

Performance Evaluation of Cross-Environment Learning Strategies for Beam Prediction in Millimeter-Wave Networks

Jan B. Schaper

Master's Thesis

September 18, 2025

Examiners

Prof. Dr.-Ing. Marina Petrova
Dr. Ljiljana Simić

Supervisors

Prof. Dr.-Ing. Marina Petrova
Enrico Tosi, M.Sc.

Mobile Communications and Computing
RWTH Aachen University



communicated by Prof. Dr.-Ing. Marina Petrova

The present work was submitted to Mobile Communications and Computing

Performance Evaluation of Cross-Environment Learning Strategies for Beam Prediction in
Millimeter-Wave Networks

Master's Thesis

presented by
Jan B. Schaper

Prof. Dr.-Ing. Marina Petrova
Dr. Ljiljana Simić

Aachen, September 18, 2025

(Jan B. Schaper)

Acknowledgements

I would like to thank my supervisor Enrico Tosi for his continuous guidance, his expertise, availability and exceptionally valuable feedback during the whole process of this work. I am grateful to Prof. Dr.-Ing. Marina Petrova and Dr. Ljiljana Simić for the opportunity to work on this thesis and their valuable feedback at crucial points during this journey. A special thanks to Dr. Ljiljana Simić for her repeated help during my academic path. I would like to thank my family for their continued unconditional support, reaching far beyond the context of this work.

Contents

Acknowledgements	II
Contents	III
Abstract	V
1 Introduction	1
2 Background	3
2.1 Millimeter Wave Communications	3
2.1.1 Propagation Properties	4
2.1.2 Directional Communication	7
2.2 Beam Management in 5G NR	9
2.3 Beam Management Leveraging Context Information	11
2.4 Beam Prediction Using Machine Learning	13
2.5 Beam Prediction Using User Equipment (UE) Position	14
2.6 Adaptability of Prediction Models Across Environments	15
2.7 Related Work	17
3 System Model	18
3.1 Network	18
3.2 Channel Model	20
3.3 Prediction Models and Performance Metrics	23
3.4 Dataset	25
3.5 UE Position Error	26
4 Cross Environment Beam Prediction	27
4.1 Preliminary Investigation on Network Parameters	27
4.1.1 Fully Connected Neural Network	27
4.1.2 Convolutional Neural Network	29
4.2 Transfer Learning	29
4.2.1 Dataset Size	30
4.2.2 Neural Network Size	31

Contents	IV
4.2.3 Noise in the UE Position	33
4.2.4 Number of Training Labels	33
4.3 Cross-Environment Generalizability	34
4.3.1 Accuracy at the Reference Next Generation Node Base Station (gNB) after Fine-Tuning on Different Target gNBs	34
4.3.2 Elastic Weight Consolidation	34
4.3.3 Dataset Mixing	35
5 Results	37
5.1 Transfer Learning	37
5.1.1 Preliminary Results	37
5.1.2 Dataset Size	39
5.1.3 Neural Network Size	47
5.1.4 Noise in the UE Position	50
5.1.5 Number of Training Labels	56
5.2 Cross-Environment Generalizability	58
5.2.1 Accuracy at the Reference gNB after Fine-Tuning on Different Tar- get gNBs	58
5.2.2 Elastic Weight Consolidation	64
5.2.3 Dataset Mixing	65
6 Conclusion and Future Work	72
A Abbreviations	74
List of Tables	77
List of Figures	78
Bibliography	82
Declaration	86

Abstract

Increasing demands on mobile communication networks lead to the exploration of the millimeter-Wave (mmWave) spectrum, providing larger bandwidths and enabling ultra-high data rates. However, the adverse propagation conditions of mmWave necessitate beamforming and thus precise beam alignment between Next Generation Node Base Station (gNB) and User Equipment (UE). Mobile UEs require constant updates of the beam alignment, leading to significant beam management overhead from beam sweeping. As a potential solution, UE position based machine learning (ML) beam prediction algorithms have shown promising results in selecting the serving beam without extensive beam sweeping. However, due to site-specific propagation environments, a model trained on one gNB can not be utilized on another gNB. Transfer learning has been shown to achieve high beam prediction accuracies by fine-tuning a neural network (NN) model trained on a reference gNB with a limited dataset from a target gNB.

This thesis provides an extensive evaluation of the impact of the factors dataset size, NN size, noise in the UE position and number of training labels on the performance of transfer learning. The cross-environment generalizability is further explored by analysing the accuracy of the NN models on the test data of the reference and target gNBs simultaneously. The potential of elastic weight consolidation (EWC) to balance the performance on reference and target gNB is investigated. Dataset mixing is proposed as an approach to achieve high beam prediction accuracies on multiple gNBs simultaneously. The investigations are performed based on ray-tracing in a 28 GHz downlink cellular Multiple-Input Multiple-Output (MIMO) network in an urban environment.

Keywords: Beam management, transfer learning, beam prediction, millimeter-wave

Introduction

The global mobile network data traffic is projected to grow by a factor of 2.3 until 2030, driven by an increase in the number of subscriptions as well as the average data volume per subscription [1]. To meet increasing demands on data rates, bandwidths and low latency capabilities, the millimeter-Wave (mmWave) spectrum offers previously unexploited potential [2]. However, mmWave communication signals experience a number of adverse propagation conditions, necessitating directional beamforming to compensate for the propagation loss. In mobile networks with environment dependent propagation paths, the required precise beam alignment between the Next Generation Node Base Station (gNB) and User Equipment (UE) has to be constantly updated, inducing a high beam management overhead [3]. One approach to reduce this overhead is employing Machine learning (ML) techniques to predict the optimal Beam Pair Link (BPL) based on the UE position [4]. However, these UE position aided beam prediction algorithms show poor performance when tested on gNBs other than their training gNB [5]. Since the training of new neural network (NN) models for every gNB is highly inefficient, it is desirable to achieve cross-environment generalizability. To reduce the amount of required training data compared to training a new NN model from scratch, the authors in [5] proposed the application of transfer learning to position based beam prediction. Fine-tuning a NN model trained on a large dataset of a reference gNB, by retraining with a small training dataset of another target gNB, is shown to reach up to 80 % accuracy in predicting the best BPL for the target gNB. While these results are promising, they do not offer an in-depth analysis of factors contributing to high prediction accuracies of transfer learning. Furthermore, the transfer learning approach still requires separate training for each target gNB, and one NN model is expected to perform only for one gNB simultaneously.

This thesis offers a performance evaluation of cross-environment learning strategies in a realistic a 28 GHz downlink cellular Multiple-Input Multiple-Output (MIMO) network. Using data obtained by the commercial ray-tracing simulator Wireless InSite [57] for 27 gNBs in a study area in Frankfurt, Germany, the performance of position based beam prediction NN models, considering fully connected neural networks (FCNNs) and convolutional neural networks (CNNs), are assessed based on the achieved accuracies of predicting the optimal BPL, as well as subsequent top BPLs. This work presents an in-depth investigation into the impact of the factors dataset size, NN size, noise in the UE position data, and the number of training labels on the performance of transfer learning in position based beam prediction. Furthermore, the cross-environment generalizability of one NN model across multiple gNB simultaneously is explored for the transfer learning approach. The potential of elastic weight consolidation (EWC) in balancing the beam prediction accuracies on the reference and target gNBs is investigated. Dataset mixing is proposed as an approach to train a NN model that reaches high beam prediction accuracies on multiple gNBs simultaneously.

The remainder of this thesis is organized as follows: Chapter 2 addresses the theoretical background of mmWave communications, thus motivating the necessity for beam management and beam prediction, and provides a baseline for position aided beam prediction with NNs. Chapter 3 presents the system model employed in the beam prediction simulations. Chapter 4 provides the investigated position based beam prediction NN models, learning strategies, and considered factors impacting the performance. In Chapter 5, the results are presented and discussed. Chapter 6 concludes this work by summarizing the results and outlining directions for future work.

2

Background

This chapter introduces the state-of-the-art approaches towards beam prediction and the corresponding fundamentals regarding mmWave and ML. In Section 2.1, the concept of mmWave for future communication networks is motivated. Emerging challenges are highlighted by expanding upon the mmWave inherent propagation properties, followed by an introduction of beamforming as a key technology to overcome these challenges. In Section 2.2, the procedures followed by the 5G New Radio (NR) standard for mmWave beam management is presented, with a special focus on beam sweeping approaches. Section 2.3 offers general remarks on beam management leveraging Context Information (CI). Section 2.4 introduces ML techniques for beam prediction and their potential to reduce the beam training overhead, with a special focus on supervised learning (SL) using NN. The concept of using beam index maps (BIMs) in combination with UE position aided beam prediction is explained in Section 2.5. In Section 2.6, the adaptability of beam prediction models across environments is motivated as a key challenge for a widespread deployment in mmWave communications. Section 2.7 gives the current status regarding works on cross-environment adaptability of SL based position aided beam prediction and thereby motivates the contribution of this thesis.

2.1 Millimeter Wave Communications

Modern radio networks have to meet steadily increasing performance requirements. Additionally to enhanced Mobile Broadband (eMBB) focused on matching the increasing demand of mobile users, several Internet of Everything (IoE) systems such as Smart Grids and autonomous driving pose new demands in the form of Ultra-Reliable, Low Latency Communications (URLLC) [6]. Technologies that have been deployed in 5G to meet these challenges include ultra-dense networking, all-spectrum access and full-duplexing. A further key contributor is identified as the utilization of mmWave frequencies, which offer an increase in the usable bandwidth at the cost of adverse propagation properties in the mmWave range. Contrary to the sub-6 GHz spectrum, which is congested by frequency bands allocated, among others, to previous generations of mobile networks, satellite and aerial communications, Bluetooth and Wireless Local Area Network (WLAN), mmWave frequencies offer larger continuous frequency blocks [7]. The 5G NR standard defines three separate bands in the mmWave range; FR 1 (410 MHz - 7.125 GHz), FR 2-1 (24.25 GHz - 52.6 GHz) and FR 2-2 (52.6 GHz - 71 GHz) with maximum transmission bandwidths of up to 100 MHz, 400 MHz and 2000 MHz, respectively [8]. To compensate for the adverse propagation properties of mmWave, MIMO technologies are employed, referring to multiple antenna elements used at the input and output of the communications channel, i.e. at the transmitter (TX) and RX, respectively. Antenna arrays consisting of multiple antenna elements can produce narrow, directed beams that offer longer transmission ranges along

their direction [2, 9]. This section gives a detailed coverage of the adverse propagation properties of mmWave transmissions in Section 2.1.1, followed by introducing beamforming as the enabling technique for directional communications 2.1.2, which is employed to compensate for the mmWave propagation properties

2.1.1 Propagation Properties

The adverse propagation conditions of mmWave compared to sub-6 GHz communications are multilayered and can be divided into those stemming from path attenuation properties that apply to the Line-of-Sight (LOS), i.e. the Free Space Path Loss (FSPL) and atmospheric attenuation, and those from interactions with objects along the path, only pertaining to the non-Line-of-Sight (NLOS), i.e. reflection, diffraction, scattering and penetration.

2.1.1.1 Free Space Path Loss

The received power P_R dependence on the transmitted power P_T , provided an unobstructed space between TX and receiver (RX), is given by the Friis free space equation [10],

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2, \quad (2.1)$$

where G_T is the transmit antenna gain, G_R is the receive antenna gain, λ is the carrier wavelength and d is the distance between TX and RX. From this, the FSPL given isotropic unity gain antennas can be derived as

$$FSPL = \frac{P_T}{P_R} = \left(\frac{4\pi d}{\lambda} \right)^2 = \left(\frac{4\pi df}{c} \right)^2, \quad (2.2)$$

where c is the speed of light and f is the carrier frequency. The dependency on d^2 arises from the energy conservation based on an isotropic source in the far field $d \gg \lambda$. The wavelength dependency arises from the effective aperture A of an isotropic antenna ($G = 1$)

$$A^E = G \frac{\lambda^2}{4\pi}, \quad (2.3)$$

which is proportional to the square of λ . Expressed in decibel and with antenna gains G_T and G_R compared to the isotropic antenna, the FSPL can be written as

$$FSPL[dB] = -G_T[dB] - G_R[dB] + 20 \log_{10}(d) + 20 \log_{10}(f) + 32.44, \quad (2.4)$$

with d in km and f in GHz. Therefore, for isotropic antennas, increasing the carrier frequency from 2.6 GHz used in LTE, to 30 GHz in the FR2 band, raises the FSPL by the non-negligible quantity of 21.2 dB.

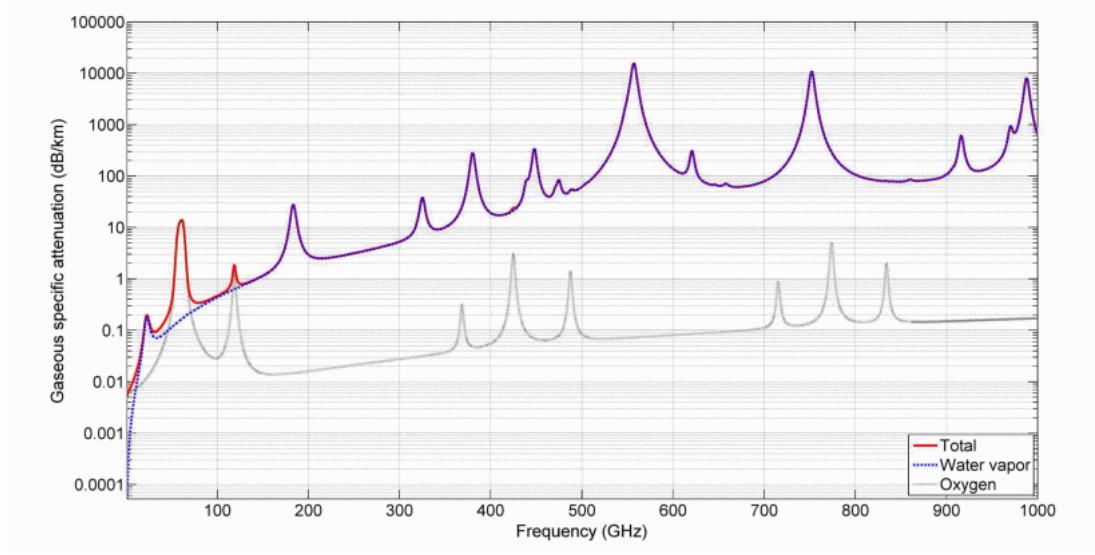


Figure 2.1: Attenuation due to atmospheric gases (in dB/km) as a function of radio frequency for an atmospheric pressure $P = 1013.25$ hPA, temperature $T = 15^\circ\text{C}$ and a water vapour density $\rho = 7.5$ g/m³ [11].

2.1.1.2 Atmospheric Attenuation

Another contributor to LOS losses is the atmospheric gas attenuation. Depending on their molecular structure, gases have a set of spectral absorption lines at different central frequencies and with varying absorption levels. The absorption lines show a general trend to attenuate more at higher frequencies, thus adversely affecting long-range communications [11]. A widely used model for the specific atmospheric attenuation γ_g given frequencies between 1 GHz and 1000 GHz, recommended by the International Telecommunication Union Radiocommunication Sector (ITU-R), known as the line-by-line calculation, describes the attenuation as the combined effect of water vapour γ_w and dry air γ_o where oxygen is the dominant contributor with $\gamma_g = \gamma_w + \gamma_o$ [12]. Frequency-independent factors are temperature T , and atmospheric pressure P . Following this, Figure 2.1 depicts the specific attenuation (in dB/km) for $P = 1013.25$ hPA, $T = 15^\circ\text{C}$ and a water vapour density $\rho = 7.5$ g/m³. As shown in Figure 2.1 for frequencies above 10 GHz, water vapour becomes the main contributor to atmospheric absorption, with exceptions around 60 GHz and 118 GHz where oxygen is predominant.

Regarding the FR2 band defined by 5G-NR, the FR2-1 band lies between the first two peaks at 23 GHz and 60 GHz, while the FR2-2 band encompasses the oxygen induced peak at 60 GHz. With an attenuation around 0.1 dB/km the atmospheric absorption in the FR2-1 band becomes negligible [13]. Frequencies lying around attenuation peaks, such as the FR2-2 band, can be beneficial in densely deployed networks to mitigate interference [14].

In addition to the atmospheric gas attenuation, the attenuation from rainfall is significantly higher in mmWave. A model for various levels of rainfall is introduced in [15]. A

precipitation rate of 25 mm/h causes an attenuation of 0.0068 dB/km at $f = 3$ GHz, negligible compared to the attenuation of 4.87 dB/km at $f = 30$ GHz, which further increases to 26.8 dB/km at a precipitation rate of 150 mm/h.

2.1.1.3 Propagation Mechanisms

The three propagation properties when electromagnetic waves interact with the environment are reflection, diffraction and scattering [10]. The interaction between objects and electromagnetic waves is dependent on the ratio of the object dimension to the wavelength λ , as the radius of the first Fresnel zone r_1 , given for distances $d_1, d_2 \gg r_1$, between the object and the TX and RX, respectively, as

$$r_1 = \sqrt{\frac{\lambda d_1 d_2}{d_1 + d_2}}, \quad (2.5)$$

becomes narrower proportional to the wavelength [16]. At distances $d_1 = d_2 = 50$ m and $\lambda = 10$ mm, corresponding to a frequency $f = 30$ GHz, the radius becomes $r_1 = 0.5$ m. As objects become larger compared to the wavelengths, diffraction lessens, resulting in higher shadowing effects [16]. Therefore, diffraction effects introduced by buildings and objects become negligible for mmWave given the shorter wavelengths.

Scattering describes the diffuse reflection on large objects with rough surfaces or on large numbers of small objects. The roughness of a surface can be tested by comparing the maximum protuberance to the critical height h_c , using the Rayleigh criterion as given in [10]:

$$h_c = \frac{\lambda}{8 \sin \rho_i}, \quad (2.6)$$

where ρ_i is the angle of incidence of the electromagnetic wave.

Given a carrier frequency of 30 GHz and an angle of incident $\rho_i = 45^\circ$, this results in a critical height of 10 mm. However, due to the high path loss in mmWave, only a few scattered components arrive at the RX, and the specific scattering environment has a large impact on the received power in NLOS [17]. Measurements at 28 [GHz] have shown a NLOS Path Loss exponent (PLE) of up to 5.76 in downtown New York City, while a NLOS PLE of 3.86 was measured at 38 GHz in Austin [18]. Measurements for 38 GHz indicate that propagation in outdoor urban scenarios is dominated by LOS and reflection [19].

2.1.1.4 Penetration

Common outdoor building materials have shown high reflection coefficients up to 0.896 and very small transmittance with a penetration loss of 40.1 dB for tinted glass at 28 GHz [20]. Furthermore, foliage losses increase with frequency and reach significant values even with single trees. For foliage depths $D_{fol} < 400$ m and frequencies $f = 200$ MHz–95000 MHz, [21] gives a formula first developed in (CCIR Rpt 236-2) for the foliage penetration loss

$$L_{fol}[db] = \frac{(f)^{0.3}(D)^{0.6}}{5}. \quad (2.7)$$

For $D_{fol} = 5$ m, this results in a foliage loss $L_{fol} = 11.6$ dB for $f = 30$ GHz compared to $L_{fol} = 5.6$ dB for $f = 2.6$ GHz.

The presence of humans has been shown to have a considerable impact on the propagation of mmWave signals. The dielectric properties of the skin allow for minimal penetration, however, diffraction can lessen the loss in parts [22]. The hand in particular can cause severe attenuation. Measurements at $f = 28$ GHz resulted in up to 18 dB attenuation for a hard hand grip, while body blockage resulted in up to 11 dB, with both effects being highly dependent on the antenna configuration [22].

2.1.2 Directional Communication

As discussed in the previous section, mmWave faces adverse propagation conditions. However, the FSPL according to equation 2.2 assumes an isotropic antenna without directivity. For a fixed effective antenna aperture A , the antenna directivity is inversely proportional to the square of the wavelength, as can be seen by reordering equation 2.3 towards the antenna gain [21]:

$$G = \frac{A^E 4\pi}{\lambda^2}. \quad (2.8)$$

Combining equations 2.1 and 2.8, the received power given the TX and RX antenna apertures A_t and A_r , respectively, the received power becomes

$$P_R = P_T \frac{A_t^E A_r^E}{\lambda^2 d^2}. \quad (2.9)$$

Therefore, given directional antennas, the additional FSPL of 21.2 dB for $f = 30$ GHz compared to $f = 2.6$ GHz from section 2.4 becomes a gain, if fixed antenna apertures A_t and A_r are assumed instead of a fixed gain $G = 1$. In practice, this can be achieved by employing antenna arrays to perform directional multi-antenna beamforming, with further proposed beam steering techniques including quasi-optical systems, leaky wave antennas, and metasurfaces [23]. Apart from a higher Signal-to-Noise Ratio (SNR), beamformers were shown to improve the Rician factor gain and the root-mean-squared (RMS) delay spread [24]. Because of a smaller size and necessary separation due to the shorter wavelengths in mmWave [25], a large number of antennas can be packed on small area arrays, which enables beamsteering with narrow beams, as well as implementation in small, portable UE devices like smartphones [21]. The total output power distribution of a phased antenna array is the vector sum of fields radiated by the individual elements and is commonly expressed by combining the individual radiation patterns with the array factor, which is composed of the array geometry and amplitude/phase shifts at individual elements [17]. These amplitude/phase shifts of individual antennas are commonly referred to as weights. Figure 2.2 shows a representation of directional control achieved in a small linear array, consisting of six elements simulating the resulting field from a phase shift applied to open-

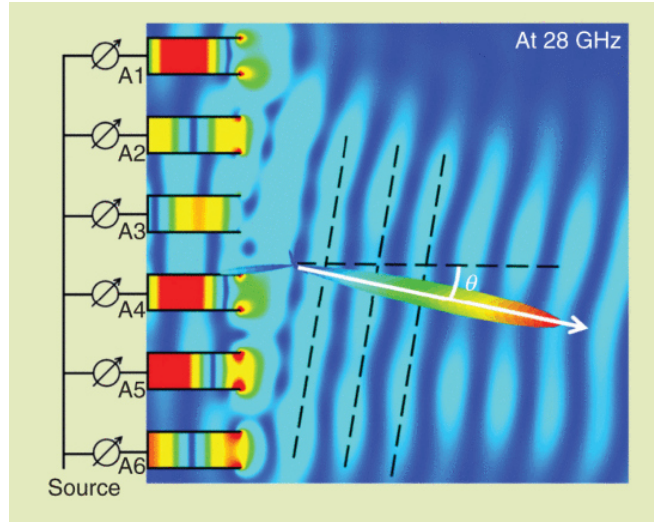


Figure 2.2: Simulated E-field plot for a phased-array antenna at 28 GHz demonstrating beam tilt [23].

ended waveguides at 28 GHz [23]. Planar arrays, used to enable 3D beamforming in the azimuth and elevation plane, follow the same principle.

There exists a number of beamforming architectures, with one differentiating characteristic being whether the weights are applied in the analog or digital domain [17]. The application in the analog domain has the advantage of low complexity and energy consumption, as only one radio frequency (RF) chain is required for the entire array. However, to support multiple users at a given time slot, additional signal processing in the form of digital precoding and combining is required. Digital beamforming uses a dedicated RF chain for each antenna element and thus provides more degrees of freedom. As a disadvantage, the increasing number of antenna elements as well as the wideband nature of mmWave results in higher complexity and power consumption. To address this issue, hybrid approaches combine digital baseband processing with a limited number of RF chains to enable multi-stream transmission, and analog beamforming to achieve antenna gain [26].

In sub-6 GHz systems, techniques to adapt the weights in real time are popular, however signal processing overhead and latency are prohibiting factors to employ similar strategies in mmWave [17]. The weights for the used beam configurations in fixed-weight beamforming can be collected into codebooks. While sub-6 GHz channels are often influenced by a rich multipath environment, the mmWave channel is dominated by a small number of dominant paths, as discussed in section 2.1.1.3. As a consequence, mmWave codebooks are often structured based on the steering angle, forming narrow beams distinguished by their direction [17].

Similarly to the TX, the weights can be applied at the RX to tune the direction of sensitivity of the antenna array. Optimal alignment of the TX and RX beams allows one to reap the full benefits of beamforming. Measurements confirmed that using a NLOS path with reflections as the propagation mechanism and optimally aligned beams can offer similar PLE in a 28 GHz setup as experienced by sub-6 GHz systems [18]. However, due to

the varying locations of dominant scattering clusters and reflection paths, achieving optimal alignment is not trivial and requires effective beam management strategies.

