Theta(t+T_c) = \Theta(t)+\Delta\ \Theta, \end{cases} \tag{3.4}$$

where $v_{\max}$ is the maximum speed; the speed variation is $\Delta v \sim \mathcal{U}[-a_{\max}\ T_c,\ a_{\max}\ T_c]$ with $a_{\max}$ being the maximum linear acceleration and $\mathcal{U}$ denoting the uniform distribution; $\Delta\Theta \sim \mathcal{U}[-\omega_{\max}\ T_c,\ \omega_{\max}\ T_c]$ is the direction variation with $\omega_{\max}$ denoting the maximum rotation speed.
In this work, we exploit this mobility model to generate different trajectories to simulate the receiver's mobility. Every update time T_c, the model parameters $v(t)$ and $\Theta(t)$ are used to determine the receiver's position with respect to the transmitter. Then using this new position, the channel parameters AoA $\theta_\ell(t)$ and AoD $\phi_\ell(t)$ are updated as shown explicitly in [101]. The other channel parameters $\rho(t)$ and $\nu_\ell(t)$ are also updated every T_c depending on the new transmitter-receiver distance and the speed respectively. Fig. 3.2 shows an example of movement pattern, within a time interval of 26 s, with $T_c = 1.3$ ms, $v_{\max} = 5$ m/s, $a_{\max} = 5$ m/s^2 and $\omega_{\max} = \pi/2$ rad/s.

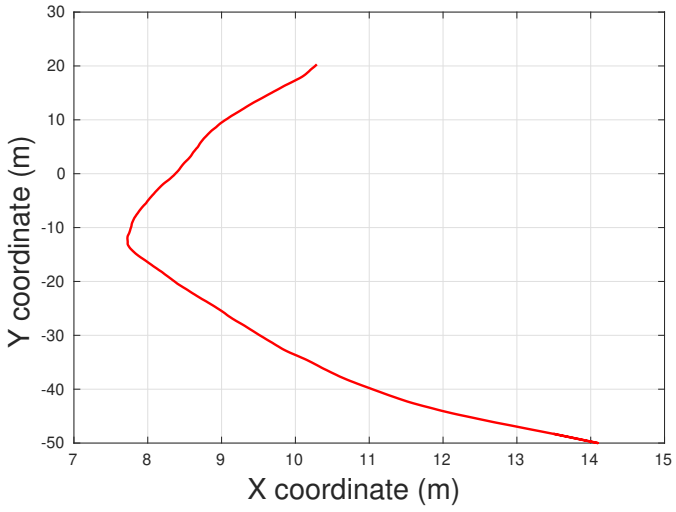

Figure 3.2: Illustration of a movement trajectory following the mobility model in [5] .

3.2.4 Received signal

The received signal $y_{i,j}$ at time t can be written as:

$$y_{i,j}(t) = \mathbf{w}_j^{\dagger}(t)\ \mathbf{H}(t)\ \mathbf{f}_i(t)\ s(t) + \mathbf{w}_j^{\dagger}(t)\ \mathbf{n}(t), \tag{3.5}$$

where i and j denote the indices of the transmit and receive beams respectively. To simplify the presentation, we drop the explicit temporal variability of the channel model parameters hereafter. During the beam alignment process, the transmitter uses a beamforming vector $\mathbf{f}_i$ to transmit its symbols $s \in \mathbb{C}$ such that $\mathbb{E}[|\ s\ |^2] = P_{\text{tr}}$, where P_{tr} is the transmit power. The receiver uses its own beamforming vector $\mathbf{w}_j$ to recover the transmitted signal. The channel noise vector, denoted by $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2)$, is a Gaussian distributed random variable. The resulting SNR at the receiver depends on the beams $\mathbf{f}_i$ and $\mathbf{w}_j$ and is expressed as

$$\text{SNR}_{i,j} = \frac{|\ \mathbf{w}_j^{\dagger}\mathbf{H}\mathbf{f}_i\ |^2\ P_{\text{tr}}}{\sigma_n^2}. \tag{3.6}$$

Assuming a stochastic channel model, we define an outage as the event in which the SNR falls below a certain threshold ξ, whose value represents the target SNR. The outage probability can be then defined as

$$\text{P}_{\text{out}}(i,j) \triangleq \text{P}[\text{SNR}_{i,j} < \xi]. \tag{3.7}$$

The choice of the threshold ξ will depend on the nature of the application. For instance, if the mmWave link is used for an application that requires high values of SNR, the value of ξ should be high as well.

The outage probability is an important performance metric in communications systems in which an average performance is less relevant than guaranteeing a minimum instantaneous quality of service [102, 103]. In 5G for instance, ultra-reliable low latency applications depend crucially on instantaneous reliability, which can be measured by the outage probability [104]. However, minimizing the outage probability above is quite a challenging problem as its explicit expression becomes intractable in practice (e.g., in our channel model with mobility or when the statistics of the channel is unknown). Indeed, even in the most simplified MIMO Rayleigh channel with perfect knowledge of the channel statistics at the transmitter, optimizing the outage probability remains an open issue [105]. Thus, we propose here to exploit the multi-armed bandit framework and sequential decision processes in an effort to approach the minimum outage.

3.3 Beamforming codebook size

Our beam alignment problem consists of finding adaptive and decoupled policies, at the transmitter and the receiver, which choose the beamforming vectors $\mathbf{f}_i$ and $\mathbf{w}_j$ that minimize the outage probability. First, we discuss an optimal size (or spatial resolution) for the beamforming codebook.

3.3.1 Codebook size problem

We assume that the beam-directions of the codebook are uniformly chosen to cover the spatial horizon between $-\pi/2$ and $\pi/2$ by dividing the angular domain by half each time we increase the codebook size. The higher the codebook size, the narrower and more directed the beams are (higher gain). Therefore, the outage probability defined in (3.7) depends on the codebook size.

Theoretically at least, as long as we increase the codebook size the received SNR also increases. The downside is that the set of candidate beamforming vectors increases, which implies a higher exploration cost. The rising question is then: *what is the codebook size that balances best the received SNR and the exploration cost?*

We define ΔSNR to quantify how much we should increase the codebook size A and still get a significant performance improvement in terms of the SNR. We use this measure as a criterion to avoid increasing the codebook size uselessly at the expense of larger exploration duration.

$$\Delta SNR = SNR_0 \; - \max_{k \in \{1,2,\ldots,A^2\}} SNR_k, \tag{3.8}$$

where the index k refers to the indices of all possible pairs (i, j) such that $k \in \{1, 2, .., A^2\}$ and $SNR_k = SNR_{i,j}$; SNR_0 represents the highest possible SNR related to the channel conditions and independent from the used beamforming codebook such that:

$$SNR_0 = \max_{\mathbf{u},\mathbf{v}} \frac{|\mathbf{u}^\dagger \mathbf{H} \mathbf{v}|^2 P_{\text{tr}}}{\sigma_n^2} = \frac{|\mathbf{u}_o^\dagger \mathbf{H} \mathbf{v}_o|^2 P_{\text{tr}}}{\sigma_n^2}, \tag{3.9}$$

where $\mathbf{u}$ and $\mathbf{v}$ are left-singular vectors and right-singular vectors of $\mathbf{H}$ respectively; $\mathbf{u}_o$ and $\mathbf{v}_o$ are the singular vectors corresponding to the largest singular value of $\mathbf{H}$.

3.3.2 Optimal size analysis

In order to optimize the codebook size, we first analyze the probability $\mathrm{P}\left(\Delta SNR \leq \epsilon\right)$ where ϵ represents the maximum allowed gap between SNR_0 and $\max_k \; SNR_k$. Then, we search for the smallest codebook size (for small exploration costs) which provides a high enough value of this probability (for high performance in terms of SNR) using numerical simulations. For simplicity, the Doppler shift in the channel model (3.1) will not be considered in this section. We distinguish two cases of single and multi-path channels.

The single-path case

We start with the particular case of a single-path mmWave channel, in which $\mathbf{H} = \sqrt{\frac{M_T M_R}{\rho}} \alpha \; \mathbf{a}_R(\theta) \; \mathbf{a}_T(\phi)^\dagger$. We can express SNR_0 and SNR_k as:

$$SNR_0 = \frac{P_T M_T M_R}{\rho \sigma_n^2} \; |\mathbf{u}_o^\dagger \mathbf{a}_R(\theta) \; \mathbf{a}_T(\phi)^\dagger \; \mathbf{v}_o|^2 \; |\alpha|^2 = B \; C_0 \; |\alpha|^2, \tag{3.10}$$

$$SNR_k = \frac{P_T M_T M_R}{\rho \sigma_n^2} \, |\mathbf{w}_j^\dagger \, \mathbf{a}_R(\theta) \, \mathbf{a}_T(\phi)^\dagger \, \mathbf{f}_i|^2 \, |\alpha|^2 = B \, C_k \, |\alpha|^2, \tag{3.11}$$

where $B = \frac{P_T M_T M_R}{\rho \sigma_n^2}$, $C_0 = |\mathbf{u}_o^\dagger \mathbf{a}_R(\theta) \, \mathbf{a}_T(\phi)^\dagger \, \mathbf{v}_o|^2$ and $C_k = |\mathbf{w}_j^\dagger \, \mathbf{a}_R(\theta) \, \mathbf{a}_T(\phi)^\dagger \, \mathbf{f}_i|^2$.

Thus, we can write ΔSNR as:

$$\Delta SNR = B \, |\alpha|^2 \left(C_0 - \max_k C_k \right). \tag{3.12}$$

Assuming that the path gain follows a complex Gaussian distribution, while the other parameters are fixed and deterministic, then its amplitude squared follows an exponential distribution, i.e, $|\alpha|^2 \sim Exp\ (\frac{1}{\sigma_\alpha})$ which results in an exponential distribution of ΔSNR such that $\Delta SNR \sim Exp\left(\frac{1}{\sigma_\alpha \ B \ \left(C_0 - \max\limits_k \ C_k \right)} \right)$. Hence, we can find a closed-form expression of $\mathrm{P}\left(\Delta SNR \leq \epsilon\right)$ as:

$$\mathrm{P}\left(\Delta SNR \leq \epsilon\right) = 1 - \exp\left(\frac{-\epsilon}{\sigma_\alpha \ B \left(C_0 - \max\limits_k \ C_k \right)} \right). \tag{3.13}$$

When the noise variance is very small ($\sigma_n^2 \to 0$), this probability approaches zero and the obtained SNR approaches the ideal value SNR_0. When the noise variance is very large ($\sigma_n^2 \to +\infty$), the obtained SNR cannot approach this ideal value.

Since the expression of the probability $\mathrm{P}\left(\Delta SNR \leq \epsilon\right)$ does not depend explicitly on the codebook size A, it is not trivial to find an analytic expression of the optimal codebook size. Instead, we exploit the obtained expression in (3.13) graphically to determine the smallest codebook size that offers a high SNR.

The multi-path case

In the multi-path mmWave channel case, we can express SNR_0 and SNR_k as follows:

$$SNR_0 = B \, \left| \sum_{\ell=1}^{L} \alpha_\ell D_\ell \right|^2 = B \ Z \tag{3.14}$$

$$SNR_k = B \, \left| \sum_{\ell=1}^{L} \alpha_\ell D_{\ell k} \right|^2 = B \ J_k, \tag{3.15}$$

where Z and J_k are defined as $Z = \left| \sum_{\ell=1}^{L} \alpha_\ell D_\ell \right|^2$ and $J_k = \left| \sum_{\ell=1}^{L} \alpha_\ell D_{\ell k} \right|^2$ for $k \in \{1, 2, ..., A^2\}$; $D_\ell = \mathbf{u}_o^\dagger \mathbf{a}_R(\theta_\ell) \, \mathbf{a}_T(\phi_\ell)^\dagger \mathbf{v}_o$ and $D_{\ell k} = \mathbf{w_j}^\dagger \mathbf{a}_R(\theta_\ell) \, \mathbf{a}_T(\phi_\ell)^\dagger \mathbf{f}_i$.

Assuming the path gains α_ℓ, $\ell \in \{1, 2, .., L\}$ to be Gaussian distributed, the sums $\sum_{\ell=1}^{L} \alpha_\ell D_\ell$ and $\sum_{\ell=1}^{L} \alpha_\ell D_{\ell k}$ follow the Gaussian distributions $\mathcal{N}\left(0, \sum_{\ell=1}^{L} \sigma_{\alpha_\ell} |D_\ell|^2\right)$ and

$\mathcal{N}\left(0, \sum_{\ell=1}^{L} \sigma_{\alpha_\ell} |D_{\ell k}|^2\right)$, respectively. Therefore, the random variables Z and J_k follow an exponential distribution such that $Z \sim Exp\left(\frac{1}{\sum_{\ell=1}^{L} \sigma_{\alpha_\ell} |D_{\ell}|^2}\right)$ and $J_k \sim Exp\left(\frac{1}{\sum_{\ell=1}^{L} \sigma_{\alpha_\ell} |D_{\ell k}|^2}\right)$.

Using (3.14) and (3.15), we obtain:

$$\Delta SNR = B\left(Z - \max_k J_k\right). \tag{3.16}$$

We denote $X = \max_k J_k$ the maximum value of A^2 exponential random variables: $J_1, J_2, ..., J_{A^2}$. Since both Z and X depend on the path gains, they are correlated random variables. Finding the joint distribution of Z and X is non trivial and, hence, we cannot obtain a closed-form expression of the distribution of ΔSNR. Consequently and as opposed to the particular single-path case, we can only compute $\mathrm{P}(\Delta SNR \leq \epsilon)$ empirically via Monte-Carlo simulations. Then, we follow the same approach as in the single-path case to determine an optimal codebook size.

3.3.3 Numerical results

Having analyzed $\mathrm{P}(\Delta SNR \leq \epsilon)$, we will use numerical simulations to find a good tradeoff between a small codebook size and a good performance in terms of ΔSNR.

In Fig. 3.3, we evaluate the probability $\mathrm{P}(\Delta SNR \leq \epsilon)$ as a function of $\log_2(A)$, where A is the codebook size, for the scenario: $M_T = 32$, $M_R = 4$, $N_T = 4$, $N_R = 2$, $P_T = 30$ dBm and $\sigma_{\alpha_\ell} = 1$ at 28 GHz carrier frequency. Both the transmitter and the receiver are equipped with ULAs of $\lambda/2$ spacing between the array elements. The path loss ρ is calculated as in [11, equation (2)]. The empirical results, for the multi-path case ($L = 3$), are obtained using Monte-Carlo simulations with $100,000$ independent channel realizations, whereas for the single-path case ($L = 1$) the closed-form expression is used.

We notice that the highest SNR obtained by the codebook is more probable to approach SNR_0 when the codebook size increases. In fact, as the size A becomes larger, the beams tend to be narrower which allows for a better alignment with the channel's best spatial path and a higher directional power gain. Moreover, as the parameter ϵ becomes smaller, the necessary codebook size to reach high probability becomes larger.

Nevertheless, we notice that ever increasing the codebook size, does not lead to a significant increase in $\mathrm{P}(\Delta SNR \leq \epsilon)$. For instance, in the case of a single-path and $\epsilon = 1$, the gain in the probability is 0.4 when we move from $A = 16$ to $A = 32$ while it is only 0.12 when we increase the size from $A = 32$ to $A = 64$.

We conclude that we do not need to continuously keep increasing the codebook size to obtain significantly higher SNR levels at the receiver. We can limit the number of beamforming vectors to reduce the beam alignment duration. For the proposed beam alignment schemes in the remaining of this chapter, we use the size $A = 32$ for the single-path channel and $A = 64$ for the multi-path channel.

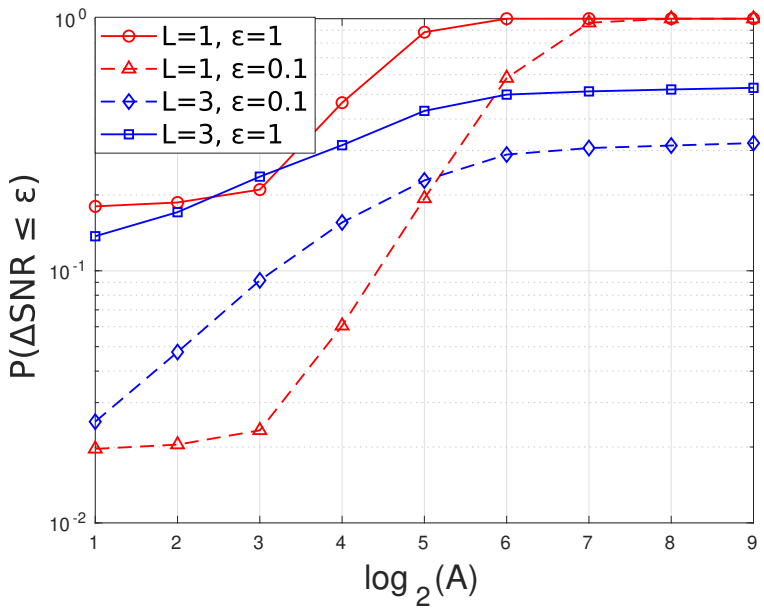

Figure 3.3: $\mathrm{P}\left(\Delta SNR \leq \epsilon\right)$ as a function of the codebook size A for the single-path and multi-path channels. Increasing the codebook size beyond a certain value, does not bring a significant performance improvement.

3.4 MAB formulation for beam alignment

Multi-armed bandit learning approaches have been considered recently to jointly tune the beam alignment vectors at the transmitter and receiver $\mathbf{f}_i$ and $\mathbf{w}_j$ in stochastic environments [26, 27, 28, 29, 30]. However, these works do not aim at minimizing the outage probability and they rely on a central authority or node that is able to compute the best pair of beams and to feedback the result to both the transmitter and the receiver.

Our main goal is to propose distributed and decoupled beam alignment policies at the transmitter and the receiver, which choose their own beamforming vectors $\mathbf{f}_{i_t}$ (beam-direction i_t at time t) and $\mathbf{w}_{j_t}$ (beam-direction j_t) independently. Furthermore, our policies do not require any knowledge on the channel state or statistics and are only based on a single bit of feedback.

Ideally, to minimize the outage probability in a time-varying environment in a decoupled way, the transmitter would like to select the best beam-direction i_t at time t solving the following problem:

$$\forall t, \quad \underset{I\in\{1,2,\ldots,A\}}{\text{minimize}} \quad P_{\text{out}}\left(I, j_t\right) \tag{3.17}$$

and the receiver would do the same:

$$\forall t, \quad \underset{J\in\{1,2,\ldots,A\}}{\text{minimize}} \quad P_{\text{out}}\left(i_t, J\right). \tag{3.18}$$

Several issues arise in this ideal formulation. First, the objective functions at each time instant are typically unknown at the transmitter and receiver. Indeed, the two

objectives are inter-dependent, which raises a causality issue, and the channel statistics may be unknown at the transmitter. Second, the definition of the outage probability becomes problematic as the environment seen by one of the nodes (either the transmitter or the receiver) depends on the decision process of the other node, which effectively results in a non-stationary environment.

All the above motivates the use of online optimization and, more specifically, the use of the adversarial multi-armed bandit framework [106] to propose adaptive beam alignment schemes that approach these goals and which do not require any assumptions on the network dynamics. Indeed, since our beam alignment problem is decoupled between the transmitter and receiver, the existing centralized approaches based on stochastic MABs [26, 27, 28, 29, 30] (which assume stochastic network dynamics) are no longer relevant. As mentioned above, even if the wireless channel is stationary the decoupled decision processes of each of the nodes may not be so.

3.4.1 Adversarial MAB formulation

The advantage of the adversarial MAB formulation described below is that, by definition [106], it does not rely on any assumptions on the network dynamics, which can vary in a completely arbitrary way including adversary or other non-stationary components. This feature is precisely what allows us to decouple the learning process between the transmitter and receiver.

In this formulation, the transmitter and the receiver are decision nodes that exploit separately an iterative online decision process as follows. At each time instant $t \in \{1, \ldots, \mathcal{T}\}$, where $\mathcal{T}$ is the time horizon or the transmission duration, a decision node chooses an action, in this case a beam-direction: $i_t \in \{1, \ldots, A\}$ at the transmitter and $j_t \in \{1, \ldots, A\}$ at the receiver. As a result of the transmission, we assume that the receiver is able to compute a binary ACK-type of reward:

$$r_{i_t,j_t}(t) \triangleq \begin{cases} 1, & \text{if } \mathrm{SNR}_{i_t,j_t}(t) \geqslant \xi, \\ 0, & \text{otherwise,} \end{cases} \tag{3.19}$$

which is then fed back to the transmitter. Based on this observed reward, the decision nodes will update their action choices and so on.

The intuition behind our chosen reward in (3.19) is that the overall average reward over the transmission horizon $\mathcal{T}$, i.e., $\frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} r_{i_t,j_t}(t)$, offers an approximation or an empirical measure of the outage probability. Moreover, assuming a stochastic channel model, the expected reward of a fixed beam pair (i, j) is directly linked to the outage probability defined in (3.7):

$$\begin{aligned} \mathbb{E}[r_{i,j}] &= \mathrm{P}[\mathrm{SNR}_{i,j} \geqslant \xi] \\ &= 1 - P_{\mathrm{out}}(i, j), \end{aligned} \tag{3.20}$$

where the expectation $\mathbb{E}[.]$ is taken over the randomness of the channel. As shown above, maximizing the expected reward is equivalent to minimizing the outage probability in the stochastic case or the centralized beam alignment problem. In our decoupled beam alignment, this average reward represents an empirical measure of the outage probability at each of the decision nodes.

3.4.2 Regret performance metric

In the MAB framework, the notion of *regret* has been considered as the relevant performance metric that evaluates the performance of an online policy [97, 26, 28]. The regret measures the gap in the average reward between the online policy and the *best fixed oracle policy in hindsight* over the time horizon $\mathcal{T}$. The latter is an ideal policy that maximizes the overall reward and relies on the non-causal knowledge of all the rewards during the entire horizon [39]. To be precise, the average regret at the transmitter side in our case writes as:

$$Reg_T = \frac{1}{\mathcal{T}} \left(\max_{I} \sum_{t=1}^{\mathcal{T}} r_{I,j_t}(t) - \sum_{t=1}^{\mathcal{T}} r_{i_t,j_t}(t) \right). \tag{3.21}$$

Similarly, the average regret at the receiver is

$$Reg_R = \frac{1}{\mathcal{T}} \left(\max_{J} \sum_{t=1}^{\mathcal{T}} r_{i_t,J}(t) - \sum_{t=1}^{\mathcal{T}} r_{i_t,j_t}(t) \right). \tag{3.22}$$

Property 1. *An online policy has the property of* no-regret *if its average regret decays to (or less than) zero asymptotically:* $\limsup_{\mathcal{T}\to\infty} Reg_Q \leq 0$, *with* $Q \in \{T, R\}$ *being the decision nodes.*

The no-regret property is an asymptotic performance guarantee ensuring that the online policy performs at least as good as the best fixed (or oracle) policy in hindsight (i.e., having perfect and non-causal knowledge of the networks dynamics throughout the horizon $\mathcal{T}$). Quite remarkably, this is achieved while relying only on strictly causal feedback amounting to a single bit of information.

3.5 Proposed beam alignment policies

As mentioned before, the beam alignment represents a crucial step in establishing a reliable link for data transmission in mmWave systems. In this section, we present three beam alignment policies exploiting adversarial MABs. The first policy is based on the original *exponential weights for exploration and exploitation* (EXP3) algorithm in [39], which will be detailed below. We then propose two novel policies by modifying the chosen actions' (or beams') rewards. Our new policies draw inspiration from the sparse nature of the mmWave channel and the correlation between successive beams, leading to better performance results.

3.5.1 EXP3-based beam alignment policy

The main idea of EXP3, introduced in [39], is to assign a probability to each possible action, and then choose an action according to this probability distribution, at each iteration t. Once an action is chosen, the decision node receives a reward and, as a result,

it updates the probability distribution following an exponential map that depends on the cumulative scores or rewards up to that instant. In our case, the beam-directions that often provide an SNR above the threshold ξ are the ones that are reinforced and have higher probabilities. In other words, EXP3 increases the probability of actions with good performance history, while not discarding completely the exploration of other actions that may perform better in the future; this effectively balances *the data exploitation versus data exploration*.

More precisely, at iteration t, the transmitter chooses a beam-direction i_t for data transmission according to the probability distribution $\hat{\mathbf{p}}_T(t)$ whose entries are defined for all $i \in \{1, 2, \ldots, A\}$ as:

$$\hat{p}_{T,i}(t) = (1-\gamma)\, p_{T,i}(t) + \frac{\gamma}{A}, \tag{3.23}$$

$$p_{T,i}(t) = \frac{\exp(\eta\, G_{T,i}(t-1))}{\sum_{k=1}^{A} \exp(\eta\, G_{T,k}(t-1))}, \tag{3.24}$$

where $G_{T,i}(t-1) = \sum_{\tau=1}^{t-1} \hat{r}_{T,i}(\tau)$ represents the cumulative score of action i. Since only the reward $r_{i_t,j_t}(t)$ of the chosen beam i_t can be observed at time t, we need to estimate other beams' rewards. For this, we define

$$\hat{r}_{T,i}(t) = \begin{cases} \dfrac{r_{i,j_t}(t)}{\hat{p}_{T,i}(t)}, & \text{if } i = i_t, \\ 0, & \text{otherwise}, \end{cases} \tag{3.25}$$

which represents an unbiased reward estimator for all beams i at time t [39].