2.2 Beam Management in 5G NR

In addition to an evolution of 4G-LTE networks, the 5G standard includes a range of new technologies, commonly referred to as 5G NR, which also include the usage of the mmWave spectrum [27], resulting in the necessity for beam alignment between mobile UEs and gNBs. Therefore, the beam management necessary to perform Initial Access (IA) and beam tracking, describing the process of following a mobile connected UE with beams, becomes more complex compared to 4G-LTE, which could rely on omnidirectional signalling. 5G NR uses codebooks at the gNB and UE to organize the antenna-weights based on their resulting beam steering direction. The goal of the beam management is to find a BPL, i.e. a combination of beams at gNB and UE that offers optimal performance based on metrics such as the SNR. As shown in Figure 2.3 the beam management in 5G NR consists of four operations [28, 29]:

- Beam sweeping: covering a spatial area with a set of transmitted and/or received beams in a predetermined procedure during a time interval. The beam selection and the beam refinement at the transmit-receive points (TRPs) of the gNB or at the UE are referred to as separate procedures.
- Beam measurement: assessment of the quality and characteristics of the received beamformed signal. The measurements are typically based on the reference signal received power (RSRP) or the Signal-to-Interference-plus-Noise Ratio (SINR) of the Synchronization Signal (SS), which is sent in bursts with a default periodicity of 20 ms. For beam tracking purposes in the connected mode, Channel State Information - Reference Signals (CSI-RSs) are sent in downlink (DL) with time and frequency offsets to their respective SS bursts, and Sounding Reference Signals (SRSs) in the uplink (UL), scheduled by the gNB.
- Beam reporting: operation used by the UE to report information gathered in the beam measurement about the set of beams that lead to the best received signal strength (RSS)/SNR. The 5G NR allows for a standalone (SA) and a non-standalone (NSA) mode, in which the control plane is managed with the help of a 4G-LTE overlay. This NSA mode is particularly helpful for beam reporting, as the 4G-LTE connection can be used to report the optimal BPL and beam failure recovery, by using the LTE link to immediately report a link failure [26].
- Beam determination: selection of the most suitable beam or beams at the UE or gNB.

A beam recovery is triggered in case of a beam failure, which occurs if the corresponding control channel degrades below a certain threshold due to blockage or beam misalignment. For this purpose, the UE maintains a candidate beam pool, which is used to send beam link failure requests and their responses.

5G NR offers a range of tunable parameters, including properties of the SS, SRS and CSI-RS, and the antenna array characteristics at UE and gNB, the beamwidth and

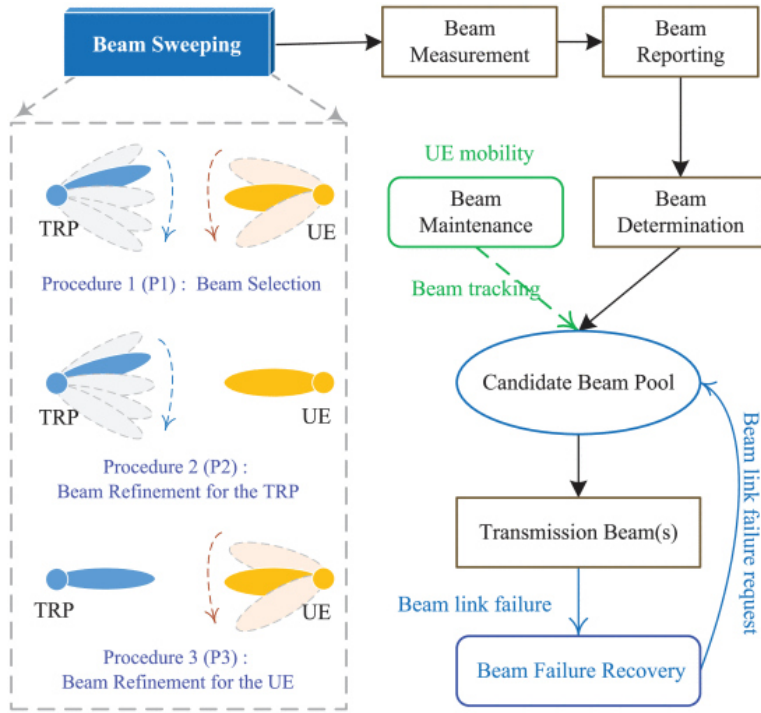


Figure 2.3: 3GPP beam management procedure for 5G NR [29].

number of simultaneously transmittable/receivable beams dependent on array size and architecture. Different performance metrics can be consulted to evaluate their impact. The accuracy of the beam management process can be assessed on the basis of a successful link establishment or the correct identification of the optimal BPL. Furthermore, the overhead due to time and frequency resources occupied by control messages, the reactivity, i.e. the time to perform IA/react to link failures, are important performance metrics. The procedure of sweeping multiple beams during beam management to find a suitable BPL is commonly referred to as beam training. The default beam training algorithm employed by 5G NR is the exhaustive search [26]. It sequentially scans all available combinations of narrow UE and gNB beams, therefore finding the optimal BPL is guaranteed. It was shown in [26] that while increasing the number of antenna elements from $M_{UE} = 4$ at UE and $M_{gNB} = 4$ at the gNB to $M_{UE} = 64$ and $M_{gNB} = 16$ can significantly increase the accuracy, it leads to an unacceptable decrease in reactivity if no simultaneous beam transmission is possible. This is because the time needed to sweep all possible beams with the exhaustive search becomes stretched over multiple SS bursts, taking up to hundreds of milliseconds to transmit and receive all possible beams, which is an intolerable delay. Mitigating factors can be an increase in available control resources, like length and frequency of the SS bursts, which increases overhead in the system, or the usage of digital/hybrid beamforming, which leads to higher power consumption and hardware requirements.

To reduce the delay induced by an exhaustive beam sweep, the *iterative* and *hierarchical* algorithms are proposed [30], which use multiple stages of exhaustive searches, starting with wider beams and subsequently scanning the narrow beams included in the previously

found wide beam. While these approaches can improve the reactivity by reducing the total amount of scanned beams, it was shown that they exhibit a lower accuracy and reduced coverage range [31]. *Agile-Link* [32] uses multi-armed beams of customized analog beamformers to gather multiple spatial directions simultaneously, combining the information of different multi-armed beams to assess each spatial direction.

Further search algorithms use the sparse, clustered nature of the mmWave channel to employ compressive sensing (CS) [33]. The greedy *Fast Neighbour Discovery* [34] algorithm chooses the first decodable link based on a random initial beacon link angle. Subsequent beacons are determined by a maximum-distance-based prediction. *BeamSpy* [35] is aimed at aiding the beam recovery by providing backup beams based on the results of previous beam searches. It uses channel impulse responses (CIRs) to model the channel using only a discrete set of signal paths and characterizes the spatial correlation by modelling how different beams share these paths.

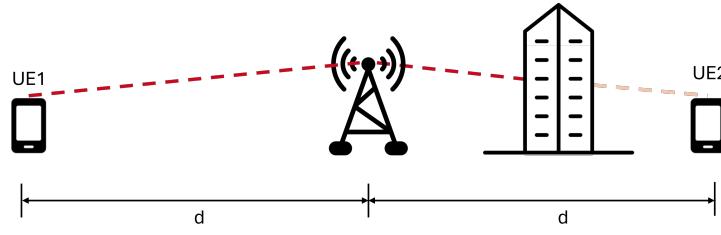
Evaluations based on realistic outdoor measurements have shown sub-optimal results for the aforementioned algorithms compared to their theoretically achievable performance, reiterating the trade-off between reactivity and required antenna gain for accuracy [36]. Consequently, future mmWave networks demand more sophisticated beam management algorithms that are able to adapt to site-specific conditions under consideration of realistic antenna patterns.

2.3 Beam Management Leveraging Context Information

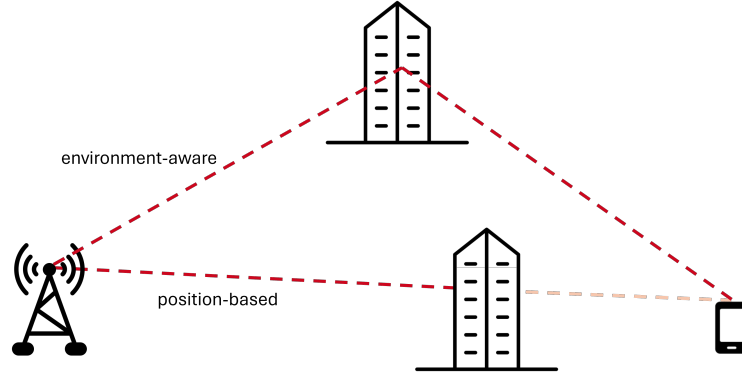
With beam codebooks constructed based on spatial directions, one of the most important information to achieve beam alignment is the relative position and orientation of the UE to gNB, particularly in LOS.

One approach to gathering this information is the usage of existing out-of-band communication signals. Future wireless systems are expected to support multi-band operations [37]. Additionally to aiding in beam reporting as in the NSA in 5G NR, this enables the exchange of CI between UE and gNB without prior establishment of a mmWave link, taking advantage of the higher range and wider beamwidth of sub-6 GHz communications. Furthermore, even though the channel gains between frequency bands outside the coherence bandwidth are uncorrelated, the signals interact with the same environments and may have a high spatial correlation [37]. In [38], a coarse angle of arrival (AoA) estimation of the sub-6 GHz channel was used to reduce the angular space of the swept beams in the 30 GHz channel and therefore limit the beam training overhead. However, the corresponding experiments have shown the spatial correlation to decrease as the LOS conditions are lost.

The UE position can be determined directly by secondary systems such as Global Positioning System (GPS). Together with data gathered by gyroscopes to account for orientation changes of the UE, the beam training can be reduced by narrowing down the potentially usable BPLs. Both systems are widespread in modern UEs such as smartphones, primary concerns are privacy-related and based on a degradation in GPS accuracy in urban environments. Measurement data from various sensors such as Radar, Lidar or Camera deployed at the gNB, can be used to estimate the AoA, in case of Radar directly infer the best BPL, or aid in accurately estimating UE position [37]. They are especially useful to distinguish between LOS and NLOS scenarios. However, while this saves resources in the RF domain, additional hardware processing, power and frequency resources are required.



(a) Channel gain in position based beam management.



(b) Example for the usage of reflections in environment-aware beamforming.

Figure 2.4: Illustration of position based versus environment-aware beam management, adapted from [37].

However, even with accurate position information, conventional distance-dependent Path Loss (PL) models are not suitable for mmWave communications, due to the effects described in Section 2.1.1. Considering a scenario like the one shown in Figure 2.4a, where two UEs are located at the same distance to a gNB, and existing position information ensure the selected beams point in the direction of the UEs and to the gNB, respectively, conventional distance-dependent PL models would assume the same channel gain for both UEs. However, as UE2 is shadowed by a building, the channel experienced by UE2 is significantly poorer in reality. To draw an accurate estimation of the PL experienced by a BPL, in addition to the position and orientation of the gNB and the UE, information about the environment is required. As shown in Figure 2.4b environment aware beam management can enable the direction of the signal towards reflection/scattering objects, thus avoiding the large shadowing losses in mmWave. The environment information can be gathered directly in the form of locations and nature of buildings and other objects, which can then be used to simulate the channel, e.g. with ray-tracing, to predict the optimal BPL at a given UE position. In an alternative approach, ML models aim to learn the function between the environment and the optimal BPL based on gathered data on optimal BPLs at UE positions in the given environment [39].

2.4 Beam Prediction Using Machine Learning

With an increase in complexity in the MIMO environment in terms of the number of antenna elements and interactions between electromagnetic waves with their environment, data-driven ML approaches can be advantageous compared to model-driven approaches. The aforementioned CI can be used to predict a selection or the best BPL, limiting or dispelling the need for beam training, respectively. Reinforcement learning (RL) techniques such as Q-Learning and Deep Q-Networks employ online learning to react to new environments and scenarios. However, there is an inherent overhead in the learning phase of online approaches, limiting the application to mobile UEs [4]. Furthermore, additional computational resources are required to perform online learning in deployed networks.

SL on the other hand is an offline approach, which is the most frequently used ML technique and has several applications in wireless communication networks even outside of beam management, such as error correction codes, data compression, mobility management, power management and channel estimation [4]. The beam selection can be categorized as a classification or regression problem, depending on whether the beam directions are in a discrete or continuous space, respectively [40]. SL uses a training dataset $D = \{(\chi_i, v_i)\}_{i=1}^N$ with N known pairs of input data vectors x and output labels $y \in C$, to train a model to find a general rule $f(\chi)$ that maps χ to v . After the training phase, the model uses the learned mapping function to predict the corresponding labels to previously unseen instances of the input χ [41]. Choosing the input vector as the set of available CI and the output as the BPL, SL enables beam prediction. Alternatively, probabilistic classifiers can estimate the probability $p(v|\chi)$ for the set of labels $v \in C$ and thus enable the prediction of a set of candidate beams ordered by the a posteriori estimate.

There are several methods to construct the mapping function $f(\chi)$. Due to their success in other fields, NNs are often considered for beam prediction. NNs are inspired by simplified models of the human brain and consist of a large number of neurons, which are able to perform basic computations that are connected in a complex network. They can be modelled as directed graphs with the neurons as nodes. Each neuron performs a basic non-linear computation referred to as its activation function $g(\cdot)$, with the input consisting of a weighted sum of the previous neurons' outputs connected to its incoming edges. Commonly used activation functions are sigmoid, step or Rectified Linear Unit (ReLU) functions [40]. Accordingly the output b_j of neuron v_j with K_j inputs a_i and weights $W_{i,j}$ is given as

$$b_j = \left(\sum_{i=0}^{K_j} W_{i,j} a_i \right). \quad (2.10)$$

During the training process, the weights are iteratively adjusted with algorithms such as the gradient descent to minimize the error between the predicted output $\hat{y}$ and the known label y according to a given loss function [41].

There exists a wide number of approaches to the structure of NN, most are based on organizing the neurons into layers with edges existing between adjacent layers. The first layer V_0 is commonly referred to as the input layer, the last layer V_T as the output layer, with a number of hidden layers in between. The number of neurons per layer can vary between layers. Figure 2.5 shows the structure of a FCNN where every neuron is connected with all neurons from adjacent layers. The series of layers with their non-linear activation

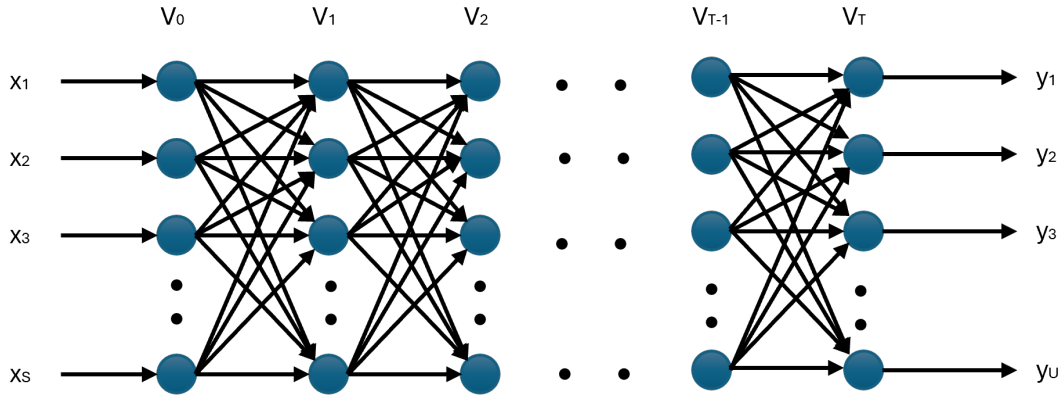


Figure 2.5: Structure of a FCNN with T layers, S inputs and U outputs. Blue dots denote the neurons, black arrows denote the connections between the neurons.

functions enable NNs to learn highly complex non-linear mapping functions. In principle, every NN that only features forward connections between adjacent layers can be expressed as a FCNN. However, it can be beneficial to reduce the number of edges and, therefore, the number of model parameters. It reduces the required computational power and training time and additionally decreases the effect of overfitting, denoting the reduced ability to accurately predict unseen data, due to an inability to identify generalized features.

Further NN architectures alter their structure according to the task and the composition of the input data. CNNs are widely used for pattern recognition on images. Convolutional layers organize their neurons into three dimensions, two of them in accordance with the image's spatial dimensions. Every neuron is connected only to a small area of these spatial dimensions of the preceding layer. CNNs often feature downsampling properties by varying the overlap between regions connected to adjacent neurons in the subsequent layer or directly in the form of pooling layers. Further information on the principles of ML and CNN can be found, for instance in [41, 42].

2.5 Beam Prediction Using UE Position

The use of relatively low-complexity NN with the UE position as input data has shown promising results in beam prediction in mmWave networks. In [43], a relatively simple FCNN, consisting of 3 hidden layers with 256 neurons each, is shown to deliver promising beam prediction accuracy on data representing a real-world scenario on the campus of the Arizona State University. The scenario consists of a gNB equipped with a beam codebook of $N_{gNB} = 64$ beams and an omnidirectional UE transmitter $N_{UE} = 1$ operating at 60 GHz. Using the UE GPS data as input, the NN is trained to predict the optimal beam, i.e. the beam delivering the highest RSS. By building a candidate set of five beams, the model was able to reach a 90% accuracy in including the optimal beam in the candidate set. To improve the robustness of models, it was proposed to apply multi-label classification by including alternative high-quality beams in the training process [44]. In a simulated indoor environment with objects mapped to 2D rectangular or circular objects, using the

UE position and orientation as input to a FCNN consisting of 5 hidden layers with 128 neurons each and $N = 64$ beams at TX and RX, it is shown that multi-label training can improve the prediction accuracy, especially for candidate sets consisting of multiple beams.

Even though NN aided beam prediction can reduce the beam training overhead, the employment of the models has an inherent cost. As the training of NNs is done offline, it does not introduce overhead or delay into the communication, nonetheless it comes with challenges further referenced to in Section 2.6. While the inference time and required computational resources for NN are considerably lower than for the training phase, they still pose a strain. Given that the UE position accuracy offered by positioning systems is limited, the average error of GPS can be in the order of meters with a significant increase in dense urban environments [45], a BIM, with an acceptable granularity can be built, providing the optimal BPL or a candidate set of BPLs for all potential UE positions in the coverage area of a gNB [46]. Such a map could be constructed online based on the data of communications taking place, resulting in a diminished offered quality during the construction phase. Alternatively, the map could be built offline based on computed data such as ray-tracing results, which have been shown to be able to accurately identify individual directional link opportunities based on publicly available 3D building models in outdoor environments [47], or on collected measurements gathered from dedicated measurement equipment. In all cases, gathering data for every possible UE location comes with costs, as even relatively small areas require a lot of measurements/simulations, e.g. an area of 100 m $\times$ 100 m sampled in a 1 m $\times$ 1 m grid results in 10000 samples. Therefore, techniques such as NNs can be applied to build the entire BIM based on finite data points, however the training of NN still requires datasets of sufficient size.

2.6 Adaptability of Prediction Models Across Environments

As described in section 2.1.1.3 the multipath components in mmWave are heavily dependent on the specific environment, therefore it is expected that beam prediction models in mmWave are environment-specific as well. As shown in Figure 2.6, an established BPL fails in providing beam alignment if the surrounding environment changes, even though the relative positions of UE and gNB are unchanged. Positions that previously had a LOS connection might be shadowed and vice versa. Small changes at the same gNB location, such as movement of cars and a dynamically changing scattering environment, that lead to an obstruction in the previously optimal BPL, can often be compensated by using the, e.g. second best BPL, as there are BPLs with acceptable performance following a different propagation path [48], and thus not being influenced by the small local change. However, entirely new environments are expected to result in an entirely new set of BPLs that offer an acceptable performance. This is confirmed in [5], where the application of a NN beam prediction model for the best BPL, trained and tested at one gNB position, at a range of new gNB positions, resulted in a drop in accuracy from 85 % to 5 %. However, as previously described, the gathering of sufficient training data in every new environment is costly, and training NN models from scratch requires time and computational resources, even though it is done offline.

As a potential solution to reduce the training data required from new environments, transfer learning has emerged as a technique that uses fine-tuning with minimal retraining to adapt ML models trained on one task to a related task [49]. Transfer learning is shown

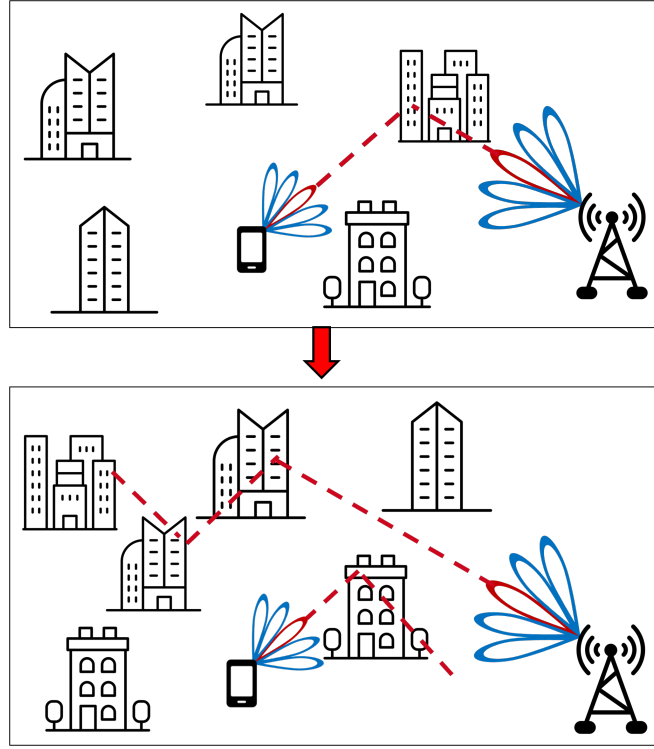


Figure 2.6: Illustration of NLOS BPL failing in a changing environment while the relative positions between UE and gNB is unchanged.

to deliver up to 80 % accuracy for predicting the best BPL in a new environment with a fine-tuning dataset size of 5 % of the covered UE positions [5].

Apart from reducing the required data from the new environment, it is desirable to generate a generalized model that is applicable across multiple environments. Due to catastrophic forgetting, the fine-tuned model after applying transfer learning is not expected to perform in the reference environment, i.e. the environment used for the original training. EWC is proposed in [50] to maintain expertise on tasks not experienced for a long time by adding a penalization to the loss function when sequentially training on new tasks, which depends on the deviation from the previously trained weights, as well as the importance of the weights for the previous tasks. The approach is shown to be effective on a number of mmWave unrelated classification tasks, such as the classification of handwritten digits included in the MNIST dataset. However, the applicability in mmWave beam selection remains open.

Although the poor performance of trained models in new environments was confirmed, a possible solution for generalized models lies in including training data from multiple gNB locations in the initial learning phase of the NN. It is expected that additional CI outside of the UE position will be required for the model to differentiate between environments. Due to their success in image-based categorization, CNNs are a potential candidate to exploit spatial relationships from feature maps.

To deploy NN based beam prediction in future communication networks, it is crucial to offer easy adaptability to new environments or generate generalized models. A protracted gathering of new training data is impractical for a widespread rollout.

2.7 Related Work

There exists a large number of works suggesting beam prediction frameworks for mmWave. The authors in [29] present a classification of papers on ML-based beam management according to their training modes, ML paradigms and models. A literature survey on beam management is included in [40]. A systematic literature review focused on beam alignment and IA frameworks for mmWave and THz was performed in [51]. However, the included works on UE position assisted SL approaches do not consider adaptability or generalization across gNB locations and environments, which was identified in [51] as a current limitation.

The authors in [52] identify generalization of ML models as a key open issue for location aided beam prediction and name dataset-mixing as a strategy used for offline learning. They state that it is difficult for the mixed dataset to cover all scenarios, but do not further investigate it.

The authors in [53] offer an overview of state-of-the-art methodologies for domain generalization problems in mmWave communications. However, while they offer existing methodologies in other related fields such as beamforming, they only identify beam prediction as an affected problem without referring to existing works on it.

The authors in [54] showed that transfer learning techniques can be leveraged to significantly reduce the required training data for location based NN aided beam alignment in changed environments. In [55], they expanded their work to build a generalized model by adding a post-processing unit after the NN and additionally exploiting sub-6 GHz Channel State Information (CSI). However, they consider an indoor environment with changes manifesting in the antenna configuration and movement of small objects instead of new outdoor environments.

The authors in [5] first investigated cross-environment transfer learning for beam prediction in mmWave and provided a model able to achieve high accuracy in unseen densely urbanized areas with limited fine-tuning data. While they also offer some insights on different dataset sizes, they do not perform an in-depth analysis on it or consider further parameters, e.g. different network architectures, noise and the number of beams used as labels. Furthermore, they do not evaluate generalization strategies.

To our knowledge, there currently exists no evaluation on the cross-environment performance of position aided SL learning strategies for beam prediction in mmWave networks. To address this, we evaluate the impact of factors such as dataset size, position noise, NN-architecture and the number of training labels on the accuracy of beam prediction in a MIMO system in realistic environments. We consider the transfer learning approach as well as the performance of dataset-mixing.

3

System Model

This chapter gives a detailed presentation of the models used in this work. Section 3.1 describes the network model in the study area in Frankfurt. In Section 3.2, the channel model that is used to generate the considered datasets is described. A description of the framework of the investigated NN beam prediction models is given in Section 3.3. Section 3.4 describes the construction of the considered datasets based on the previously introduced channel model. In Section 3.5 the method used for modelling UE positioning error is described.

3.1 Network

The mmWave cellular network we consider is located in an urban area close to the main station in Frankfurt, Germany. The study area consists of an area of 500 m x 500 m shown in Figure 3.1, bounded by the red square. The study area consists of multiple narrow streets, as well as a wider open area in front of the main station. The red triangle represents the location and direction of the plane depicted in Figure 3.2. As shown in Figure 3.2, the buildings are mostly townhouses with six to eight floors, resulting in an urban-dense environment. We consider 27 gNBs roughly uniformly distributed in the study area, placed at building edges and corners at a height of 10 m. The locations of the gNBs are depicted as blue dots in the scheme representation of the study area in Figure 3.3, also giving the corresponding indices. The study area in Figure 3.3 is undercoated in yellow.

We consider a downlink cellular network operating with a carrier frequency $f_c = 28$ GHz and a nominal gNB transmit power of $P_{tx} = 20$ dBm. The gNBs are equipped with phased antenna arrays of size 8×8 , the UE uses a 4×4 antenna array, resulting in a total number of antenna elements of $M_{gNB} = 64$ and $M_{UE} = 16$, respectively. The UE and gNBs employ analog beamforming using 3D antenna patterns obtained from lab measurements with commercially available mmWave phased antenna arrays [36]. Figure 3.4a and 3.4b show antenna patterns in azimuth cut for some example ID entries in UE and gNB codebooks, respectively. As shown in Figure 3.4, there exists a main lobe in the direction assigned to the ID, exhibiting a significant gain. Owing to the larger number of antenna elements, the lobes of the patterns at the gNB in Figure 3.4b are narrower and have a higher gain in their main lobe than at the UE in Figure 3.4a. Additionally, the gain ratio between main lobe and side lobe is higher at the gNB, especially for beams at larger ϕ_k . We use predefined codebooks to span the entire angular space, with a size of $N_{UE} = 16$ at the UE and $N_{gNB} = 64$ at the gNB, corresponding to the array sizes. This gives a beamwidth of 22.5° at the UE and 5.6° at the gNBs. Schemes for the resulting beams spanning the angular space at the UE and gNBs, are depicted in Figure 3.5a and Figure 3.5b, respectively, with each beam having a corresponding ID. A directional BPL between the UE and a given gNB

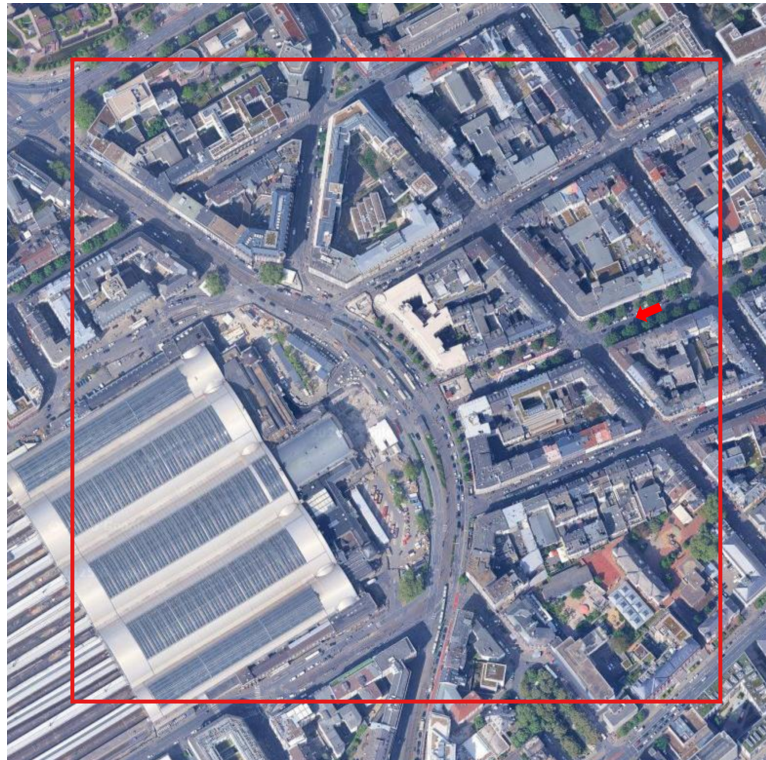


Figure 3.1: Aerial image of the considered study area near the main station in Frankfurt, Germany. The red square represents the area border, the red arrow shows the location and direction of the pane in Figure 3.2.



Figure 3.2: Exemplary street level image of a street in the study area, as an illustration for the building density.

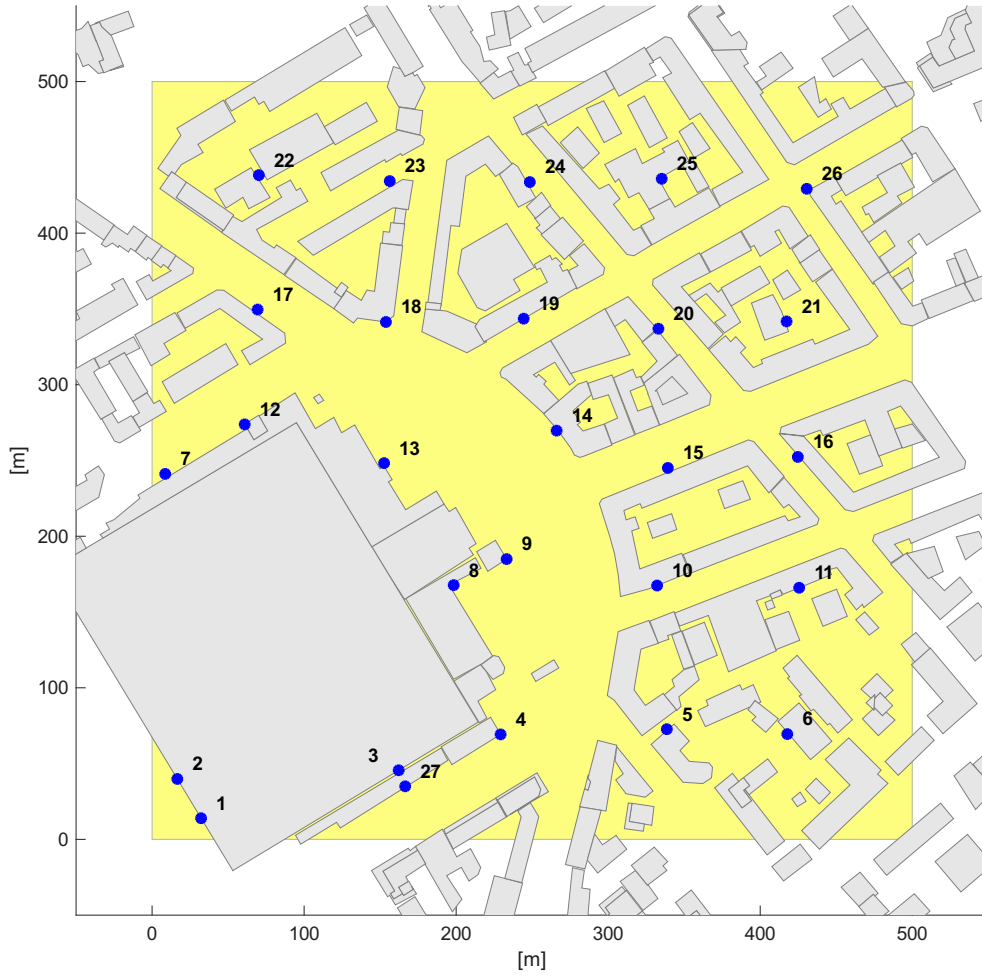
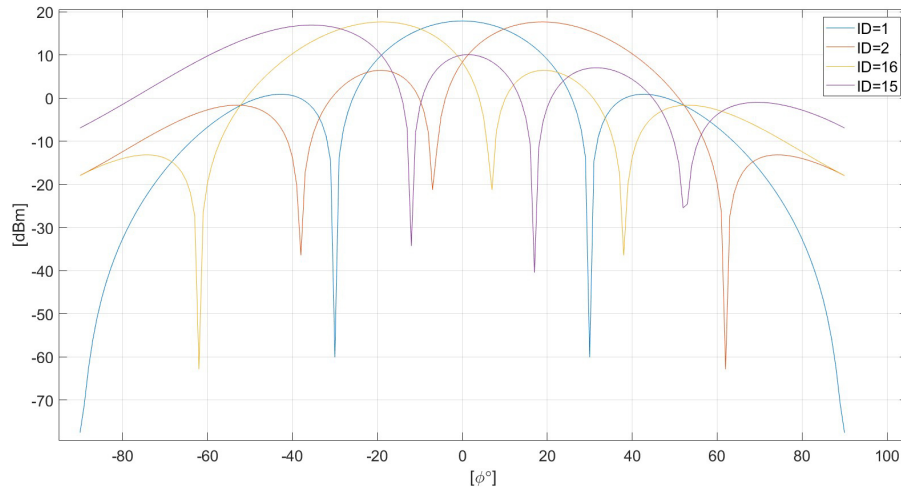


Figure 3.3: Building layout (grey) and gNB locations (blue dots) with their indices in the study area.

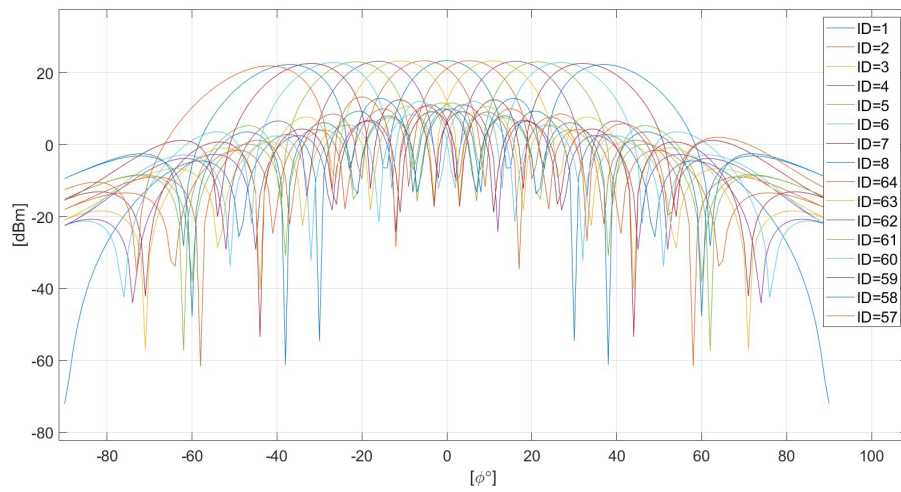
is then defined as $l_{i,j}$, with i denoting the UE beam ID and j denoting the gNB beam ID. With $N_{UE} = 16$ and $N_{gNB} = 64$ there are a total of $N_{BPL} = 1024$ potential BPLs.

3.2 Channel Model

We use site-specific channel data, obtained by the commercial ray-tracing simulator Wireless InSite [57], with a ray-launching granularity of 0.1° . The data is based on publicly available 3D building models shown in Figure 3.3, with the standardized International Telecommunication Union (ITU) glass as the building material. Given that the propagation characteristics of mmWave show a domination of reflection and LOS, as discussed in Section 2.1.1.3, scattering and diffraction are neglected, and the maximum number of reflections is set to four [47]. The channel information is collected for all potential UE positions inside the study area, with a grid resolution of $1\text{ m} \times 1\text{ m}$, and for each gNB shown in Figure 3.3. We use a gNB height of 10 m and an UE height of 1.7 m.



(a) UE.



(b) gNB.

Figure 3.4: UE and gNB antenna patterns in azimuth cut for a selected set of some example ID entries in (a) UE and (b) gNB codebooks [56].

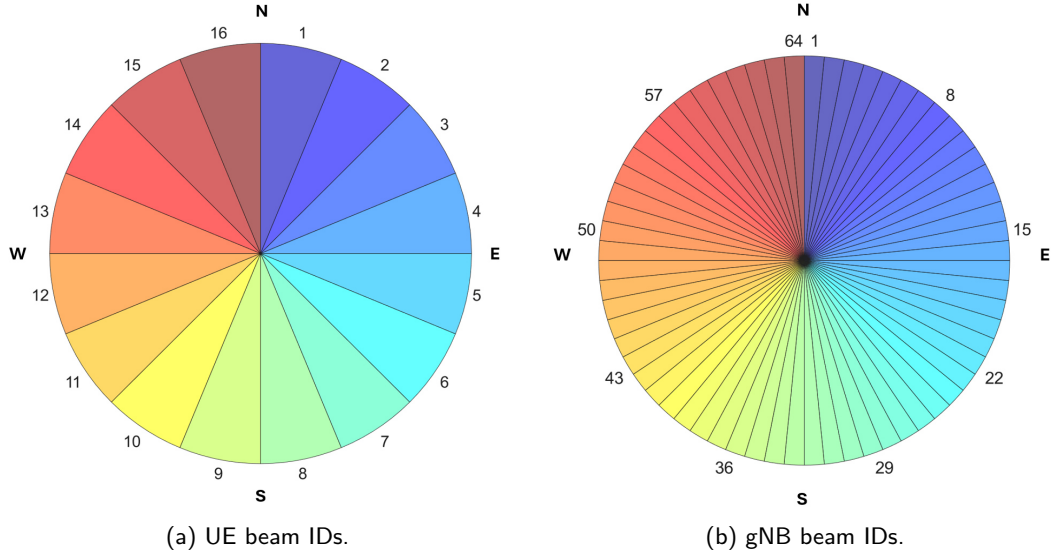


Figure 3.5: Beam direction of the codebook entries of (a) UE and (b) gNB.

The ray-tracing provides all propagation paths from each gNB to the potential UE positions, with the following information for each propagation path k :

- A distinction between LOS and NLOS
- The angle of arrival (AoA), $\{\phi_k^{UE}, \theta_k^{UE}\}$, at the UE in the azimuth and elevation, respectively
- The angle of departure (AoD), $\{\phi_k^{gNB}, \theta_k^{gNB}\}$, at the gNB in the azimuth and elevation, respectively
- The total PL, including reflection losses for NLOS paths

From the omnidirectional path data produced by the ray-tracing simulator, we determine the directional RSS of the specific BPLs by aligning the antenna patterns of the UE and the gNB according to the ID given in the codebook entries. From the antenna patterns, the gains in the direction of the propagation path k , for the beam configuration given by the UE ID i and gNB ID j , can be given as

$$G_{i,k}^{UE} = G_i^{UE}(\phi_k^{UE}, \theta_k^{UE}), \quad (3.1)$$

$$G_{j,k}^{gNB} = G_j^{gNB}(\phi_k^{gNB}, \theta_k^{gNB}), \quad (3.2)$$

The $RSS_{i,j}$ corresponding to the BPL $l_{i,j}$ at an UE position results from the sum of all K propagation paths. In dependence on the transmit power P_{TX} , the RSS is given by

$$RSS_{i,j} = P_{TX} \left(\sum_{k=1}^K PL_k G_{i,k}^{UE} G_{j,k}^{gNB} \right), \quad (3.3)$$

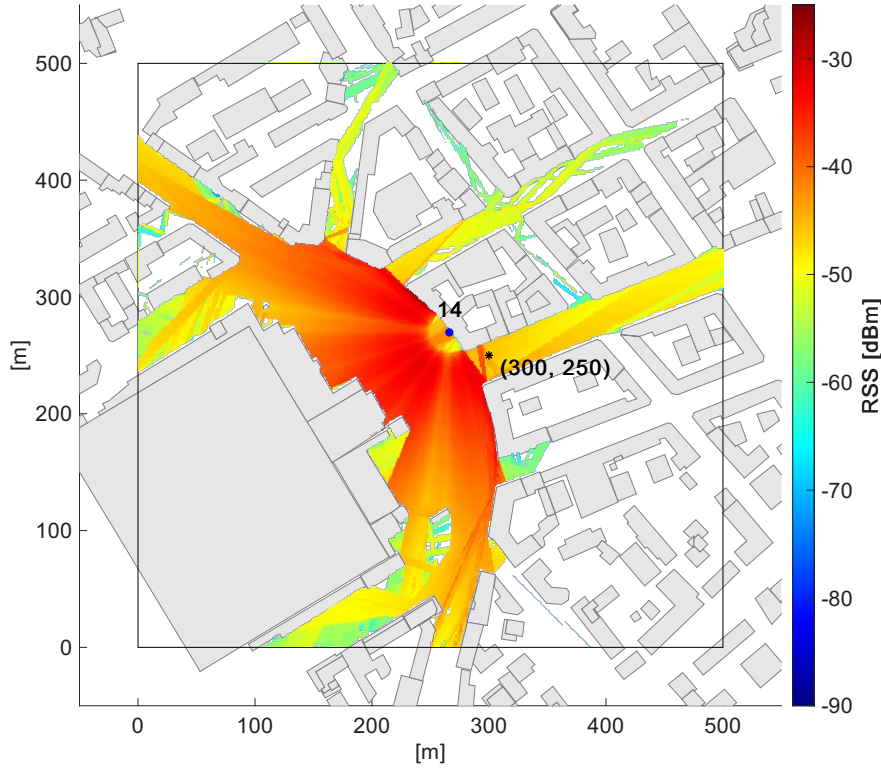


Figure 3.6: Coverage map of gNB_{14} , showing the maximum RSS (in dBm), corresponding to the best BPL selection, at each UE position as a heat map.

with the path loss PL_k of path k .

The RSS data of every candidate BPL $l_{i,j}$ for a given gNB_z and UE position $UE(x, y)$ is saved in a matrix $R_{gNB_z, UE(x, y)}$ of size $N_{gNB} \times N_{UE}$. This BPL matrix is built for every combination of UE position and gNB, however we only consider UE positions to be covered by a gNB if the maximum RSS of all BPLs exceeds the minimum RSS threshold, chosen as $RSS_{thresh} = -174$ dBm. Additionally, positions inside the buildings are not covered as we consider outdoor propagation. As an illustration, Figure 3.6 shows the maximum RSS values, corresponding to the optimal BPL selection, of gNB_{14} as a heat map. The PL incurred by the distance from the gNB, as well as the transition from LOS to NLOS conditions, can be observed in the RSS pattern. Figure 3.7 shows a representation of an exemplary BPL matrix with corresponding RSS values in form of a heat map at UE position (300, 250), served by gNB_{14} , omitting the representation of RSS below RSS_{thresh} . It shows the presence of a few dominant propagation paths that result in high RSS values at a limited number of clusters of spatially adjacent BPLs $l_{i,j}$.

3.3 Prediction Models and Performance Metrics

In the considered communication system, there is an optimal BPL l_{i^*, j^*} for every UE position, i.e. a BPL achieving the maximum RSS from the corresponding BPL matrix

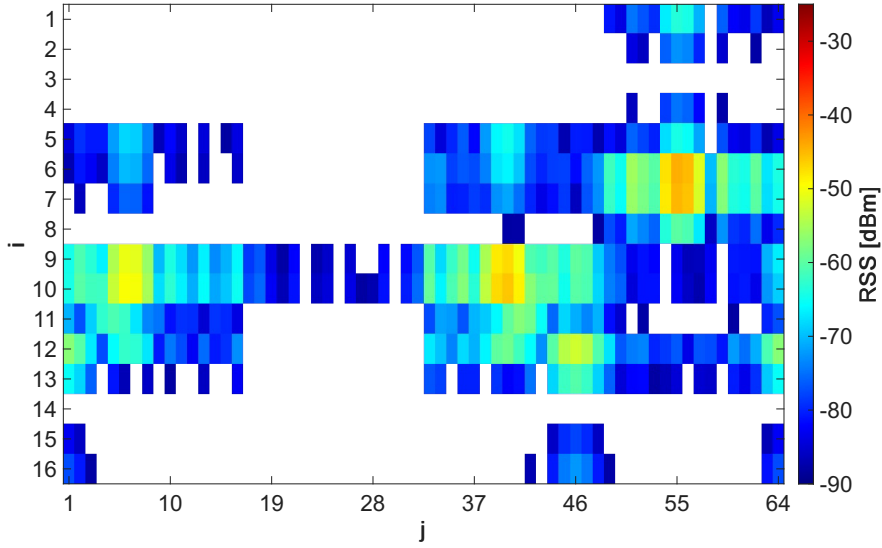


Figure 3.7: Heat map representation of exemplary BPL matrix, with i denoting the UE beam ID and j denoting the gNB beam ID, at UE position (300,250), with gNB_{14} as TX.

$R_{gNB_z, UE(x,y)}$. We map the BPLs from the two dimensional representation $l_{i,j}$ to the one dimensional l_k with the mapping function

$$l_k(l_{i,j}) = ((i-1)N_{gNB} + j) - 1, \quad (3.4)$$

thus we get the codebook $K = \{k_n\}_{n=1}^{N_{gNB} \cdot N_{UE}}$ and the optimal BPL becomes l_{k^*} .

The goal of the beam prediction is to estimate the BPL $l_{\hat{k}}$ that maximizes the probability of selecting the optimal BPL solely based on the UE position $UE(x, y)$:

$$l_{\hat{k}} = \arg \max_{k \in K} P(k = k^* | \{x, y\}). \quad (3.5)$$

We consider the top-1 accuracy Ac_{t1} as the accuracy of correctly predicting the optimal BPL, i.e. $l_{\hat{k}} = l_{k^*}$, furthermore we use the top-x accuracy Ac_{tx} , as the accuracy of having the best BPL l_{k^*} in the top x predicted.

We use the *TensorFlow* toolbox version 2.3.0 on *Python* version 3.7 to implement the investigated prediction models. The specific architectures are given in Chapter 4, however some general points can be made. While the input vector χ changes depending on the architecture, the output layer of all NNs contains $N_{UE} \cdot N_{gNB} = N_{BPL} = 1024$ neurons with a softmax activation function, where each element of the output vector is in the range [0,1] and all elements sum to 1. Therefore the results can be interpreted as a probability distribution, with each neuron representing one BPL. Note that the BPLs mapping in Equation 3.4 is shifted by -1 to fit the output layer's indexing. The hidden layers use the ReLU activation function. We use the *GlorotNormal* initializer with the seed 42 for all layers. To train the models, we use the *Adam* optimizer with a multistep learning rate with

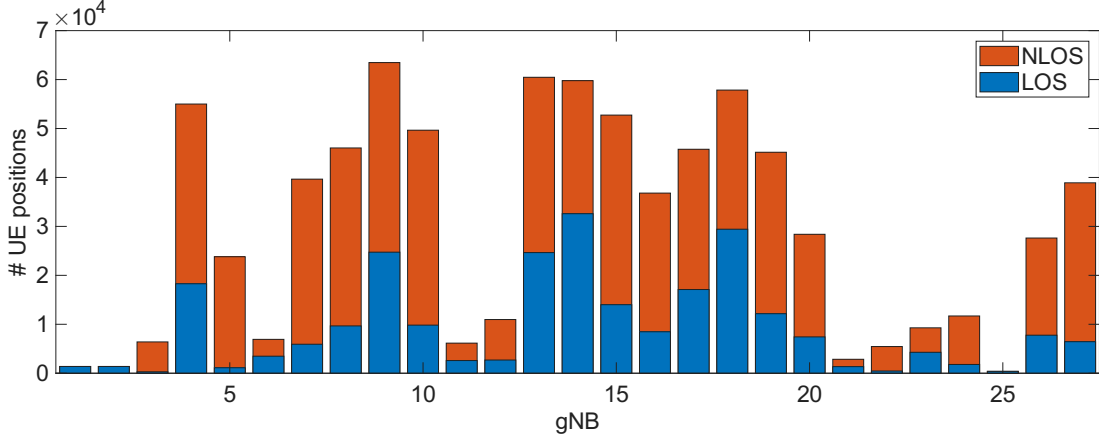


Figure 3.8: Number of UE positions included in the datasets D_{gNB_z} , with the portion of LOS positions given in blue, and the portion of NLOS positions given in red.

steps after 20 and 40 epochs. After the first step, the initial learning rate l_{ri} gets multiplied by the factor γ , and at the second step by the factor γ^2 , therefore γ is referred to as the multistep learning rate multiplier. As loss function we use the weighted cross-entropy loss function given by

$$L = \sum_{n=1}^{N_{label}} -w_n \log \hat{p}_n, \quad (3.6)$$

where N_{label} denotes the number of training labels, consisting of the corresponding number of best BPLs, $\hat{p}_n$ denotes the predicted probability that the n^{th} best BPL is selected, and w_n is the corresponding weight. Apart from our investigation into the impact of the number of training labels (see Section 4.2.4), we consider $N_{label} = 5$, and we use the weights $[0.5, 0.2, 0.15, 0.075, 0.075]$, to prioritize the selection of the best BPL while also promoting the selection of one of the subsequent top-5 BPL [5].

3.4 Dataset

Based on the BPL matrix $R_{gNB_z, UE(x,y)}$, we built our dataset D_{gNB_z} for every gNB, each containing data for every covered UE position. Given that we do not build any BPL matrix for positions where the RSS corresponding to the best BPL is below RSS_{thresh} , D_{gNB_z} varies in size between gNBs. The resulting number of UE positions included in each gNB dataset is given in Figure 3.8, also showing the portions of NLOS and LOS positions.