The parameters $\eta > 0$ and $\gamma \in (0, 1]$ are tuning or learning parameters that tradeoff between data exploration and exploitation. Increasing the value of γ draws the probability distribution away from the exponential Gibbs distribution in (3.24) towards the uniform distribution, and hence moves away from data exploitation towards more exploration. An opposite behaviour is observed for the parameter η. When increasing η, the exponential Gibbs distribution moves away from the uniform distribution (i.e., when $\eta = 0$) towards a Dirac or a deterministic pure exploitation policy. Notice that both parameters have to be very carefully tuned to optimize the tradeoff exploration vs. exploitation.

After the transmission, the transmitter receives 1-bit of feedback or the value of the reward $r_{i_t,j_t}(t)$ from the receiver and updates the cumulative scores as follows:

$$G_{T,i}(t) = G_{T,i}(t-1) + \hat{r}_{T,i}(t), \quad \forall i \in \{1, 2, \ldots, A\}. \tag{3.26}$$

The new cumulative rewards will be then exploited to update the transmitter's probability distribution $\hat{\mathbf{p}}_T(t+1)$ for the next round and so on. These different steps are summarized in the algorithm BA-EXP3.

Remark that the BA-EXP3 online policy consists of two equally important ingredients: i) the exponential mapping in (3.24) that reinforces the probabilities to choose beams that have performed well in the past, while still exploring new beams; and ii) the

estimated rewards $\hat{r}_{T,i}(t)$ based on which the cumulative score in (3.26) is computed and which effectively evaluates the performance of past explored beams.

The receiver runs a similar algorithm independently from the transmitter. The two nodes' learning processes are linked via the feedback signaling. More precisely, the receiver uses its own probability distribution $\hat{\mathbf{p}}_R(t)$, defined similarly as in (3.23), to choose a beam-direction j_t at round t. Then, the receiver evaluates the binary reward for the chosen beam-directions i_t and j_t by comparing the received SNR with the threshold ξ as in (3.19), and then updates its probability distribution for the next round. We further assume that the obtained reward is sent back to the transmitter via a reliable control channel as a 1-bit feedback information [1].

BA-EXP3: Exponential Weight for Beam Alignment at Tx

Parameters: $\eta > 0$ and $\gamma \in (0, 1]$

Initialization: $G_i(0) = 0$ and $p_{T,i}(1) = 1/A$, $\forall i$

Repeat for $t = 1, 2, ..., \mathcal{T}$

- Select action i_t with probability distribution $\hat{\mathbf{p}}_T(t)$
- Receive feedback $r_{i_t,j_t}(t) \in \{0, 1\}$ from Rx
- Update the cumulative rewards as in (3.26)
- Update the distribution $\hat{\mathbf{p}}_T(t+1)$ via (3.23)

The following theoretical result from [39] indicates that the expected average regret of the algorithm BA-EXP3 decays to zero as $\mathcal{O}(1/\sqrt{\mathcal{T}})$. This decay rate is optimal and cannot be improved in the absence of strong stationarity assumptions regarding the underlying network dynamics [111, 32]. As argued in Sec. 3.4, in our *distributed* beam alignment problem, the wireless environment depends on the other node's decisions and, hence, does not evolve following a stochastic stationary process.

Corollary 1 (Theorem 1 in [39])**.** *If the BA-EXP3 beam alignment policy is run at both the transmitter and receiver with the parameters* $\eta = \dfrac{\gamma}{A}$ *and* $\gamma = \min\left\{1, \sqrt{\dfrac{A \log A}{(e-1)\,\mathcal{T}}}\right\}$ *for a horizon* $\mathcal{T}$*, then the expected average regret is upper bounded as:*

$$\mathbb{E}[Reg_Q] \leq 2\sqrt{e-1}\sqrt{\frac{A \log A}{\mathcal{T}}}, \tag{3.27}$$

with $e = \exp(1)$*, and the expectation is taken over the randomness of the BA-EXP3 policy.*

The upper bound of the expected average regret in (3.27) shows that the BA-EXP3 policy is asymptotically optimal when $\mathcal{T}$ grows large. Also, when the transmission horizon $\mathcal{T}$ is finite or small, this bound also provides a worst-case guarantee in terms of the gap between the empirical outage of BA-EXP3 compared with the ideal oracle policy, which depends only on $\mathcal{T}$ and the number of available beams A.

[1] The control channel can be either a microwave channel as proposed in the ECMA 387 standard [107], or a mmWave channel as in the IEEE 802.15.3c [108] and IEEE 802.11ad [109] standards. On one hand, the directional mmWave channel is low-cost, but may suffer from poor reliability due to difficult propagation characteristics. On the other hand, the omni-directional microwave channel is more reliable at the expense of additional hardware and energy consumption [110].

3.5.2 Modified exponential weights algorithm (MEXP3)

Measurement campaigns have demonstrated the existence of only a few multi-path components in the mmWave propagation environment, which leads to a limited number of available propagation paths providing a high enough SNR for data transmission. We exploit this channel sparsity to adapt the BA-EXP3 algorithm and identify faster the good beam-directions.

To do so, let us denote the global reward matrix $\mathbf{R}(t) \in \{0,1\}^{A\times A}$ such that

$$\mathbf{R}(t) = [r_{i,j}(t)]_{\substack{1\leq i\leq A\\ 1\leq j\leq A}} \tag{3.28}$$

where the rewards are defined in (3.19). The matrix $\mathbf{R}(t)$ contains the rewards of all possible pairs (i,j) at time t and is not fully available at any of the two nodes. A typical example of a reward matrix $\mathbf{R}(t)$ is illustrated in Fig. 3.4 for a particular mmWave channel setting and $A = 16$. Due to the characteristics of the mmWave channel, the matrix $\mathbf{R}(t)$ has a particular sparse structure. The few non-zero entries are all grouped in one or a few clusters and correspond to the set of good beam-directions.

Hence, the goal of our online policies is to identify the indices i (beam-directions at Tx) and j (beam-directions at Rx) that correspond to a unit value in this matrix (to avoid an outage event and guarantee a minimum SNR at the receiver). Based on this observation, we leverage the structure of the reward matrix to define a modified reward:

$$\tilde{r}_{T,i}(t) = \begin{cases} \dfrac{-1}{1-\hat{p}_{T,i}(t)}, & \text{if } \ i = i_t \text{ and } r_{i_t,j_t}(t) = 0, \\ \dfrac{\beta}{\hat{p}_{T,i}(t)}, & \text{if } \ i = i_t \text{ and } r_{i_t,j_t}(t) = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{3.29}$$

where $\beta \geq 1$ is a weighting parameter which affects the beam selection probabilities and, hence, represents another parameter that tradeoffs between data exploration and exploitation and which needs to be carefully tuned.

The intuition behind the above modified reward is to reinforce good beam-directions (i.e., the ones that provide $r_{i_t,j_t}(t) = 1$) by associating them a reward β-times higher than the original BA-EXP3. Moreover, the poor beams ($r_{i_t,j_t}(t) = 0$) are penalized by associating them a strictly negative reward as opposed to zero. Dividing by the quantity $1 - \hat{p}_{T,i_t}(t)$ leads to an important and fast penalization of a past good beam that has accumulated a high probability to be chosen, but which has become obsolete because of changes in the mmWave environment. Also, dividing by $1 - \hat{p}_{T,i_t}(t)$ insures a soft penalization of a beam with low probability to avoid discarding it completely as it may become a future good beam. To sum up, this denominator penalizes the poor beams according to their past performance and not randomly by just assigning a negative reward. Therefore, the modified reward encourages the algorithm to adapt faster to the changes in the channel and to keep track of good beam-directions.

MEXP3: Modified Exponential Weight for Beam Alignment at Tx

Parameters $\eta > 0$, $\beta \geq 1$ and $\gamma \in (0, 1]$

Initialization: $G_i(0) = 0$ and $p_{T,i}(1) = 1/A$, $\forall i$

Repeat for $t = 1, 2, ..., \mathcal{T}$

- Select action i_t with probability distribution $\hat{\mathbf{p}}_T(t)$
- Receive feedback $r_{i_t,j_t}(t)$ from Rx
- Construct the modified rewards $\tilde{r}_{T,i}(t)$ as in (3.29)
- Update the cumulative rewards:

$$G_{T,i}(t) = G_{T,i}(t-1) + \tilde{r}_{T,i}(t), \ \forall i$$

- Update the distribution $\hat{\mathbf{p}}_T(t+1)$ via (3.23)

Although the new algorithm MEXP3 may seem quite similar to the original algorithm BA-EXP3 at first, our new modified reward $\tilde{r}_{T,i_t}(t)$ in (3.29) results in a very different behavior with respect to the regret and other performance metrics. This modified reward changes one of two key ingredients of the original EXP3 algorithm: the cumulative scores $G_{T,i}(t) = \sum_{\tau=1}^{t} \tilde{r}_{T,i}(\tau)$ that evaluate the performance of the past explored beams, which are then mapped on the probability simplex (via the exponential map). In particular, the no-regret proof and showing that the expected cumulative regret of MEXP3 grows sub-linearly with respect to the time horizon is very different than the proof in [39]. One of the main challenges we have overcome is that $\tilde{r}_{T,i}(t)$ no longer represents an unbiased reward estimator for beam i. All the details behind our proof are presented in the Appendix.

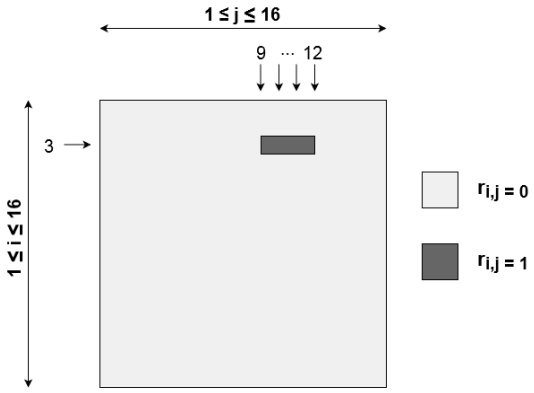

Figure 3.4: Illustration of a typical reward matrix in mmWave channels, for the setting: $M_T = 32$, $M_R = 4$, $L = 1$, $A = 16$, $\xi = 6$ dB and a carrier frequency $f_c = 28$ GHz.

Theorem 1. *If the MEXP3 policy is run at both the transmitter and receiver with the parameters* $A \geq 3$, $\eta = \dfrac{\gamma}{\beta A}$, $\gamma = \min\left\{1 - \frac{2}{A}, \sqrt{\frac{2A \ln A}{e\ \mathcal{T}}}\right\}$ *and* $\beta \geq \max\left\{1, \sqrt{\dfrac{2}{A}}\ \left(\gamma - \dfrac{\gamma}{A}\right)^{-1}\right\}$, *then the expected average regret is bounded by:*

$$\mathbb{E}[Reg_Q] \leq \sqrt{2\ e}\ \sqrt{\frac{A \ln A}{\mathcal{T}}}. \tag{3.30}$$

Notice that the expectation of the regret in Theorem 1 is taken only with respect to the random sequence of the chosen beam-directions. This means that the no-regret property holds irrespective from the underlying system dynamics, which can be arbitrary and even non stationary.

The above result shows that MEXP3 provides an optimal asymptotic regret performance similarly to the original BA-EXP3; the regret decays as $\mathcal{O}(1/\sqrt{\mathcal{T}})$. Even though we cannot improve this decay rate under arbitrary and non-stationary network dynamics, the upper bound we obtain for MEXP3 is tighter than the one for BA-EXP3 in the multiplicative constant, which may be important in the finite horizon regime. This indicates that MEXP3 outperforms BA-EXP3 in terms of regret and that the gap between the two algorithms is larger for relatively short transmissions (finite $\mathcal{T}$).

3.5.3 Nearest neighbour-aided beam tracking (NBT-MEXP3)

Here, we propose an additional modification by using an additional information to accelerate further the beam search and tracking. Our empirical observations of the temporal evolution of the reward matrix $\mathbf{R}(t)$ indicates a temporal correlation between the locations of its unit-valued clusters that depends on the mobility of the receiver (speed, orientation, etc.) and on the wireless characteristics (blockage, LOS, NLOS, etc.). The location of the clusters does not change abruptly or randomly but rather smoothly following the mobility of the receiver.

Concretely, this means that future good beam-directions are more likely to be among the neighboring directions that have performed well in the past. Therefore, we can exploit this intuition to keep track of good beams with the aid of their nearest neighbours. This new feature increases the tracking speed of the good beams by adapting to the user's mobility and other changes in the channel. For this, we modify the rewards of the non-chosen beams as follows:

$$\tilde{r}_{T,k}(t) = \frac{\beta' \, r_{i_t,j_t}(t)}{\hat{p}_{T,i_t}(t)}, \quad \forall k \in V_{i_t}, \tag{3.31}$$

where $V_{i_t} = \{i_t - 1, i_t + 1\}$ is the set of the nearest neighbors[2] of the chosen beam-direction i_t at time t and parameter $\beta' \in [1, \beta]$, which plays a similar role as β for the neighbouring beams.

Combining the modified reward in (3.29) for the chosen action i_t with the reward in (3.31) for its neighbors, we construct a new reward vector $\tilde{\mathbf{r}}_T(t) = [\tilde{r}_{T,k}(t)]_{k\in\{1,\ldots,A\}}$,

[2]We focus only on the two nearest neighbors for simplicity reasons and also based on our empirical observations. Choosing a larger (or optimized) size for the neighbors' set could be of interest for future research.

which is used to update the cumulative rewards for each beam-direction, as follows

$$\tilde{r}_{T,k}(t) = \begin{cases} \dfrac{(-1)^{1+r_{i_t,j_t}(t)}\ \beta^{\ r_{i_t,j_t}(t)}}{1 - r_{i_t,j_t}(t) + (-1)^{1+r_{i_t,j_t}(t)}\hat{p}_{T,k}(t)}, & \text{if } k = i_t, \\ \dfrac{\beta'\ r_{i_t,j_t}(t)}{\hat{p}_{T,k-1}(t)}, & \text{if } k = i_t + 1, \\ \dfrac{\beta'\ r_{i_t,j_t}(t)}{\hat{p}_{T,k+1}(t)}, & \text{if } k = i_t - 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.32}$$

The resulting NBT-MEXP3 algorithm is detailed below.

NBT-MEXP3: Nearest Neighbour-aided Beam Tracking with MEXP3 at Tx

Parameters $\eta > 0$, $1 \leq \beta' \leq \beta$ and $\gamma \in (0, 1]$

Initialization: $G_i(0) = 0$ and $p_{T,i}(1) = 1/A$, $\forall i$

Repeat for $t = 1, 2, ..., \mathcal{T}$

- Select action i_t with probability distribution $\hat{\mathbf{p}}_T(t)$
- Receive feedback $r_{i_t,j_t}(t)$ from Rx
- Construct the reward vector $\tilde{\mathbf{r}}_T(t)$ as in (3.32)
- Update the cumulative rewards:

$$G_i(t) = G_i(t-1) + \tilde{r}_{T,i}\,(t)\,,\ \forall i$$

- Update the distribution $\hat{\mathbf{p}}_T(t+1)$ via (3.23)

Although finding a sub-linear upper bound for the regret of NBT-MEXP3 is not trivial, our extensive numerical simulations indicate that the NBT-MEXP3 policy holds the no-regret property asymptotically.

Conjecture 1. *The proposed NBT-MEXP3 beam alignment policy has the no-regret property and the average regret decays to zero as $\mathcal{O}(1/\sqrt{\mathcal{T}})$, similarly to BA-EXP3 and MEXP3.*

The proof of the above conjecture is left open for future work. By following a similar approach as in the proof of Corollary 1 and Theorem 1, an encountered difficulty comes from the ratio $\dfrac{p_{T,k}(t)}{\hat{p}_{T,i_t}(t)}$, $k \in V_{i_t}$, which appears in the expectation of the regret and which cannot be bounded appropriately. This term is due to our modified reward $\tilde{r}_{T,i}(t)$ in (3.32) assigning a non-zero reward to the neighbouring beam of a good direction.

3.6 Numerical results

In this section, we evaluate the performance of the proposed beam alignment algorithms in terms of regret, outage, throughput and delay in a typical mmWave setting

described in Sec. 3.2 and specified here. Notice that our online beam alignment policies and their theoretical guarantees do not rely on any assumptions on the underlying network dynamics. This implies that the conclusions drawn below carry over many other mmWave settings incorporating various practical aspects and specifications.

The plotted curves are averaged over $10,000$ scenarios or time-varying channel realizations over the horizon $\mathcal{T}$. Our online policies do not rely on any initial knowledge of the wireless environment (i.e., the beam search starts with a random choice following the uniform distribution). The channel is assumed to remain constant during a transmission frame T_c, which represents the channel coherence time. The duration T_c consists of several sub-frames such that each sub-frame represents one iteration of the online beam alignment policies at both Tx and Rx or, more precisely, the time interval between two successive feedback signals. In our simulations, we consider a channel coherence time $T_c = 1.3$ ms [112] and a sub-frame duration of 250 μs [113]. This results in 5 sub-frames per coherence interval, meaning that the channel conditions change every 5 iterations of our online policies, because of device mobility, time-varying wireless characteristics, etc.

We compare our policies with existing ones in the literature but also with several relevant benchmarks, which we briefly described below.

- **Centralized-UCB**: the centralized beam alignment policy proposed in [28] based on stochastic MABs and the upper-confidence bound (UCB) algorithm.
- **Exhaustive search**: the brute-force policy that tries all A^2 beam pairs (one at each iteration) in a round-robin fashion and selects the best one after A^2 iterations; this optimal beam is then exploited until the channel changes and the process is reinitialized.
- **Rand**: the random beam-direction is drawn following the uniform distribution.

3.6.1 System parameters

We consider the mmWave MIMO point-to-point link of Fig. 3.1 with $M_T = 32$, $N_T = 4$, $M_R = 4$ and $N_R = 2$. Both nodes are equipped with ULAs with $\lambda/2$ spacing between their elements. The transmission power is $P_{\text{tr}} = 37$ dBm. The size of the beamforming codebook is $A = 32$ for single-path channels, which ensures a good tradeoff between beam alignment accuracy and exploration cost. The threshold ξ for the SNR at the receiver is fixed at 6 dB.

Regarding the wireless channel matrix, the commonly used geometric model in (3.1) is adopted with $\alpha_\ell \sim \mathcal{N}(0,1)$, the carrier frequency $f_c = 28$ GHz and a bandwidth of 1 MHz, which meets the narrowband channel assumption according to the maximum delay spread measurements in [114]. The noise power density equals -174 dBm/Hz. The path loss ρ is calculated following the close-in free space model in [100, 114] as follows

$$\rho = 20 \log \frac{4\pi f_c}{c} + 10\ n_p \log D \quad [\text{dB}], \tag{3.33}$$

where $c = 3 \times 10^8$ is the speed of light, D is the distance between the transmitter and receiver and n_p is the path loss exponent, which equals 2.1 for LOS and to 3 for NLOS [114, Table 3]. The Doppler shift is updated every T_c such that $\nu_\ell = \frac{v\ f_c}{c}$ as in [13]. Unless stated otherwise, we consider a single-path channel ($L = 1$). For the multi-path channel $L = 3$, we consider a LOS path combined with two NLOS paths determined by two reflectors positioned randomly between the transmitter and the receiver for each new channel realization.

The location of the transmitter is fixed (e.g., a base station). The receiver (a mobile user) is assumed to move in the area covered by the transmitter's ULA within a distance less than 200 m. The mobility model in (3.4) is used with the typical parameters: speed $v_{max} = 30$ km/h, acceleration $a_{max} = 2$ m/s^2 and rotation speed $\omega_{max} = \pi/4$ rad/s. The position of the receiver is updated every transmission frame of duration 1.3 ms, which corresponds to the channel coherence time under our dynamic conditions [112]. For each receiver position, a new channel matrix $\mathbf{H}$ is computed by updating its parameters as detailed in Section 3.2. In other words, the channel conditions change (implying that the good beam-directions that meet the SNR requirement also change) every 5 iterations of our algorithms in the figures below.

The learning parameters of our online policies are chosen empirically based on extensive numerical simulations. Here, we set $\eta = 0.02$, $\gamma = 0.001$ for BA-EXP3; $\eta = 0.023$, $\gamma = 0.03$ and $\beta = 10$ for MEXP3; $\eta = 0.01$, $\gamma = 0.001$, $\beta = 10$ and $\beta' = 5$ for NBT-MEXP3. Naturally, we exploit the values and the ranges obtained in Corollary 1 and Theorem 1 as starting point. Notice that these values are optimal only with respect to the upper-bounds of the regret and are not necessarily optimal in terms of the actual regret. A more efficient way to fine tune these parameters is a non-trivial issue to be investigated in future work.

3.6.2 Average regret

We start by comparing our policies to the original BA-EXP3 and the Centralized-UCB at the transmitter side. The average regret is plotted in Fig. 3.5. We also plot the upper bounds of the expected average regret of BA-EXP3 and MEXP3 given in Corollary 1 [39] and Theorem 1.

We first notice that all policies based on exponential learning: BA-EXP3, MEXP3 and NBT-MEXP3, yield an average regret lower and decaying faster compared with Centralized-UCB. This can be explained by the fact that the Centralized-UCB policy has a larger set of choices, the A^2 beam-direction pairs, whereas the distributed policies have a set of only A beam-directions. The larger search set of Centralized-UCB requires more data exploration, which leads to more regret.

Also, both our modified policies outperform the original BA-EXP3. They are more adapted to the varying mmWave channel since they are inspired from its particular structure. The upper-bounds validate our theoretical results: the obtained bound for MEXP3 in Theorem 1 is tighter than the bound for the BA-EXP3.

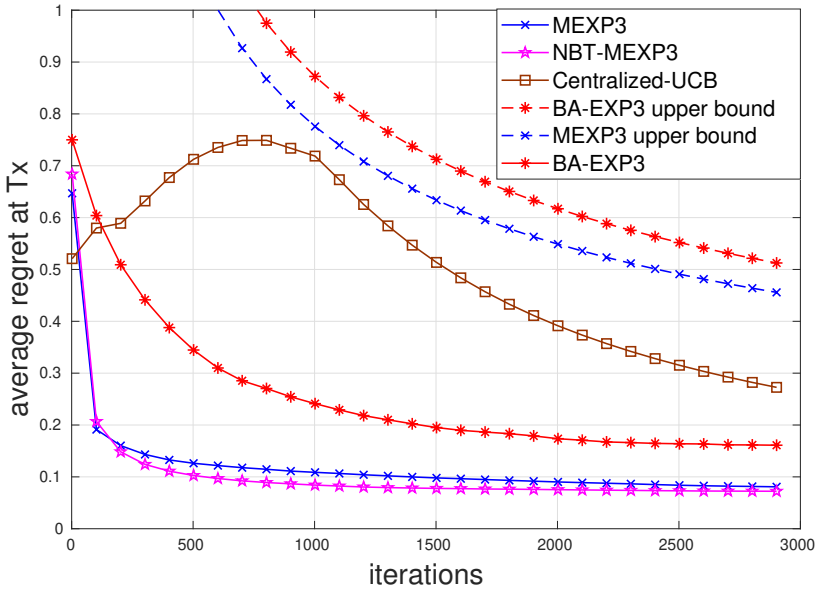

Figure 3.5: The average regret at the transmitter decays faster for our proposed policies MEXP3 and NBT-MEXP3, with a slight advantage for the latter. The three distributed policies based on exponential learning clearly outperform the Centralized-UCB policy.

3.6.3 Outage

Fig. 3.6 illustrates the empirical outage of the different beam alignment policies. Our new algorithms, MEXP3 and NBT-MEXP3, outperform clearly the original BA-EXP3 and the other policies. This highlights the interest of exploiting the special structure of the mmWave channel and modifying the rewards as in MEXP3. The nearest neighbours additional reward modification in the NBT-MEXP3 policy provides a slight performance improvement compared to MEXP3. Also, the exhaustive search policy results in high outage similarly to the random policy. This is mainly caused by the fact that only 5 beam pairs can be explored from the total of $A \times A = 1024$ possibilities before the change in the channel conditions occurs. In turn, this effectively renders the gathered information about those 5 trials outdated and irrelevant.

Regarding the number of iterations needed to reach an outage below 10% (around 2000 iterations for MEXP3), it is equivalent to a duration of 500 ms. Although 500 ms may seem long at first, it is a low price to pay for the entire transmission duration. Indeed, once these early learning stages have passed, our method is capable to adapt to the network changes and track good beams while reliably transmitting data. At the opposite, traditional methods have to perform dedicated training and identify good beam-directions *every time the channel has changed* (every T_c) before transmitting any data at all, having a crucial impact on the effective performance.