An example section of the resulting dataset is shown in Table 3.1 for $D_{gNB_{14}}$. In the first two columns, we give the two coordinates of $UE(x,y)$, the exact interpretation of them is further described in Section 4.1. Columns three and four contain the position of gNB_{14} according to the coordinate system shown in Figure 3.3. Note that they are constant inside the dataset $D_{gNB_{14}}$ and are only used for learning with mixed datasets in accordance with

$UE(x, y)$ (m)		$gNB(x, y)$ (m)		Top BPLs		
x	y	x	y	Top 1	...	Top 30
1	236	266	270	154	...	211
1	237	266	270	154	...	14
.	.	.	.	.	...	.
264	56	266	270	248	...	254
.	.	.	.	.	...	.
500	340	266	270	615	...	742
500	341	266	270	616	...	608

Table 3.1: Exemplary section of the dataset D_{gNB_z} .

Section 4.3.3. The next columns are filled by the best performing BPLs, ordered based on their RSS. The number of BPLs used as training labels is one investigated parameter (see Section 4.2.4), we include up to the Top 30 BPLs in the dataset to be flexible. To build the training and test datasets, we randomly shuffle the rows of D_{gNB_z} and then use 80 % as training data and 20 % as test data.

3.5 UE Position Error

We do not specify the method used for gathering UE position information, however positioning systems are susceptible to positioning errors, e.g. the average error in GPS can be in the order of meters [45]. To investigate the impact of UE position errors, our investigation includes a section (see Section 4.2.3), in which we model UE positioning errors as a noise in the UE position data, with the components x_{err} and y_{err} independently normally distributed, as given for x_{err} as:

$$p_{x_{err}}(x_{err}) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x_{err}-\mu_x)^2}{2\sigma_x^2}}, \quad (3.7)$$

with $\mu_x = x$ and $\sigma_x = 5.93$ m, as the standard deviation in longitude direction. The noise in latitude direction y is constructed in the same way, with the standard deviation $\sigma_y = 1.73$ m. We derive latitude and longitude variance values from the global navigation satellite system (GNSS) measurement dataset in [58] collected in Frankfurt.

4

Cross Environment Beam Prediction

In this chapter, the investigated models for cross-environment beam prediction are presented in Section 4.1. Section 4.2 gives the selected parameters for the investigation of the performance of transfer learning. Section 4.3 describes our approaches towards cross-environment generalizability of the investigated models.

4.1 Preliminary Investigation on Network Parameters

In this section, we describe the process performed to get the parameters of the employed FCNN in Section 4.1.1 and the CNN in Section 4.1.2. In our investigations on the FCNN, the given results are the mean of five simulations, due to the long training times, the results for CNN come from one single simulation.

4.1.1 Fully Connected Neural Network

We initially use a FCNN as introduced in [5] for UE position based beam prediction. Figure 4.1 shows the scheme of the NN architecture. Before starting our investigation on cross-environment learning performance, we executed a range of tests to adjust our model hyperparameters for the task of position-based beam prediction. For this purpose, we investigated the performance of five different methods to preprocess the UE position:

- S1: The absolute coordinates (x,y) according to the coordinate system in Figure 3.3
- S2: The absolute coordinates (x,y) according to the coordinate system in Figure 3.3 scaled to the range $[0,1]$ with min-max normalization
- S3: The relative UE position $UE_{rel}(x,y)$ in relation to gNB_z , with $UE_{rel}(x_{rel}, y_{rel}) = UE(x,y) - gNB_z(x,y)$ element wise
- S4: The relative UE position in relation to gNB_z , scaled to the range $[-1,1]$ with min-max normalization
- S5: The relative UE position in relation to gNB_z , transformed into polar coordinates $UE_{rel}(r, \alpha)$

S1 represents the UE position format as it exists without any further processing. In S3-S5, we try to leverage the relationship between the AoD and AoA with the BPL, by giving the UE position relative to the gNB location. Previous investigations on FCNNs have shown an advantage offered by the normalization of input data [59], which we use in S2 and S4 by performing min-max normalization.

Since the structures differ in their scaling, we investigate a range of initial learning rates on two example gNBs (gNB_{11} and gNB_{14}) without transfer learning, i.e. we train the

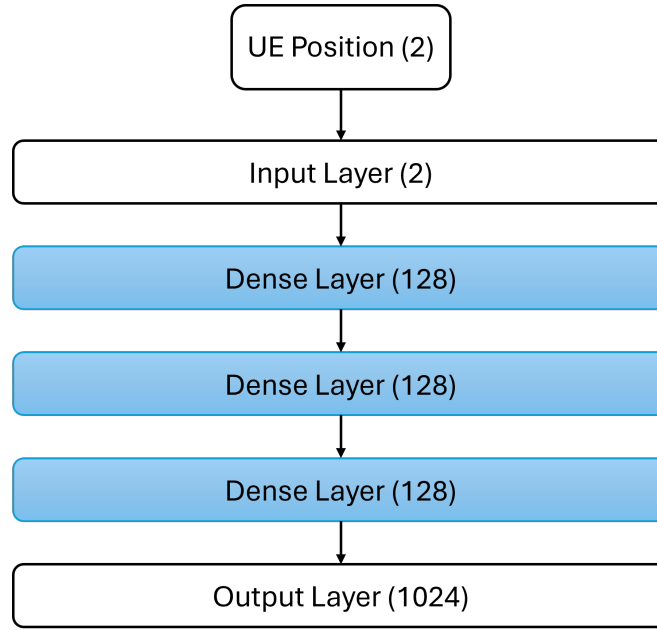


Figure 4.1: Architecture of the FCNN used for beam prediction [5]. In Dense Layers, every neuron is connected to every neuron in the adjacent layers. The input consists of the two dimensional UE position, the dimensions of the input and each layer are given in brackets.

models with the training dataset of one gNB and subsequently test on the testing dataset of the same gNB. For each input data preprocessing modality S1-S5, we test the top-1 accuracy Ac_{t1} for the initial learning rates $l_{ri} = [0.2, 0.02, 0.002, 0.0002, 0.00002]$ and the multistep learning rate multiplier $\gamma = [0.5, 0.3, 0.1]$, based on the average of five repeated simulations. The list of the FCNN parameters is provided in Table 4.1.

Parameters	Values
Input size	2 ([S1-S5])
Label size	5
Hidden layers	3 layers, 128 nodes each
Output size	1024 (number of potential BPLs)
Activation	ReLU on hidden layers
Loss function	Weighted cross-entropy loss
Percentage of D_{gNBz} for Training/Test	[80%, 20%]
Learning batch size	128
Learning epochs	60
Initial learning rate l_{ri}	[0.2, 0.02, 0.002, 0.0002, 0.00002]
Multistep learning rate factor γ	[0.5, 0.3, 0.1]

Table 4.1: NN architecture and hyperparameters for investigating UE position preprocessing.

Based on the results without transfer learning, we choose $l_{ri} = 0.002$ for S1, S3 and S5, $l_{ri} = 0.02$ for S2, S4, and $\gamma = 0.3$ for S1-S3. We then compare the achieved accuracies of the input preprocessing modalities S1-S5 in transfer learning, by training the model with the full dataset $D_{gNB_{ref}}$ on a reference gNB gNB_{ref} and subsequently fine-tuning the model with the training data of a target gNB gNB_{tar} , using just 5 % of the dataset $D_{gNB_{tar}}$ for the fine-tuning at the target gNB. We evaluate every combination of reference and target gNBs.

4.1.2 Convolutional Neural Network

Alternatively to using the UE position $UE(x, y)$ as a tuple as an input for a FCNN, we use a one-hot map of the study area, resulting in a 500×500 matrix filled with zeros except at the entry (x, y) , where it has the value one. The one-hot maps of each position are generated during the training process, because generating them in advance introduces memory constraints on the used hardware. We use this map as an input for a CNN with the idea to exploit the capabilities of CNN in image-related tasks. Initially, we tested different CNN architectures on the task of beam prediction, leveraging the whole dataset of one gNB. Due to the large size of the input data in form of the 500×500 input picture, we decide to use a low-complexity NN that is still able to reach a high Ac_{t1} . Based on the tries on gNB_{11} , we choose the CNN described in Figure 4.2 and Table 4.2, further referred to as CNN-S, as its Ac_{t1} was among the best reached results, while keeping limited complexity in the architecture. However, the training of the CNN with the given input size still takes considerable time, therefore in contrast to the FCNN we do not take the mean from multiple repeated simulations.

Parameters	Values
Input size	$(500 \times 500 \times 1)$
Label size	5
Hidden layers	2 convolutional layers
Output size	1024 (number of potential BPLs)
Activation	ReLU on hidden layers
Loss function	Weighted cross-entropy loss
Percentage of D_{gNB_z} for Training/Test	[80%, 20%]
Learning batch size	128
Learning epochs	10
Initial learning rate	0.002
Multistep learning rate factor γ	0.3

Table 4.2: CNN-S architecture and hyperparameters.

4.2 Transfer Learning

In this Section, the investigated parameters impacting the performance of transfer learning are presented and discussed. In Section 4.2.1, we present our approach for evaluating the advantage of transfer-learning compared to training a new model. In Section 4.2.2, the NN that was used to investigate possible advantages of using larger NNs is presented. In

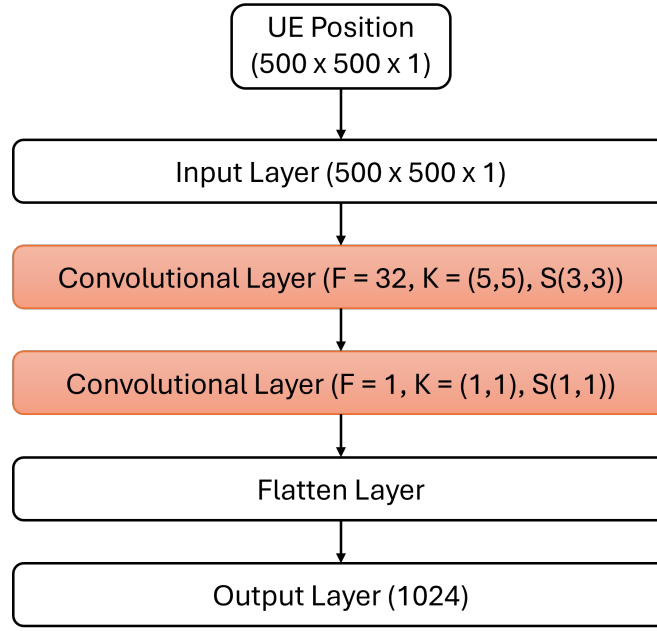


Figure 4.2: Architecture of the CNN referred to as CNN-S. The input consists of the UE position in form of the 500×500 one-hot map. In convolutional layers, the value F gives the number of filters, K gives the size of the filters, and S gives the stride. The flatten layer reduces the dimensions to fit the output layer.

Section 4.2.3, the investigation on UE position noise is introduced, and Section 4.2.4 gives details on the investigation on the number of training labels.

4.2.1 Dataset Size

Based on the results of the hyperparameter investigation described in Section 4.1.1, we choose the FCNN architecture and hyperparameters given in Table 4.3. We refer to this architecture as FCNN-S. To investigate the performance of transfer learning, we consider two scenarios for the investigated models:

- B1: In this baseline scenario we vary the percentages pd of the datasets D_{gNB_z} we use. We train the model with the training data of one gNB and test with the testing data of the same gNB. If not stated otherwise, we choose $pd = 100\%$.
- B2: In the scenario B2 we use transfer learning. We use the complete datasets $D_{gNB_{ref}}$ ($pd = 100\%$) split into training and test data for training on the reference gNB gNB_{ref} , and $pd = 5\%$ of the dataset $D_{gNB_{tar}}$ of the target gNB_{ref} to fine-tune the model on the target scenario.

We investigate the impact of the dataset size by varying the percentage pd in B1. To evaluate the gain achieved by transfer learning, we compare the accuracies of B2 with the accuracies of B1 with $pd = 5\%$ on the target gNBs, i.e. using the same portion of the dataset used for fine-tuning in B2 to train a new model.

Parameters	Values
Input size	2 (S3)
Label size	5
Hidden layers	3 layers, 128 nodes each
Output size	1024 (number of potential BPLs)
Activation	ReLU on hidden layers
Loss function	Weighted cross-entropy loss
Percentage of D_{gNB_z} for Training/Test	[80%, 20%]
Learning batch size	128
Learning epochs	60
Initial learning rate	0.002
Multistep learning rate factor γ	0.3

Table 4.3: Parameters of the FCNN referred to as FCNN-S.

4.2.2 Neural Network Size

To investigate potential room for improvement in the beam prediction accuracy with FCNNs, we modify the NN given in Figure 4.1 and Table 4.3 by varying the number and size of hidden layers. Based on the Ac_{t1} results of scenario B1 with gNB 10 and gNB 25, $pd = 100\%$, and B2 with (gNB_{ref}, gNB_{tar}) combinations (8, 9) and (15, 19), we choose the larger FCNN given in Figure 4.3 and Table 4.4, which we refer to as FCNN-L. Due to additional hidden layers and layers with more neurons, it is able to reach higher Ac_{t1} values. Our tests have shown no improvement by further adding hidden layers or neurons in the hidden layers in FCNN-L. The analysis of the results obtained with FCNN-L is presented in Section 5.1.3.

Parameters	Values
Input size	2 (S3)
Label size	5
Hidden layers	8 layers, increasing number of neurons
Output size	1024 (number of potential BPLs)
Activation	ReLU on hidden layers
Loss function	Weighted cross-entropy loss
Percentage of D_{gNB_z} for Training/Test	[80%, 20%]
Learning batch size	128
Learning epochs	60
Initial learning rate	0.002
Multistep learning rate factor γ	0.3

Table 4.4: Parameters of FCNN-L, identical to the FCNN in Table 4.3 with the exception of the structure of hidden layers.

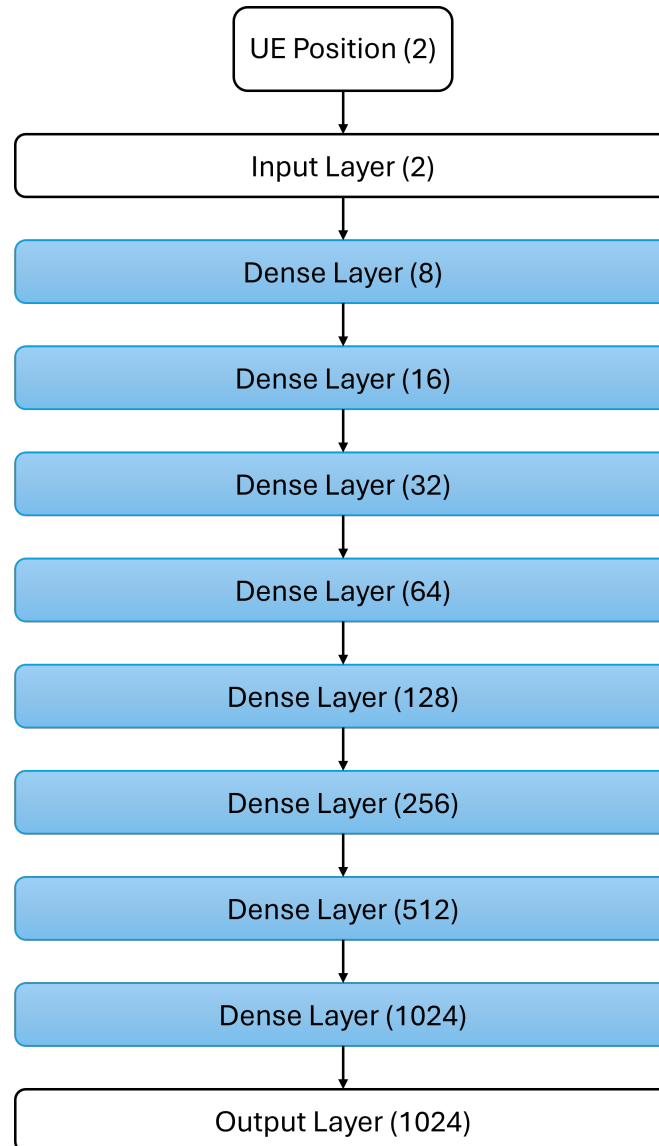


Figure 4.3: Architecture of the FCNN-L. In Dense Layers, every neuron is connected to every neuron in the adjacent layers. The input consists of the two dimensional UE position, the dimensions of the input and each layer are given in brackets.

4.2.3 Noise in the UE Position

We consider three scenarios where the UE position is noisy as per Equation 3.7:

- N1: Both in the complete reference $D_{gNB_{ref}}$ and target $D_{gNB_{target}}$ datasets
- N2: Both in the test data of the reference $D_{gNB_{ref}}$ and target $D_{gNB_{target}}$ datasets
- N3: As N2 and additionally in the training data of dataset $D_{gNB_{target}}$

The reasoning for considering N2 is that dedicated measurement equipment used for gathering training data might have more precise positioning capabilities than the UEs. Similarly, the positioning data of ray tracing models does not suffer from the same noise, e.g. it is unaffected by GPS noise but depends on the accuracy of the building modelling. As we try to reduce the costs of gathering data in new environments, the data gathered at the target gNB might not be of the same quality as at the reference gNB. Therefore, in N3 we consider UE position noise in all the data except in the training data of the reference gNB. For all three scenarios N1-N3, we compare the achieved top-1 and best-in-top-5 accuracies with the baseline results B2, which are obtained with a dataset not affected by position error.

The representation of the UE position as a map in the CNN allows us to introduce uncertainty to the input data. We built a two dimensional normal distribution according to the assumed noise parameters given in Section 3.5. We use min-max normalization on the resulting probability density function (pdf) and limit its tails to five values, corresponding to five meters, in both dimensions, as the pdf values decline by over one order of magnitude when exceeding this range. Therefore, instead of using a one-hot map for the UE position, we centre the pdf cut-out at the UE position. We refer to this method of structuring the input data for position uncertainty in the CNN as G1. The results of our investigation on noisy UE position data are given in section 5.1.4

4.2.4 Number of Training Labels

The goal of the employed weighted cross-entropy function is to prioritize the selection of the best BPL, while also promoting the selection of subsequent top-5 BPL [5]. To evaluate the effectiveness of this approach, we compare the top-x accuracy not only for the best BPL l_{k^*} , but additionally for the second to fifth best BPL, while also varying the number of training labels N_{label} . We denote this probability of the true n^{th} best BPL being among the x BPLs with the highest predicted probabilities $\hat{p}_k$ as Ac_{tx}^n . Using higher N_{label} is expected to be particularly advantageous for Ac_{tx}^n with corresponding higher n . However, higher N_{label} might be disadvantageous for Ac_{tx}^1 , especially for low x as the 2^{nd} to n^{th} best BPLs might be predicted over the best BPL l_{k^*} . To investigate the impact of N_{label} we simulate the Ac_{tx}^n for $N_{label} = 1$ to $N_{label} = 5$, with the weights shown in Table 4.5 used in the weighted cross-entropy loss function given in equation 3.6. The weights are structured in a descending way, with the best BPL weighted with the highest factor. Corresponding to the maximum investigated number of training labels, we investigate Ac_{tx}^n for $n = 1$ to $n = 5$. To adequately display the possible advantages of every combination, we consider the wider range Ac_{t1}^n to Ac_{t10}^n . The results of this investigation can be found in Section 5.1.5.

N_{label}	weights
1	[1.0]
2	[0.7, 0.3]
3	[0.6, 0.25, 0.15]
4	[0.5, 0.25, 0.15, 0.1]
5	[0.5, 0.2, 0.15, 0.075, 0.075]

Table 4.5: Weights employed in the weighted cross-entropy loss function given N_{label} , according to equation 3.6. The weights of $N_{\text{label}} = 5$ are equal to the weights used in other sections of this work.

4.3 Cross-Environment Generalizability

In this section, the investigated approaches towards cross-environment generalizability are presented, starting in Section 4.3.1 with the evaluation of the accuracy at the reference gNB after fine-tuning on different target gNBs. The approach of using EWC to retain accuracy on the reference gNB after fine-tuning on different target gNBs is presented in Section 4.3.2. In Section 4.3.3, we present our approach to use dataset mixing to produce a model that achieves high accuracies at multiple gNBs simultaneously.

4.3.1 Accuracy at the Reference gNB after Fine-Tuning on Different Target gNBs

As an initial step towards cross-environment generalizability, we investigate the accuracies of the investigated models on the reference gNB after the fine-tuning stage in B2. We test models trained following B2 with $gNB_{\text{ref}} = 14$, and compare the accuracies achieved on gNB 14 with the achieved accuracies before the fine-tuning. The results are given in Section 5.2.1

4.3.2 Elastic Weight Consolidation

As we observe in our investigations in Section 5.2.1, introducing new training data of a task B to an existing FCNN model trained for task A often leads to a severe loss in performance on task A, commonly referred to as catastrophic forgetting. In our application, the fine-tuning stage of transfer learning with data from a target gNB is task B and leads to low beam prediction accuracies on the test data of the reference gNB, which is task A. To limit catastrophic forgetting, elastic weight consolidation (EWC) is proposed in [50]. It tries to estimate the importance of the tunable NN parameters for task A with the help of the Fisher information matrix, and anchors the important parameters by applying a quadratic penalty to the loss function. The resulting loss function for task B is given by

$$L(\theta) = L_B(\theta) + \sum_i \frac{\lambda}{2} F_i (\theta_i - \theta_{A,i}^*)^2, \quad (4.1)$$

where λ sets how important task A is compared to task B (with higher lambda we give more importance to task A), F_i is the Fisher information matrix, θ_i are the tunable parameters, $\theta_{A,i}^*$ are the tunable parameters after the training on task A, and $L_B(\theta)$ is the loss for task

B, in our case the weighted cross-entropy loss function given in Equation 3.6. Thus EWC penalizes a larger difference $(\theta_i - \theta_{A,i}^*)$ between the current parameters and the parameters after training on task A, and assigns parameters important for task A a higher weight, estimated by F_i . We refer to the accuracy on the test data of task A after first training on task A followed by training on task B as $Ac^{A,B-A}$, and the corresponding accuracy on the test data of task B as $Ac^{A,B-B}$.

One performance evaluation of EWC is in [50] on the MNIST dataset composed of handwritten digits [60]. To validate our implementation, which we adapted from [61], and get a baseline for the performance of EWC, we perform categorization on the MNIST dataset using the hidden layer structure of FCNN-S, with task A being the classification of digits between even digits and task B being the classification of digits between uneven digits. Our investigation on the MNIST dataset has shown that adding a dropout layer as the last hidden layer, which randomly sets input units to zero, significantly improves the accuracy of the FCNN. Additionally, we found that reducing l_{ri} significantly increases $Ac^{A,B-A}$. Therefore, we investigate a modified version of FCNN-S, which we refer to as FCNN-SD. FCNN-SD adds a dropout layer with a dropout factor of 0.5 as the last hidden layer to the FCNN-S and changes the initial learning rate to $l_{ri} = 0.0002$. To isolate the effect of EWC, we test FCNN-SD without EWC, employing the loss function from Equation 3.6, and with EWC, employing the loss function from Equation 4.1. We found $\lambda = 25$ as a suitable weight.

We investigate the performance of EWC for retaining BPL prediction accuracy on the test data of the reference gNB 14 after fine-tuning on a different target gNB following B2. The results can be found in Section 5.2.2.

4.3.3 Dataset Mixing

Mixing datasets of multiple tasks into one training dataset is another way to avoid catastrophic forgetting [52]. We test the generalizability across gNBs that are included in the mixed dataset, as well as to gNBs not included in the mixed training dataset, to investigate the accuracy of our model on unseen scenarios. To reduce the required training time, we used $pd = 60\%$ of the datasets while maintaining the $[80\%, 20\%]$ train/test split. We included every gNB except gNB 10 and 13 into the mixed dataset used for training. We investigated the following accuracies:

- M1: The accuracy on the test data of gNBs included in the mixed dataset
- M2: The accuracy on the test data of gNB 10 and 13, which are not included in the mixed dataset

Additionally to FCNN-S, FCNN-L and CNN-S, we test the accuracies for the NNs FCNN-SG and FCNN-LG, in which we add the gNB position in absolute coordinates (x_{gNB}, y_{gNB}) as two additional dimensions to the input data of FCNN-S and FCNN-L. Furthermore, we modify the input data of CNN-S such that it represents the UE position relative to the gNB position, i.e. the one-hot map has a one at the position of S3, shifted such that the centre of the input matrix corresponds to $UE(0, 0)$. To be able to represent every possible relative UE position, the input matrix has to be extended to the size 1000×1000 . We refer to the CNN-S with the relative input data format as CNN-SR.

To investigate possible advantages of a mixed dataset for the reduction of the required training data, we test the M1 accuracies of FCNN-SG and FCNN-LG using $pd = 5\%$, which corresponds to the pd used in B2. The corresponding results are given in Section 5.2.3.

5

Results

In this chapter, the results corresponding to the simulation setups introduced in Chapter 4 are presented and discussed. Section 5.1 contains the results of the investigations on the performance of transfer learning. Section 5.2 presents the results on cross-environment generalizability of the investigated models.

5.1 Transfer Learning

At the start of this section, the preliminary results regarding the initial learning rate and the preprocessing of the UE position used as input data are presented. In the following sections, the results of the investigation on the impact of the respective parameters are presented, with the dataset size in Section 5.1.2, the NN size in Section 5.1.3, the UE position noise in Section 5.1.4, and the number of training labels in Section 5.1.5.

5.1.1 Preliminary Results

The top-1 accuracy results on the NN in Table 4.1 for the training on gNB 11 and 14 are given in Figure 5.1a and Figure 5.1b, respectively. The top-1 accuracy Ac_{t1} is plotted for each UE position preprocessing S1-S5, given in different colours. The tested initial learning rates l_{ri} are plotted on the x-axis, while the tested learning rate multipliers γ are given as solid, dashed and dotted lines. As shown in Figure 5.1, all the preprocessing modalities S1-S5 show the best accuracies either at $l_{ri} = 0.02$ or $l_{ri} = 0.002$. The two scaled UE position preprocessing modalities S2 and S4 show a clear advantage for $l_{ri} = 0.02$, especially at gNB 11, where $l_{ri} = 0.02$ with preprocessing modality S2 outperforms $l_{ri} = 0.002$ by 40 percentage points (pp). By comparing the two gNBs, we can see that $l_{ri} = 0.02$ performs better in comparison to $l_{ri} = 0.002$ at gNB 11 over all preprocessing modalities S1-S5. A possible explanation for this is the difference in the size of the datasets $D_{gNB_{11}}$ and $D_{gNB_{14}}$. As shown in Figure 3.8, $D_{gNB_{11}}$ is around 10 times smaller than $D_{gNB_{14}}$ and therefore only 10 % of the training batches are available. A larger initial learning rate is necessary to reach a sufficient adaptation of the NN weights for these small datasets. An additional observation is that the authors in [5] used $l_{ri} = 0.2$ for the given NN with an unspecified γ , while the results in Figure 5.1 show a significantly worse accuracy over all investigated UE position preprocessing modalities for $l_{ri} = 0.2$. This indicates that the implementation and framework used for the FCNN can have a significant impact on the performance. We can also see that the chosen γ is shown to have a comparably minor impact on Ac_{t1} over all combinations, with a maximum deviation of around 6 pp.

Based on the result shown in Figure 5.1, we perform transfer learning over all possible combinations of reference and target gNBs using the best parameters found. We choose $l_{ri} = 0.02$ for preprocessing modalities S2 and S4, and $l_{ri} = 0.002$ for preprocessing modalities S1, S3 and S5. We choose $\gamma = 0.3$ as it performs best at multiple combinations of the

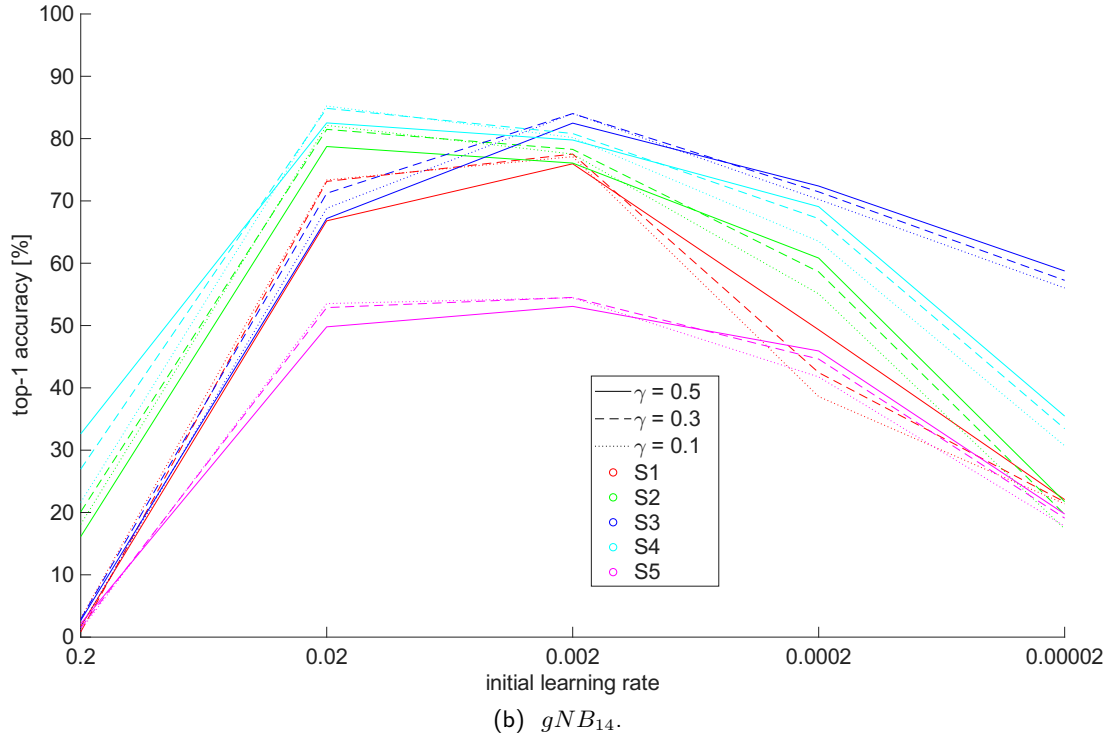
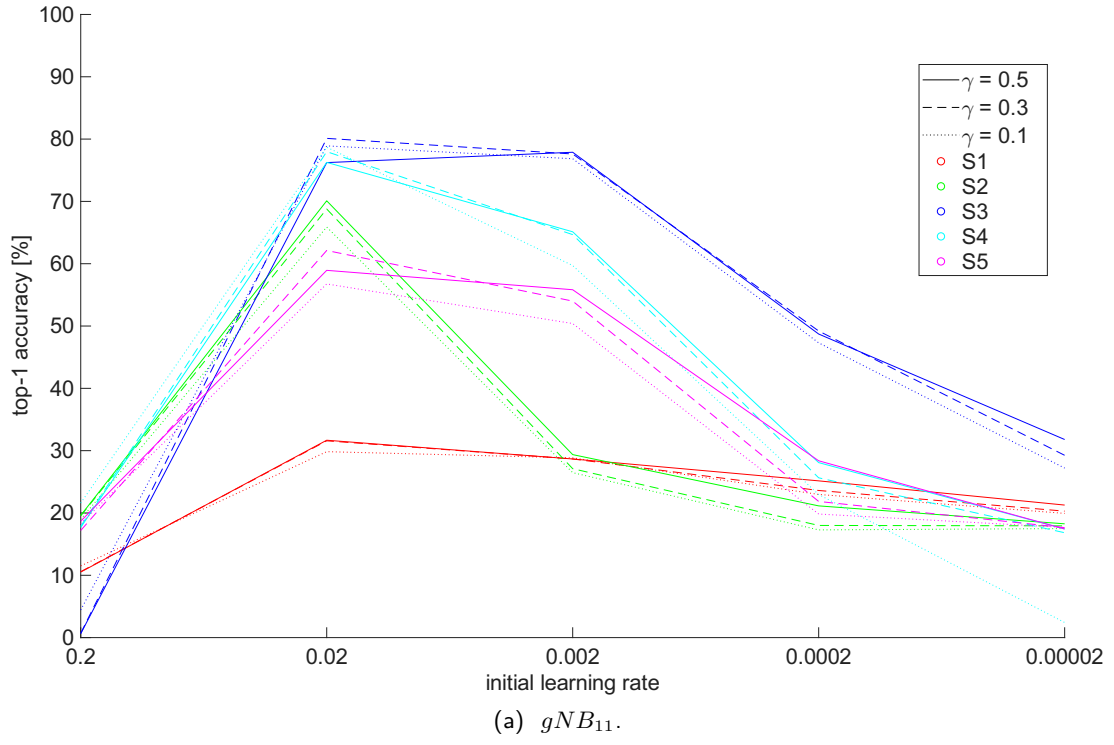


Figure 5.1: Top-1 accuracy plots on initial learning rates and γ for the FCNN-S given in Table 4.1 on gNB_{11} and gNB_{14} without transfer learning. The l_{ri} are given on the x-axis, the multistep learning rate multipliers γ are given with solid, dashed and dotted lines, and the UE position preprocessing modalities S1-S5 are given in different colours.

chosen initial learning rates and UE position preprocessing modalities. The corresponding transfer learning top-1 accuracies Ac_{t1} are given in Figure 5.2 in the form of accuracy colour maps. The reference gNB gNB_{ref} is given on the y-axis, and the target gNB gNB_{ref} is given on the x-axis. In Figure 5.2a, depicting S1, we can observe a strong dependency of Ac_{t1} on the size of the target and reference datasets $D_{gNB_{tar}}$ and $D_{gNB_{ref}}$. By comparing the combinations with low accuracy with the dataset sizes shown in Figure 3.8, we can see that gNBs with smaller datasets achieve considerably lower Ac_{t1} values. An exception exists for target gNBs 1 to 3, which offer good Ac_{t1} results despite a small dataset size. The results for gNB 1 and gNB 2 can be explained by their location in the corner of the study area, as depicted in Figure 3.3. As they are located on a straight wall, their coverage area consists only of LOS transmission, therefore easier propagation pattern to learn for the model. The comparatively good performance of gNB 3 on the other hand might be explained by consulting the coverage map of gNB 3, depicted in Figure 5.3. The location of gNB 3 in a narrow space between two buildings leads to coverage on a limited number of narrow propagation paths. Resulting from this, there are only 25 different best BPLs l_{k^*} in D_{gNB_3} , compared to 157 at gNB_{22} , which simplifies the beam prediction task significantly.

In comparison to S1, S3 shows a significantly improved Ac_{t1} over all combinations of gNB_{tar} and gNB_{ref} . Since the BPL is dependent on the AoD and AoA at UE and gNB, respectively, representing the position of the UE relative to the gNB offers an intuitive advantage. In LOS situations, the same relative position will result in the same l_{k^*} , independent of the position with respect to the study area. The representation in polar coordinates in S5 shows worse results compared to S3 in our simulation. A possible reason might be the different scales in the angular and distance components.

In our simulation, the data normalization in S2 and S4 does not offer an advantage in transfer learning compared to S1 and S3, respectively. A possible initial explanation is that the higher initial learning rate l_{ri} leads to a higher adjustment in the trained weights during the initial steps of the fine-tuning, leading to a loss of previously trained information. However, investigations using an initial learning rate $l_{ri} = 0.02$ during training on gNB_{ref} and $l_{ri} = 0.002$ during fine-tuning on gNB_{tar} do not show any improvement.

Given the clearly favourable performance of S3 in our preliminary simulations, we choose it as the preprocessing modality for the further investigations on the FCNN.

5.1.2 Dataset Size

To show the general dependence of Ac_{t1} on the dataset sizes, Figure 5.4a shows Ac_{t1} for B1 using different percentages pd of the datasets $D_{gNB_{ref}}$. It can be seen that with a fixed gNB, using a higher percentage of the dataset pd improves the prediction accuracy Ac_{t1} , as expected. However, at one point the accuracy improvement saturates, as can be seen e.g. for gNB 15. Reasons for the variations in the accuracy improvement between gNBs can be found in the varying total sizes of $D_{gNB_{ref}}$, as well as the varying portions of LOS and NLOS, shown in Figure 3.8. In LOS, there exists a direct relation between the relative UE position S3 and the best BPL l_{k^*} , which is not the case for NLOS. Even though their dataset sizes are comparatively small, gNB 1 and gNB 2 show high Ac_{t1} from $pd = 40\%$ as they only include LOS positions. The narrow paths covered by gNB 3, shown in Figure 5.3, result in a high portion of NLOS and consequently lower Ac_{t1} . Comparing gNB 15 and gNB 14, gNB 15 reaches higher Ac_{t1} at $pd = 40\%$ as the portion of LOS positions is

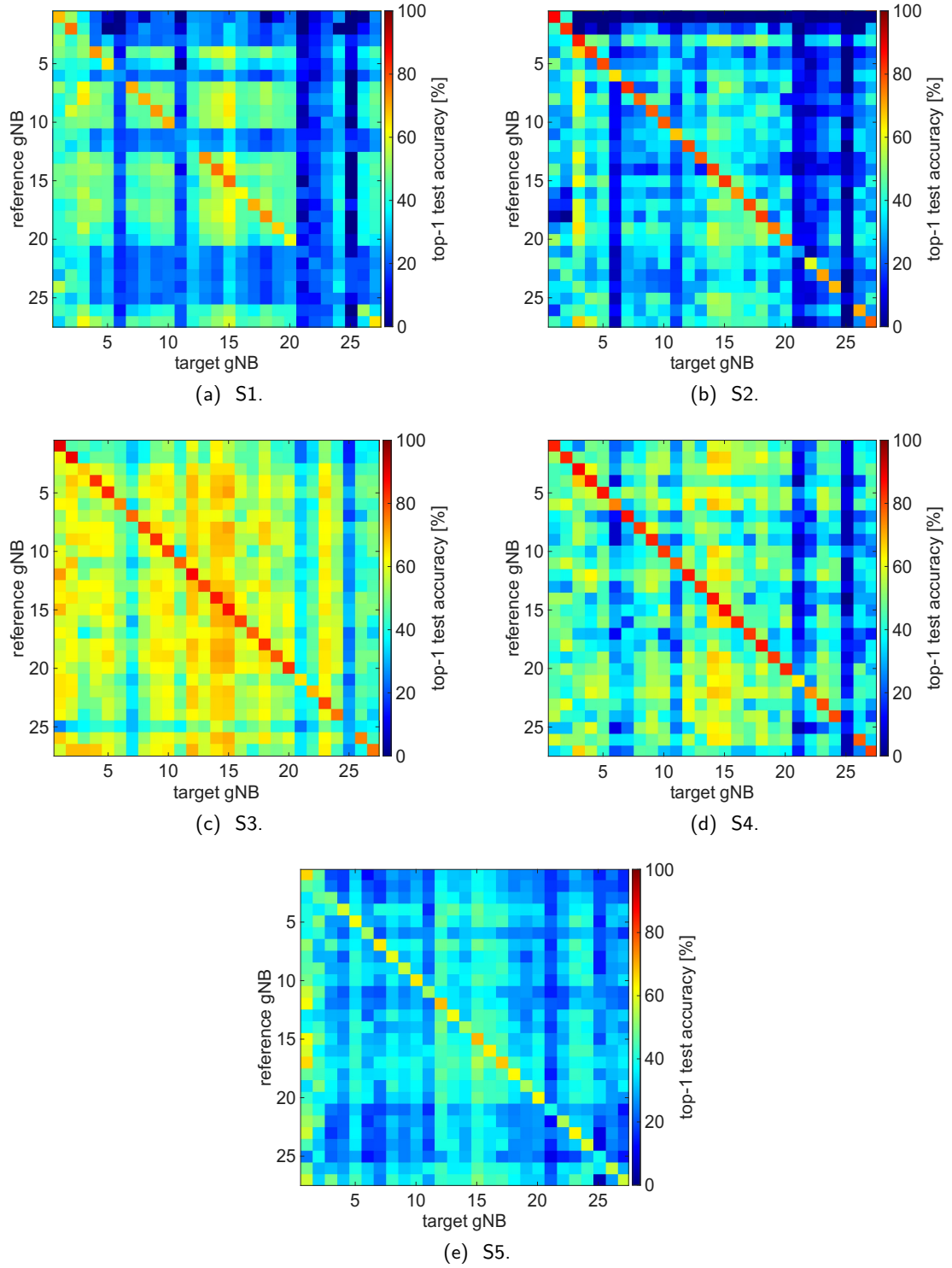


Figure 5.2: Top-1 accuracies Ac_{t1} for all combinations of gNB_{ref} and gNB_{tar} , according to B2, with the diagonal $gNB_{tar} = gNB_{ref}$ corresponding to B1, using the UE position preprocessing modalities S1-S5, with $\gamma = 0.3$, $l_{ri} = 0.02$ for preprocessing modalities S2 and S4, and $l_{ri} = 0.002$ for preprocessing modalities S1, S3 and S5.

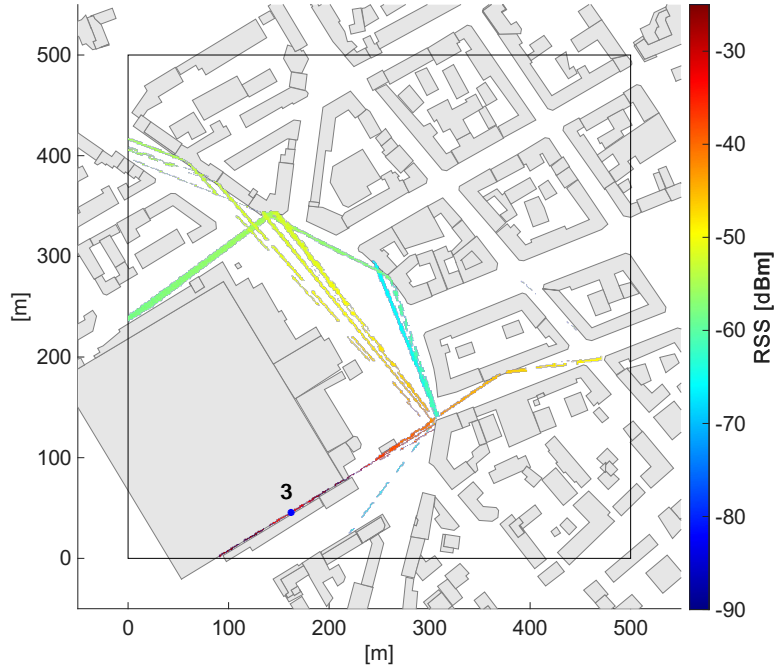


Figure 5.3: Coverage map of gNB_3 , representing the maximum RSS at each UE position as a colour map, showing a narrow propagation path, resulting in a low number of distinctive optimal BPLs in the dataset.

higher. For that gNB the improvement in Ac_{t1} between $pd = 5\%$ and $pd = 100\%$ is up to 50 pp.

We now compare Ac_{t1} B1 with $pd = 5\%$ and B2, therefore we compare the accuracy obtained by using a small dataset to train a new model from scratch, and the accuracy obtained by using the same limited amount of data to fine-tune a model trained elsewhere. The comparison is aimed at highlighting the benefits of transfer learning. Figure 5.5 shows the difference between Ac_{t1} of B1 with $pd = 5\%$, shown in Figure 5.4a, and B2, given in Figure 5.2c. For each combination of gNB_{ref} and gNB_{tar} in B2, the difference in Ac_{t1} to the corresponding $gNB^{B1} = gNB_{tar}$ is given in form of an accuracy difference colour map. Values $Ac_{t1} > 0$ represent a performance advantage of the transfer learning in B2. The diagonal $gNB_{ref} = gNB_{tar}$ show the difference in Ac_{t1} between B1 with $pd = 5\%$ and B1 with $pd = 100\%$, with positive values representing higher values for the latter. The data points shown in Figure 5.5 are calculated based on the mean of 5 repeated simulations.

The accuracy difference colour map in Figure 5.5 shows that the choice of a suitable gNB_{ref} is essential to extract an advantage from employing transfer learning. While most combinations can reach a significant improvement, e.g. $gNB_{ref} = 14$ and $gNB_{tar} = 10$ improves the Ac_{t1} by approximately 15 pp compared to B1 with $gNB^{B1} = 10$, some combinations perform worse with transfer learning in B2 than in B1. Most of the combinations which show a disadvantage can be explained by a small size of $D_{gNB_{ref}}$. The average difference using $gNB_{ref} = 14$ over all gNB_{tar} excluding the combination $gNB_{tar} = gNB_{ref}$ is approximately 14 pp, while $gNB_{ref} = 21$ only reaches approximately 8 pp on average compared to B1, following the differing sizes of $D_{gNB_{14}}$ and $D_{gNB_{21}}$ shown in Figure 3.8.

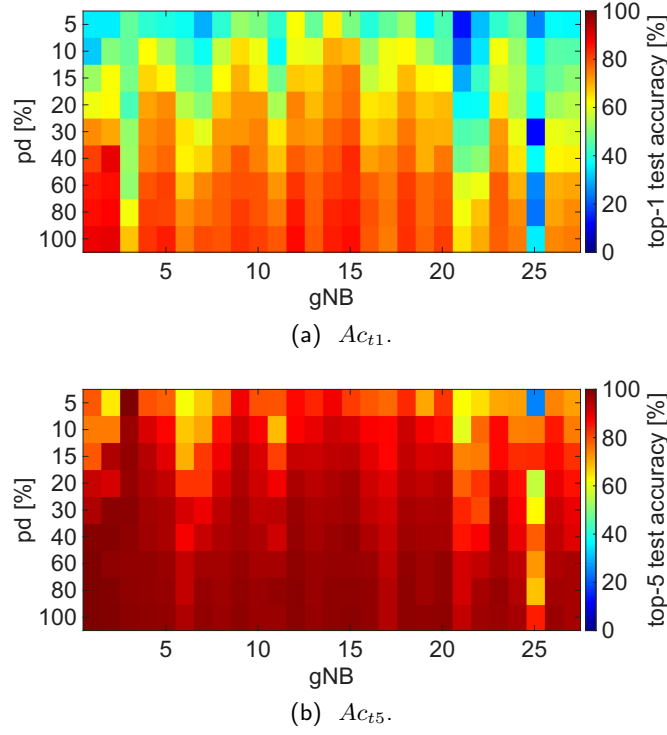


Figure 5.4: Colour map of prediction accuracies (a) Ac_{t1} and (b) Ac_{t5} for B1, using different percentages pd of the datasets D_{gNB} .

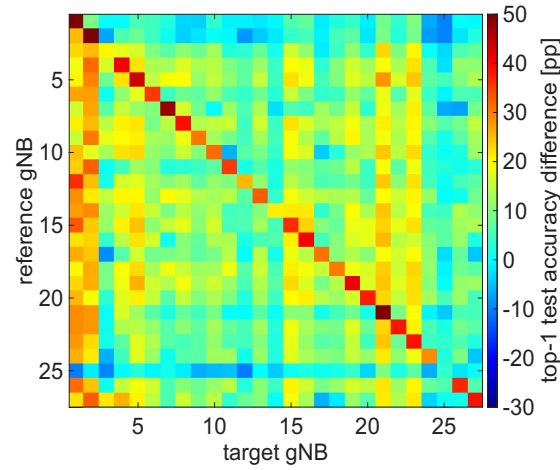


Figure 5.5: Ac_{t1} accuracy difference colour map, giving the difference between B1 with $pd = 5\%$ and B2 in pp, with corresponding $gNB^{B1} = gNB_{tar}$ for each combination of gNB_{ref} and gNB_{tar} in B2. $Ac_{t1} > 0$ represents $Ac_{t1}^{B2} > Ac_{t1}^{B1}$. Values on the diagonal $gNB_{ref} = gNB_{tar}$ show the difference in Ac_{t1} between B1 with $pd = 5\%$ and B1 with $pd = 100\%$, with positive values representing higher values for the latter.

However, $gNB_{ref} = 25$ is the only reference gNB which has a negative mean difference across all target gNBs, with a value of -1 pp. Focussing now on the effect of the target gNBs, the Ac_{t1} values at $gNB_{tar} = 7$, $gNB_{tar} = 12$ and $gNB_{tar} = 14$ show comparatively low improvement. The limited improvements at $gNB_{tar} = 7$ can be explained by the location of gNB 7, which is very close to the left edge of the study area, shown in Figure 3.3, therefore it might present unique propagation characteristics that are hard to derive from other areas. The low Ac_{t1} improvement at $gNB_{tar} = 12$ and $gNB_{tar} = 14$ can be attributed to the good B1 performance with $pd = 5\%$. From the dataset sizes, NLOS portions and coverage maps, we cannot make out a reason for the high Ac_{t1} values in B1 of gNB 14. The dataset attributes of gNB 14 are very similar to gNB 18, and both gNBs are placed on the same side of the large free area in the study area. The reason for the good B1 performance of $gNB_{tar} = 14$ with $pd = 5\%$ might be a favourable random selection of the used $pd = 5\%$ of $D_{gNB=14}$, as B1 in Figure 5.4a are based on one simulation result. The target gNBs $gNB_{tar} = 1$ and $gNB_{tar} = 2$ show the highest improvements with a mean improvement over all reference gNBs of approximately 24 pp. The high portion of LOS positions leads to good results in fine-tuning, while the low dataset size leads to low Ac_{t1} values for B1 with $pd = 5\%$. The mean improvement for Ac_{t1} over all combinations $gNB_{ref} \neq gNB_{tar}$ in B2 is approximately 11 pp.

It has to be noted that the Ac_{t1} results show considerable variance over repeated simulations when gNBs with small dataset sizes are included, e.g. for the case $gNB_{ref} = gNB_{tar} = 3$ in Figure 5.5, corresponding to B1 with $pd = 100\%$, the five Ac_{t1} values obtained from the five repeated simulations range from 28% to 56%, and in B2 for the combination $gNB_{ref} = 17$ and $gNB_{tar} = 3$, Ac_{t1} ranges from 32% to 45%. In contrast, the combination $gNB_{ref} = 17$ and $gNB_{tar} = 14$ only varies by 2 pp over the 5 repeated simulations.