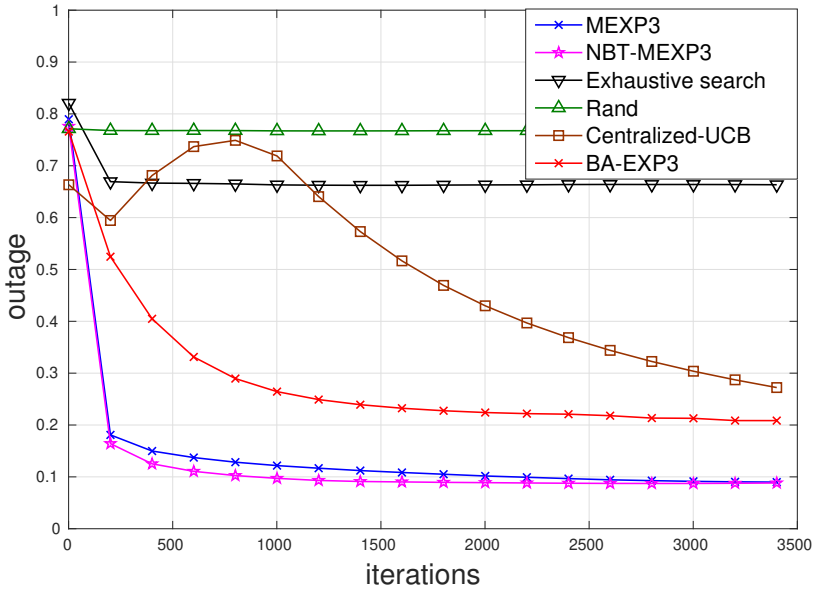

Figure 3.6: Our novel policies, MEXP3 and NBT-MEXP3, outperform the original BA-EXP3, Centralized-UCB as well as the other benchmarks. This shows the importance of exploiting the structure of the mmWave channel and of our modified rewards to reach lower outage.

3.6.4 Effective throughput

Fig. 4.4 illustrates the evolution of the average achievable rate as a function of the iterations. In our proposed policies, we do not separate the communication in two distinct phases: beam alignment training and data transmission. Instead, the transmitter and receiver communicate effectively during the whole frame while adjusting the beam-directions (at the cost of higher outage levels in the early learning stages). In Fig. 4.4, we make the same assumption for Centralized-UCB and exhaustive search policies for comparison purposes. Our novel policies MEXP3 and NBT-MEXP3 outperform the original BA-EXP3 and the other benchmarks in terms of the speed in reaching higher data rates.

3.6.5 Average delay

Here, we compare the average beam alignment delay of the three exponential learning policies, which represents the average time interval required to identify good beam-directions that provide an SNR above the threshold for a given channel. Fig. 3.8 depicts the average delay as a function of the SNR threshold for two different codebook sizes $A = \{8, 32\}$. We notice that reaching higher SNR thresholds require more exploration time to find good beams. This highlights the *latency vs. reliability tradeoff* between the delay and the SNR at the receiver.

Regarding the impact of the codebook size A, Fig. 3.8 shows that the average delay increases with the codebook size. Indeed, the beams of a larger codebook are narrower and induce more delay given that the search set of beams is larger. However, using a

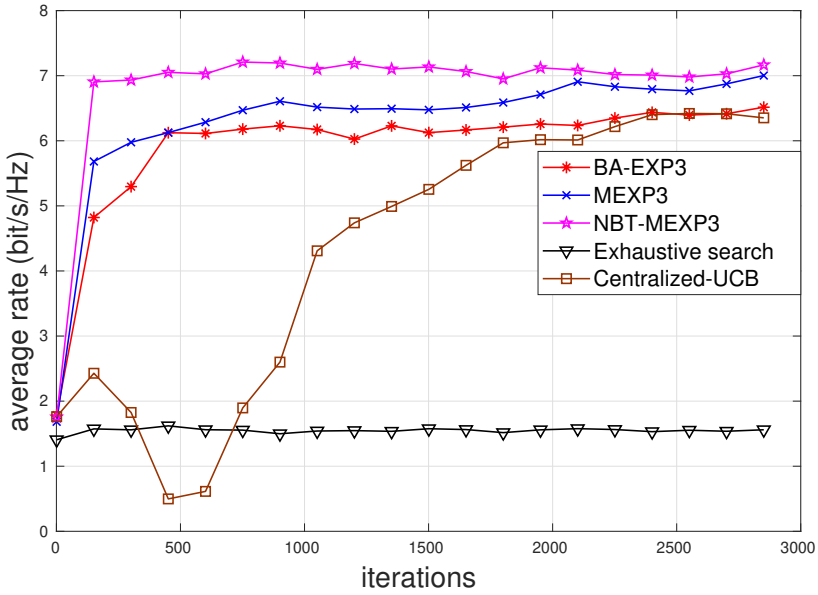

Figure 3.7: Exploiting neighbouring beams (NBT-MEXP3) is beneficial in achieving higher average rate.

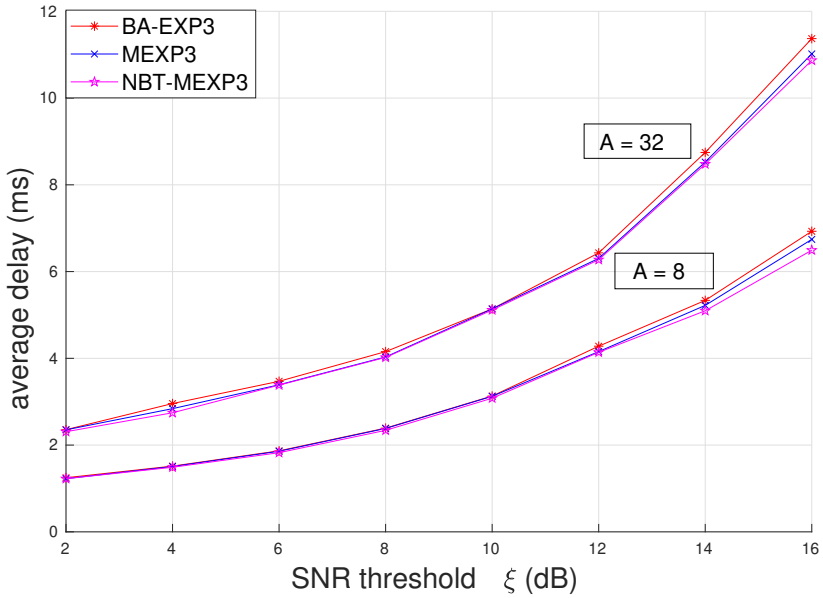

Figure 3.8: The average delay as function of the SNR threshold ξ and for codebook sizes $A = \{8, 32\}$. All exponential learning policies lead to similar performance with a slight advantage for the NBT-MEXP3 policy.

large codebook allows to focus the signal's energy in a more compact angular domain to reach higher beamforming gains, which illustrates again the latency vs. reliability tradeoff.

3.6.6 Impact of user mobility

We now investigate the ability of our NBT-MEXP3 and MEXP3 algorithms to support high-mobility conditions and their impact on the empirical outage. We compare the outage performance obtained with the following mobility parameters: $v_{max} = 30$ km/h, $a_{max} = 2$ m/s^2 and $\omega_{max} = \pi/4$ rad/s (low-mobility); and with more dynamic parameters: $v_{max} = 110$ km/h, $a_{max} = 5$ m/s^2 and $\omega_{max} = \pi/2$ rad/s (high-mobility). In Fig. 3.9, we can see that increasing the mobility of the receiver leads to higher outage levels as expected. Higher mobility implies more frequent changes in the mmWave channel which affects the quality of the beam alignment and results in lower SNR. Moreover, the proposed algorithms need more iterations to reach low outage levels compared to the low-mobility setting. Fig. 3.9 shows that our proposed policies may be suitable for high-mobility mmWave applications with an increased delay cost.

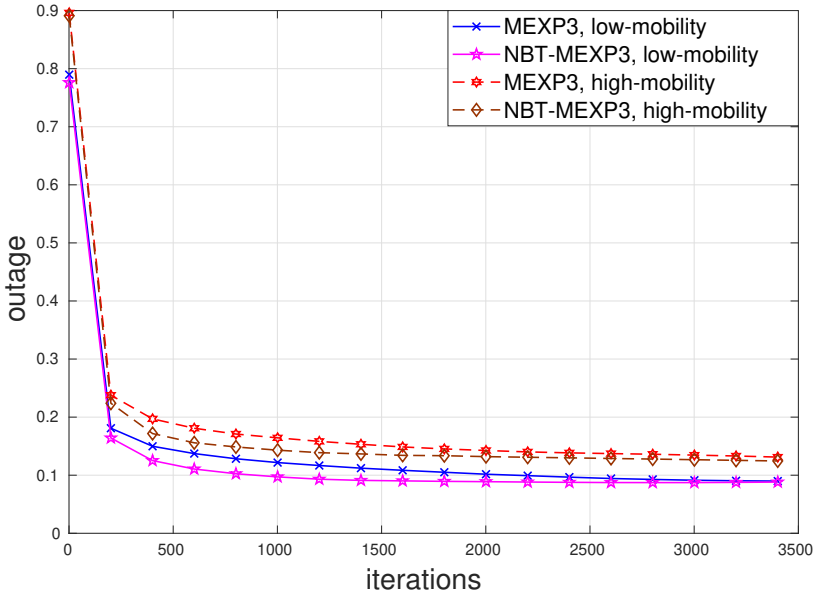

Figure 3.9: Impact of the user's speed: higher mobility leads to higher outage levels.

3.6.7 Impact of multi-path channels

We compare the outage performance of the proposed beam alignment policies, MEXP3 and NBT-MEXP3, in a multi-path channel (when $L = 3$) composed of one LOS path and two NLOS components and the single LOS channel (when $L = 1$). Fig. 3.10 shows the ability of the proposed policies to adjust the beams even in a multi-path channel with an additional exploration cost, as it takes longer to reach lower outage levels. This can be explained by the less favorable propagation conditions (involving higher path loss for NLOS paths combined with possible destructive combination of multi-path components).

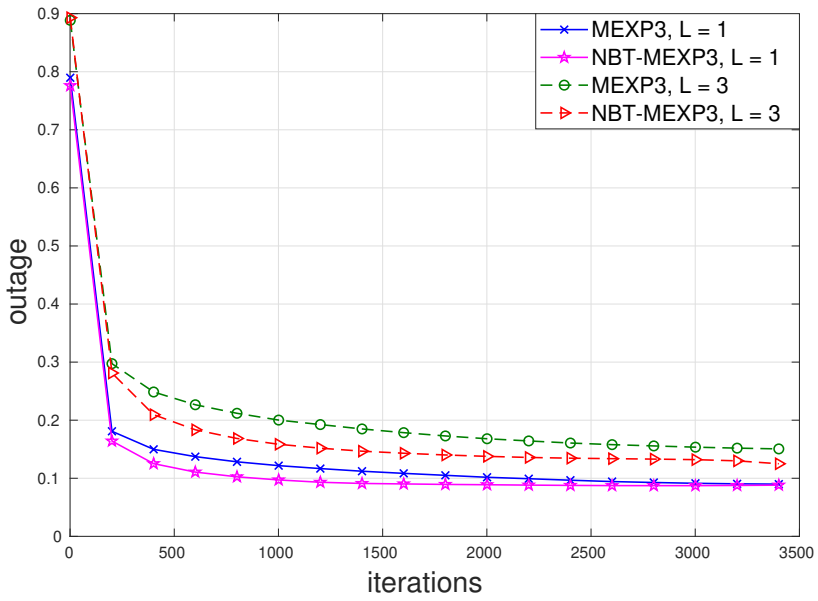

Figure 3.10: multi-path components and NLOS paths lead to higher outage and increase the exploration cost

3.6.8 Complexity vs. performance

Regarding the complexity of the various policies, we discuss here the scalability of each iteration in function of the codebook size A. Exhaustive search and the random policies have a constant cost, $\mathcal{O}(1)$, as only one pair of beams can be tested at each sub-frame or iteration. For Centralized-UCB policy, the complexity of each iteration scales linearly with the number of available choices, in this case the number of joint beamforming pairs, $\mathcal{O}(A^2)$. The complexity of all distributed online policies based on the adversarial MAB framework: BA-EXP3, MEXP3 and NBT-MEXP3, also scales linearly with the number of choices, which in this case represents the number of individual beams at each decision node, i.e., $\mathcal{O}(A)$ (the size of the probability distributions updated at each iteration).

The above highlights the tradeoff between complexity and performance. The least complex policies are the ones which perform quite poorly in terms of performance (random and exhaustive search). Remarkably, our online policies allows one to distribute the complexity between the transmitter and receiver, resulting in relatively low complexity policies that are also capable of adapting to the dynamic and unpredictable changes in the network. Centralized-UCB suffers from the larger number of joint beam pairs, A^2, and, because it inherently relies on stochastic and stationary channel assumptions, it might not be suitable for multi-user scenarios, in which the network dynamics will depend on other decision nodes and will hence be non-stationary.

On the contrary, our distributed online policies rely on no assumptions on the underlying network dynamics and can be extended to multi-user scenarios as discussed in the next section.

3.7 Extensions to wideband, multi-user mmWave networks

For the sake of simplicity and clarity of presentation, we have focused on a narrowband single point-to-point mmWave link. The extension to wideband mmWave networks (multi-carrier or single-carrier) involves adapting the codebook design (specifically the digital part of the beams) as in [115, 116]. Once this is done, our online policies can be easily exploited. The amount of feedback bits over the control channel would equal the number of carriers (one bit per carrier) in the multi-carrier case.

The extension of the proposed policies to multi-link networks is straightforward. The adversarial MAB framework and our online policies based on the exponential weights algorithm rely on no assumptions on the underlying network dynamics, which can easily incorporate the interference from other transmitter-receiver pairs. The one-bit feedback mechanism would operate in a similar manner for each individual pair, under the mild assumption that each one has access to an interference-free control channel. The definitions of the outage and the feedback reward in (3.7) and (3.19) can be also adapted using the signal-to-interference-plus-noise ratio (SINR) instead of the SNR.

For illustration purposes, we consider two closely located pairs such that each receiver may experience interference from the beams of the neighbouring transmitter. We use the same system and mobility parameters for the two pairs, which are identical to the ones described in Sec. 3.6. The locations of the transmitters are fixed and separated by a distance of 100 m. Fig. 3.11 represents the overall system outage. We consider that the system is in outage if at least one of the two links is in outage (SNR below the threshold). We remark that the same conclusions hold for a two-link network similarly to the single link case in Fig. 3.6.

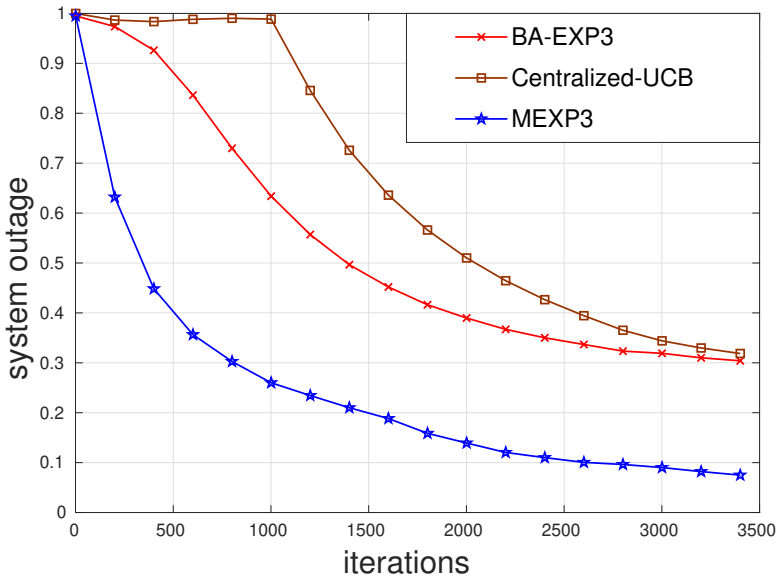

Figure 3.11: System outage as a function of iterations in a two-link network. MEXP3 leads to lower outage.

In the uplink (multiple access channels) or the downlink (broadcast channels), the design of the multi-user beamforming codebooks is much more involving and requires non trivial interference management, allocation of the multiple antennas over the served users, etc. Nevertheless, once the codebooks have been properly designed and each transmitter and receiver has access to its own finite set of actions (beams), the multi-armed bandits framework and our online policies based on the exponential weights algorithm can be easily adapted. The feedback mechanism would require a number of control channels equal to the number of users (as in the multi-link networks) knowing that only a single bit required to/from each of the users. At last, in the downlink, the transmitter would have to wait for all one-bit feedback signals to arrive from the receivers before transmitting new data.

3.8 Conclusion

In this chapter, we address the beam alignment problem in dynamic mmWave networks. We exploit the adversarial multi-armed bandit framework to design distributed policies, in which the transmitter and the receiver choose their beams separately while relying only on a one-bit feedback. Building on the well known exponential weights algorithm (EXP3), we propose two novel beam alignment policies, MEXP3 and NBT-MEXP3, that exploit the mmWave characteristics and lead to tracking optimal beam-directions more efficiently. We prove rigorously that our MEXP3 online policy has the no-regret property, while a conjecture is provided for NBT-MEXP3 (validated via extensive simulations). The performance of the proposed algorithms is demonstrated via numerical results in terms of regret, outage, throughput and average delay in a practical mmWave setting. We show that our policies outperform the original BA-EXP3 and other existing centralized policies by being capable to adapt to the rapid and unpredictable changes of the mmWave channel. In the following chapter, we investigate more a machine learning tools from the deep learning framework to steer the beams at the transmitter.

4

Deep learning for mmWave beam prediction

4.1 Introduction

The second part of the thesis contributions is presented in this chapter. In chapter 3, we presented distributed MAB-based policies, to align the beams of both the transmitter and the receiver with one-bit feedback. Here, we exploit the deep learning framework to tackle the beam alignment problem in mmWave networks exploiting the available CSI at the sub-6 GHz band. First, we employ a neural network to map sub-6 GHz channels into mmWave beamformers leveraging the unsupervised learning. The universal approximation capabilities of DNNs allow to learn such complex mapping. Then, we propose to use federated learning to predict the beams of a multiple-links network to avoid local data exchange with a central server and overcome training data scarcity. Numerical results are presented to illustrate the performance of the proposed methods. At the end, we conduct a comparison between the proposed methods in this chapter and the previous one particularly in term of communication rate and computation complexity.

4.2 System Model and Problem Formulation

We consider a wireless network composed of J access point – user links. For simplicity, we assume that each access point (AP) serves a unique user [117]. Each AP is equipped with a sub-6 GHz receive-array of M antennas (uplink) and a mmWave transmit-array of N antennas (downlink) as illustrated in Fig. 4.1. Each user is equipped with a single mmWave receive antenna and a single sub-6 GHz transmit antenna. The communication in each link is performed via multiple carrier frequencies, i.e., orthogonal frequency-division multiplexing (OFDM).

We focus on a link between an AP and its intended user, under the assumption that only one link is active at a given moment or that the multi-link interference is considered as additive noise. The justification for this simplifying assumption is provided at the end of this section.

The AP aims at predicting the downlink mmWave beamforming vector for its user based on the uplink received signal at the sub-6 GHz band, which can be written in

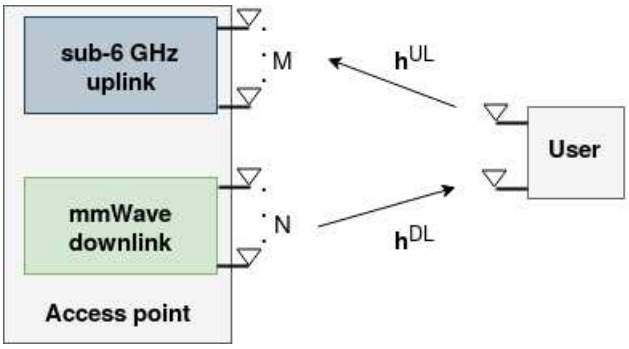

Figure 4.1: An access point – user link.

the ℓ^{th} subcarrier with $\ell \in \{1, \ldots, L\}$ as follows:

$$\mathbf{y}^{\text{UL}}[\ell] = \mathbf{h}^{\text{UL}}[\ell]\; x^{\text{UL}}[\ell] + \mathbf{n}^{\text{UL}}[\ell], \tag{4.1}$$

where $\mathbf{h}^{\text{UL}}[\ell] \in \mathbb{C}^{M\times 1}$ is the sub-6 GHz uplink channel vector; $x^{\text{UL}}[\ell]$ is the uplink pilot symbol having an average power P^{UL} such that $\mathbb{E}[\ |x^{\text{UL}}[\ell]|^2\ ] = P^{\text{UL}}/L$; and $\mathbf{n}^{\text{UL}}[\ell]$ is the additive Gaussian noise vector at sub-6 GHz.

In the downlink, an analog transceiver is used to transmit data via the mmWave antenna array. The received signal, in the ℓ^{th} subcarrier, can be written as:

$$y^{\text{DL}}[\ell] = \mathbf{h}^{\text{DL}\dagger}[\ell]\; \mathbf{f}\; x^{\text{DL}}[\ell] + n^{\text{DL}}[\ell], \tag{4.2}$$

where $\mathbf{h}^{\text{DL}}[\ell] \in \mathbb{C}^{N\times 1}$ is the mmWave downlink channel vector; $\mathbf{f} \in \mathbb{C}^{N\times 1}$ is the normalized downlink beamforming vector ($\|\mathbf{f}\|^2 = 1$); $x^{\text{DL}}[\ell]$ is the transmitted symbol with average power P^{DL} such that $\mathbb{E}[|\ x^{\text{DL}}[\ell]\ |^2] = P^{\text{DL}}/L$; and $n^{\text{DL}}[\ell] \sim \mathcal{N}(0, (\sigma^{\text{DL}})^2)$ is the additive Gaussian noise in the mmWave band.

The main idea here is to exploit the sub-6 GHz uplink channels to predict locally the downlink mmWave beamforming vector at each AP (link) using deep neural networks and federated learning. The rationale is that the uplink channels capture information regarding the wireless environment that is invariant with the frequency band (e.g., geometry of the various obstacles and buildings, higher order channel statistics, etc). This information can then be exploited to construct beamforming vectors in the mmWave band. Moreover, estimating the sub-6 GHz uplink channels requires less training overhead and exploits an already acquired technology compared to mmWave channels.

Multi-link interference The uplink channel vectors $\{\mathbf{h}^{\text{UL}}[\ell]\}_\ell$ are used in the dataset for training our neural network and also as inputs for mmWave downlink beam prediction when the neural network is exploited. In the uplink, the multi-link interference affects the uplink channels' estimation quality. We investigate this effect by considering the multi-link interference as additive noise in Fig. 4.13.

In the downlink, the idea is to design a neural network taking as input only the direct sub-6 GHz channel of the served user for mmWave beam prediction and no other information regarding multi-link interference. Hence, designing a neural network that predicts the mmWave beam accounting for the downlink multi-link interference without

additional input information about this interference is far from trivial. At the opposite, including such information at the input of the neural network, would require a larger dataset, a more complex network architecture, coupled with a more advanced (computationally complex) joint estimation technique of the direct channel and interference terms in the uplink.

Finally, the effect of the downlink multi-link interference on the performance of the predicted beams is expected to be limited due to the mmWave beamforming itself [7]. Otherwise stated, the performance improvement when taking the interference term explicitly into account, may not justify the additional complexity that it involves.

For all these reasons, we ignore the downlink multi-link interference in building our federated learning approach. Nevertheless, in Fig. 4.14, we evaluate the robustness of our approach in a multi-link setting with mobile users, in which the downlink interference is taken into account.

4.3 Channel-beam mapping via DL

We start by investigating the special case of a single AP – user link. We propose a novel unsupervised deep learning scheme to approximate the complex and non-linear mapping function between $\{\mathbf{h}^{\mathrm{UL}}[\ell]\}_{\ell=1}^{L}$ and the downlink mmWave beamforming vectors $\mathbf{f}$. Since there is no known parametric model able to capture such a relationship, data-driven approaches become essential. The used dataset contains only pairs of the channels at sub-6 GHz and mmWave and does not contain information regarding the best beams at mmWave (the ground-truth), which motivates the unsupervised learning approach to learn relevant features for beam prediction.

In the following, we explain in details the key ingredients of our proposed channel-beam mapping solution. We then evaluate its performance via extensive numerical experiments and compare it with existing methods based on supervised deep learning [6].

4.3.1 Dataset construction

Our learning approach relies on the available *DeepMIMO* dataset [118] and is composed of channel pairs of the form: $(\{\mathbf{h}[\ell]^{\mathrm{UL}}\}_{\ell=1}^{L}, \{\mathbf{h}[\ell]^{\mathrm{DL}}\}_{\ell=1}^{L})$, generated for different user positions within the predefined grid around the fixed access points. *DeepMIMO* employs the accurate 3D ray-tracing simulator *Wireless Insite* [119] to generate the uplink and downlink channels. It is an open access dataset supporting various carrier frequencies in both the sub-6 GHz and mmWave band, which fits the requirements of our problem.