Due to the long training times of the CNN-S, we investigate the impact of the dataset size only for the selected gNB 14. Figure 5.6a depicts the Ac_{t1} of case B1 at gNB 14 of the FCNN-S in blue and the CNN-S in red. Both models follow a similar trend of improving Ac_{t1} with increasing pd , with the CNN-S comparatively performing better, especially at lower pd . Figure 5.7a depicts the B2 Ac_{t1} results with fixed $gNB_{ref} = 14$ and variable gNB_{tar} spanning the x-axis, for the CNN-S in blue and the FCNN-S in red. The CNN-S has less fluctuation in the Ac_{t1} over different gNB_{tar} , even though the results are not calculated as a mean of five simulations. An outlier compared to the FCNN-S lies in $gNB_{tar} = 3$, where the Ac_{t1} of the CNN-S is comparatively low. Since the dataset for gNB 3 is largely NLOS, this might be an indication that the CNN-S has a disadvantage in NLOS conditions. The yellow and purple bars at $gNB_{tar} = 14 = gNB_{ref}$ represent B1 with $pd = 100\%$ of the CNN and FCNN, respectively. The obtained Ac_{t1} values are in a similar range for the CNN-S and FCNN-S.

We now focus on analysing the top-5 accuracy results, i.e. Ac_{t5} . Figure 5.4b shows the Ac_{t5} for B1 using different percentages pd of the datasets $D_{gNB_{ref}}$. As expected, compared to the Ac_{t1} in Figure 5.4a the Ac_{t5} values are significantly higher, reaching values up to 99% for $pd = 100\%$. Most effects seen for the top-1 accuracy can be observed in the Ac_{t5} values as well. An exception lies in gNB 3, which is amongst the gNB with the lowest Ac_{t1} but shows high Ac_{t5} values. As previously discussed gNB 3 has a large portion of NLOS, but its coverage consists of narrow paths, resulting in a low number of best BPLs l_{k^*} , as depicted in Figure 5.3.

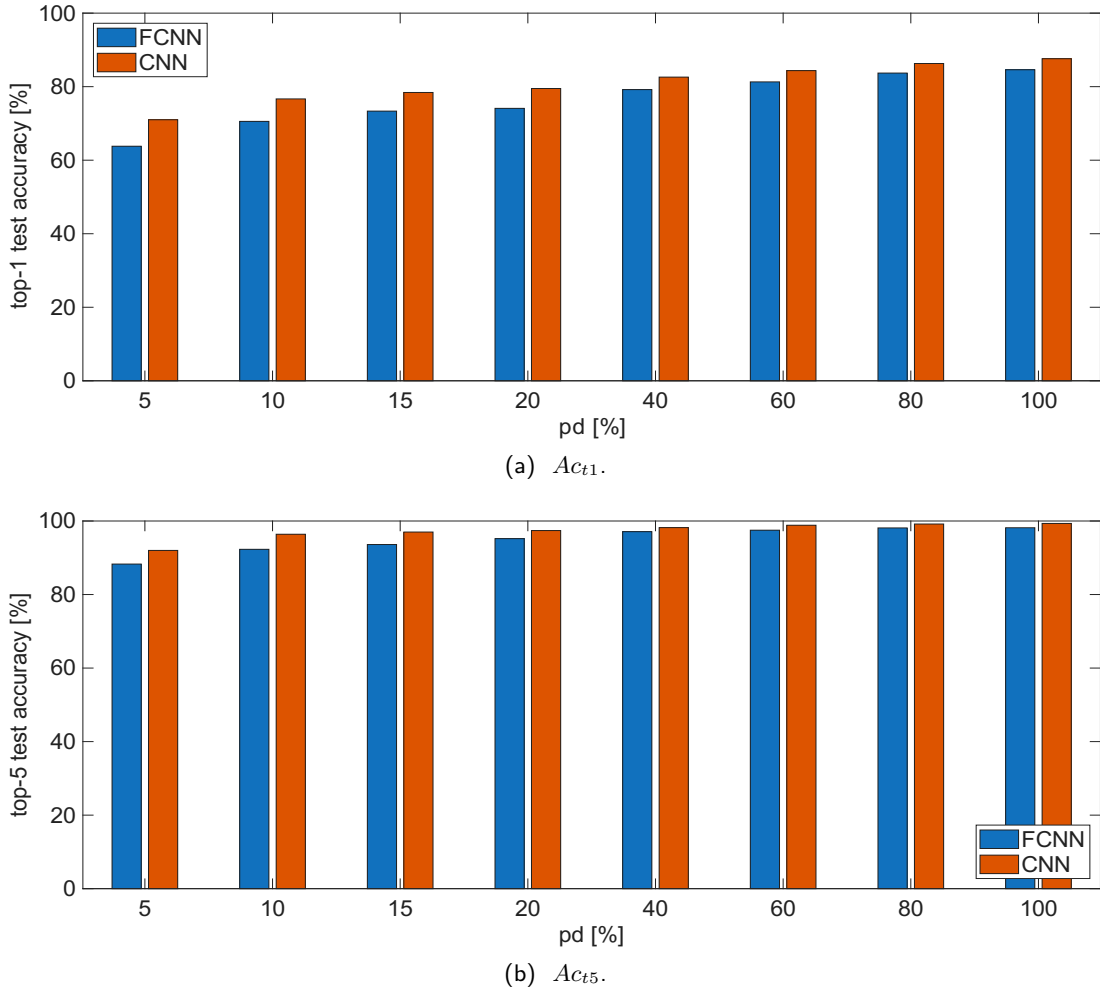


Figure 5.6: B1 prediction accuracies (a) Ac_{t1} and (b) Ac_{t5} at gNB 14 of the FCNN in blue and the CNN in red, over different percentages pd of the datasets $D_{gNB=14}$.

The accuracy colour map for Ac_{t5} of the FCNN-S in B2 setup with the diagonal $gNB_{ref} = gNB_{tar}$ representing B1 with $pd = 100\%$, corresponding to the Ac_{t1} in Figure 5.2c, is given in Figure 5.8. The improvement in the Ac_{t5} values observed in B1 is also present in B2. As an example, the mean B2 Ac_{t5} of $gNB_{ref} = 14$ over all gNB_{ref} is approximately 86 %, compared to Ac_{t1} being approximately 55 %. Similar to Ac_{t1} in Figure 5.2c, the low values for $gNB_{tar} = 21$, $gNB_{tar} = 25$ and $gNB_{ref} = 25$ might be explained by the low dataset sizes, and the low values for $gNB_{tar} = 7$ by its location at the edge of the study area. The high Ac_{t5} at gNB 3, observed in Figure 5.4b for B1, transfers to $gNB_{tar} = 3$ in B2. The mean accuracy over all combinations in B2 of Ac_{t5} is approximately 84 %, compared to 53 % for Ac_{t1} .

Figure 5.9 depicts the improvement of Ac_{t5} from B1 with $pd = 5\%$ from Figure 5.4b to the B2 results shown in Figure 5.8 as an accuracy difference colour map. As for Ac_{t1} , reference gNBs 1,2 and 25 show the lowest increase, $gNB_{ref} = 25$ being the only reference gNB with a negative mean improvement over all target gNBs, with approximately -2 pp. As previously discussed, the Ac_{t5} of gNB 3 in B1 with $pd = 5\%$ is already high due to

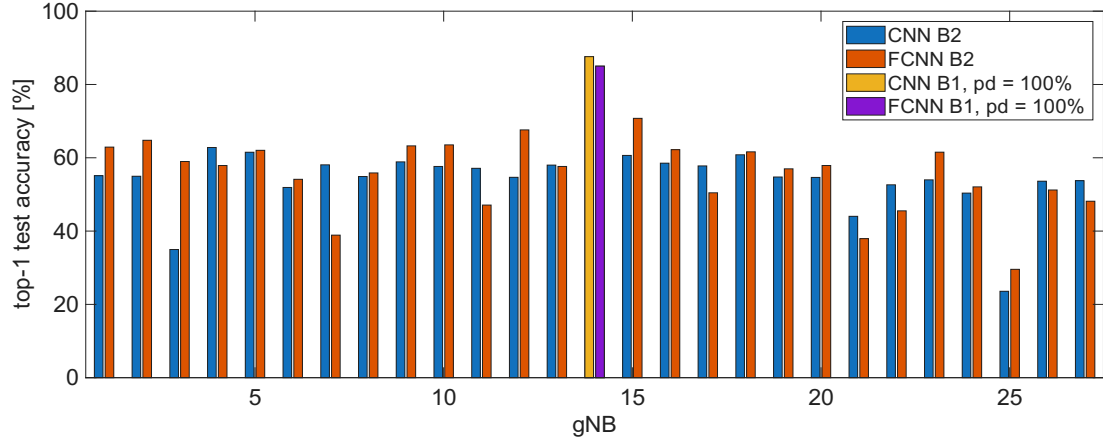
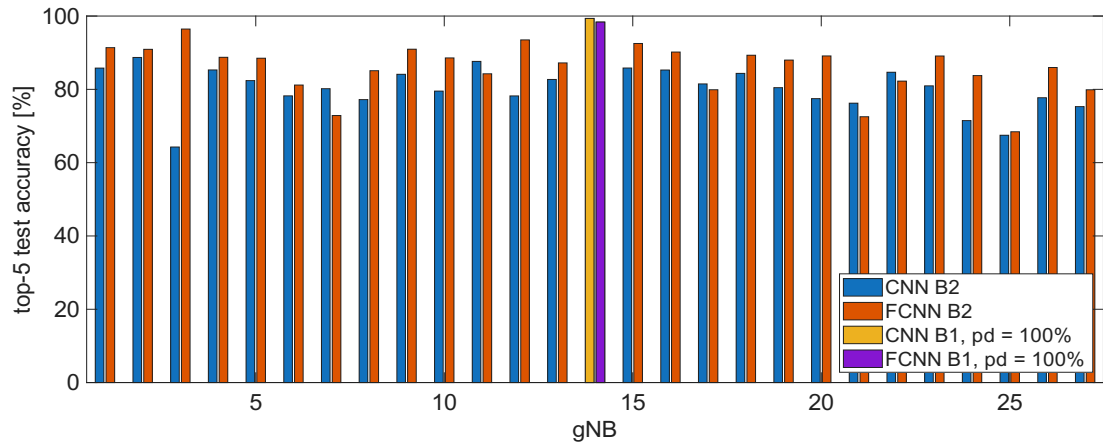
(a) Ac_{t1} .(b) Ac_{t5} .

Figure 5.7: B2 (a) Ac_{t1} and (b) Ac_{t5} with $gNB_{ref} = 14$ and gNB_{tar} spanning the x-axis. The blue bars are the accuracy results of the CNN, and the red bars depict the FCNN results. The yellow and purple bars at $gNB_{tar} = 14 = gNB_{ref}$ represent B1 with $pd = 100\%$ of the CNN and FCNN, respectively.

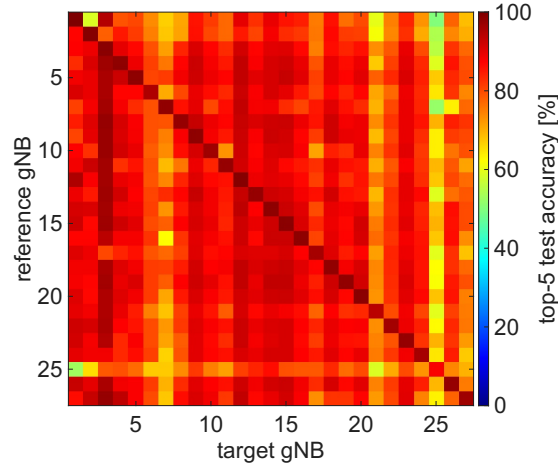


Figure 5.8: Top-5 accuracy (i.e. Ac_{t5}) of FCNN-S for all combinations of gNB_{ref} and gNB_{tar} , according to B2, with the diagonal $gNB_{tar} = gNB_{ref}$ corresponding to B1

its limited number of best BPLs in the covered area. This special case leads to a negative improvement of Ac_{t5} of approximately 6 pp over all reference gNBs for target gNB 3. The low value of Ac_{t5} for B1 with $pd = 5\%$ in Figure 5.4b leads to a high improvement of B2 Ac_{t5} for $gNB_{tar} = 25$. We observe varying B1 results on B1 for gNB 25, with Ac_{t5} for $pd = 20\%$ being significantly lower than for $pd = 15\%$, with 56 % and 83 %, respectively, therefore the low $pd = 5\%$ results are likely due to a disadvantageous selection of data. In comparison, the 5 repeated simulations in B2 for $gNB_{ref} = 14$ and $gNB_{tar} = 25$ only vary by 10 pp between 64 % and 74 %. This indicates that the selection of the data for the fine-tuning in B2 has a lower impact than in B1, especially for small dataset sizes. The mean improvement for Ac_{t5} over all combinations $gNB_{ref} \neq gNB_{tar}$ in B2 is approximately 9 pp.

Figure 5.6b depicts Ac_{t5} of case B1 at gNB 14 for the FCNN-S in blue and the CNN-S in red. Similar to the Ac_{t1} results, the CNN performs slightly better, the difference is more pronounced at lower pd . The Ac_{t5} results of the CNN-S for B2 are depicted in Figure 5.7 in blue, with the FCNN-S Ac_{t5} in red as a comparison. The trends observed for the Ac_{t1} results can be observed for Ac_{t5} as well, with the CNN-S values at $gNB_{tar} = 3$ being significantly lower than the FCNN-S values.

For both accuracies Ac_{t1} and Ac_{t5} of FCNN-S, the standard deviation of the mean accuracy per gNB_{tar} across all gNB_{ref} , referred to as Std_{tar}^{Ac} , is higher than the standard deviation of the mean accuracy per gNB_{ref} across all gNB_{tar} , referred to as Std_{ref}^{Ac} . We can also observe this in the accuracy colour maps shown in Figure 5.2c and 5.8 in form of clearer contours between the columns compared to the contours of the rows. For the top-1 accuracy, they are approximately $Std_{tar}^{Ac_{t1}} = 9.8$ pp and $Std_{ref}^{Ac_{t1}} = 4.2$ pp, for the top-5 accuracy, they are approximately $Std_{tar}^{Ac_{t5}} = 7.0$ pp and $Std_{ref}^{Ac_{t5}} = 2.9$ pp. Therefore, the impact of the target gNB on the prediction accuracy is higher than the impact of the reference gNB. The standard deviations Std_{tar}^{Ac} and Std_{ref}^{Ac} are lower for the top-5 accuracy, therefore the characteristics of both reference and target gNB have a lower impact on the top-5 accuracies than on the top-1 accuracies.

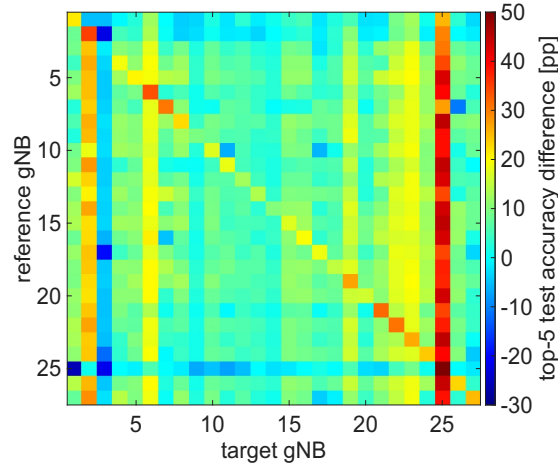


Figure 5.9: Top-5 accuracy difference colour map of FCNN-S between B1 with $pd = 5\%$ and B2, with corresponding $gNB^{B1} = gNB_{tar}$. $Ac_{t5} > 0$ represents $Ac_{t5}^{B2} > Ac_{t5}^{B1}$.

5.1.3 Neural Network Size

By modifying the number and size of hidden layers of the FCNN-S, we can improve Ac_{t1} . The accuracy colour map for Ac_{t1} of FCNN-L is given in Figure 5.10a. The previous observations for the FCNN-S hold, with Ac_{t1} being still dependent on the dataset size and on the portion of NLOS positions. However, FCNN-L shows higher Ac_{t1} at most combinations. The difference between $Std_{tar}^{Ac_{t1}} = 11.5$ pp and $Std_{ref}^{Ac_{t1}} = 3.6$ pp is higher in FCNN-L than in FCNN-S, therefore the influence of the target gNB compared to the reference gNB on Ac_{t1} increases compared to FCNN-S. Figure 5.11a depicts the difference in Ac_{t1} from the FCNN-S results shown in Figure 5.2c and the FCNN-L results from Figure 5.10a, with positive values representing higher Ac_{t1} of FCNN-L. The larger NN is able to handle the complexity stemming from the placement of gNB 7 at the edge of the study area. The highest gain in Ac_{t1} is achieved for $gNB_{tar} = 7$, with a mean improvement of approximately 24 pp across all gNB_{ref} . The lowest improvement is achieved at $gNB_{tar} = 25$. Target gNBs with a small dataset sizes show a low improvement, with $gNB_{tar} = 2$, $gNB_{tar} = 6$, $gNB_{tar} = 21$ and $gNB_{tar} = 23$ having a mean improvement of 2 pp to 3 pp across all gNB_{ref} , while the Ac_{t1} value of $gNB_{tar} = 21$ across all gNB_{ref} improves by approximately 11 pp. The mean improvement over all combinations gNB_{ref} and gNB_{tar} in B2 is approximately 10 pp.

The accuracy colour map for Ac_{t5} of FCNN-L is given in Figure 5.10b. It shows a similar distribution of low Ac_{t5} as observed in the Ac_{t1} values, with $gNB_{tar} = 3$ showing a high improvement from Ac_{t1} to Ac_{t5} as observed for the FCNN-S in Figure 5.8. The Ac_{t5} FCNN-L improvement as an accuracy difference colour map is given in Figure 5.11b, with positive values denoting higher accuracies for FCNN-L. The overall improvement, with a mean of approximately 4 pp across all gNB_{ref} and gNB_{tar} B2 combinations, is lower than for Ac_{t1} . This can be explained in part by the higher overall Ac_{t5} values, which leave less room for improvement. However, while the mean for a given gNB_{ref} across all gNB_{tar} is positive for all gNB_{ref} , the mean for $gNB_{tar} = 2$, $gNB_{tar} = 6$, $gNB_{tar} = 21$, $gNB_{tar} = 23$ and $gNB_{tar} = 25$ across all gNB_{ref} in this case is negative up to -2 pp. Therefore, the larger FCNN-L does not achieve a consistent improvement in Ac_{t5} for low

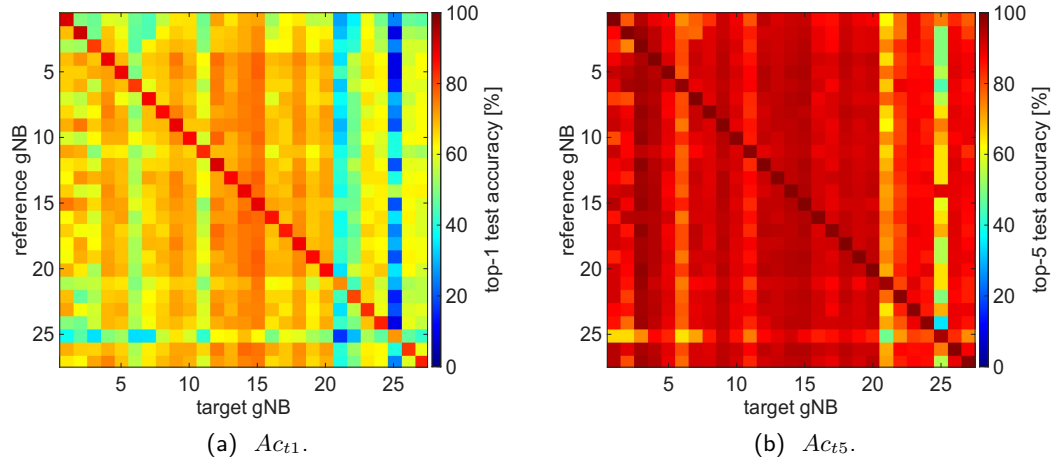


Figure 5.10: FCNN-L top-1 accuracies Ac_{t1} (a) and top5 accuracies Ac_{t5} (b) for all combinations of gNB_{ref} and gNB_{tar} , according to B2, with the diagonal $gNB_{tar} = gNB_{ref}$ corresponding to B1.

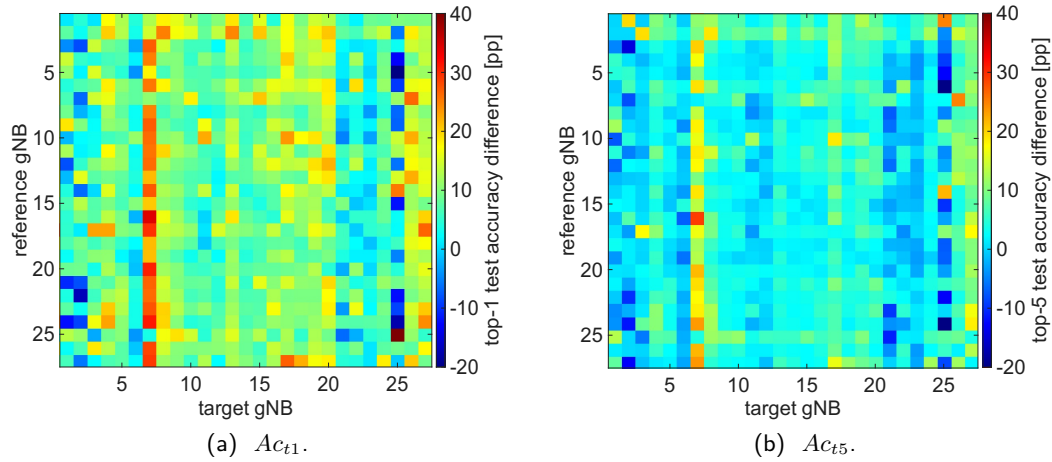


Figure 5.11: Ac_{t1} and Ac_{t5} difference colour maps between the FCNN-S results given in Figure 5.2c and Figure 5.8, and the FCNN-L results in Figure 5.10. Positive values represent higher accuracies of FCNN-L.

dataset size gNB_{tar} , a possible reason being overfitting promoted by a larger NN size. For both Ac_{t1} and Ac_{t5} , the improvement achieved by the FCNN-L for the cases B2 and B1 with $pd = 100\%$ is on a similar level, as the diagonal $gNB_{ref} = gNB_{tar}$ does not stand out in the accuracy difference colour maps.

In the top-1 and top-5 accuracy difference between B2 and B1 with $pd = 5\%$, shown in the accuracy difference colour maps in Figure 5.12a and 5.12b, respectively, we can observe similar gNB_{tar} that improve thanks to transfer learning, with top-5 improvements being considerably lower, because the B1 Ac_{t5} results already reach values into the 90% range, limiting the possible improvement. For both Ac_{t1} and Ac_{t5} using $gNB_{ref} = 25$ in B2 reduces the accuracy compared to B1 with $pd = 5\%$. The difference in improvement

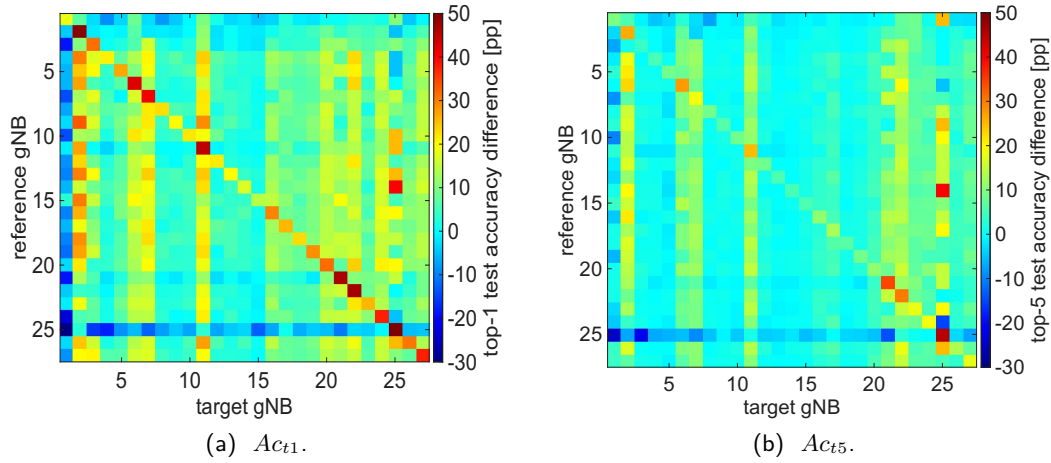


Figure 5.12: Ac_{t1} and Ac_{t5} difference colour maps of FCNN-L between B1 with $pd = 5\%$ and B2, with corresponding $gNB^{B1} = gNB_{tar}$. Positive values represent $Ac_{tx}^{B2} > Ac_{tx}^{B1}$.

between $gNB_{tar} = 1$ and $gNB_{tar} = 2$ stands out in Ac_{t1} and Ac_{t5} . Even though they show very similar characteristics in dataset size and environment, as shown in Figure 3.8 and Figure 3.3, respectively, the B1 $pd = 5\%$ mean accuracy results over 5 simulations show a top-1 accuracy of $Ac_{t1} = 73\%$ at gNB 1, compared to $Ac_{t1} = 43\%$ at gNB 2. The high B1 results at gNB 1 make $gNB_{tar} = 1$ the only target gNB with a negative mean improvement over all gNB_{ref} .

Even though the offline training nature of the investigated models, and the possibility to build BIMs to perform offline inference, reduces the relevance, we want to briefly comment on the training and inference times of the investigated models. Table 5.1 gives the number of tunable parameters of the models FCNN-S, FCNN-L and CNN-S, as well as the median training and inference time of five simulations for case B1 at gNB 14 with $pd = 100\%$, resulting in 47,821 positions used in the training and 11956 used in the testing. The higher number of tunable parameters in FCNN-L and CNN-S compared to FCNN-S leads to longer training and inference times with a factor in the same order of magnitude. The impact on the training time is higher than the impact on the inference time. The training time of CNN-S is particularly large because in addition to a higher number of tunable parameters, the transformation of the two dimensional UE position $UE(x, y)$ to the one-hot position map used as input data for the CNN-S is done during the training process, as otherwise the large sizes of the position maps would introduce memory constraints. In practice, the times are dependent on the employed hardware, however the inference times per position, especially for FCNN-S and FCNN-L, indicate that online inference would be viable. E.g. the inference times per position are low compared to the default SS burst periodicity of 20 ms in 5G NR. However, given that the largest covered area in the investigated dataset contains fewer than 65,000 UE positions, compared to 165,504 tunable parameters of FCNN-S, we suggest the usage of BIMs in combination with the investigated models.

Properties	FCNN-S	FCNN-L	CNN-S
# tunable parameters	165,504	1,750,664	26,876,513
ratio of tunable parameters with respect to FCNN-S	1	10.6	162.4
training time [s]	24.5	186	6737
training time per UE position [s]	5.1×10^{-4}	3.9×10^{-3}	0.141
ratio of training time with respect to FCNN-S	1	7.6	275.0
inference time [s]	0.158	0.703	12.9
inference time per position [s]	1.3×10^{-5}	5.9×10^{-5}	1.1×10^{-3}
inference time compared to FCNN-S	1	4.4	81.6

Table 5.1: Number of tunable parameters and median training- and inference-time of five simulations of FCNN-S, FCNN-L and CNN-S for case B1 at gNB 14 with $pd = 100\%$, resulting in 47,821 positions used in the training and 11956 positions used in the testing.

5.1.4 Noise in the UE Position

To investigate the impact of noisy UE position data, the Ac_{t1} prediction accuracy results of the FCNN-S model with the noise scenarios N1-N3 are given in Figure 5.13, with the accuracy colour maps of the scenarios shown on the left and the accuracy difference colour maps to the baseline FCNN-S results shown in Figure 5.2c on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3. Comparing N1-N3, we observe very similar behaviours regarding accuracy differences between target gNBs, however the mean loss in N1 over all B2 combinations is the lowest with 12.0 pp compared to 12.9 pp for and 13.4 pp for N3. For N1, this results in a mean top-1 accuracy over all B2 combinations of 40.7%. The target gNBs $gNB_{tar} = 6$ and $gNB_{tar} = 23$ show particularly high losses, with a mean over all gNB_{ref} of 20.9 pp and 21.8 pp in N1, respectively. As shown in Figure 3.8, the dataset sizes and NLOS portions of these gNBs are very similar to each other. The target gNB $gNB_{tar} = 3$ shows the lowest mean loss over all gNB_{ref} with 1.3 pp in N1. As previously discussed and shown in Figure 5.3, the coverage of gNB 3 consists of narrow paths with few distinct optimal BPLs l_{k^*} , leading to a lower impact of noise in the UE position. Compared to the standard deviation of the mean accuracy per gNB_{tar} across all gNB_{ref} without noise $Std_{tar}^{Ac_{t1}} = 9.8$ pp, the standard deviation in N1 is only $Std_{tar}^{Ac_{t1}} = 8.9$ pp, therefore there are lower accuracy differences between different target gNBs. Additionally, the accuracy difference between B1 with $pd = 100\%$ on the diagonal $gNB_{ref} = gNB_{tar}$ to B2 in the accuracy colour maps in Figure 5.13 is significantly less pronounced than without noise shown in Figure 5.2c, therefore in presence of UE position noise, the accuracy difference between using a small target dataset to fine-tune a pre-existing model and training a new model with a large dataset is significantly lower than without UE position noise.

The corresponding Ac_{t5} prediction accuracy results are given in Figure 5.14. As for the Ac_{t1} results, the cases N1-N3 show a similar structure, with N1 having the lowest mean loss over all combinations in B2 of 7.3 pp, compared to 8.9 pp and 8.6 pp for N2 and N3, respectively. For N1, this results in a mean top-5 accuracy over all B2 combinations of 76.7%. The Ac_{t5} loss is lower than the Ac_{t1} loss, indicating a reduced impact of UE

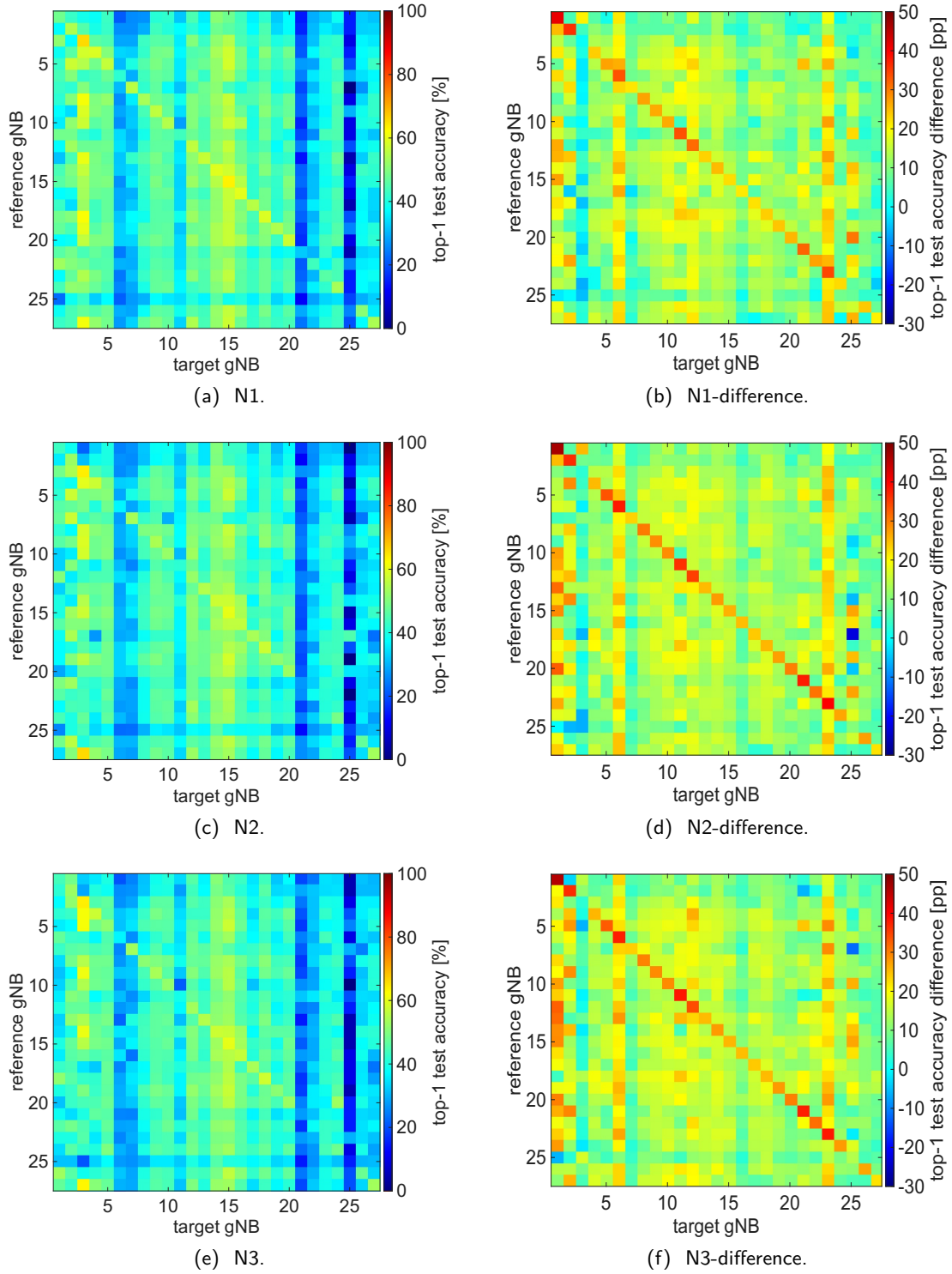


Figure 5.13: Top-1 Ac_{t1} accuracy colour maps with noise N1-N3 of the FCNN-S on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-S, shown in Figure 5.2c, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3.

position noise on the top-5 accuracies. Compared to $Std_{tar}^{Ac_{t5}} = 7.0$ pp without noise, the standard deviation in N1 is $Std_{tar}^{Ac_{t5}} = 10.3$ pp, therefore, in contrast to the Ac_{t1} results, the target gNB has a higher impact on the top-5 accuracy with noise in the UE position. The top-5 accuracy difference between B1 with $pd = 100\%$ on the diagonal $gNB_{ref} = gNB_{tar}$ to B2 in the accuracy colour maps in Figure 5.14 is significantly different between N1 and N2,N3. In N1, which includes noise in the training data, B1 shows a significant Ac_{t5} advantage, while for N2 and N3 the diagonal is hardly cognizable, further indicating a general advantage of considering noise in the training process.

Figure 5.15 depicts the Ac_{t1} of FCNN-L under noisy scenarios N1-N3. Interestingly, the mean losses over all combinations in B2, shown in the right column of Figure 5.15, are higher than in FCNN-S and show less difference between N1-N3. They result in 17.2 pp for N1, 17.3 pp for N2, and 17.7 pp for N3, amounting to an average Ac_{t1} of FCNN-L in B2 with N1 of 45.4 % compared to 40.7 % of FCNN-S. A possible explanation for the comparatively low accuracies of FCNN-L compared to the scenario without noise is that the effect of possible overfitting of the FCNN-L on the position data is increased by noise. Another factor might be the existence of a natural limit in achievable accuracy in the presence of noise. As was observed for FCNN-L without noise in Figure 5.10a, the accuracy for $gNB_{tar} = 7$, with gNB 7 being located at the edge of the study area, is significantly improved compared to FCNN-S. Similar to the FCNN-S, $Std_{tar}^{Ac_{t1}} = 10.9$ pp in N1 is slightly lower than $Std_{tar}^{Ac_{t1}} = 11.5$ pp without noise. The difference between Ac_{t1} on the diagonal representing B1 with $pd = 100\%$ and B2 is only observable for N1, as was observed for the top-5 accuracy with FCNN-S.

Figure 5.16 shows the Ac_{t5} of FCNN-L in the presence of UE position noise N1-N3 corresponding to Figure 5.14 for FCNN-S, with the resulting top-5 accuracies in the left column. As for Ac_{t1} , the diagonal is also clearly cognizable in the N1 Ac_{t5} colour map in Figure 5.16a. In contrast to the previously discussed cases, we observe a discernible difference in the mean losses, shown in the right column of Figure 5.16, between the cases which include noise in the training process N1 and N3, with respect to N2, which features the UE position noise only in the test data. The mean Ac_{t5} losses in B2 of N1 and N3 are 9.5 pp and 9.9 pp, respectively, while N2 shows a mean loss of 11.7 pp. Thus, including noise modulation in the training data offers a notable advantage. Due to the lower B2 losses together with the effects pertaining to the B1 scenario on the diagonal, i.e. $gNB_{ref} = gNB_{tar}$, we can observe higher top-5 accuracies for B2 than for B1 with $pd = 100\%$ uniquely in the N3 Ac_{t5} colour map shown in Figure 5.16e. In this scenario, the fine-tuning data in B2 include noise, while the training data in B1 does not. The mean top-5 accuracy over all the transfer combinations of FCNN-L in N1, shown in Figure 5.16a, results in 78.1 %, compared to 76.7 % in FCNN-S shown in Figure 5.14a. Similar to the FCNN-S, the impact of the target gNB on the top-5 accuracy is higher in N1 than in the case without noise, resulting in $Std_{tar}^{Ac_{t5}} = 12.8$ pp in N1 compared to $Std_{tar}^{Ac_{t5}} = 8.2$ pp in the noise free scenario. In both FCNN-S and FCNN-L, and for both the top-1 and top-5 accuracies, the N1 scenario, which included UE position noise in all datasets, has shown the least losses in B2. Therefore, it is advisable to consider noise in all stages, even if noise-free data is available for training the models.

Figure 5.17 shows a comparison between the top-1 and top-5 accuracy results for B2 with $gNB_{ref} = 14$ of the CNN-S without UE position noise, using a one-hot map for the UE position and G1, in blue and orange, respectively. The corresponding accuracy results for the noisy scenario N1 are given in yellow and purple, respectively. The case

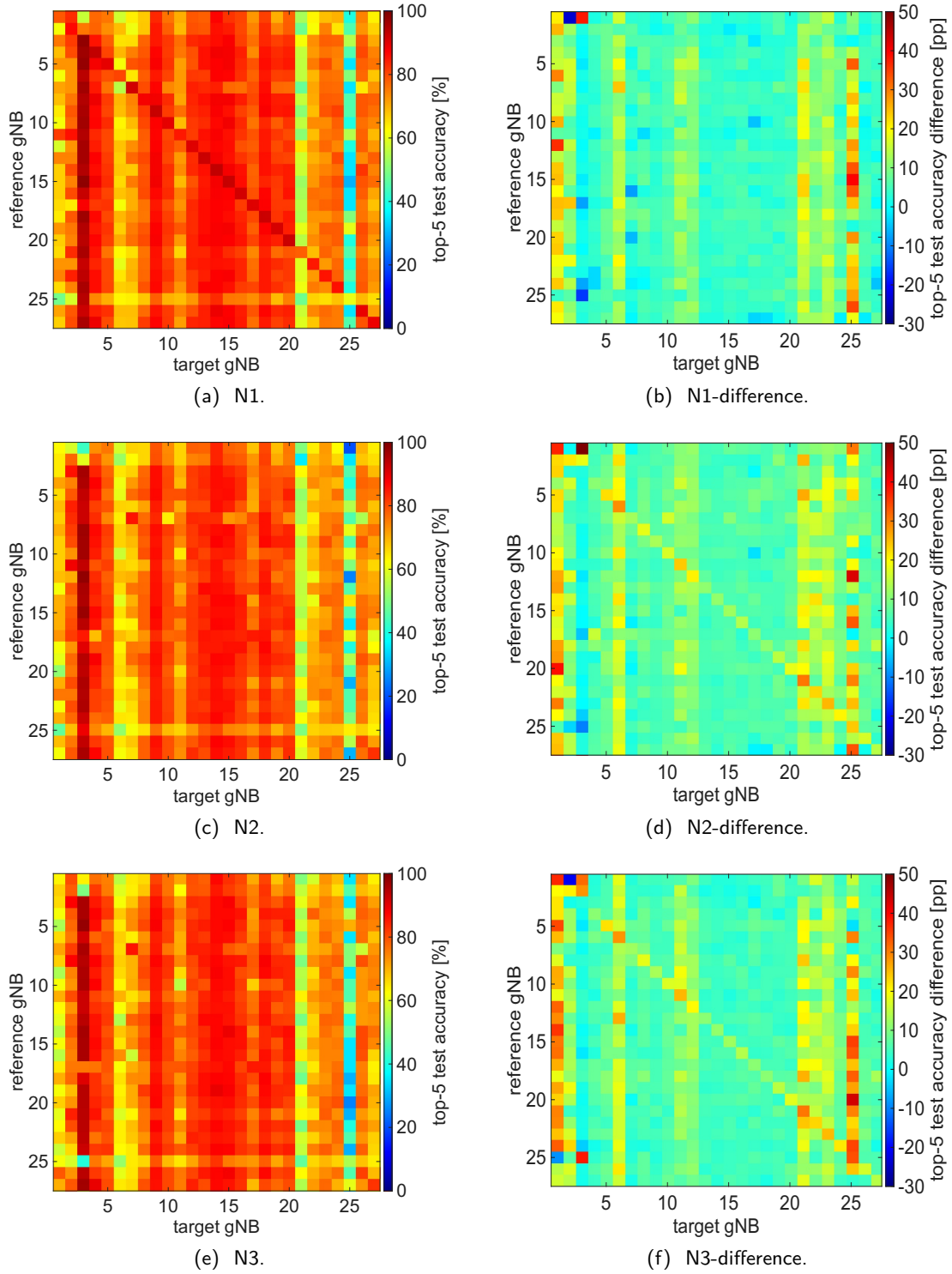


Figure 5.14: Top-5 Ac_{t5} accuracy colour maps with noise N1-N3 of the FCNN-S on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-S, shown in Figure 5.8, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3.

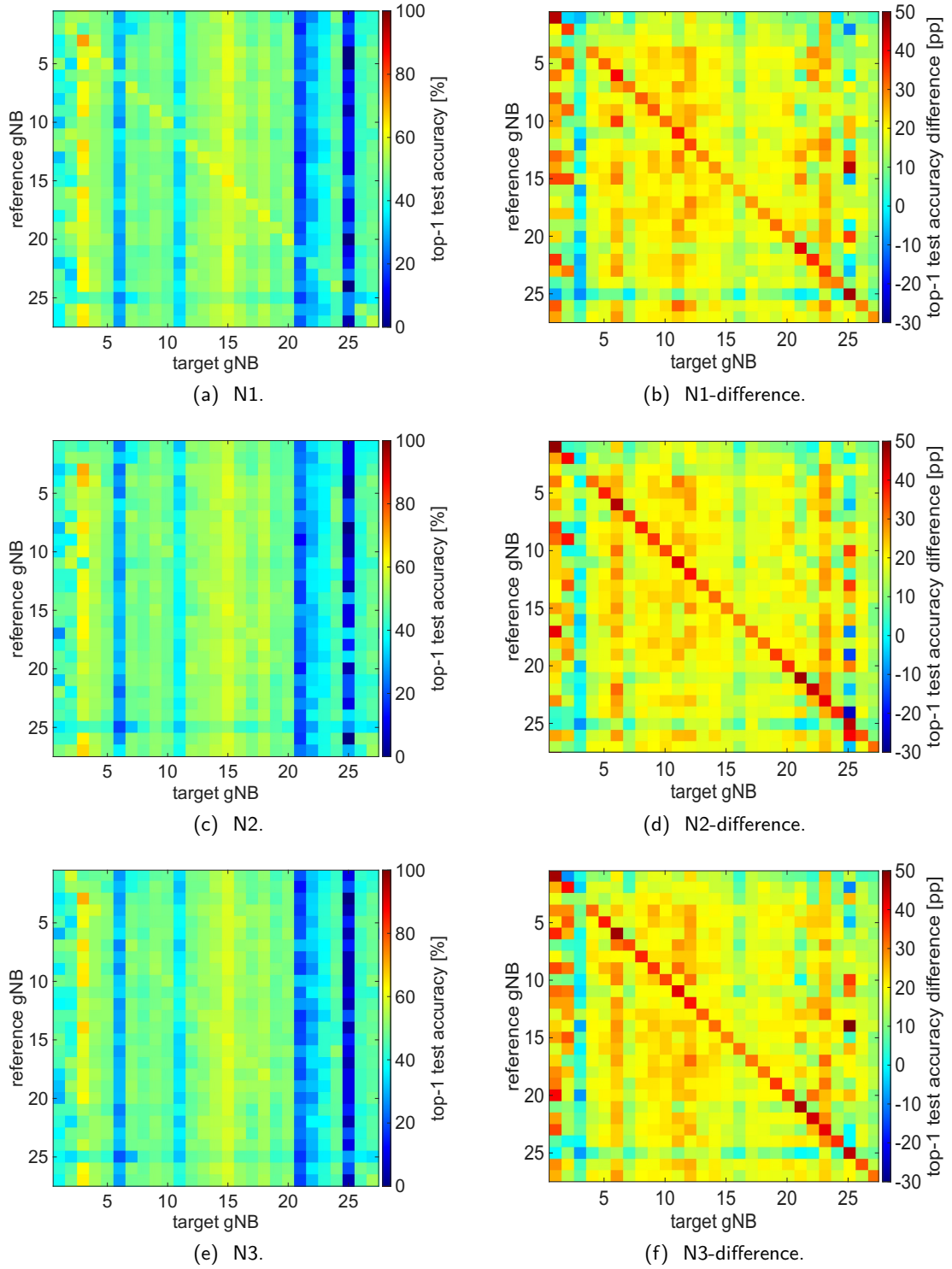


Figure 5.15: Top-1 Ac_{t1} accuracy colour maps with noise N1-N3 of the FCNN-L on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-L, shown in Figure 5.10a, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3.

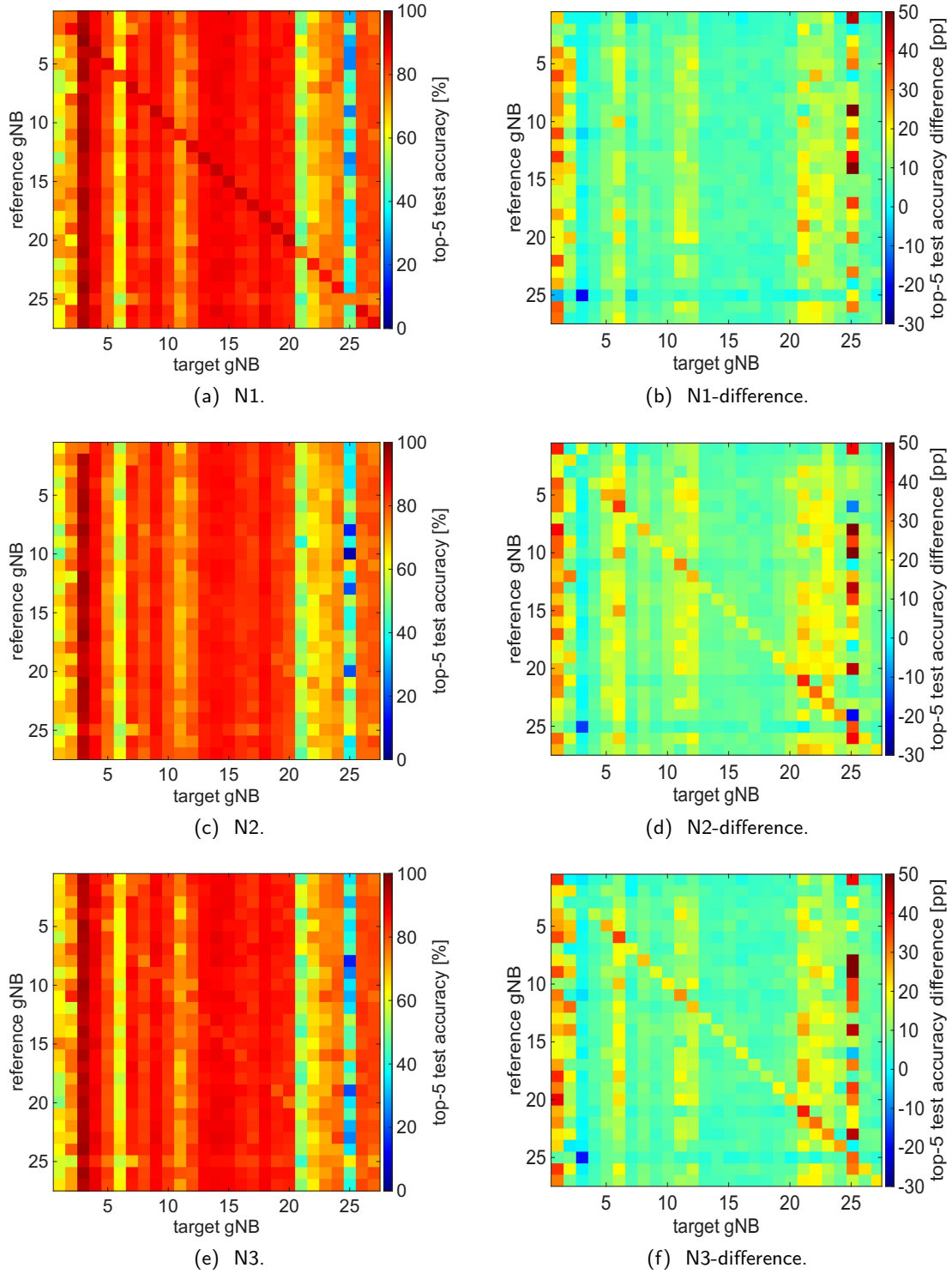


Figure 5.16: Top-5 Ac_{t5} accuracy colour maps with noise N1-N3 of the FCNN-L on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-L, shown in Figure 5.10b, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3.