More specifically, the uplink and downlink channels are generated using the outdoor ray-tracing scenario '$O1$', which is available at 3.5 GHz and 28 GHz carrier frequencies [118]. The user positions are sampled every 20 cm in the 2D row grid composed of 2751 rows and 181 columns (called User Grid 1) as shown in Fig. 4.2. In the single

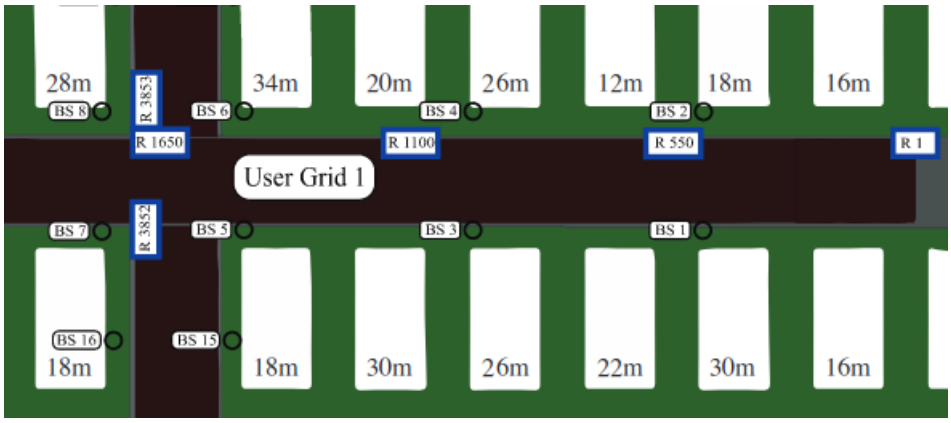

Figure 4.2: Top view of the '$O1$' scenario of the *DeepMIMO* dataset in [6].

link case, we choose the third AP (called BS 3) serving a user inside its cell, placed between the rows $700 - 1300$ of the grid. Unless otherwise specified, the benchmark system parameters are chosen as in table 4.1 , similarly to the work [6] for fairness of comparison. The parameter λ denotes the wavelength.

Once generated, each complex entry of the channel vectors $\mathbf{h}^{\mathrm{UL}}[\ell]$ is decomposed into real and imaginary parts, which are stacked into a $2LM$ real-valued vector containing the uplink channel information of all L subcarriers. This operation is done for simplicity of implementation, given that most existing deep learning libraries operate on real numbers. Similarly, the mmWave channels $\mathbf{h}^{\mathrm{DL}}[\ell]$ are partitioned into real and imaginary parts to form one real valued vector of dimension $2LN$. The obtained dataset is divided into a training dataset (80% of the total size) and a test set (the remaining 20%). The training dataset is further split into a training set (85% of its size) and a validation set (the remaining 15% of the initial training dataset). The same repartition is used throughout the simulations.

Parameters	uplink	downlink
AP	BS 3	BS 3
Users rows	700-1300	700-1300
Carrier frequency	3.5 GHz	28 GHz
Number of antennas	M = 4	N = 64
Antenna spacing	$\lambda/2$	$\lambda/2$
Bandwidth (GHz)	0.02	0.5
OFDM user subcarriers L	32	32
Number of paths	15	5
Transmit power (dBm)	–	34
Noise power (dBm/Hz)	-174	-174

Table 4.1: System parameters for dataset construction.

4.3.2 Neural network architecture

We design a fully-connected neural network where each neuron is connected to all the neurons of the preceding and following layers as illustrated in Fig. 4.3. These networks are *structure agnostic*, making no particular assumptions about the inputs and serving a general purpose. Furthermore, such networks guarantee the flow of information between the inputs and outputs of each layer, which makes it able to capture any kind of dependencies between the layers (provided appropriate data and training). These characteristics make fully-connected networks suitable for our problem, since we do not have specific knowledge about the complex relationship between sub-6 GHz channels and mmWave beamformers, coupled with the different wireless parameters impacting it.

The mini-batch data is passed through a batch normalization layer to standardize its data and stabilize the learning. The uplink sub-6 GHz channel vector of dimension $2LM$ represents the input of a deep neural network composed of 4 hidden fully-connected layers of S, $2S$, $2S$, S neurons respectively with rectified linear unit (ReLu) as an activation function[1]. Every layer employs an $\mathbf{L}2$-norm regularization with weight decay equal to 10^{-7}. The output layer is a fully-connected one of size $2N$, which provides directly the real and imaginary parts of the mmWave beamforming vector. It is associated with an L2-normalization layer to ensure that the predicted beamforming vector is of unit norm ($\|\mathbf{f}\|^2 = 1$).

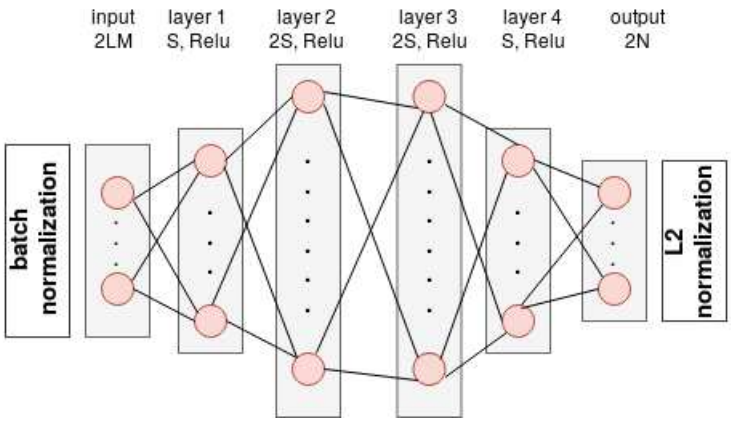

Figure 4.3: Architecture diagram of the fully-connected neural network employed at the access point.

4.3.3 Model-based loss function

Since the main purpose is to predict *good* mmWave beams in terms of communication rate, we define a model-inspired loss function $\mathcal{L}$, which allows to take advantage of the

[1]This four-layer architecture and the specific number of neurons S has been chosen as a result of empirical trials to best tradeoff the training and validation performance. We investigate in details the optimal choice of S as a function of the number of antennas (M and N) in Sec. 4.3.5.

known rate function in wireless OFDM systems jointly with the capability of neural networks to learn the complex and unknown channel-beam mapping function:

$$\mathcal{L} = -\frac{1}{\mathcal{B}} \sum_{i=1}^{\mathcal{B}} \mathcal{R}_i, \tag{4.3}$$

where $\mathcal{B}$ is the size of the mini-batch and $\mathcal{R}_i$ is the average data rate over the L subcarriers for the i^{th} sample: $(\{\mathbf{h}_i[\ell]^{\text{UL}}\}_{\ell=1}^{L}, \{\mathbf{h}_i[\ell]^{\text{DL}}\}_{\ell=1}^{L})$ of the mini-batch, and which can be written as

$$\mathcal{R}_i = \frac{1}{L} \sum_{\ell=1}^{L} \log_2 \left(1 + \frac{P^{\text{DL}}}{L(\sigma^{\text{DL}})^2} \mid \mathbf{h}_i^{\text{DL}\dagger}[\ell] \ \mathbf{f}_i \mid^2 \right), \tag{4.4}$$

with $\mathbf{f}_i$ denoting a normalized beamforming vector predicted by the neural network for the i^{th} sample (the output elements of the neural network are re-shaped into an N-dimension complex vector).

The above model-based loss function sets our work apart from existing ones [6, 35, 34], in which first a loss function based on some average prediction error is minimized, and then the performance of the prediction is evaluated in terms of its communication performance. Our neural network is trained and optimized to maximize directly the communication rate and skip the intermediary step. Choosing a communication-tailored loss as opposed to a generic data-driven one can only improve the communication performance of our method.

Further motivation is that computing a data-oriented prediction error is not possible with the available *DeepMIMO* dataset, which is only composed of channel pairs $(\{\mathbf{h}[\ell]^{\text{UL}}\}_{\ell=1}^{L}, \{\mathbf{h}[\ell]^{\text{DL}}\}_{\ell=1}^{L})$ and does not contain the corresponding optimal beamforming vectors $\mathbf{f}$ (or the ground-truth). Creating a different dataset composed of pairs of the type $(\{\mathbf{h}[\ell]^{\text{UL}}\}_{\ell=1}^{L}, \mathbf{f})$ as in [6] may be quite problematic in our regression problem because there are an infinite number of optimal beam vectors $\mathbf{f}$ maximizing the communication rate. Indeed, the rate function above is invariant to a multiplication of $\mathbf{f}$ by a complex scalar of unit-norm and an arbitrary selection might hinder the generalization capability of the neural network.

4.3.4 Evaluation of the channel-beam mapping

We evaluate here the performance of our proposed channel-beam mapping method in terms of communication rate in the case of a single link. The presented results are obtained with the neural network in Fig. 4.3 with $S = 1024$ after 100 training epochs [2] using the adaptive moment estimation (ADAM) optimizer [120] with a learning rate of 10^{-4} and a batch size of $\mathcal{B} = 256$ samples. The different learning models are implemented and trained using *TensorFlow* [121].

[2] An epoch is reached when the entire training set is passed forward and backward through the neural network once.

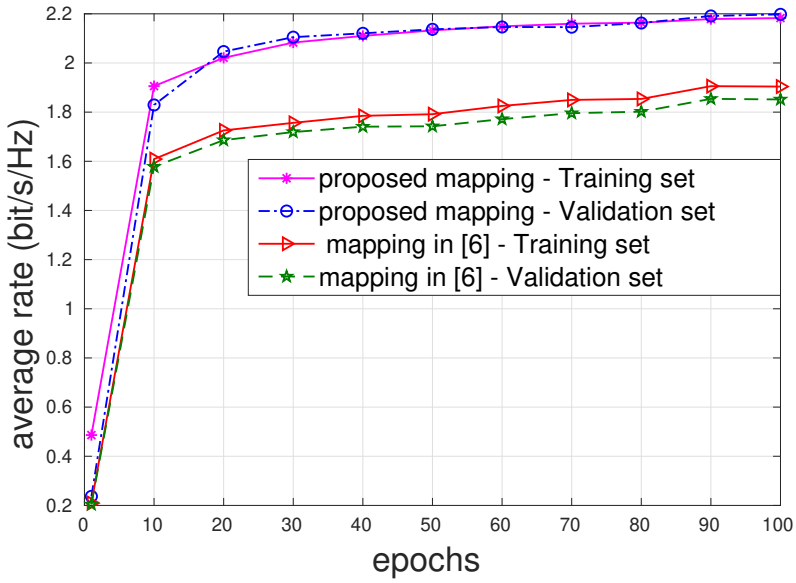

Figure 4.4: Average rate on the training and validation sets. Our method yields higher rates and better generalization.

Training performance In Fig. 4.4, we evaluate the communication rate, computed on the training and validation sets. Compared to the existing method in [6], our method achieves higher rates and yields smaller gap between the training and validation rates, which indicates a better generalization performance. Indeed, combining a regression formulation with our communication-tailored loss function allows our method to have better beam prediction quality and higher rates.

Prediction performance In Fig. 4.5, we plot the empirical cumulative distribution function (CDF) of the average rate obtained on the test set containing samples unseen by the neural network during its training. In addition, we evaluate the performance of our neural network trained with uplink and downlink channels in the mmWave band at 28 GHz (instead of sub-6 GHz channels in the uplink), in the same setting otherwise. We also compare the results with the perfect downlink CSI case (ideal benchmark), assuming perfect and instantaneous knowledge of the mmWave channels, which are used to construct the beamforming vectors for each subcarrier such that $\mathbf{f}^{\star}[\ell] = \mathbf{h}^{\mathrm{DL}}[\ell]/||\mathbf{h}^{\mathrm{DL}}[\ell]||$ to maximize the received power at the receiver.

First, it can be seen that our method performs better than [6] and closer to the ideal case, indicating better generalization performance, which confirms the results in Fig. 4.4. Second, Fig. 4.5 shows that training with uplink mmWave channels yields much lower rates (98% of the achieved rates are below 1 bit/s/Hz). This can be explained by the poor quality of mmWave uplink channels (whose difficult propagation conditions can not be overcome by the low number of uplink antennas: $M = 4$) and/or by the fact that our neural network architecture may not be suitable in this case.

mmWave frequency robustness In Fig. 4.6, we evaluate the generalization capability (or robustness) of our trained neural network with respect to the downlink mmWave frequency. The rising question is: *can the optimized neural network trained*

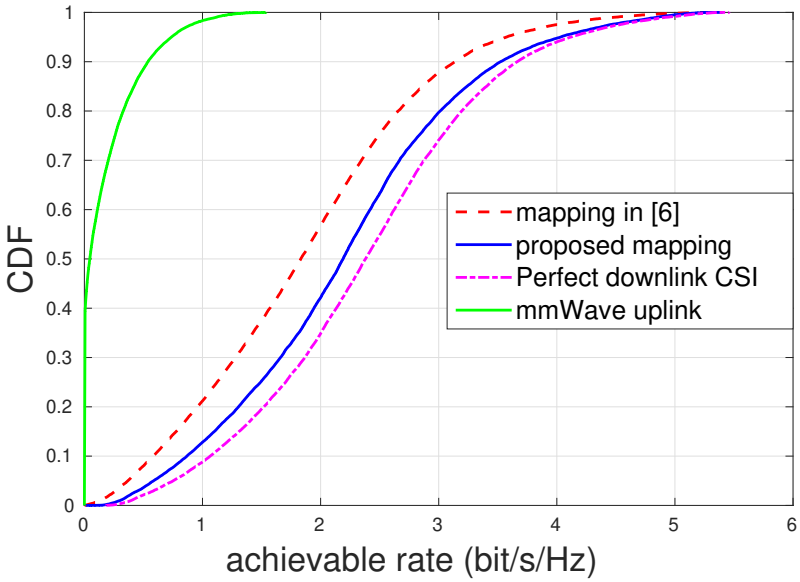

Figure 4.5: Empirical CDF of the achievable rate over the test set. Our proposed channel-beam mapping is closer to the ideal case.

at 28 *GHz still be used to predict good beams at a different mmWave frequency?* To answer this question, we exploit two differently trained neural networks with datasets of the type (UL channels at 3.5 GHz, DL channels at 28 GHz) and (UL channels at 3.5 GHz, DL channels at 60 GHz), respectively, to predict beamforming vectors at 60 GHz, in the same setting otherwise.

Fig. 4.6 demonstrates that the beam predictions at 60 GHz (downlink mmWave) provide almost identical rates irrespective from the mmWave carrier frequency of the training data. The root mean square difference between the predicted beams by both approaches is around 0.11, while the difference between their achieved rates over the test set is 0.003 bit/s/Hz. This means that, once trained, our neural network can be used to predict beams at different mmWave frequencies (for the same wireless environment) with almost identical rate performance and no re-training required. Note that the rate values in Fig. 4.6 at 60 GHz are smaller than the ones in Fig. 4.5 at 28 GHz because the propagation loss increases for higher frequencies.

4.3.5 Optimal dimension of the neural network

In the previous subsection, the results were obtained for a carefully tuned neural network size: $S = 1024$. Below, we justify this choice. More precisely, we investigate how to optimally tune the size S of the neural network as a function of the number of receive (uplink) and transmit (downlink) antennas: M and N. Intuitively, S has to be sufficiently large to capture the complex relationship between the inputs (sub-6 GHz channels) and outputs (mmWave beams) of the neural network. However, the larger the S, the higher the computational complexity (for both the training phase and the prediction or running phase) and also the larger the training dataset has to be. The idea is to choose the smallest value of S providing the best rate performance.

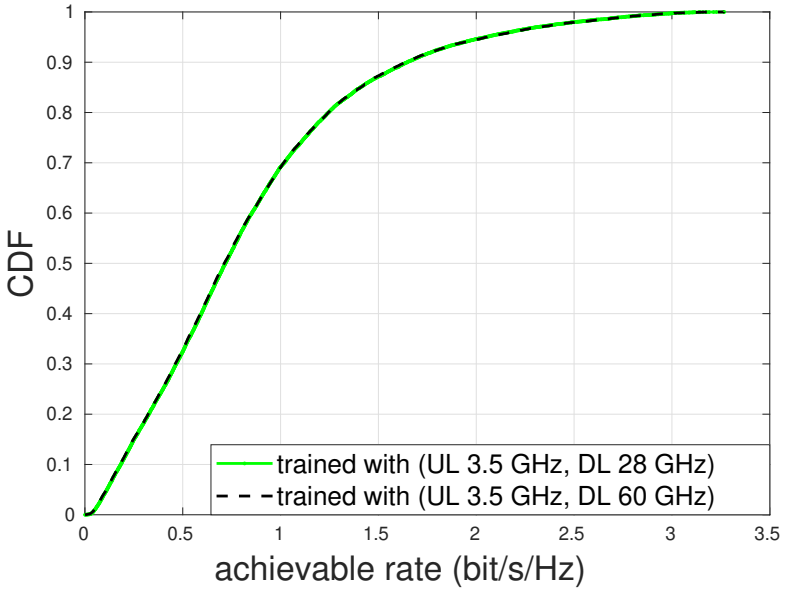

Figure 4.6: Empirical CDF of the achievable rate over the test set at 60 GHz when the network is trained with downlink channel samples at both 28 GHz and 60 GHz. Our approach is robust to changes in the mmWave frequency.

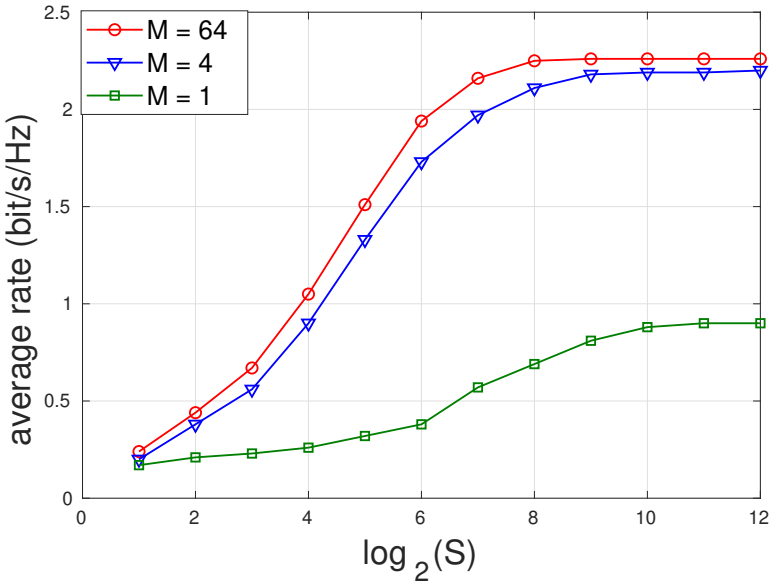

Figure 4.7: Impact of the input size M and the neural network size S, for $N = 64$ and $L = 32$. When the input size M increases, the optimal size decreases.

In Fig. 4.7, we evaluate the average rate over the test set for different sizes of the neural network $S \in [2, 2^{12}]$ and different number of uplink antennas $M \in \{1, 4, 64\}$, for fixed $N = 64$ and $L = 32$. For any value M, the rate increases as the network size becomes larger until hitting a ceiling, from which no more significant rate improvement is observed. To further increase the rate performance, the number of uplink antennas M has to be increased to provide more information at the input of the neural network. The small gap between the rate obtained for $M = 4$ and $M = 64$ indicates that $M = 4$ provides already sufficient information to reach high rates. Increasing M further does not result in a significant rate improvement.

The optimal size (required to reach the rate ceiling) decreases with the input dimension M. Indeed, Fig. 4.7 shows that the rate ceiling is reached for $S = 2^{10}$ when $M = 1$, for $S = 2^9$ when $M = 4$, and for $S = 2^8$ when $M = 64$. Therefore, increasing the input size M provides more information at the input of the network, which results in a reduction of the size of the neural network.

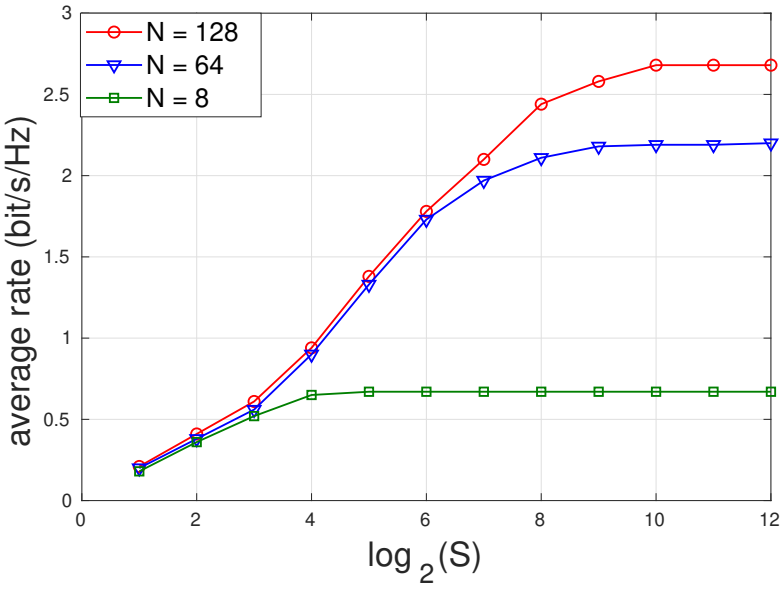

Figure 4.8: Impact of the output size N and the neural network size S, for $M = 4$ and $L = 32$. When the output size N increases, the optimal size increases as well.

In Fig. 4.8, we evaluate the average rate over the test set for different sizes of the neural network $S \in [2, 2^{12}]$ and different numbers of downlink antennas at the AP $N \in \{8, 64, 128\}$, for fixed $M = 4$ and $L = 32$. The rate increases as a function of S until hitting a ceiling as in Fig. 4.7. Similarly to the above, increasing the number of transmit downlink antennas N allows to increase the beams resolution (higher beamforming gain) and, hence to achieve higher rates.

The optimal size of the neural network increases with the output size N. More specifically, the rate ceiling is hit for $S = 2^4$ when $N = 8$, for $S = 2^9$ when $N = 64$, and for $S = 2^{10}$ when $N = 128$. This is intuitive as a more complex architecture is required to capture the relationship between a fixed-size input and an increasing size output.

At last, this analysis explains the good performance of our channel-beam mapping presented in Sec. 4.3.4 for the chosen neural network size $S = 1024$.

4.4 Multi-link beam prediction via federated learning

In this section, we investigate the more general case of multiple AP–user links. There are several ways in which our channel-beam method can be applied. One naive idea would be for each AP to perform individual training of their neural network based on

locally available data and completely independently from one another. Such a fully distributed approach might perform quite poorly when local data is scarce and when cooperation would be beneficial.

At the opposite, a centralized approach would require the APs to send all their training data to a central node, which performs the neural network training and then sends the resulting neural network parameters back to each AP. Of course, this also raises several issues in terms of prohibitive communication overhead and privacy.

We propose a distributed learning approach in between the two extremes by exploiting federated learning [122]. In this framework, the APs send to a server only their neural network parameters and not their actual data, leading to less signaling overhead and privacy issues. The server computes a global and more informed model (based on the knowledge acquired by all APs), which is then fed back to the APs. Federated learning also enables parallel computation speeding up the training process by splitting the computational load among the APs.

Our FL approach consists of a training phase and an exploitation phase, which are detailed hereafter.

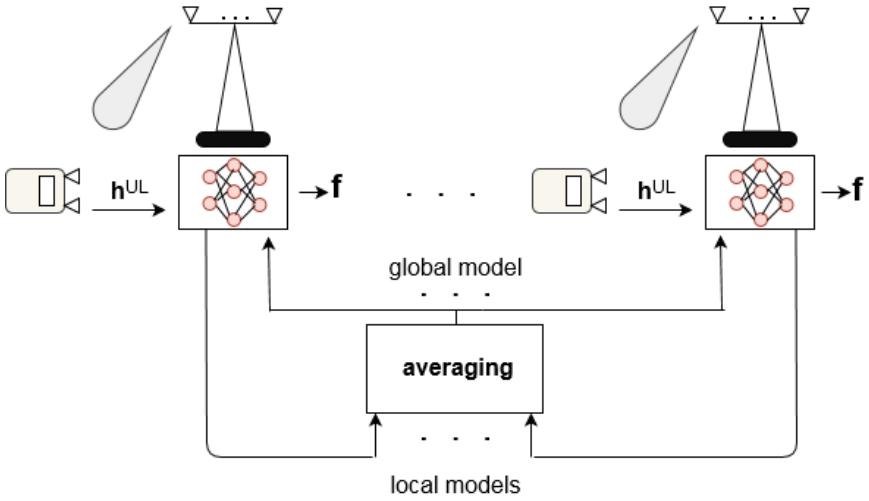

Figure 4.9: Federated learning for beam prediction based on sub-6 GHz channels.

4.4.1 Federated learning: training phase

In federated learning, the training phase is performed collaboratively and two different methods are proposed here: the synchronous FL (SFL) and asynchronous FL (AFL). In SFL, a training epoch corresponds to a local epoch at every AP, whereas in AFL, it corresponds to a local epoch at only one AP (the updating AP). In the centralized or individual approaches, a training epoch is reached when all the batches (or all training samples) have been visited by the loss minimization method.

In SFL, all APs optimize locally their neural network based on their available data as detailed in Sec. 4.3, at each training epoch. An epoch is reached when the loss minimization method at each AP has visited all its available local samples. Then, the

parameters or weights of the local networks are uploaded to a server for aggregation (a simple average operation), as illustrated in Fig. 4.9. On the contrary, in AFL, the APs take turns and only one AP optimizes its neural network based on their local available data as in Sec. 4.3 at each training epoch. This AP then uploads its optimized network parameters to the server, which are then averaged with the previous global network parameters to obtain the new network update. For both FL methods, the updated global network parameters are downloaded by each AP and the process repeats until the end of the training phase.

Compared with SFL, the AFL method reduces the communication overhead at each training epoch at the cost of having to perform more training epochs to reach the same performance levels.

4.4.2 Federated learning: exploitation phase

After the training phase, the final global neural network is exploited by all the APs locally. Each AP runs this trained network to predict the mmWave beamforming vectors by feeding it with locally estimated sub-6 GHz uplink channels.

During the exploitation, the APs can collect new data to refine and update the global model in case of major changes in the wireless environment.

4.4.3 Dataset construction

Unless stated otherwise, the system parameters in the *DeepMIMO* dataset are the same as in Sec. 4.3.4. We consider a system of 4 AP–user links. Following the nomenclature in *DeepMIMO*, the chosen access points are: BS 1, BS 4, BS 6, and BS 7; their users are positioned within their specific cells characterized by the grid rows: $1-599$, $600-1200$, $1201-1550$ and $1551-2200$, respectively.

4.4.4 Evaluation of the federated learning approach

Here, we evaluate the performance of our proposed distributed mmWave beam prediction in terms of overall network average rate, which is computed as the average of the rates of all AP–user links.

We start by evaluating the average rate of our federated learning approach for both synchronous (SFL) and asynchronous (AFL) updates of the global model. We compare both methods to the perfect downlink CSI method and the following benchmarks:
Centralized learning (CL): performed by a central authority using a neural network trained on the global dataset including all local datasets of the APs;
Individual learning (IL): each AP trains its neural network using its local dataset independently from the others (fully distributed learning).

FL training performance In Fig. 4.10, we illustrate the evolution of the average rate computed on the training and validation sets for CL, SFL and AFL methods. First, notice that the values of the average rate on the training and validation sets are close for all methods, which implies that the proposed neural network performs well without data underfitting or overfitting and after only 20 training epochs. The rates achieved by the FL schemes approach the centralized performance, while relieving the system from the overhead signaling cost caused by sharing the local data with the central entity. As expected, AFL method is a little slower than SFL, but reaches eventually comparable performance. AFL reduces the neural network parameter exchange between APs and the server by 75% (only one out of the four APs performs local training and sends the network parameters at each training epoch) at the cost of a relatively small performance loss, highlighting the *signaling vs. performance tradeoff*. Aside from simplifying the training, AFL also reduces the power consumption by 75%.

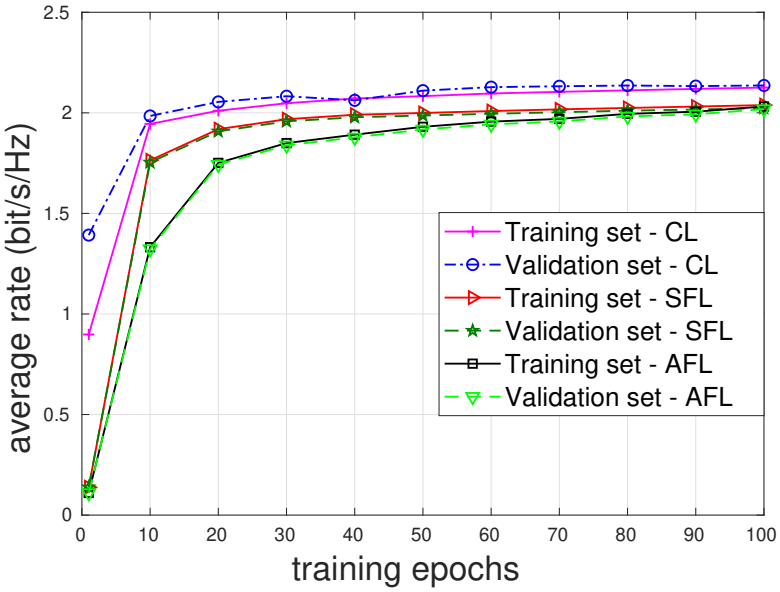

Figure 4.10: Average rate evaluated on the training and the validation sets. AFL and SFL rates are close to centralized learning.

FL prediction performance Fig. 4.11 represents the empirical cumulative distribution function (CDF) of the average rate over the test sets for the CL, AFL and SFL methods and also for the perfect downlink CSI beams. For each learning method, the already trained neural networks are exploited to predict downlink mmWave beamforming vectors from uplink sub-6 GHz channels.

We remark that both our FL methods (AFL and SFL) perform close to CL and are not far from the ideal performance, indicating a good generalization performance on unseen test data. This means that, our proposed FL methods can provide almost the same average rate performance as CL, while maintaining the advantages of the FL framework. Moreover, the AFL method performs almost as good as SFL after the same number of training epochs, while enjoying 75% less signaling overhead and power consumption during the training process.

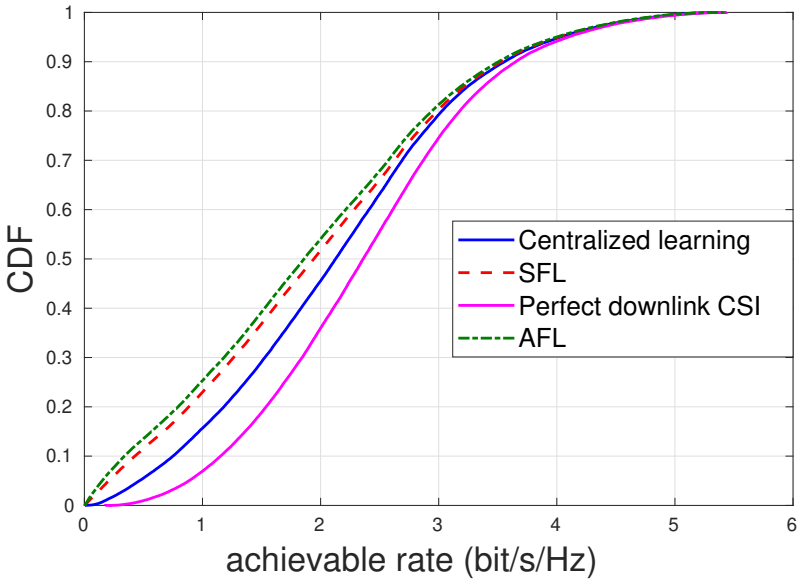

Figure 4.11: Empirical CDF of the achievable rate over the test set. Our FL methods perform close to the centralized learning.

Impact of scarce available data In Fig. 4.12, we evaluate the average rate over the test sets for different training set sizes relative to the total training set size of each access point. We also evaluate here the individual learning scheme. First, increasing the training set size improves the average rate for all schemes, because of the better generalization performance. When the local available data is scarce, our FL methods clearly outperforms individual learning and reaches by up to 50 % higher rates with SFL and 41% with AFL, showing the interest of sharing the network parameters in such cases. When the amount of local available data is sufficiently large, federated learning does not bring any advantage and the parameter averaging operation (performed at the central node during the training phase) leads to a sub-optimal performance compared with individual learning.

Impact of imperfect available data In order to test the robustness of our FL schemes to imperfect data, we train the various methods when the uplink channels are contaminated with different levels of noise and then we test the resulting neural networks on unseen input data.

Fig. 4.13 illustrates the evolution of the average rate over the test sets as a function of different noise variance levels (in between 10^{-11} and 10^{-8}) corrupting the uplink sub-6 GHz channels (at the input of neural networks) during the training phase. The empirical variance of the uplink sub-6 GHz channels is of around 10^{-9}. The SFL scheme outperforms centralized and individual learning at high noise levels when the quality of the uplink sub-6 GHz channel estimations is poor. For a noise variance of 10^{-8}, the relative gain is of 14% compared to centralized learning. This result can be explained by the averaging step of the SFL scheme, which acts as a regularization and noise smoothing operation, thus improving its generalization performance in this regime. The same advantage is not observed with AFL, as it does not average all local models at every training epoch (lower noise smoothing effect).

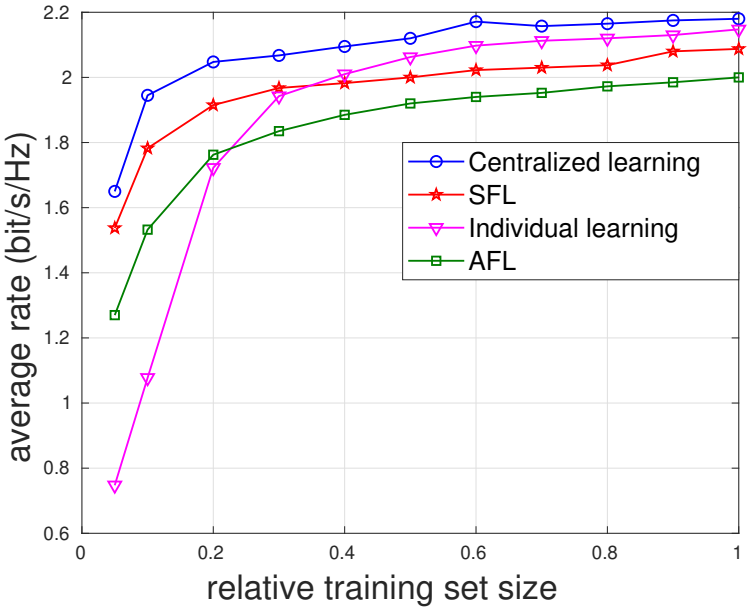

Figure 4.12: Impact of the training set size. For scarce training data, our FL schemes outperform the individual learning and approach centralized learning.

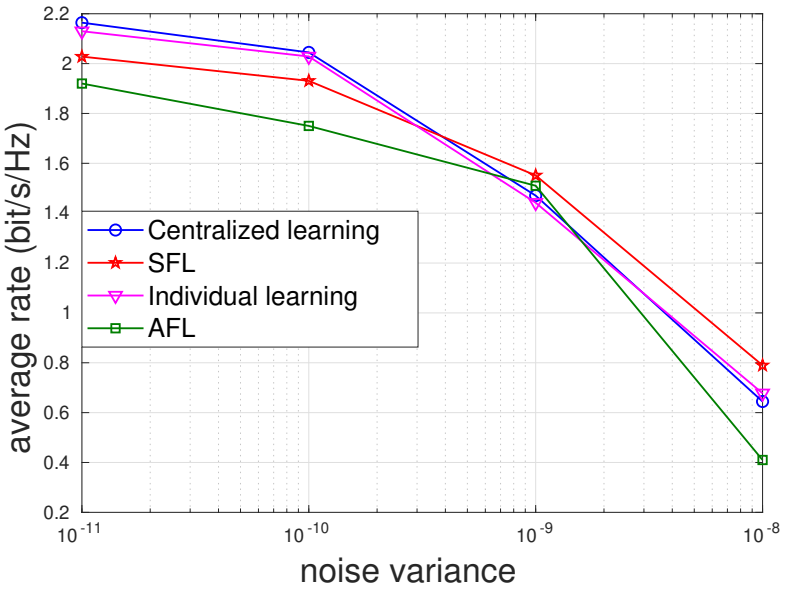

Figure 4.13: Impact of noisy training data (sub-6 GHz channel estimation quality). Our SFL scheme outperforms centralized and individual learning in the high noise regime.

Interfering multiple links and user mobility We evaluate our FL approaches in the presence of downlink multi-link interference and user mobility. The downlink interference for an AP – user link is defined as the mmWave signals received from other APs. The mobility of the four users is modeled as in [5] with a maximum speed of 80 km/h, maximum acceleration of 2 m/s^2 and maximum rotation speed of $\pi/4$ rad/s. Their positions are restricted to the ranges of the APs. The APs exploit similar neural networks, which are trained offline, having little available data: only 10% of the training set, to predict their beams during each time slot of 20 ms. Note that the

offline training phase has been performed without taking into account the downlink inter-link interference as discussed in Sec. 4.2. The results are compared to the fully centralized, individual learning and are illustrated in Fig. 4.14. The rate of the perfect downlink CSI curve in Fig. 4.14 is evaluated in an ideal scenario without multi-link interference representing a performance upper bound. The curves are averaged over 1000 random and independent trajectories for each user.

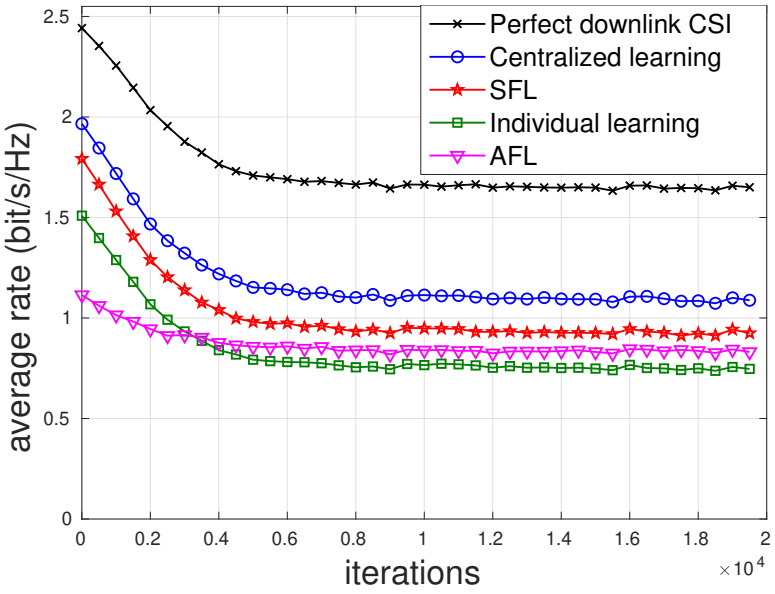

Figure 4.14: For scarce training data, SFL outperforms individual learning and reduces the rate gap with centralized learning in the presence of downlink interference and user mobility.

We can see that the SFL method outperforms individual learning and is close to the centralized one when the training is done with little data (10% of the available training set) even in the presence of downlink interference. This illustrates the utility of the SFL method to improve the rate performance when the available training data is scarce and when sharing the local model with other APs is beneficial.

4.5 Deep learning vs. multi-armed bandits

In this section, the objective is to compare deep learning methods with classic reinforcement learning, and more specifically, multi-armed bandits (MABs) from the previous chapter in terms of communication rate and computation complexity. For this, we consider a mobile user served by a fixed AP; the user's mobility follows the model in [5] with a maximum speed of 80 km/h, maximum acceleration of 2 m/s^2 and maximum rotation speed of $\pi/4$ rad/s. The initial user position is chosen randomly (via uniform distribution) and it is restricted to its AP range: rows $700 - 1300$ of the 2D grid described in Sec. 4.3.1. The channels are generated with *DeepMIMO* based on the user time-varying positions. Aside from the deep learning method in [6] and the ideal downlink CSI case described earlier, we consider the reinforcement learning algorithms: EXP3, MEXP3 and UCB.

In all MAB algorithms, the AP computes the beams *on-the-fly* via an online process relying only on a strictly causal 1-bit (ACK/NACK) reward mechanism. At each iteration, the AP chooses a beam from a predefined discrete codebook of size A for transmission. At the end of the transmission, the value 1 is fed back to the transmitter if the achieved rate in all carriers were above the threshold 1 bit/s/Hz, and zero otherwise. Based on this feedback, the beam selection process is updated and a new beam is chosen. For comparison reasons, the MAB algorithms use the same discrete beamforming codebook as in [6] composed of $A = 64$ beams.

4.5.1 Average communication rate

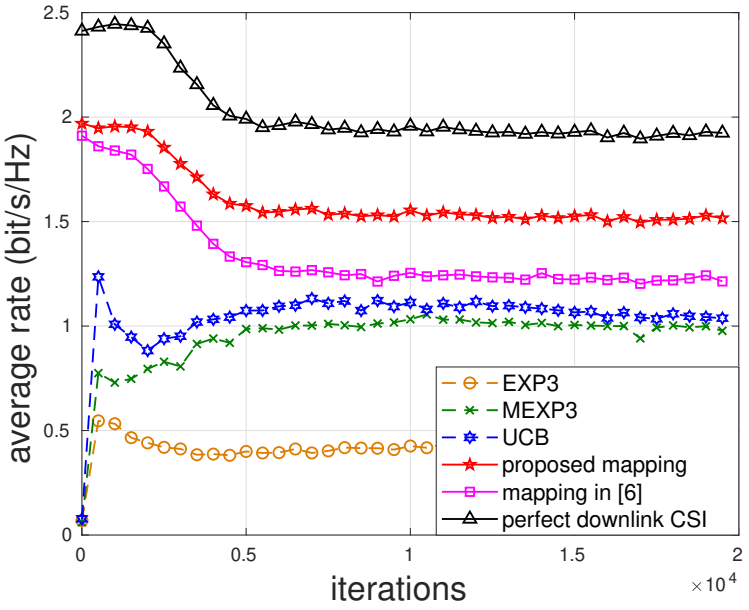

Figure 4.15: Average rate of a single mobile user. Our unsupervised learning method remains the closest to the ideal case despite the channel variations.

In Fig. 4.15, we plot the average rate obtained with the various methods as a function of iterations, where each iteration corresponds[3] to 20 ms. The curves are averaged for 1000 independent user trajectories. We remark that our proposed channel-beam mapping outperforms all other methods and that the deep learning approaches outperform MAB ones, obtaining up to 50 % higher rates. The rate decay of deep learning methods and of the ideal case is only related to the wireless environment and the user mobility. The average distance between the user and its access point increases with time (iterations) over the different trajectories as illustrated in Fig. 4.16, which explains the rate decrease.

The gap between our method and MAB algorithms can be explained by the fact that deep learning methods have the advantage of being pre-trained and optimized for this problem, as opposed to MAB algorithms, which learn the good beams in an online and adaptive manner (via trial and error). This gap is also due to the quantization and

[3]This duration corresponds to the beam coherence time during which the beams remain valid according to our mobility conditions [16] and the 5G new radio standard[123].

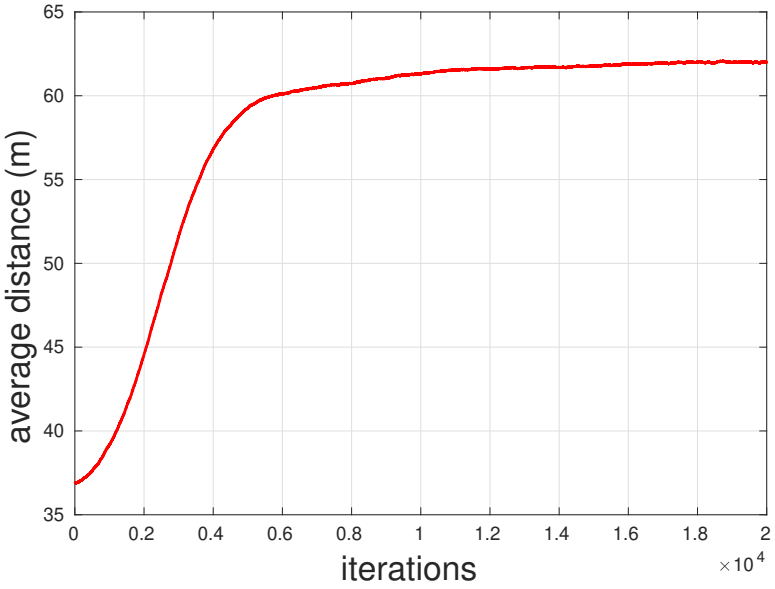

Figure 4.16: Average distance between the mobile user and its AP over different trajectories.

its implied sub-optimality for MABs. Indeed, we see that the gap between the deep learning method in [6], which also suffers from this quantization, and MABs is much smaller.

At last, the best performance of UCB among the three MAB algorithms is explained by the fact that, in this particular single-link setting, there is no adversarial or non-stationary component and the wireless environment is stochastic, in which UCB is known to be optimal in terms of regret.

In Fig. 4.17, we compare MAB algorithms with the federated learning ones (DL-based) in the same multi-link setting with downlink interference and limited (10%) training data as described in Fig. 4.14. The same conclusions hold, as in the single-link case, except for the order of performance among MAB-based methods. The UCB policy is outperformed by MEXP3 in this setting. This can be explained by the fact that, in the multi-link setting, the wireless environment is no longer stochastic as every link chooses its beam simultaneously in a distributed manner. In such non-stationary environments, UCB has no guarantees for optimal performance as opposed to MEXP3. We can also note that MEXP3 achieves similar rate levels as the AFL method after 6000 iterations.

4.5.2 Computational complexity

MAB algorithms have a relatively low complexity and come with worst-case theoretical guarantees in terms of regret. As mentioned above, they are not pre-trained but learn the best beams on the fly, which implies that a certain amount of exploration time (or number of iterations) is required before achieving good performance results. Roughly speaking, UCB attains its ε-optimal performance (ε regret level) after $1/\varepsilon$ iterations (up to logarithmic factors), whereas EXP3 and MEXP3 attain it after $1/\varepsilon^2$ iterations.

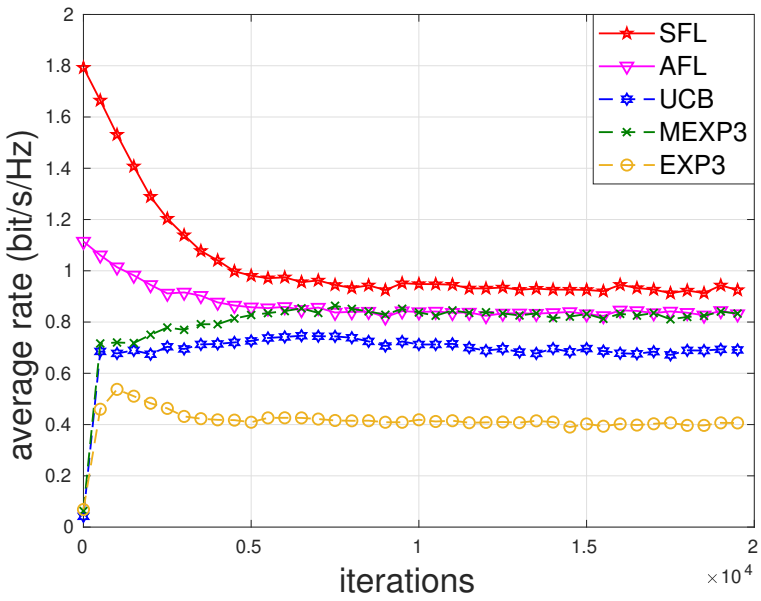

Figure 4.17: Average rate of mobile users. SFL outperforms MAB algorithms in the multi-link setting.

The complexity of one iteration of these algorithms scales linearly with the codebook size: $\mathcal{O}(A)$. Regarding the complexity of our deep learning method during the running phase, predicting a mmWave beam has a complexity of $\mathcal{O}(LMS + S^2 + SN)$ at each iteration.

For a more detailed complexity comparison between the two approaches (DL and RL), we can focus on the number of floating point operations performed during one running iteration. For deep learning methods, the number of operations per iteration is proportional to the number of neurons in the neural network (eg. multiplication by the weights, adding bias, ...) and is about 10^6 operations in our neural network and 10^7 in [6]. For MAB methods, the number of operations is proportional to the codebook size A (much smaller than the number of neurons in neural networks), and is roughly around 10^2 operations. The number of operations during one iteration can also be used as an indicator of the energy consumption of the various methods: more operations per iteration implies more power consumption.

To sum up, deep learning methods outperform classical reinforcement learning ones in terms of communication rate (up to 50% higher rates) at the cost of being more computationally complex (10^4 times more complex) and of requiring offline training based on relevant and sufficient data. On the flip side, reinforcement learning and MABs do not require offline training, they rely on simple feedback information and low-cost iterations, and also come with performance guarantees in terms of regret. Nevertheless, the explosion of deep learning applications can be explained by the advances of the electronics industry and the exponential growth of Internet traffic resulting in new devices (at the edge of the network) with high processing capabilities and large amounts of locally available data. Choosing the appropriate solution will depend on the target application and on its characteristics and requirements in terms of energy consumption, processing capabilities, data availability, latency, etc.

4.6 Conclusion

In this chapter, we leverage unsupervised deep learning to exploit the channel knowledge at sub-6 GHz and predict beamforming vectors in the mmWave band. We evaluate the performance of the proposed channel-beam mapping and compare it to an existing supervised learning method. We also show how to choose an optimal size of our neural network depending on the number of transmit and receive antennas at the access point. Then, we investigate the impact of training data availability and introduce a federated learning approach to predict the beams of multiple access point – user links. We consider both synchronous and asynchronous FL methods. Our numerical simulations show the high potential of our approach, especially when the local available data is scarce or imperfect (noisy). Eventually, we compare the deep learning methods and the classic reinforcement learning algorithms from the previous chapter. Our analysis and simulations show that choosing an appropriate beam alignment method depends on the target application and is a tradeoff between rate and computational complexity. The next chapter summarizes our conclusions of the thesis and provides some perspectives for future work.

5

Conclusions and Perspectives

In this chapter, we conclude the manuscript and propose several perspectives and potential leads for future research.

5.1 Conclusions

The main objective of this thesis is to exploit machine learning tools to tackle a fundamental problem for future mmWave networks: the beam alignment. mmWave communications provide a promising solution to the spectrum gridlock by taking advantage of large frequency bands available in the range of $20 - 300$ GHz. These communications suffer from severe path loss, weak diffusion and high penetration loss. Highly directional beams, enabled through large antenna arrays and beamforming techniques, are a promising solution against these difficulties. As the wave-length is very short, the integration of several antennas on a small-size device becomes feasible. Thus, mmWave communications can benefit from the massive number of antennas at both the access point level and even the user level.

Before any useful data can be transmitted in mmWave, beam alignment is a crucial step to guarantee a reliable communication link. We have seen that traditional approaches are not capable of coping with users mobility and time variations of the channel. Hence, we propose dynamic and adaptive beam allocation policies exploiting online optimization and machine learning techniques. These tools go beyond classic optimization and allow developing algorithms that are efficient in spite of the network dynamics that can be non-stationary and unpredictable due to users' mobility and connectivity patterns.

In chapter 3, we propose online beam alignment policies relying on one-bit feedback to steer the beams of both ends of the communication link in a distributed manner. We formulate the beam alignment problem via the adversarial MAB framework, which copes with arbitrary network dynamics including non-stationary or adversarial components. The more general adversarial MAB setting (compared to stochastic bandits) allows us to decouple the beam alignment problem and split the learning between the transmitter and receiver. Both nodes use an online learning algorithm to choose their own beam-direction, in a distributed manner without knowing each others choices in advance. Hence, the learning is carried out at both the transmitter and receiver without relying on a central node and using only a one-bit feedback. Immediate advantages are that each node explores only its own set of beam-directions and not the set of beam

pairs which is much larger besides reducing the amount of exchanged data during the beam alignment process.

First, we investigate the optimal size of the beamforming codebook to be used for the beam alignment to reduce the exploration cost. We provide the minimal codebook size that guarantees a certain quality of service related to the empirical outage. In the case of single-path mmWave channels, a closed-form expression is obtained, while for multi-path channels we use numerical experiments to compute this performance measure. We show that increasing the codebook resolution beyond a certain point does not offer a better outage performance but implies an increasing exploration cost. Hence, restricting the number of beamforming vectors offers a good tradeoff between the outage performance and the exploration cost.

Then, we exploit the well-known EXP3 algorithm for a distributed search of the best beams in a point-to-point mmWave MIMO system. The performance of the proposed method is evaluated in terms of the notion of regret, outage probability and compared with existing algorithms. Our simulation results show a decreasing average regret and outage as the learning proceeds, which implies more accuracy in the beam alignment.

Building on the original EXP3 algorithm and by exploiting the structure and sparsity of the mmWave channel, we propose a modified policy (MEXP3). Our MEXP3 uses a modified reward that reinforces the exploitation of good beam-directions and penalizes the poor ones. We then prove that the new MEXP3 has the no-regret property and that the average regret decays to zero optimally as $\mathcal{O}(1/\sqrt{\mathcal{T}})$ similarly to the original EXP3 where $\mathcal{T}$ denotes the time horizon. Moreover, for fixed and finite horizons, our regret upper-bound for MEXP3 is tighter (smaller multiplicative constant factor) than the original EXP3 bound, suggesting better performance in practical settings. We then introduce an additional modification that accounts for temporal correlation between successive beams and propose another beam alignment policy namely NBT-MEXP3. The property of no-regret is conjectured for NBT-MEXP3 and validated via extensive numerical simulations.

Although the asymptotic regret performance of the proposed algorithms, $\mathcal{O}(1/\sqrt{\mathcal{T}})$, is optimal and cannot be improved under arbitrary network dynamics, our two novel policies MEXP3 and NBT-MEXP3 offer significant performance improvements in practical mmWave settings. Numerical simulations show that the proposed policies offer better practical performance especially in terms of outage and throughput for both single and multi-path channels. Our modified rewards lead to online learning algorithms that adapt better and faster to the varying mmWave channel, which results in lower outage and higher data rates compared to other existing policies.

In chapter 4, we consider other machine learning tools going beyond the MAB framework. First, we propose an unsupervised deep learning method to design a channel-beam mapping, which exploits the channel knowledge at sub-6 GHz to predict mmWave beamforming vectors using a trained neural network. We illustrate, via extensive numerical results, the performance of the proposed mapping in terms of communication rate and compare it to existing methods. We also show that the optimal size of our neural network decreases, when increasing the network input size (number of uplink receive antennas), since the information about the sub-6 GHz uplink channels becomes

richer. On the contrary, when the output of the neural network increases (number of transmit downlink antennas), the optimal neural network size has to increase, since the input-output relationship becomes more complex.

Furthermore, we employ our proposed channel-beam mapping in a federated learning framework to predict the beams of multiple AP – user links in a distributed manner. Federated learning consists in the APs sharing the parameters of their locally trained neural networks to aggregate them into a more informed global model for all APs. The main advantages are three fold: i) it spares each AP from sharing its local data, which may pose privacy issues, while still pooling on the knowledge acquired by other APs; ii) it reduces the signaling overhead; and iii) it distributes the computation load, compared to a fully centralized approach (the APs send their local data to the central server in charge of training the neural network). We investigate both synchronous and asynchronous sharing approaches. The asynchronous approach reduces the signaling cost (and power consumption) during the training at a cost of relatively small rate degradation.

We evaluate our proposed FL schemes and compare them to a fully centralized and a fully distributed (no cooperation between the APs, each neural network is trained independently using local data) benchmarks. In the case of scarce available training data, the relative rate gains of our synchronous and asynchronous FL approaches can reach up to 50% and 41% compared to the fully distributed approach, respectively. The relative rate gain reaches 14% for synchronous FL compared to centralized learning, when the quality of the training data is poor (when uplink sub-6 GHz channel estimations suffer from noise). At last, we evaluate our federated learning methods in the case of mobile users and taking into account the downlink multi-link interference. Our results demonstrate the efficiency of our approaches in the case of scarce training data, which makes it practical in this setting.

Finally, we compare the deep learning methods with our MAB-based methods (from chapter 3) in terms of rate performance and computational complexity. Compared to MAB algorithms, deep learning methods provide higher communication rates at the additional cost of offline training and higher online (running) computation complexity. Therefore, technical requirements and application characteristics, such as latency, computation power, energy consumption and data availability for training, can be taken into account to leverage one approach over the other.

5.2 Perspectives

In this last section, we present several possible perspectives of our research work during the thesis. First, we discuss some short-term perspectives and then long-term perspectives.

5.2.1 Short-term perspectives

As we have seen in chapter 3, the no-regret property of the proposed NBT-MEXP3 algorithm is only conjectured. An immediate possible perspective is to prove the no-regret property and provide an upper bound for the average regret. To do so, an appropriate bound for the ratio $\frac{p_{T,k}(t)}{\hat{p}_{T,i_t}(t)},\ k \in V_{i_t}$ needs to be found. This term appears in the expectation of the regret and is a result of assigning non-zero rewards to the neighbouring beams. Using an appropriate bound for this term, a similar approach can be followed as in the proof of Corollary 1 and Theorem 1.

Regarding the performance gap between our two novel algorithms, numerical results have shown that NBT-MEXP3 only slightly outperforms MEXP3. Here, we propose two leads that could improve further the NBT-MEXP3 performance. First, the mobility model can be combined with learning techniques to predict the users immediate future position and orientation to help refine the beams search set and adapt faster to the dynamic variations of the channel. Second, the size of the neighbors set can be optimized and timely-adapted using an online learning technique instead of being fixed to the two adjacent neighbors.

Another perspective related to chapter 3 concerns the choice of the learning parameters for the different exponential learning algorithms. In our simulations, we set their values following an empirical trial basis. The tuning of theses parameters is very important to improve the algorithms' performance. As a result, an inappropriate choice could lead to very poor performance. That is why it becomes interesting to explore the possibility of exploiting online optimization algorithms to provide a more efficient method for fine tuning these parameters in an adaptive manner.

In the second part of the thesis, we designed a neural network taking as input only the sub-6 GHz channel of the served user for mmWave beam prediction. No additional information regarding multi-link interference is provided. We argued that incorporating such information could lead to a more complex network architecture. Future work may consider including the multi-link interference in the neural network design, architecture and training. Once this is done, it can be of interest to compare both approaches (with and without interference) and discuss the impact of the downlink multi-link interference on the performance of the proposed channel-beam mapping to see whether it is justified or not to complexify the architecture.

Finally, deep reinforcement learning techniques that combine the deep learning advantage of modeling complex relationships with the online adaptability of reinforcement learning (no offline training) could be investigated for the distributed beam steering problem. Multi agent deep reinforcement learning [124] is a promising framework for the multi-link case.

5.2.2 Long-term perspectives

More advanced machine learning tools can be leveraged for the beam alignment problem. One relevant technique is meta-learning (or learning to learn) [125]. Meta-learning

aims to mitigate conventional ML inefficiencies in terms of data and training time requirements. It exploits domain knowledge to automatize the choice of inductive bias (i.e. a learning model class with a specific learning procedure) using data from tasks that can be related to future tasks of interest. Specifically, meta-learning can be leveraged for our beam selection problem to learn a global model for multi-link networks or different related beam selection tasks such as beam selection in indoor and outdoor environments, beam selection at different carrier frequencies, etc.

Another possible long-term extension would be to consider the other two communication challenges identified in chapter 2 (full duplex communications and network densification), which were not treated in this thesis. These challenges have been studied recently as we have reported in chapter 2. However, some remaining issues can be of future interest such as exploiting advanced online learning algorithms for novel self interference mitigation techniques for full duplex systems or interference management for highly dense networks.

At last, moving up in the electromagnetic spectrum, terahertz frequencies (higher than 300 GHz) [126] are attracting much attention for beyond mmWave networks. Given the similarities of propagation conditions with the current mmWave band, the beam alignment problem remains highly relevant and challenging in the terahertz band. An interesting approach could be using meta-learning and transfer learning [127] to exploit the domain knowledge acquired at mmWave frequencies to propose novel and adapted beam alignment policies for terahertz frequencies.

6

Bibliography

[1] I. Geneva, "Attenuation by atmospheric gases in the frequency range 1-350 GHz," *Tech. Rep. ITU-R*, pp. 676–2, 1995.

[2] Z. Qingling and J. Li, "Rain attenuation in millimeter wave ranges," in *IEEE 7th International Symposium on Antennas Propagation and EM Theory*, 2006, pp. 1–4.

[3] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE communications magazine*, vol. 52, no. 2, pp. 106–113, 2014.

[4] G. R. MacCartney and T. S. Rappaport, "73 GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York city," in *IEEE ICC*, 2014, pp. 4862–4867.

[5] L. De Nardis and M.-G. Di Benedetto, "Momo: a group mobility model for future generation mobile wireless networks," *arXiv preprint arXiv:1704.03065*, 2017.

[6] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun*, vol. 68, no. 9, pp. 5504–5518, 2020.

[7] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in *IEEE ICC*, 2013, pp. 5163–5167.

[8] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.

[9] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave (mmWave) communications for 5G: Opportunities and challenges." arxiv preprint," *arXiv*, vol. 1502, 2015.

[10] H. T. Friis, "A note on a simple transmission formula," *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, 1946.

[11] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—with a focus on propagation models," *IEEE Trans. Antennas Propag*, vol. 65, no. 12, pp. 6213–6230, 2017.

[12] I. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions and link-budget," *IEEE Communications Surveys and Tutorials*, 2017.

[13] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE JSTSP*, vol. 10, no. 3, pp. 436–453, 2016.

[14] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[15] X. Song, S. Haghighatshoar, and G. Caire, "Efficient beam alignment for mmWave single-carrier systems with hybrid MIMO transceivers," *arXiv preprint arXiv:1806.06425*, 2018.

[16] 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*, vol. 6, pp. 37 328–37 348, 2018.

[17] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE JSAC*, vol. 32, no. 6, pp. 1164–1179, 2014.

[18] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE JSTSP*, vol. 8, no. 5, pp. 831–846, 2014.

[19] P. . IEEE P802.11 ad, "Wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 GHz band," *IEEE Computer Society*, 2012.

[20] D. E. Berraki, S. M. Armour, and A. R. Nix, "Application of compressive sensing in sparse spatial channel recovery for beamforming in mmWave outdoor systems," in *IEEE WCNC*, 2014, pp. 887–892.

[21] J. Choi, "Beam selection in mm-wave multiuser MIMO systems using compressive sensing," *IEEE Trans. Commun*, vol. 63, no. 8, pp. 2936–2947, 2015.

[22] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[23] R. Wang, O. Onireti, L. Zhang, M. A. Imran, G. Ren, J. Qiu, and T. Tian, "Reinforcement learning method for beam management in millimeter-wave networks," in *2019 UCET*. IEEE, 2019, pp. 1–4.

[24] M. Elsayed, M. Erol-Kantarci, and H. Yanikomeroglu, "Transfer reinforcement learning for 5G-NR mm-Wave networks," *IEEE Trans. Wireless Commun*, 2020.

[25] V. Raj, N. Nayak, and S. Kalyani, "Deep reinforcement learning based blind mmWave MIMO beam alignment," *arXiv preprint arXiv:2001.09251*, 2020.

[26] M. Hashemi, A. Sabharwal, C. E. Koksal, and N. B. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," *arXiv preprint arXiv:1712.00702*, 2017.

[27] A. Asadi, S. Müller, G. H. A. Sim, A. Klein, and M. Hollick, "FML: Fast Machine Learning for 5G mmWave Vehicular Communications," in *IEEE INFOCOM*, 2018.

[28] J.-B. Wang, M. Cheng, J.-Y. Wang, M. Lin, Y. Wu, H. Zhu, and J. Wang, "Bandit inspired beam searching scheme for mmWave high-speed train communications," *arXiv preprint arXiv:1810.06150*, 2018.

[29] V. Va, T. Shimizu, G. Bansal, and R. W. Heath Jr, "Online learning for position-aided millimeter wave beam training," *arXiv preprint arXiv:1809.03014*, 2018.

[30] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang *et al.*, "Fast mmWave beam alignment via correlated bandit learning," *arXiv preprint arXiv:1909.03313*, 2019.

[31] J. Zhang, Y. Huang, Y. Zhou, and X. You, "Beam alignment and tracking for millimeter wave communications via bandit learning," *IEEE Trans. Commun*, vol. 68, no. 9, pp. 5519–5533, 2020.

[32] E. V. Belmega, P. Mertikopoulos, R. Negrel, and L. Sanguinetti, "Online convex optimization and no-regret learning: Algorithms, guarantees and applications," *arXiv preprint arXiv:1804.04529*, 2018.

[33] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun*, vol. 67, no. 10, pp. 7331–7376, 2019.

[34] M. S. Sim, Y.-G. Lim, S. H. Park, L. Dai, and C.-B. Chae, "Deep learning-based mmWave beam selection for 5G NR/6G with sub-6 GHz channel information: Algorithms and prototype validation," *IEEE Access*, vol. 8, pp. 51 634–51 646, 2020.

[35] A. M. Elbir and S. Coleri, "Federated learning for hybrid beamforming in mm-wave massive MIMO," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2795–2799, 2020.

[36] T. Peken, R. Tandon, and T. Bose, "Unsupervised mmWave beamforming via autoencoders," in *IEEE ICC*, 2020, pp. 1–6.

[37] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2018.

[38] D. Jagyasi and M. Coupechoux, "DNN based beam selection in mmw heterogeneous networks," *arXiv preprint arXiv:2102.02672*, 2021.

[39] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "Gambling in a rigged casino: The adversarial multi-armed bandit problem," in *36th Annual Symposium on Foundations of Computer Science*. IEEE, 1995, pp. 322–331.

[40] NI, "mmWave: The battle of the bands [white paper]," https://www.ni.com/fr-fr/innovations/white-papers/16/mmwave–the-battle-of-the-bands.

[41] Y. Azar, G. N. Wong, K. Wang, R. Mayzus, J. K. Schulz, H. Zhao, F. Gutierrez, D. Hwang, and T. S. Rappaport, "28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York city," in *IEEE ICC*, 2013, pp. 5143–5147.

[42] H. Xu, T. S. Rappaport, R. J. Boyle, and J. H. Schaffner, "Measurements and models for 38-GHz point-to-multipoint radiowave propagation," *IEEE JSAC*, vol. 18, no. 3, pp. 310–321, 2000.

[43] J. N. Murdock, E. Ben-Dor, Y. Qiao, J. I. Tamir, and T. S. Rappaport, "A 38 GHz cellular outage study for an urban outdoor campus environment," in *IEEE WCNC*, 2012, pp. 3085–3090.

[44] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag*, vol. 61, no. 4, pp. 1850–1859, 2012.

[45] A. M. Al-Samman, T. Abd Rahman, and M. H. Azmi, "Indoor corridor wideband radio propagation measurements and channel models for 5G millimeter wave wireless communications at 19 GHz, 28 GHz, and 38 GHz bands," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.

[46] S. Nie, M. K. Samimi, T. Wu, S. Deng, G. R. MacCartney Jr, and T. S. Rappaport, "73 GHz millimeter-wave indoor and foliage propagation channel measurements and results," *NYU WIRELESS Technical Report, TR-2014–003*, 2014.

[47] C. J. Hansen, "Wigig: Multi-gigabit wireless communications in the 60 GHz band," *IEEE Wireless Communications*, vol. 18, no. 6, pp. 6–7, 2011.

[48] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE JSAC*, vol. 32, no. 6, pp. 1152–1163, 2014.

[49] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[50] C.J.GIBBINS, "Radiowave propagation in the millimetric bands," *Business Opportunities in the Millimetric Wavebands*, 1990.

[51] "Attenuation due to clouds and fog," *ITU-R Recommendation 840*, 1992.

[52] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE JSAC*, vol. 32, no. 6, pp. 1164–1179, 2014.

[53] C. C. Tan and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. on Commun*, vol. 48, no. 12, pp. 2032–2040, 2000.

[54] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wirel. Commun*, vol. 13, no. 3, pp. 1499–1513, 2014.

[55] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[56] F. Gholam, J. Vía, and I. Santamaría, "Beamforming design for simplified analog antenna combining architectures," *IEEE Trans. Veh Technol*, vol. 60, no. 5, pp. 2373–2378, 2011.

[57] J. Wang, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE JSAC*, vol. 27, no. 8, 2009.

[58] P. Xia, S.-K. Yong, J. Oh, and C. Ngo, "Multi-stage iterative antenna training for millimeter wave communications," in *IEEE GLOBECOM*, 2008, pp. 1–6.

[59] I. 802.11ad, "Ieee 802.11ad standard draft d0.1." in *[Online]. Available: www.ieee802.org/11/Reports/tgadupdate.htm*. IEEE.

[60] C. Kim, T. Kim, and J.-Y. Seol, "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in *IEEE Globecom Workshops*, 2013, pp. 61–65.

[61] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process*, vol. 53, no. 11, pp. 4091–4103, 2005.

[62] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Trans. Wirel. Commun*, vol. 5, no. 12, 2006.

[63] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, 2014.

[64] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer, "Steering with eyes closed: mmWave beam steering without in-band measurement," in *IEEE INFOCOM*, 2015, pp. 2416–2424.

[65] S. K. Sharma, T. E. Bogale, L. B. Le, S. Chatzinotas, X. Wang, and B. Ottersten, "Dynamic spectrum sharing in 5G wireless networks with full-duplex technology: Recent advances and research challenges," *IEEE Communications Surveys and Tutorials*, 2017.

[66] Z. Xiao, P. Xia, and X.-G. Xia, "Full-duplex millimeter-wave communication," *arXiv preprint arXiv:1709.07983*, 2017.

[67] X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia, and X.-G. Xia, "Beamforming based full-duplex for millimeter-wave communication," *Sensors*, vol. 16, no. 7, p. 1130, 2016.

[68] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE JSTSP*, vol. 10, no. 3, pp. 501–513, 2016.

[69] M. Jain, J. I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proceedings of the 17th ICMCN*. ACM, 2011, pp. 301–312.

[70] K. Miura and M. Bandai, "Node architecture and MAC protocol for full duplex wireless and directional antennas," in *IEEE PIMRC*. IEEE, 2012, pp. 369–374.

[71] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, 2014.

[72] S. Talwar, D. Choudhury, K. Dimou, E. Aryafar, B. Bangerter, and K. Stewart, "Enabling technologies and architectures for 5G wireless," in *IEEE IMS*. IEEE, 2014, pp. 1–4.

[73] R. Mudumbai, S. Singh, and U. Madhow, "Medium access control for 60 GHz outdoor mesh networks with highly directional links," in *IEEE INFOCOM*, 2009, pp. 2871–2875.

[74] A. Khandekar, N. Bhushan, J. Tingfang, and V. Vanghi, "LTE-advanced: Heterogeneous networks," in *European Wireless Conference*. IEEE, 2010, pp. 978–982.

[75] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Trans. Wireless Commun*, vol. 18, no. 3, 2011.

[76] I. Muhammad and Z. Yan, "Supervised machine learning approaches: A survey." *ICTACT Journal on Soft Computing*, vol. 5, no. 3, 2015.

[77] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–36, 2018.

[78] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless communications*, vol. 24, no. 5, pp. 175–183, 2017.

[79] F. D. Calabrese, L. Wang, E. Ghadimi, G. Peters, and P. Soldati, "Learning radio resource management in 5G networks: Framework, opportunities and challenges," *arXiv preprint arXiv:1611.10253*, 2016.

[80] L. Sanguinetti, A. Zappone, and M. Debbah, "A deep-learning framework for energy-efficient resource allocation in massive MIMO systems," in *Asilomar Conference on Signals, Systems, and Computers*, 2018.

[81] M. E. Celebi and K. Aydin, *Unsupervised learning algorithms*. Springer, 2016.

[82] S. Laha, N. Chowdhury, and R. Karmakar, "How can machine learning impact on wireless network and IOT?–a survey," in *11th IEEE ICCCNT*, 2020, pp. 1–7.

[83] N. Ye, X. Li, J. Pan, W. Liu, and X. Hou, "Beam aggregation-based mmWave MIMO-NOMA: An AI-enhanced approach," *IEEE Trans. Veh Technol*, vol. 70, no. 3, pp. 2337–2348, 2021.

[84] J. Cui, Z. Ding, and P. Fan, "The application of machine learning in mmWave-NOMA systems," in *IEEE VTC*, 2018, pp. 1–6.

[85] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[86] B. Hussain, Q. Du, and P. Ren, "Semi-supervised learning based big data-driven anomaly detection in mobile wireless networks," *China Communications*, vol. 15, no. 4, pp. 41–57, 2018.

[87] M. Camelo, A. Shahid, J. Fontaine, F. A. P. de Figueiredo, E. De Poorter, I. Moerman, and S. Latre, "A semi-supervised learning approach towards automatic wireless technology recognition," in *IEEE DYSPAN*, 2019, pp. 1–10.

[88] V. Radu, P. Katsikouli, R. Sarkar, and M. K. Marina, "A semi-supervised learning approach for robust indoor-outdoor detection with smartphones," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, 2014, pp. 280–294.

[89] Y. Long, Z. Chen, and S. Murphy, "Broad learning based hybrid beamforming for mm-wave MIMO in time-varying environments," *IEEE Communications Letters*, vol. 24, no. 2, pp. 358–361, 2019.

[90] A. Gosavi, "Reinforcement learning: A tutorial survey and recent advances," *INFORMS Journal on Computing*, vol. 21, no. 2, pp. 178–192, 2009.

[91] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[92] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.

[93] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[94] L. Zhou, "A survey on contextual multi-armed bandits," *arXiv preprint arXiv:1508.03326*, 2015.

[95] D. Bouneffouf and I. Rish, "A survey on practical applications of multi-armed and contextual bandits," *arXiv preprint arXiv:1904.10040*, 2019.

[96] I. Aykin, B. Akgun, M. Feng, and M. Krunz, "Mamba: A multi-armed bandit framework for beam tracking in millimeter-wave systems," in *IEEE INFOCOM*, 2020, pp. 1469–1478.

[97] A. Marcastel, E. V. Belmega, P. Mertikopoulos, and I. Fijalkow, "Online power optimization in feedback-limited, dynamic and unpredictable IOT networks," *IEEE Trans. Signal Process*, vol. 67, no. 11, pp. 2987–3000, 2019.

[98] Á. López-Raventós and B. Bellalta, "Concurrent decentralized channel allocation and access point selection using multi-armed bandits in multi bss wlans," *Computer Networks*, vol. 180, p. 107381, 2020.

[99] J. Song, J. Choi, and D. J. Love, "Codebook design for hybrid beamforming in millimeter wave systems," in *IEEE ICC*, 2015, pp. 1298–1303.

[100] S. Sun, T. A. Thomas, T. S. Rappaport, H. Nguyen, I. Z. Kovacs, and I. Rodriguez, "Path loss, shadow fading, and line-of-sight probability models for 5G urban macro-cellular scenarios," in *IEEE Globecom Workshops*, 2015, pp. 1–7.

[101] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. Wireless Commun*, vol. 17, no. 3, pp. 1822–1835, 2018.

[102] C.-H. Yao, Y.-Y. Chen, B. P. Sahoo, and H.-Y. Wei, "Outage reduction with joint scheduling and power allocation in 5G mmWave cellular networks," in *IEEE PIMRC*, 2017, pp. 1–6.

[103] J. N. Murdock, E. Ben-Dor, Y. Qiao, J. I. Tamir, and T. S. Rappaport, "A 38 GHz cellular outage study for an urban outdoor campus environment," in *IEEE WCNC*, 2012, pp. 3085–3090.

[104] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[105] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecommun*, vol. 10, no. 6, pp. 585–595, 1999.

[106] S. Shalev-Shwartz *et al.*, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2012.

[107] E. TC48, "High rate 60 GHz phy, mac and hdmi pal," *ECMA standard*, vol. 387, 2008.

[108] J. P. GILB, "Part 15.3: Wireless medium access control and physical layer specifications for high rate wireless personal area networks: Amendment 2: Millimeter-wave based alternative physical layer extension," *P802-15-3c-DF3_Draft_Amendment*.

[109] I. S. Association, "IEEE std 802.11 ad-2012, part 11: Wireless Lan Medium Access Control and physical layer specifications, amendment 3: Enhancements for very high throughput in the 60 GHz band," *IEEE Computer Society*, 2012.

[110] H. S. Ghadikolaei, "Mac aspects of millimeter-wave cellular networks," in *Wireless Mesh Networks-Security, Architectures and Protocols*. IntechOpen, 2019.

[111] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth, "How to use expert advice," *Journal of the ACM*, vol. 44, no. 3, pp. 427–485, 1997.

[112] F. Khan, Z. Pi, and S. Rajagopal, "Millimeter-wave mobile broadband with large scale spatial processing for 5G mobile communication," in *IEEE Allerton*, 2012, pp. 1517–1523.

[113] C. Herranz, M. Zhang, M. Mezzavilla, D. Martin-Sacristán, S. Rangan, and J. F. Monserrat, "A 3 GPP NR compliant beam management framework to simulate end-to-end mmWave networks," in *ACM MSWIM*, 2018, pp. 119–125.

[114] J. Lee, J. Liang, M.-D. Kim, J.-J. Park, B. Park, and H. K. Chung, "Measurement-based propagation channel characteristics for millimeter-wave 5G Giga communication systems," *ETRI Journal*, vol. 38, no. 6, pp. 1031–1041, 2016.

[115] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, 2016.

[116] W. Huang, Y. Huang, R. Zhao, S. He, and L. Yang, "Wideband millimeter wave communication: Single carrier based hybrid precoding with sparse optimization," *IEEE Trans. Veh Technol*, vol. 67, no. 10, pp. 9696–9710, 2018.

[117] K. Shen and W. Yu, "Fractional programming for communication systems — part I: Power control and beamforming," *IEEE Trans. Signal Process*, vol. 66, no. 10, pp. 2616–2630, 2018.

[118] A. Alkhateeb, "Deepmimo: A generic deep learning dataset for millimeter wave and massive MIMO applications," *arXiv preprint arXiv:1902.06435*, 2019.

[119] Remcom, "Wireless Insite," http://www.remcom.com/wireless-insite.

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

[121] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th OSDI*, 2016, pp. 265–283.

[122] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Federated learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 13, no. 3, pp. 1–207, 2019.

[123] 3GPP, "NR; physical channels and modulation (release 15)," *3rd Generation Partnership Project (3GPP)*, vol. 38.211, p. V15.2.0, 2018.

[124] P. Hernandez-Leal, B. Kartal, and M. E. Taylor, "A survey and critique of multiagent deep reinforcement learning," *Autonomous Agents and Multi-Agent Systems*, vol. 33, no. 6, pp. 750–797, 2019.

[125] O. Simeone, S. Park, and J. Kang, "From learning to meta-learning: Reduced training overhead and complexity for communication systems," in *2020 2nd 6G Wireless Summit*. IEEE, 2020, pp. 1–5.

[126] Z. Chen, X. Ma, B. Zhang, Y. Zhang, Z. Niu, N. Kuang, W. Chen, L. Li, and S. Li, "A survey on terahertz communications," *China Communications*, vol. 16, no. 2, pp. 1–35, 2019.

[127] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.

[128] D. L. Donoho *et al.*, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[129] J. Karjalainen, M. Nekovee, H. Benn, W. Kim, J. Park, and H. Sungsoo, "Challenges and opportunities of mm-wave communication in 5G networks," in *IEEE CROWNCOM*, 2014, pp. 372–376.

[130] IEEE, "IEEE 802.11 ad, Amendment 3: Enhancements for very high throughput in the 60 GHz band," 2012.

[131] I. Chafaa, E. V. Belmega, and M. Debbah, "One-bit feedback exponential learning for beam alignment in mobile mmWave," *IEEE Access*, vol. 8, pp. 194 575–194 589, 2020.

A

Proof of Theorem 1

We start by the following lemma which will be exploited in the main proof [1].

Lemma 1. *For the parameters* $A \geq 3$, $\eta = \dfrac{\gamma}{\beta A}$, $\beta \geq \max\left\{1, \sqrt{\dfrac{2}{A}}\ \left(\gamma - \dfrac{\gamma}{A}\right)^{-1}\right\}$ *and* $0 < \gamma \leq 1 - \dfrac{2}{A}$, *we have*

$$p_{T,i_t}(t)\ \tilde{r}^2_{T,i_t}(t) \leq \frac{\beta^2\ A}{2\ (1-\gamma)},\ \forall t \geq 1.$$

Since $0 \leq p_{T,i_t}(t) \leq 1$, to prove this result, it suffices to show that for all $t \geq 1$

$$0 \leq \frac{2\ (1-\gamma)\ \beta^{\ 2r_{i_t,j_t}(t)}}{\beta^2\ A\ \left(1 - r_{i_t,j_t}(t) + (-1)^{1+r_{i_t,j_t}(t)}\hat{p}_{T,i_t}(t)\right)^2} \leq 1, \tag{A.1}$$

We distinguish the two cases depending on the value of the reward of the chosen actions. a) If $r_{i_t,j_t}(t) = 1$, the inequalities in (A.1) are met for $\gamma \leq 1 - \dfrac{2}{A}$. b) If $r_{i_t,j_t}(t) = 0$, the inequalities in (A.1) are true for
$\beta \geq \max\left\{1, \sqrt{\dfrac{2}{A}}\ \left(\gamma - \dfrac{\gamma}{A}\right)^{-1}\right\}$.

For the main proof of Theorem 1, we define the sum $S_t \triangleq \sum_{i=1}^{A} \exp\left(\eta\ G_{T,i}\left(t-1\right)\right)$, $\forall t \geq 1$. Then, for any $A > 1$, $\eta > 0$, $\beta \geq 1$ and $0 < \gamma < 1$, we have the following ratio

$$\frac{S_{t+1}}{S_t} \quad = \quad \sum_{i=1}^{A} p_{T,i}(t) \exp\left(\eta\ \tilde{r}_{T,i}(t)\right). \tag{A.2}$$

From Lemma 3.3 in [39] and the inequality $\tilde{r}_{T,i}(t) \leq \dfrac{\beta A}{\gamma}$, the following holds:

$$\exp\left(\eta\ \tilde{r}_{T,i}(t)\right) \leq 1 + \eta\ \tilde{r}_{T,i}(t) + \Phi_M(\eta)\ \tilde{r}^2_{T,i}(t)$$

with $\Phi_M(\eta) = \dfrac{\exp(M\eta) - 1 - M\ \eta}{M^2}$ and $M = \dfrac{\beta A}{\gamma}$.
Using this inequality, the ratio in (A.2) can be expressed as

$$\frac{S_{t+1}}{S_t} \leq 1 + \eta\ p_{T,i_t}(t)\ \tilde{r}_{T,i_t}(t) + \Phi_M(\eta)\ p_{T,i_t}(t)\ \tilde{r}^2_{T,i_t}(t), \tag{A.3}$$

[1] We provide a proof to Theorem 1 at the transmitter side as similar steps hold for the regret bound at the receiver.

since the unchosen beams at time t have a zero reward.

Next, the idea is to upper bound the last two terms of the inequality (A.3) as follows:

$$p_{T,i_t}(t)\ \tilde{r}_{T,i_t}(t) \leq \frac{\beta\ r_{i_t,j_t}(t)}{1-\gamma},\ \forall t \geq 1, \tag{A.4}$$

$$p_{T,i_t}(t)\ \tilde{r}^2_{T,i_t}(t) \leq \frac{\beta^2\ A}{2\ (1-\gamma)}, \forall t \geq 1, \tag{A.5}$$

the latter follows from Lemma 1.
Since $\eta > 0$ and $\Phi_M(\eta) > 0$, combining (A.3), (A.4), (A.5) and the fact that $1 + x \leq \exp(x),\ \forall x \in \mathbb{R}$ leads to

$$\frac{S_{t+1}}{S_t} \ \leq\ \exp\left(\frac{\eta\ \beta\ r_{i_t,j_t}(t)}{1-\gamma} + \frac{\Phi_M(\eta)\ \beta^2\ A}{2\ (1-\gamma)}\right). \tag{A.6}$$

Now, by first taking the logarithm and then summing over $t = 1, ..., \mathcal{T}$ in the above, we further obtain

$$\ln \frac{S_{\mathcal{T}+1}}{S_1} \leq \frac{\eta\ \beta}{1-\gamma}\left(\sum_{t=1}^{\mathcal{T}} r_{i_t,j_t}(t)\right) + \frac{\Phi_M(\eta)\ \beta^2\ A}{2\ (1-\gamma)}\mathcal{T}. \tag{A.7}$$

Since $S_1 = A$ and $S_{\mathcal{T}+1} \geq \exp\left(\eta\ G_{T,k}(\mathcal{T})\right)$, for an arbitrary and fixed $k \in \{1, \ldots, A\}$, we get

$$\ln \frac{S_{\mathcal{T}+1}}{S_1} \geq \eta\ G_{T,k}(\mathcal{T}) - \ln A,\ \forall k \tag{A.8}$$

with $G_{T,k}(\mathcal{T}) = \sum_{t=1}^{\mathcal{T}} \tilde{r}_{T,k}(t)$.

Combining (A.7) and (A.8), we can bound the overall rewards of the algorithm, denoted by G_{Alg}, as follows

$$\begin{aligned} G_{\text{Alg}} &\triangleq \sum_{t=1}^{\mathcal{T}} r_{i_t,j_t}(t) \\ &\geq \frac{1-\gamma}{\beta}\sum_{t=1}^{\mathcal{T}} \tilde{r}_{T,k}(t) - \frac{1-\gamma}{\eta\ \beta}\ln A - \frac{\Phi_M(\eta)\ \beta\ A\ \mathcal{T}}{2\eta} \end{aligned}$$

Taking the expectation with respect to the distribution of the chosen beams $\langle i_1, ..., i_{\mathcal{T}}\rangle$ in the random online policy, we obtain

$$\mathbb{E}[G_{\text{Alg}}] \geq \frac{1-\gamma}{\beta}\sum_{t=1}^{\mathcal{T}} \mathbb{E}[\tilde{r}_{T,k}(t)] - \frac{1-\gamma}{\eta\ \beta}\ln A - \frac{\Phi_M(\eta)\ \beta\ A\ \mathcal{T}}{2\eta},$$

with

$$\mathbb{E}[\tilde{r}_{T,k}(t)] \ = \ \begin{cases} \beta, & \text{if } \ r_{k,j_t}(t) = 1, \\ \dfrac{-\hat{p}_{T,k}(t)}{1-\hat{p}_{T,k}(t)}, & \text{if } \ r_{k,j_t}(t) = 0. \end{cases}$$

We can now show that

$$\mathbb{E}[G_{\text{Alg}}] \geq (1-\gamma)\mathcal{T} - \frac{1-\gamma}{\eta\ \beta}\ln A - \frac{\Phi_M(\eta)\ \beta\ A\ \mathcal{T}}{2\eta},\ \forall t \geq 1.$$

Next, let $EG_{\max} \triangleq \max_{I,J} \sum_{t=1}^{\mathcal{T}} \mathbb{E}[r_{i_t,j_t}(t)]$, denote the expected cumulative rewards of the oracle best solution in hindsight. Knowing that it can not be higher than $\mathcal{T}$, as all rewards are either 0 or 1, we show that

$$EG_{\max} - \mathbb{E}[G_{\text{Alg}}] \leq \gamma\ \mathcal{T} + \frac{1-\gamma}{\eta\ \beta}\ln A + \frac{\Phi_M(\eta)\ \beta\ A\ \mathcal{T}}{2\eta}. \tag{A.9}$$

Substituting $\Phi_M(\eta) = \frac{\gamma^2}{\beta^2\ A^2}\left(\exp\left(\frac{\beta\eta A}{\gamma}\right) - 1 - \frac{\beta\eta A}{\gamma}\right)$ and $\eta = \frac{\gamma}{\beta\ A}$ in (A.9) yields to

$$EG_{\max} - \mathbb{E}[G_{\text{Alg}}] \leq \frac{(1-\gamma)\ A\ln A}{\gamma} + \frac{\exp\{(1)\}\ \gamma\ \mathcal{T}}{2}.$$

Hence, the expected average regret is upper bounded as

$$\frac{EG_{\max} - \mathbb{E}[G_{\text{Alg}}]}{\mathcal{T}} \quad \leq \quad \frac{A\ln A}{\gamma\ \mathcal{T}} + \frac{\exp\{(1)\}\ \gamma}{2}.$$

The upper bound above is a convex function with respect to γ. We can thus minimize it and obtain the optimal step-size $\gamma = \sqrt{\frac{2\ A\ln A}{\exp\{(1)\}\ \mathcal{T}}}$ and the following optimal bound of the expected average regret

$$\frac{EG_{\max} - \mathbb{E}[G_{\text{Alg}}]}{\mathcal{T}} \leq \sqrt{2\ \exp\{(1)\}}\sqrt{\frac{A\ln A}{\mathcal{T}}},$$

which completes our proof.

B

French summary

B.1 Introduction générale

La bande des ondes millimétriques est considérée comme une alternative à la bande fréquentielle traditionnelle au dessous de 6 GHz [8, 9] qui souffre d'une congestion et ne permet plus d'atteindre les hauts débits nécessaires pour les applications des nouveaux réseaux de communications comme la transmission des vidéos de très haute définition qui nécessitent une large bande spectrale.

B.1.1 Contexte et motivations

La propagation des ondes millimétriques est particulièrement affectée par de pertes importantes de l'énergie du signal transmis à cause du pathloss en espace libre [10] (qui augmente avec la fréquence) et autres pertes dues à la pénétration des autres objets de l'environnement ou à l'absorption par des particules de l'atmosphère [11, 12]. Ces conditions difficiles de propagation ont conduit à l'utilisation des faisceaux hautement directif, en exploitant des larges réseaux d'antenne avec des techniques de formation de voies (beamforming) [13, 14], pour focaliser l'énergie du signal transmis sur son récepteur désiré et compenser les pertes de propagation. De plus, les faibles longueurs d'onde dans la bande millimétrique permettent d'incorporer un nombre important des antennes dans des réseaux de petite taille afin d'obtenir un large gain du beamforming [15]. Tous ces caractéristiques spécifiques à la bande des ondes millimétrique ont mené à un ensemble de défis pratiques concernant le déploiement des systèmes sans fil à onde millimétrique comme nous l'avons expliqué dans le deuxième chapitre. Dans cette thèse, nous focalisons sur un défi fondamental qui est le problème de gestion de ces faisceaux dans des systèmes dynamiques opérant dans la bande des ondes millimétriques.

En effet, ce type de communication directive conduit au problème d'alignement des faisceaux ou les faisceaux de l'émetteur et ceux du récepteur doivent être ajustés et alignés constamment pour garantir un lien de communication fiable. Le d'alignement des faisceaux pose aussi le défi d'une large charge de signalisation et d'entraînement sur le réseaux qui s'intensifient avec l'augmentation du nombre des antennes et la taille du dictionnaire du beamforming.

En outre, les méthodes d'alignement des faisceaux doivent supporter la mobilité des utilisateurs et les variations imprévisibles et même non stochastiques du réseau (par exemple, à cause du comportement des utilisateurs et leurs collectivité intermittente). Dans ces conditions dynamiques, l'ajustement des faisceaux et l'identification des vecteurs

optimaux de beamforming impliquent une charge additionnelle de signalisation et d'entraînement qui affecte les performances du système. Par conséquent, il est indispensable de concevoir des méthodes en ligne et adaptatives pour l'alignement des faisceaux afin de permettre le déploiement des ondes millimétriques pour des applications mobiles à très haut débit comme la réalité virtuelle, véhicules autonomes, etc. Ceci représente la problématique principale de cette thèse: *Comment identifier les meilleures vecteurs de beamforming dans des systèmes dynamiques à onde millimétriques ?*

B.1.2 État de l'art

Le problème d'alignement des faisceaux a été largement étudié dans la littérature. Les premières approches classiques peuvent être divisées en deux catégories: l'entraînement des faisceaux et l'acquisition comprimée [16]. La première approche consiste à entraîner un ensemble candidat prédéfini des vecteurs de beamforming par une recherche exhaustive [17] ou par une recherche hiérarchique adaptative [18, 19] afin d'identifier des bons beamformers en terme d'une certaine métrique (par exemple le rapport signal bruit). Les principales limitations de cette approche réside dans la charge significative d'entraînement et de signalisation qui les rendent inutiles pour des applications mobile.

Les méthodes basées sur l'acquisition comprimée [20, 21] estiment les paramètres du canal millimétrique (gain des trajets de propagation, angles de départ, angle d'arrivée), en se basant sur la nature parcimonieuse du canal, pour construire convenablement des vecteurs de beamforming. Ceci réduit le coût d'entraînement par rapport aux méthodes précédentes mais nécessite des hypothèses sur les variations temporelles du canal (statique, stochastique) durant la phase d'estimation imposant des limitations lorsque le canal est hautement dynamique et possiblement non stochastique.

Récemment, les outils de l'apprentissage automatique ont été considérés grâce à leurs capacité à apprendre des modèles inconnus et complexes. L'apprentissage automatique permet d'apprendre approximativement la fonction complexe qui relie des éléments de l'environnement sans fil (position, fréquence, ...) au choix optimal des vecteurs de beamforming avec moins d'hypothèses sur la dynamique du réseau. Dans ce contexte, deux approches principales ont été exploitées pour l'alignement des faisceaux: l'apprentissage par renforcement et l'apprentissage profond.

L'apprentissage par renforcement (RL) [22] repose sur des interactions en ligne entre l'agent d'apprentissage et son environnement sans fil, ce qui donne lieu à un feedback utilisé pour déterminer une direction convenable du faisceau. Les auteurs de [23] ont utilisé le Q-learning pour sélectionner les faisceaux qui répondent aux exigences de qualité de service à l'utilisateur. Une autre politique d'alignement des faisceaux basée sur le Q-learning est proposée dans [24] pour optimiser la capacité du réseau. Dans [25], l'apprentissage par renforcement profond est utilisé pour effectuer l'alignement de faisceaux pour les systèmes multi-utilisateurs en ondes millimétriques. Dans cette thèse, nous nous concentrons sur l'application d'un cas particulier d'apprentissage par renforcement : les bandits manchot (multi-armed bandits: MAB) pour l'alignement de faisceau, qui a reçu un intérêt significatif dans la littérature récente.

Dans le cadre du MAB, l'alignement des faisceaux est reformulé en un problème de prise de décision séquentielle, dans lequel les dispositifs (par exemple, un nœud central ou l'émetteur/récepteur) choisissent à chaque étape un bras (c'est-à-dire une direction de faisceau), parmi un ensemble fini de choix, et observent ensuite une récompense (par exemple, le rapport signal bruit résultant au niveau du récepteur). Les dispositifs apprennent alors la meilleure direction de faisceau en exploitant conjointement les récompenses passées observées et en explorant de nouvelles directions de faisceau. Les auteurs de [26] ont proposé l'algorithme d'alignement de faisceau qui restreint l'ensemble de recherche des meilleures directions en utilisant la corrélation entre les faisceaux consécutifs et l'unimodalité de la puissance du signal reçu. Un algorithme d'alignement de faisceau en ligne pour les communications véhiculaires à ondes millimétriques a été présenté dans [27], qui utilise la direction d'arrivée du véhicule comme information contextuelle. Dans [28], une autre méthode d'alignement des faisceaux a été étudiée, qui nécessite une connaissance parfaite des débits de données pour toutes les directions de faisceau choisies. La méthode de [29] incorpore l'emplacement du récepteur comme une information supplémentaire hors bande pour améliorer l'alignement du faisceau. Dans [30], la technique proposée vise à réduire le délai d'alignement du faisceau en exploitant les connaissances acquises précédemment sur le canal. Un autre algorithme, basé sur les MAB, est également proposé dans [31] pour l'alignement et le suivi du faisceau.

Toutes les méthodes existantes basées sur les MABs citées ci-dessus dépendent d'une autorité centrale, qui choisit d'abord conjointement la meilleure paire de directions de faisceau de l'émetteur et du récepteur, puis renvoie le résultat aux deux dispositifs, ce qui entraîne une surcharge de signalisation élevée. De plus, ces approches sont déterministes et exploitent l'algorithme upper-confidence bound (UCB) des bandits stochastiques [32]. Cela implique qu'elles ne sont pertinentes que dans des environnements sans fil stochastiques et stationnaires et qu'elles ne peuvent pas prendre en compte d'autres composantes éventuellement non stationnaires, telles que le comportement imprévisible et la connectivité des dispositifs interférants.

En général, les techniques RL ne nécessitent pas de phase d'apprentissage hors ligne et ont une complexité de calcul relativement faible. Cependant, elles nécessitent un certain temps d'exploration pour identifier les bonnes directions de faisceau (au début de la communication ou après des variations significatives du canal), ce qui peut affecter leurs performances dans les applications sensibles à la latence.

En s'appuyant sur le large succès des réseaux neuronaux, les approches basées sur les données ont récemment trouvé leur place dans les communications sans fil grâce à l'apprentissage profond supervisé et non supervisé [33]. Plusieurs travaux existants exploitent les réseaux neuronaux comme des approximateurs universels pour apprendre la relation entre l'environnement sans fil et les directions optimales des faisceaux. Dans l'article [16], un ensemble de points d'accès se coordonne pour servir un utilisateur en apprenant un faisceau approprié via un apprentissage profond supervisé, qui exploite un signal de liaison montante omnidirectionnel dans la bande millimétrique. La nature centralisée de cette approche implique une signalisation forte entre les différents points d'accès et une entité centrale. De plus, le signal de liaison montante omnidirectionnel mmWave transmis par une seule antenne peut être très limité en portée et en puissance.

Dans l'article [6, 34], un réseau neuronal est entraîné pour projeter les informations sur les canaux sub-6 GHz en vecteurs de formation de faisceau en ondes millimétriques pour une liaison émetteur-récepteur unique. L'approche proposée est basée sur la classification, dans laquelle les vecteurs de formation de faisceau sont sélectionnés à partir d'un dictionnaire de beamforming discret prédéfini, ce qui donne une solution sous-optimale.

Les auteurs de [35] ont utilisé le cadre de l'apprentissage fédéré pour lier les canaux d'ondes millimétriques aux faisceaux analogiques dans un réseau de liaison descendante multi-utilisateurs. Par conséquent, la politique proposée nécessite la connaissance des matrices des canaux à ondes millimétriques, qui sont plus difficiles à estimer et nécessitent un coût d'apprentissage plus important que les canaux à moins de 6 GHz. Un réseau neuronal profond est formé dans [36] suivant l'apprentissage non supervisé pour choisir les vecteurs de formation de faisceau pour les communications en ondes millimétriques. Le travail dans [37] propose une méthode de conception de formation de faisceau en utilisant l'apprentissage non supervisé pour maximiser le débit du réseau. Un travail récent [38] utilise un réseau neuronal profond pour sélectionner conjointement une station de base mmWave et un faisceau offrants les meilleures performances de communication dans des réseaux cellulaires hétérogènes.

Bien que prometteuses, toutes ces méthodes d'apprentissage profond reposent sur la disponibilité d'une quantité suffisante de données pertinentes et sur une conception optimale de l'architecture du réseau neuronal. L'acquisition de données d'entraînement n'est actuellement pas triviale, car elle est à la fois coûteuse et longue, sans oublier les problèmes de confidentialité et de sécurité des données qu'elle peut soulever.

B.1.3 Contributions

Les contributions de cette thèse peuvent être divisées en deux parties principales. La première partie est consacré à l'exploitation des bandits manchots pour l'alignement des faisceaux de l'émetteur et du récepteur d'une manière distribuée. La deuxième partie consiste à utiliser des outils de l'apprentissage profond pour le problème d'alignement des faisceaux et à établir une comparaison entre les deux approches des deux parties.

Dans la première partie de la thèse, nous nous concentrons sur les politiques d'alignement de faisceaux distribués qui ne nécessitent pas l'existence d'un nœud central et ne reposent sur aucune hypothèse concernant la dynamique du réseau, contrairement aux travaux existants [26, 27, 28, 29, 30]. Pour cela, nous nous appuyons sur l'algorithme des poids exponentiels pour l'exploration et l'exploitation (EXP3) [39] pour définir de nouvelles politiques d'alignement de faisceaux capables de s'adapter à de tels *environnements arbitraires et imprévisibles*.

À notre connaissance, notre travail est le premier à exploiter les poids exponentiels pour l'alignement des faisceaux dans les réseaux à ondes millimétriques. Par rapport aux méthodes traditionnelles, nos politiques visent à apprendre les meilleures directions de faisceau d'une manière adaptative et en ligne sans dépendre d'un entraînement prédéployé à chaque fois que le canal change. En effet, il est possible de transmettre

simultanément des données tout en suivant les bons faisceaux dès le début de la transmission. Bien sûr, cela a un coût en termes d'une mauvaise qualité du lien dans les premières étapes du processus d'apprentissage. Le principal avantage de nos politiques adaptatives est que ce coût n'intervient qu'au début de la transmission ou lorsque le canal subit des changements importants, et qu'elles ne nécessitent pas d'entraînement dédié à chaque temps de cohérence du canal (ni d'optimisation de la durée de la phase d'entraînement, qui a un impact crucial sur l'efficacité de la transmission des données). Enfin, nos politiques d'alignement de faisceau en ligne ne nécessitent pas une connaissance parfaite du canal et reposent uniquement sur un bit de retour (de type ACK/NACK) qui permet essentiellement de savoir si le SNR cible a été atteint au niveau du récepteur.

Dans la deuxième partie, nous proposons une méthode d'apprentissage profond non supervisée pour projeter les canaux de liaison montante sous-6 GHz en vecteurs de formation de faisceau d'ondes millimétriques de liaison descendante. Le réseau neuronal que nous proposons prend en entrée les canaux de liaison montante sous 6 GHz et produit directement les vecteurs de formation de faisceau en ondes millimétriques correspondants, en exploitant les modèles et les caractéristiques des canaux d'entrée (statistiques des canaux et caractéristiques de l'environnement).

Contrairement aux méthodes d'apprentissage profond supervisés, notre approche ne nécessite pas de données avec étiquette, par conséquent, ne requiert pas le calcul du faisceau descendant optimal pour chaque échantillon d'entraînement. Contrairement aux travaux existants [6, 34], nous formulons le problème de la fonction canal-faisceau comme une régression et non une classification, ce qui implique que les faisceaux prédits ont une résolution angulaire continue (et donc plus élevée) et peuvent surmonter la sous-optimalité causée par un dictionnaire de beamforming quantifié. Par rapport à [6], nous concevons une architecture de réseau neuronal plus simple, avec moins de paramètres d'apprentissage, optimisée sur la base d'une fonction objective adaptée à la communication, de sorte que les vecteurs de formation de faisceau prédits maximisent directement le débit de communication. Ainsi, notre approche combine à la fois les ingrédients basés sur un modèle (via la fonction du débit de communication) et les ingrédients orientés données (via le réseau neuronal) et tire parti des deux mondes. Contrairement à la méthode [35], notre méthode ne requiert que l'information sur l'état du canal disponible à moins de 6 GHz pour prédire les faisceaux en ondes millimétriques, qui est beaucoup plus facile à acquérir que celui en ondes millimétriques.

En outre, nous étudions le cas plus général d'un réseau composé de plusieurs liaisons point d'accès-utilisateur. Nous proposons une méthode d'apprentissage fédéré pour prédire les vecteurs de formation de faisceau en ondes millimétriques localement, à chaque point d'accès, afin de préserver la confidentialité de leurs données et d'éviter la signalisation lourde de l'apprentissage centralisé. Nous étudions la mise à jour synchrone et asynchrone des modèles locaux. L'approche asynchrone permet de réduire le coût de communication des mises à jour pendant la phase d'apprentissage et contribue à l'économie d'énergie puisqu'un seul point d'accès entraîne son réseau neuronal à chaque itération. Ces avantages se font au détriment d'une certaine dégradation des performances en termes de débit, ce qui illustre le compromis entre *coût de formation vs. performance du débit*. Enfin, nous comparons les méthodes par renforcement clas-

siques aux méthodes par apprentissage profond en termes de performance de débit de communication et de complexité de calcul.

B.2 Liste des publications

Cette thèse a mené aux publications suivantes :

- **I. Chafaa**, R. Negrel, E.V. Belmega, and M. Debbah, "Unsupervised deep learning for mmWave beam steering exploiting sub-6 GHz channels", soumis à IEEE Trans. on Wireless Commun., Apr. 2021.

- **I. Chafaa**, E.V. Belmega, and M. Debbah, "One-bit Feedback exponential learning for beam alignment in mobile mmWave", IEEE Access, pp.194575-194589, Oct. 2020.

- **I. Chafaa**, R. Negrel, E.V. Belmega, and M. Debbah, "Federated channel-beam mapping: from sub-6GHz to mmWave", IEEE WCNC, Workshop on Distributed Machine Learning for Future Communications and Networking, Mar. 2021.

- **I. Chafaa**, E.V. Belmega, and M. Debbah, "Exploiting Channel Sparsity for Beam Alignment in mmWave Systems via Exponential Learning", IEEE ICC, Open Workshop on Machine Learning for Communications (ML4COM), Dublin, Ireland, Jun. 2020.

- **I. Chafaa**, E. V. Belmega, and M. Debbah, "Adversarial Multi-armed Bandit for mmWave Beam Alignment with One-bit Feedback", ACM ValueTools 2019, Palma de Mallorca, Spain, Mar. 2019.

B.3 Liste des posters

[P3] **I. Chafaa**, "Adversarial Multi-armed Bandits for mmWave Beam Alignment with One-Bit Feedback", IEEE Training School: Machine learning for communications, Paris, France, 2019.

[P2] **I. Chafaa**, "Online Exponential Learning for Beam-Alignment in dynamic millimeter wave Systems", Meet-up Doctorants & Industrie, Paris, France 2019.

[P1] **I. Chafaa**, "Adversarial Multi-armed Bandits for mmWave Beam Alignment with One-Bit Feedback", Journée des doctorants, ETIS, Cergy-Pontoise, France 2019.

B.4 Autres activités durant la thèse

- Co-encadrement d'un étudiant en Master pour son projet d'initiation en recherche: Bandits Manchots pour un alignement des faisceaux efficace dans les canaux à onde millimétrique.

- Participation comme rapporteur dans des conférences et revues scientifiques IEEE.

- Participation à l'équipe d'organisation de la conférence "IEEE International Symposium on Information Theory (ISIT)" à Paris, 2019.

- Assister aux séminaires organisés par l'équipe ICI (Information Communication Imagerie) au laboratoire ETIS.

- Suivi des cours et formations proposés par l'école doctorale (134 heures).

- Assister aux différent webinaires et séminaires proposés par l'école doctorale.

Dans le reste de ce document, nous allons résumer le contenu de chaque chapitre du manuscrit de la thèse. Enfin, nous présentons une conclusion générale de la thèse suivie par une liste des références bibliographiques et des cours suivis durant cette thèse.

B.5 Résumé du premier chapitre

Le premier chapitre du manuscrit représente une introduction générale pour la thèse. Dans ce chapitre, nous présentons la problématique de la thèse qui est le problème d'alignement des faisceaux dans les réseaux à onde millimétrique. Nous expliquons le contexte et les motivations derrières ce travail notamment la nécessité de concevoir des méthodes d'alignement des faisceaux compatibles avec des réseaux dynamiques et des variations temporelles du canal à onde millimétrique.

De plus, nous présentons une étude détaillée sur l'état de l'art du problème d'alignement des faisceaux dans les réseaux à onde millimétrique en présentant les différents travaux qui précèdent notre travail et en précisant, pour chacun, les limitations et les différences avec nos méthodes proposées. Notre étude couvre les différents méthodes proposées pour le problème d'alignement des faisceaux dans les réseaux à onde millimétrique en commençant par les approches classiques basées sur l'entraînement des faisceaux et l'acquisition comprimée et ensuite les nouvelles approches basées sur des outils de l'apprentissage automatique comme l'apprentissage par renforcement et l'apprentissage profond.

Nous présentons également nos différents contributions par rapport aux travaux précédents. Ensuite, nous donnons un aperçu général sur le contenu des chapitres du manuscrit. Au final, nous présentons une liste des différentes publications issues de cette thèse et des posters présentés durant différents évènements.

B.6 Résumé du deuxième chapitre

Le deuxième chapitre du manuscrit est constitué de deux parties principales. La première partie est consacrée à la présentation de la bande des ondes millimétriques. Nous expliquons les motivations derrière le passage à cette bande de très hautes fréquences et nous détaillons les trois bandes candidates principales pour les réseaux futurs en explicitant leurs occupations spectrales et leurs intérêts pour les réseaux sas fil.

Ensuite, nous discutons les différents propriétés de propagation des ondes millimétriques en se basant sur les résultats des différents compagne de mesures [7, 41, 42, 43, 44, 45, 4, 46] effectués récemment dans des milieux urbains et ruraux en environnements intérieurs et extérieurs afin de comprendre le comportement des signaux transmis dans cette bande. Nous regroupons les différentes caractéristiques de propagation en deux propriétés principales: la portée de transmission limitée et la nature parcimonieuse du canal à onde millimétrique. Ceci est du aux différents pertes subits par le signal transmis notamment le pathloss, l'absorption par des particules en atmosphères, perte par pénétration des différents objets et perte par précipitations (pluie, neige, brouillard).

Nous mettons également l'accent sur la viabilité de la bande des ondes millimétriques pour les futurs réseaux sans fil, malgré ses conditions de propagation généralement défavorables par rapport à la bande traditionnelle inférieure à 6 GHz. L'atténuation importante subit par le signal peut être gérée et n'exclut pas l'utilisation de cette bande dans les réseaux sans fil, à condition d'utiliser des réseaux d'antennes avec des techniques de formation de voies. Cependant, il est nécessaire d'utiliser de nouveaux modèles pour le canal à ondes millimétriques, car le modèle de Rayleigh, largement utilisé, ne peut être appliqué en raison du caractère parcimonieux du canal. Des travaux approfondis sur les nouveaux modèles des canaux à ondes millimétriques ont été menés par différents groupes de recherche sur la cinquième génération (5G) au cours des dernières années [11]. De plus, de nouveaux algorithmes adaptés à la matrice de canal de faible rang sont nécessaires pour différentes opérations telles que le précodage hybride et l'estimation

du canal, comme le démontrent respectivement [54] et [18]. En fait, la nature parcimonieuse du canal peut même être exploitée pour proposer des algorithmes efficaces, comme nous le proposons dans le troisième chapitre du manuscrit.

Le déploiement pratique des réseaux à ondes millimétriques fait encore face à des difficultés qui découlent essentiellement des caractéristiques spécifiques de cette large bande du spectre par rapport aux systèmes traditionnels. Nous présentons quatre défis majeurs pour les communications à ondes millimétriques: les techniques de beamforming, le problème d'alignement des faisceaux, communications en full duplex et la densification du réseau. Tous ces éléments représentent encore des pistes de recherche possible pour cete bande. Cependant, le travail de cette thèse se concentre sur un défi particulier : le problème d'alignement de faisceau. Le full duplex et la densification du réseau sont des technologies émergentes pour les systèmes à ondes millimétriques, qui présentent plusieurs défis qui ne sont pas traités dans cette thèse et qui sont laissés pour de futur travaux.

Dans la deuxième partie de ce chapitre, nous résumons les outils d'apprentissage automatique les plus pertinents pour cette thèse. Nous présentons un aperçu général des trois paradigmes principaux d'apprentissage automatique: l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement. Nous expliquons leurs avantages et leurs inconvénients ainsi que leur utilisation potentielle pour les communications sans fil et plus particulièrement pour l'alignement des faisceaux dans les réseaux à ondes millimétriques. Les détails techniques des algorithmes spécifiques, employés dans cette thèse, sont détaillés dans le reste du manuscrit. Enfin, nous comparons les méthodes classiques de l'apprentissage par renforcement avec celles de l'apprentissage profond.

B.7 Résumé du troisième chapitre

Ce chapitre représente la première partie des contributions de la thèse. Il est dédié à l'exploitation des bandits manchots pour le problème d'alignement de faisceaux en ondes millimétriques. Nous abordons le problème de l'alignement des faisceaux dans les réseaux dynamiques à ondes millimétriques. Nous exploitons le cadre du bandit manchot pour concevoir des politiques distribuées, dans lesquelles l'émetteur et le récepteur choisissent leurs faisceaux séparément en se basant uniquement sur un retour d'information d'un bit.

D'abord, nous présentons le modèle du système considéré dans cette étude en détaillant le dictionnaire de beamforming, le modèle du signal reçu, le modèle de mobilité utilisé et le modèle géométrique du canal à onde millimétrique. Nous formulons également le problème de l'alignement de faisceau comme un problème de bandits manchots adversatif qui permet d'avoir moins d'hypothèse sur la dynamique de l'environnement.

Ensuite, nous étudions la taille optimale du dictionnaire de beamforming qui représente le nombre optimal des vecteurs de beamforming à utiliser. Nous montrons, via une étude empirique, qu'il n'est pas nécessaire d'augmenter continuellement la taille du

dictionnaire de beamforming pour obtenir des niveaux de rapport signal/bruit sensiblement plus élevés au niveau du récepteur. Par conséquent, nous pouvons limiter le nombre de vecteurs de formation de faisceaux pour réduire la durée de l'alignement des faisceaux.

Nous proposons trois politiques d'alignement de faisceaux exploitant les bandits manchots. La première politique est basée sur l'algorithme original *poids exponentiels pour l'exploration et l'exploitation* (EXP3) de [39]. Nous proposons ensuite deux nouvelles politiques (MEXP3 et NBT-MEXP3) en modifiant les récompenses des actions (ou faisceaux) choisies. Nos nouvelles politiques s'inspirent de la nature parcimonieuse du canal à onde millimétrique et de la corrélation entre les faisceaux choisis successifs, ce qui permet d'obtenir de meilleures performances. Nous discutons également de la propriété de non-retour pour les politiques proposées. Nous prouvons rigoureusement que notre politique en ligne MEXP3 possède la propriété de non-regret, tandis qu'une conjecture est fournie pour NBT-MEXP3 (validée par des simulations).

La démonstration est fournie en détail en annexe à la fin du manuscrit. Les performances des algorithmes proposés sont démontrées par des résultats numériques en termes de regret et de débit dans un contexte pratique d'ondes millimétriques. Nous montrons que nos politiques surpassent la politique originale BA-EXP3 et d'autres politiques centralisées existantes en étant capables de s'adapter aux changements rapides et imprévisibles du canal à onde millimétrique. Enfin, nous discutons l'extension des méthodes proposées pour des systèmes à plusieurs utilisateurs et aux larges bande spectrale. [3mm]

B.8 Résumé du quatrième chapitre

La deuxième partie des contributions de la thèse est présentée dans ce chapitre. Dans le chapitre 3, nous avons présenté des politiques distribuées basées sur les bandits manchots, pour aligner les faisceaux de l'émetteur et du récepteur avec seulement un bit de feedback . Ici, nous exploitons le cadre de l'apprentissage profond pour résoudre le problème de l'alignement des faisceaux dans les réseaux à ondes millimétriques en exploitant l'information du canal disponible dans la bande inférieure à 6 GHz.

D'abord, nous employons un réseau neuronal pour projeter les canaux sous 6 GHz en vecteurs de beamforming dans la bande des ondes millimétriques en exploitant l'apprentissage non supervisé. Les capacités d'approximation universelles d'un réseau neuronal profond permettent d'apprendre la fonction complexe qui relie les canaux de la bande sous 6 GHz aux vecteurs de beamforming dans la bande millimétrique. L'apprentissage non supervisé permet de faciliter l'obtention des données d'entraînement. Nous expliquons en détails les différentes composantes de la solution proposée: la construction du dataset, l'architecture du réseau neuronal proposé et la fonction objective utilisée pour l'entraînement du réseau neuronal. Nous évaluons la méthode proposée par plusieurs simulations numérique en la comparant avec autre méthodes existantes. Nous discutons aussi le choix de la taille du réseau neuronal en fonction des paramètres du système notamment le nombre d'antennes au point d'accès.

Ensuite, nous proposons d'utiliser l'apprentissage fédéré pour prédire les faisceaux d'un réseau à liaisons multiples afin d'éviter l'échange de données locales avec un serveur central et de surmonter la pénurie de données d'entraînement. Nous discutons les méthodes synchrone et asynchrone de l'apprentissage fédéré. Des résultats numériques sont présentés pour illustrer la performance des méthodes proposées. Nos simulations numériques montrent le fort potentiel de notre approche, basée sur l'apprentissage fédéré en particulier lorsque les données locales disponibles sont rares ou imparfaites (bruités).

Enfin, nous effectuons une comparaison entre les méthodes proposées dans ce chapitre et le précédent, notamment en termes de taux de communication et de complexité de calcul. Nous comparons les méthodes d'apprentissage profond et les algorithmes d'apprentissage par renforcement classiques. Notre analyse et nos simulations montrent que le choix d'une méthode d'alignement de faisceau appropriée dépend de la nature de l'application, ses propriétés techniques et représente un compromis entre le débit de communication et la complexité de calcul.

B.9 Conclusion générale

L'objectif principal de cette thèse est d'exploiter les outils d'apprentissage automatique pour un problème fondamental dans les futurs réseaux à ondes millimétriques : l'alignement des faisceaux. Les communications à ondes millimétriques offrent une solution prometteuse à l'engorgement du spectre en tirant parti des larges bandes de fréquences disponibles dans la gamme des $20-300$ GHz. Ces communications souffrent d'une forte atténuation, d'une faible diffusion et d'une perte de pénétration élevée. Les faisceaux hautement directionnels, rendus possibles par de grands réseaux d'antennes et des techniques de formation de faisceaux, constituent une solution prometteuse à ces difficultés. Comme la longueur d'onde est très courte, l'intégration de plusieurs antennes sur un dispositif de petite taille devient possible. Ainsi, les communications en ondes millimétriques peuvent bénéficier du nombre massif d'antennes au niveau du point d'accès et même au niveau des équipements des utilisateurs.

Avant de pouvoir transmettre des données utiles dans la bande des ondes millimétriques, l'alignement des faisceaux est une étape cruciale pour garantir un lien de communication fiable. Nous avons vu que les approches traditionnelles ne sont pas capables de faire face à la mobilité des utilisateurs et aux variations temporelles du canal. Alors, nous proposons des politiques d'allocation de faisceaux dynamiques et adaptatives en exploitant des techniques d'optimisation en ligne et d'apprentissage automatique. Ces outils vont au-delà de l'optimisation classique et permettent de développer des algorithmes efficaces malgré la dynamique du réseau qui peut être non stationnaire et imprévisible en raison de la mobilité des utilisateurs et des modèles de connectivité.

Nous proposons des politiques d'alignement de faisceaux en ligne qui s'appuient sur un feedback limité en un bit pour diriger les faisceaux des deux noeuds du lien de communication de manière distribuée. Nous formulons le problème de l'alignement des faisceaux par le biais du cadre des bandits manchots adversatif, qui permet de faire face à une dynamique de réseau arbitraire comprenant des composants non stationnaires ou

adversaires. Ce cadre plus général (comparé aux bandits stochastiques) nous permet de découpler le problème d'alignement des faisceaux et de répartir l'apprentissage entre l'émetteur et le récepteur. Les deux nœuds utilisent un algorithme d'apprentissage en ligne pour choisir leur propre directions de faisceau, de manière distribuée, sans connaître les choix des autres à l'avance. L'apprentissage s'effectue donc à la fois au niveau de l'émetteur et du récepteur sans dépendre d'un nœud central et en utilisant uniquement un feedback d'un bit. Les avantages immédiats sont que chaque nœud n'explore que son propre ensemble de directions de faisceau et non l'ensemble des paires de faisceaux qui est beaucoup plus grand, en plus de réduire la quantité de données échangées pendant le processus d'alignement du faisceau.

D'abord, nous étudions la taille optimale du dictionnaire de beamforming à utiliser pour l'alignement du faisceau afin de réduire le coût d'exploration. Nous fournissons la taille minimale qui garantit une certaine qualité de service. Dans le cas des canaux à ondes millimétriques à trajet unique, nous obtenons une formule explicite, tandis que pour les canaux à trajets multiples, nous utilisons des expériences numériques pour calculer cette mesure de performance. Nous montrons que l'augmentation de la résolution du dictionnaire de beamforming au-delà d'un certain point n'offre pas une meilleure performance mais implique un coût d'exploration croissant. Par conséquent, la restriction du nombre de vecteurs de formation de faisceau offre un bon compromis entre la performance et le coût d'exploration.

Ensuite, nous exploitons l'algorithme EXP3 bien connu pour une recherche distribuée des meilleurs faisceaux dans un système MIMO en ondes millimétriques. Les performances de la méthode proposée sont évaluées en termes de notion de regret, de probabilité de coupure et comparées aux algorithmes existants. Nos résultats de simulation montrent une diminution du regret moyen et de la probabilité de coupure au fur et à mesure de l'apprentissage, ce qui implique une plus grande précision dans l'alignement du faisceau.

En se basant sur l'algorithme EXP3 et en exploitant la structure parcimonieuse du canal à ondes millimétriques, nous proposons une politique modifiée (MEXP3). Notre MEXP3 utilise une récompense modifiée qui renforce l'exploitation des bonnes directions de faisceau et pénalise les mauvaises. Nous prouvons ensuite que la nouvelle MEXP3 possède la propriété de non-regret et que le regret moyen décroît jusqu'à zéro de manière optimale en $\mathcal{O}(1/\sqrt{\mathcal{T}})$ de manière similaire à EXP3 où $\mathcal{T}$ désigne l'horizon temporel. De plus, pour les horizons fixes et finis, notre borne supérieure de regret pour MEXP3 est plus serrée (facteur constant multiplicatif plus petit) que celle de l'algorithme EXP3, ce qui suggère une meilleure performance dans des contextes pratiques. Nous introduisons ensuite une modification supplémentaire qui tient compte de la corrélation temporelle entre les faisceaux successifs et proposons une autre politique d'alignement des faisceaux, appelée NBT-MEXP3. La propriété de non-regret est conjecturée pour NBT-MEXP3 et validée par des simulations numériques étendues.

Bien que la performance asymptotique du regret des algorithmes proposés, $\mathcal{O}(1/\sqrt{\mathcal{T}})$, soit optimale et ne puisse pas être améliorée dans le cadre d'une dynamique de réseau arbitraire, nos deux nouvelles politiques MEXP3 et NBT-MEXP3 offrent des améliorations de performance significatives dans des paramètres pratiques d'ondes millimétriques. Les simulations numériques montrent que les politiques proposées offrent de meilleures

performances pratiques, notamment en termes de probabilité de coupure et de débit pour les canaux mono-trajets et à multiples trajets. Nos récompenses modifiées conduisent à des algorithmes d'apprentissage en ligne qui s'adaptent mieux et plus rapidement aux variations du canal à ondes millimétriques, ce qui permet d'augmenter les débits par rapport aux autres politiques existantes.

Nous considérons d'autres outils d'apprentissage automatique avancés allant au-delà du cadre des bandits manchots. Tout d'abord, nous proposons une méthode d'apprentissage profond non supervisée pour apprendre une fonction canal-faisceau, qui exploite la connaissance du canal à moins de 6 GHz pour prédire les vecteurs de formation de faisceau en ondes millimétriques à l'aide d'un réseau neuronal entraîné. Nous illustrons, à l'aide de nombreux résultats numériques, les performances de la fonction proposée en termes de débit de communication et la comparons aux méthodes existantes. Nous montrons également que la taille optimale de notre réseau neuronal diminue lorsque la taille de l'entrée du réseau (nombre d'antennes de réception de la liaison montante) augmente, car les informations sur les canaux de liaison montante en dessous de 6 GHz deviennent plus riches. Au contraire, lorsque la sortie du réseau neuronal augmente (nombre d'antennes de transmission en liaison descendante), la taille optimale du réseau neuronal doit augmenter, car la relation entrée-sortie devient plus complexe.

En outre, nous employons la fonction canal-faisceau proposée dans un cadre d'apprentissage fédéré pour prédire les faisceaux de plusieurs liaisons point d'accès-utilisateur de manière distribuée. L'apprentissage fédéré consiste à ce que les points d'accès partagent les paramètres de leurs réseaux neuronaux entraînés localement afin de les agréger dans un modèle global plus informé pour tous les points d'accès . Les principaux avantages sont de trois ordres : i) il évite à chaque point d'accès de partager ses données locales, ce qui peut poser des problèmes de confidentialité, tout en mettant en commun les connaissances acquises par les autres points d'accès ; ii) il réduit le coût de signalisation ; et iii) il distribue la charge de calcul, par rapport à une approche entièrement centralisée (les points d'accès envoient leurs données locales au serveur central chargé d'entraîner le réseau neuronal). Nous étudions les approches de partage synchrone et asynchrone. L'approche asynchrone réduit le coût de la signalisation (et la consommation d'énergie) pendant l'entraînement au détriment d'une dégradation relativement faible du débit.

Nous évaluons les schémas d'apprentissage fédéré que nous proposons et les comparons à des méthodes entièrement centralisés et entièrement distribués (aucune coopération entre les points d'accès, chaque réseau neuronal est entraîné indépendamment en utilisant des données locales). Dans le cas où les données d'entraînement disponibles sont rares, les gains de débit de nos approches synchrone et asynchrone peuvent atteindre respectivement 50% et 41% par rapport à l'approche entièrement distribuée. Le gain de débit relatif atteint 14% pour la méthode synchrone par rapport à l'apprentissage centralisé, lorsque la qualité des données d'apprentissage est médiocre (lorsque les estimations des canaux sub-6 GHz de la liaison montante sont bruitées). Enfin, nous évaluons nos méthodes d'apprentissage fédéré dans le cas d'utilisateurs mobiles et en tenant compte de l'interférence inter-lien en liaison descendante. Nos résultats démontrent l'efficacité de nos approches dans le cas où les données d'apprentissage sont rares, ce qui les rend pratiques dans ce contexte.

Enfin, nous comparons les méthodes d'apprentissage profond avec nos méthodes basées sur les bandits manchots en termes de débit et de complexité de calcul. Par rapport aux algorithmes bandits manchots, les méthodes d'apprentissage profond fournissent des taux de communication plus élevés au détriment d'un entraînement hors ligne et d'une complexité de calcul en ligne (en cours d'exécution) plus élevée. Par conséquent, les exigences techniques et les caractéristiques des applications, telles que la latence, la puissance de calcul, la consommation d'énergie et la disponibilité des données pour l'entraînement, peuvent être prises en compte pour favoriser une approche plus que de l'autre.