$gNB_{ref} = gNB_{tar} = 14$ represents B1 with $pd = 100\%$. We observe that employing G1 as an input format has only a very limited impact on the achieved accuracies, with and without the presence of noise. The results stem from one random selection of the used dataset portions, therefore the generalizations of the results should be done with caution. The mean difference across all target gNBs of G1 compared to the one-hot map has an absolute value below 1 pp. Nevertheless, the comparably low accuracies of the G1 input format at $gNB_{tar} = 3$ stands out, as it is present without noise and in N1. Some target gNBs, such as $gNB_{tar} = 25$ which has the smallest dataset size among all gNBs, show an advantage for G1, however other gNBs with small dataset sizes do not confirm that observation as an overall trend. The mean N1 loss of the CNN induced by noisy position data is approximately 18 pp for the top-1 accuracy and 15 pp for the top-5 accuracy, resulting in significantly lower mean accuracies compared to FCNN-S and FCNN-L, with $Ac_{t1} \approx 37\%$ and $Ac_{t5} \approx 66\%$.

5.1.5 Number of Training Labels

We now investigate the impact of the number of training labels on the model's ability to prioritize the selection of the best BPL, while also predicting subsequent top BPLs. The Ac_{tx}^n results of the scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 10$ using FCNN-S are given on the y-axis in Figure 5.18 for $n = 1$ to $n = 5$ best BPL in (a) to (e). The number of labels in every dataset entry line N_{label} is given in different colours. The values on the x-axis denote the x in the top- x accuracy Ac_{tx}^n . The target gNB $gNB_{tar} = 10$ represents a gNB with a large dataset size and a good transfer learning performance in previous investigations. The results are calculated as the mean of five repeated and therefore separately trained FCNN-S models for each N_{label} , and thus five different random selections of the portion of the datasets used for training and testing. As shown in Figure 5.18a, the top- x accuracy for the best BPL l_{k^*} is similar for different N_{label} values, in fact the difference at the top-1 value is approximately 7 pp. While the lowest top-1 value belongs to $N_{label} = 5$, possibly indicating a disadvantage of using more training labels to predict the best BPL, the highest value belongs to $N_{label} = 2$. As there is no continuous trend dependent on the number of labels, the observed differences are most likely due to variations in the accuracy caused by the random selection of the used dataset portions, as the highest variation in the top-1 results among the five trained FCNN-S models is around 7 pp. Similarly, the variation between top- x accuracies at higher x is low in Figure 5.18a, the difference between the top-10 accuracies Ac_{t10}^1 being under 2 pp. Regarding the top-1 accuracies of the second best BPL Ac_{tx}^2 , depicted in Figure 5.18b, we can observe very limited variation between the N_{label} . Regarding the top- x accuracies with $x \geq 2$, we can observe that using $N_{label} \geq 2$ significantly improves the accuracy. The effect is slightly increasing with x , ranging from approximately 19 pp for the top-2 accuracies to 22 pp for the top-10 accuracies. This trend continues for Ac_{tx}^3 to Ac_{tx}^5 in the Figures 5.18c - 5.18e. In these cases, the Ac_{tx}^n values for $x < n$ are similar, while for $x > n$ using $N_{label} \geq n$ significantly improves the accuracy results. Even for $N_{label} \leq n$, increasing the number of labels N_{label} increases the accuracy for $x > n$, e.g. the Ac_{t5}^3 of $N_{label} = 1$ is approximately 25 % compared to 39 % using $N_{label} = 2$ and 59 % with $N_{label} \geq n = 3$. Regarding the results Ac_{tx}^5 for $N_{label} = 5$, we observe that it offers a noticeable advantage at $x = 4 \leq n$ of around 5 pp. Contrary to the other weights given in Table 4.5, which are descending with an increase in n , the weights for $n = 4$ and $n = 5$ are equal for $N_{label} = 5$. An unexpected observation is that the

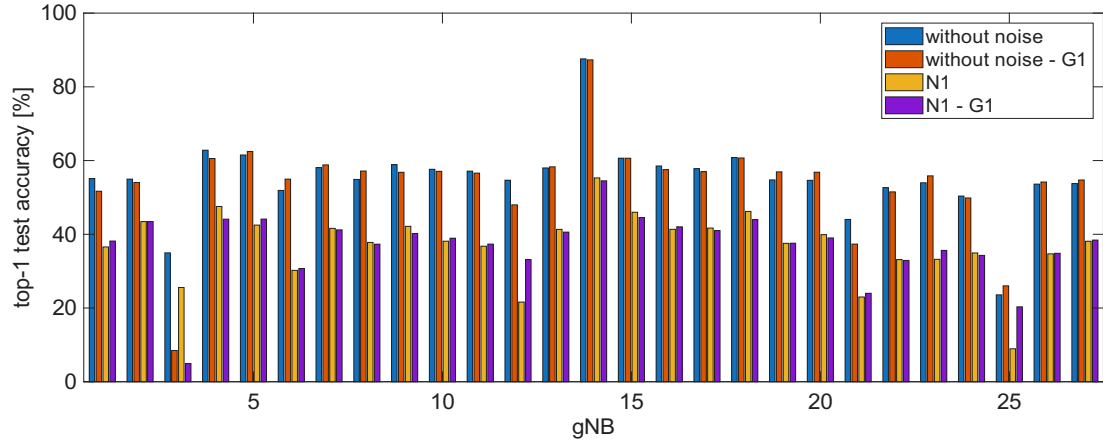
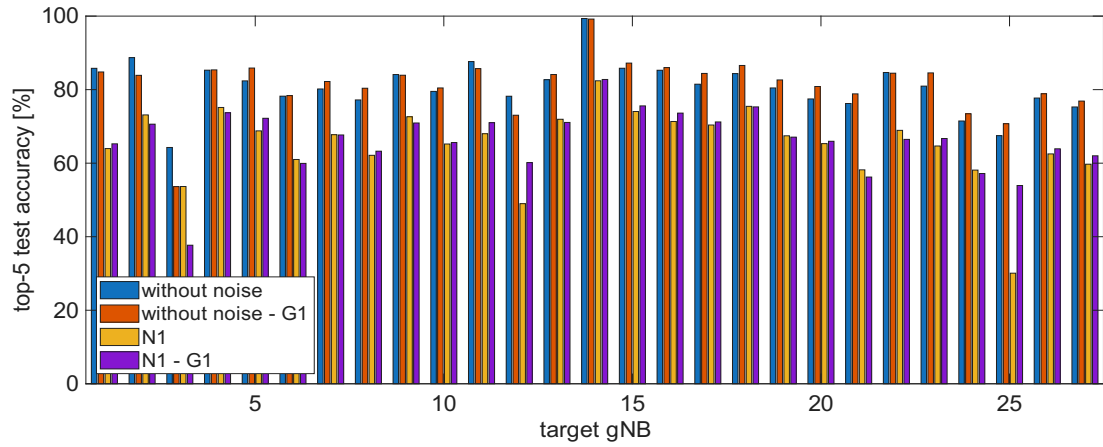
(a) Ac_{t1} .(b) Ac_{t5} .

Figure 5.17: CNN-S B2 Ac_{t1} and Ac_{t5} with $gNB_{ref} = 14$ and gNB_{tar} given on the x-axis. The results at $gNB_{ref} = gNB_{tar} = 14$ represent B1 with $pd = 100\%$. The blue and orange bars show the results without UE position noise, using a one-hot map and G1 as input, respectively. The yellow and purple bars show the corresponding results with N1 as UE position noise.

Ac_{tx}^5 values using $N_{label} = 5$ are lower than the Ac_{tx}^4 values, even though the weights of the fourth and the fifth best beam in the cross-entropy loss function are equal. A possible explanation is that the effects that improve the accuracy for higher N_{label} in cases with $x \geq n > N_{label}$ are more advantageous for lower n . Such effects might originate in the clustered nature of the BPL RSS shown in Figure 3.7. With small deviations between the learned relation of UE positions and the top- n BPLs and the actual top- n BPLs at these positions, the predicted best BPL might still be included in the cluster of good performing BPLs around l_{k^*} .

The Ac_{tx}^n results of the scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 6$ using FCNN-S, representing a target gNB with a low dataset size and low performance in the previous investigations, are shown in Figure 5.19. We observe the same trends as for $gNB_{tar} = 10$ in Figure 5.18, with $N_{label} = 2$ having the highest Ac_{t1}^1 accuracy and $N_{label} = 5$ having the lowest. This may indicate that the chosen weights of $N_{label} = 2$ offer an advantage, however the difference of approximately 10 pp, compared to variations of up to 20 pp across the results of the five random selections of the dataset portions, does not allow conclusions. As previously observed, the accuracies of $gNB_{tar} = 6$ are overall lower compared to $gNB_{tar} = 10$, however the difference varies across n and x , e.g. the difference of $N_{label} = 5$ at Ac_{t5}^5 is approximately 2 pp compared to 9 pp at Ac_{t10}^5 and Ac_{t5}^1 .

The Ac_{tx}^n results for $gNB_{tar} = 10$ including noise according to N1 in Figure 5.20 show the same trends as observed in Figure 5.18 without noise, with the expected lower overall results. The Ac_{t1}^1 results in Figure 5.20a show the highest value for $N_{label} = 5$, further reinforcing that the previously observed lower accuracies were due to the random selections of dataset portions.

Across the Figures 5.18 - 5.20 we observe a saturation of the Ac_{tx}^1 values with an increase in x , while the saturation for higher n is setting in at higher x . We observe a significant percentage of the predicted best BPL being the actual second best BPL across all N_{label} , with Ac_{t1}^2 of approximately 8 % to 12 %. In addition to the top-1 accuracy Ac_{t1}^1 widely used as a metric in this work, a significant percentage of the predictions failing to match l_{k^*} predict the second best BPL instead. The Ac_{t1}^n for $n > 2$ are around 2 %.

5.2 Cross-Environment Generalizability

Besides reducing the required training data from new gNB locations, it is desirable to find a model that can effectively perform beam prediction at multiple locations, eliminating the need to handle dedicated models for each location. As an initial step towards a generalized model, Section 5.2.1 investigates the beam prediction accuracy on the reference gNB after the fine-tuning stage in different target gNBs. In Section 5.2.2, the performance of EWC on retaining prediction accuracy on the reference gNB after fine-tuning on a target gNB is evaluated. In Section 5.2.3, the accuracies achieved by employing mixed training datasets, containing training data from multiple gNBs, are investigated.

5.2.1 Accuracy at the Reference gNB after Fine-Tuning on Different Target gNBs

Figure 5.21 shows the top-1 and top-5 BPL prediction accuracies on the test data of $gNB_{ref} = 14$ after the fine-tuning stage of B2 on the target gNBs given on the x-axis, using CNN-S, FCNN-S and FCNN-L, in blue, orange and yellow, respectively. The values for $gNB = 14 = gNB_{ref}$ give the accuracies achieved on gNB 14 before the fine-tuning on

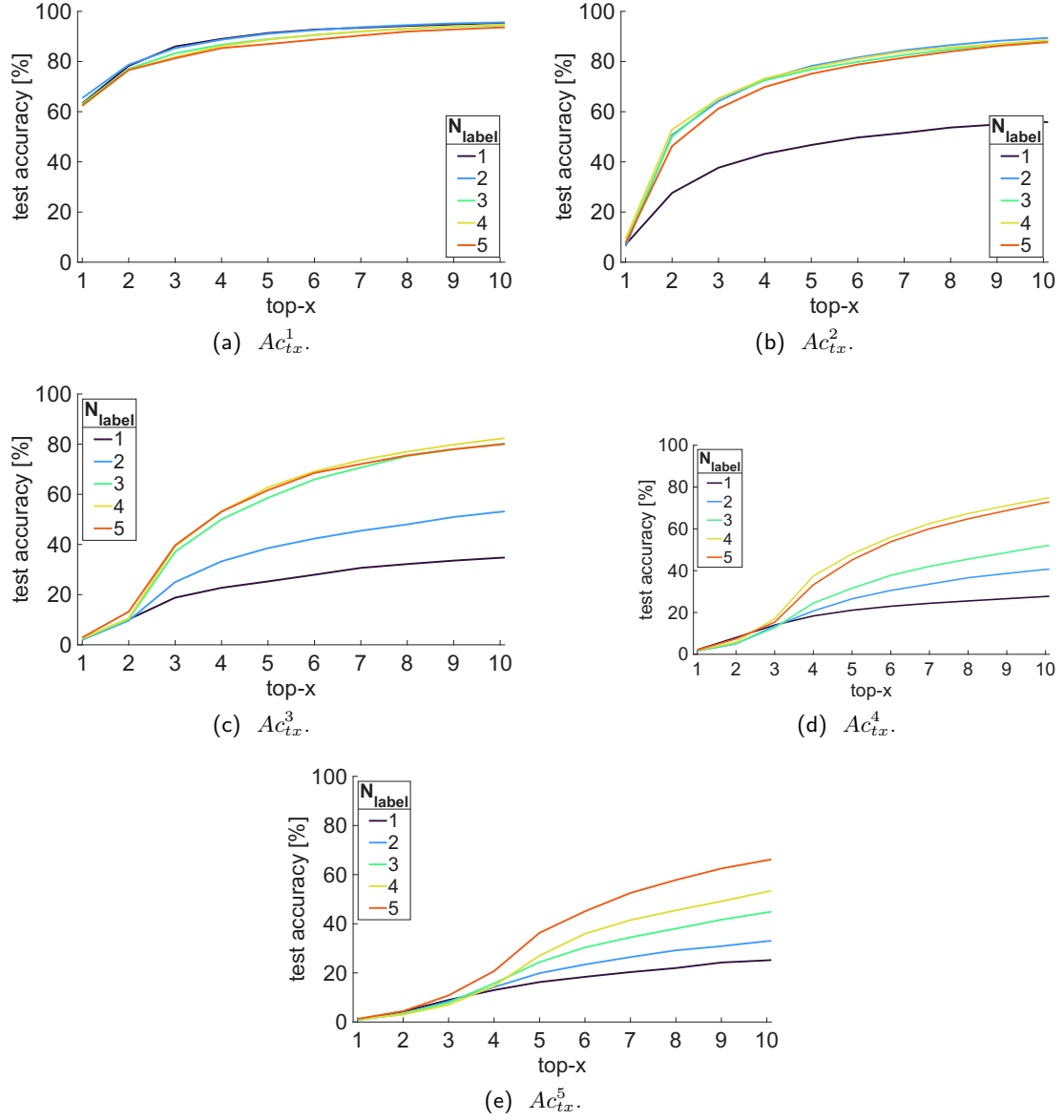


Figure 5.18: Top- x accuracies from the best BPL in (a) to the fifth best BPL in (e) using scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 10$ on FCNN-S. The values on the x-axis denote the x in the top- x accuracy Ac_{tx}^n , and the number of labels N_{label} used in the weighted cross-entropy loss function is given in different colours.

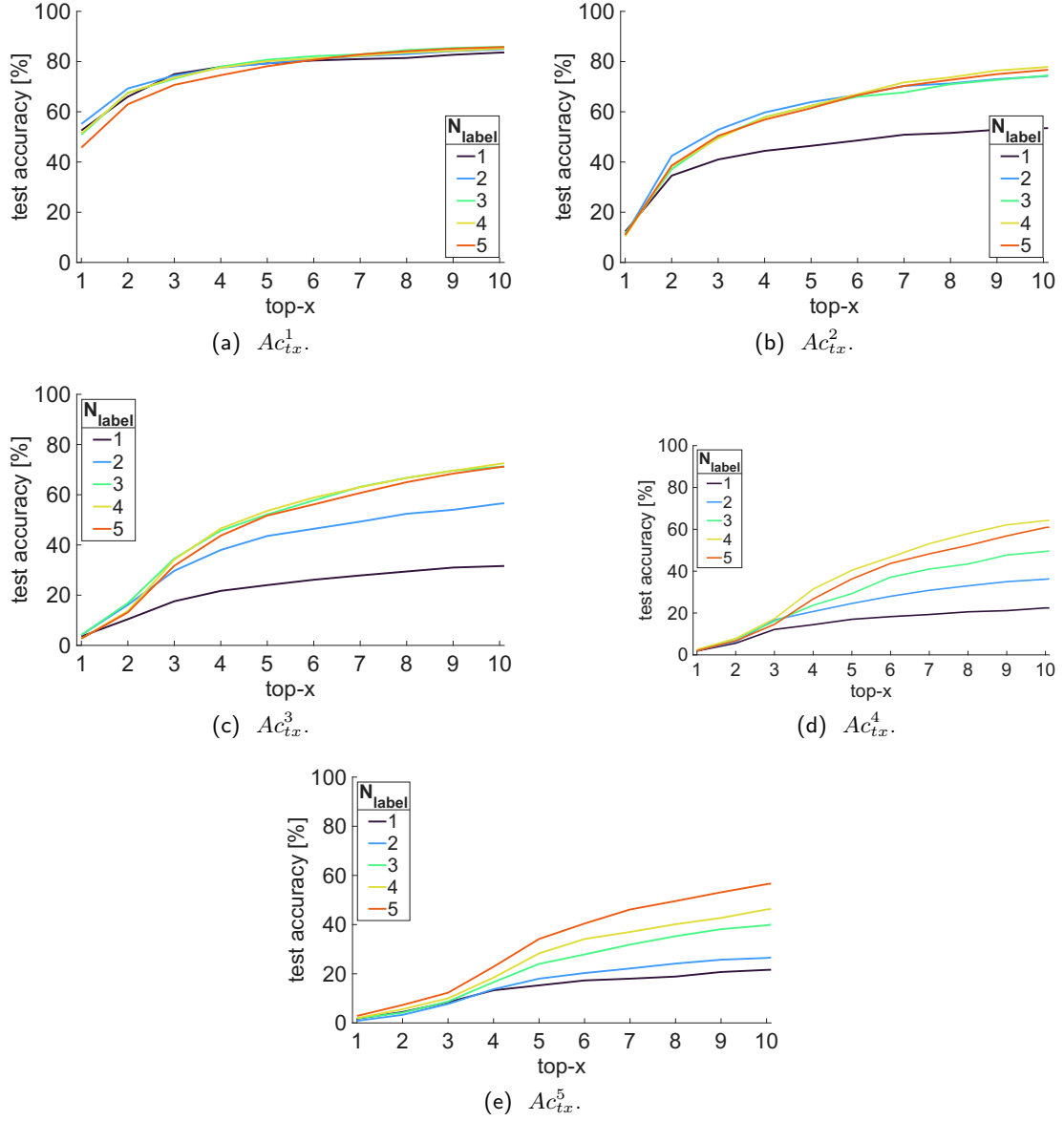


Figure 5.19: Top- x accuracies from the best BPL in (a) to the fifth best BPL in (e) using scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 6$ on FCNN-S. The values on the x-axis denote the x in the top- x accuracy Ac_{tx}^n , and the number of labels N_{label} used in the weighted cross-entropy loss function is given in different colours.

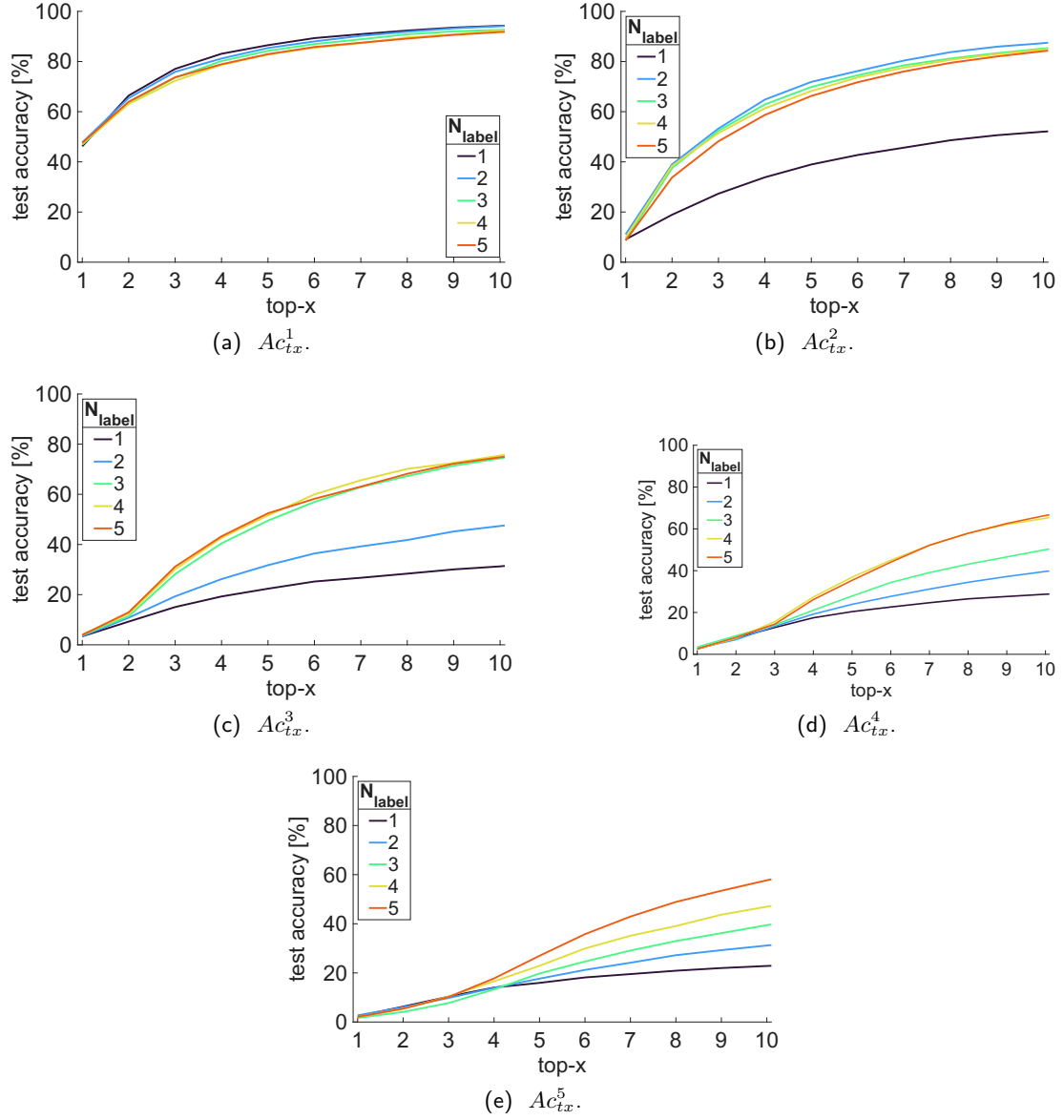


Figure 5.20: Top- x accuracies from the best BPL in (a) to the fifth best BPL in (e) using scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 10$, and noise according to N1 on FCNN-S. The values on the x-axis denote the x in the top- x accuracy Ac_{tx}^n , and the number of labels N_{label} used in the weighted cross-entropy loss function is given in different colours.

a target gNB, which is equivalent to B1 with $pd = 100\%$. We observe significantly higher values on the reference gNB after training on the target gNBs for the CNN-S compared to FCNN-S and FCNN-L for most target gNBs in the top-1 accuracies, and for all target gNBs in the top-5 accuracies. However, the differences in the achieved accuracies after fine-tuning compared to the values before fine-tuning range from under 1 pp for fine-tuning at gNB 1, to 75 pp for the top-1 accuracy for gNB 13. A reason for the higher values of the CNN-S might be found in its higher number of tunable parameters and the property of neurons being only connected to the limited number of neurons in their filter window and thus only being influenced by a small local area of the input space, leading to a limited impact of the new training data on the previously trained weights. The differences between the target gNBs might be explained by the different dataset sizes of the target gNBs. Additionally, due to the filter windows of the CNN, the portion of shared UE positions between $gNB_{ref} = 14$ and the target gNB might have an impact. Therefore, in addition to the total dataset sizes given in Figure 3.8, Figure 5.22 gives the ratio of the dataset sizes with respect to the dataset size of the considered reference gNB 14, as well as the portion of the gNB 14 UE positions that are included in the datasets of the gNB on the x-axis in blue and red, respectively. Note that the shown values represent $pd = 100\%$, the training sets used for fine-tuning are composed of 5% randomly selected from this data. The central position and large coverage area of gNB 14, shown in Figure 3.6, lead to a close relation between the dataset size and the portion of gNB 14 positions included in other datasets. We can observe a clear inverse relationship between dataset sizes along with the portions of gNB 14 positions included in the other datasets and the accuracies of CNN-S, e.g. by comparing the accuracies in Figure 5.21 obtained by gNB 9 and 21 that have opposite dataset sizes and portions of gNB 14 positions in their datasets. Comparing gNB 3 and gNB 11, we observe that with nearly the same dataset size, being 10.7% and 10.3% of the dataset size of gNB 14, respectively, the accuracies of gNB 3 are significantly lower than those of gNB 11, with $Ac_{t1} = 60.2\%$ and $Ac_{t5} = 91.1\%$ after fine-tuning on gNB 3 compared to $Ac_{t1} = 84.7\%$ and $Ac_{t5} = 99.3\%$ after fine-tuning on gNB 11. Given the portions of gNB 14 positions of 9.7% at gNB 3 and 0% at gNB 11, this indicates that the portion of overlapping positions is a deciding factor. By comparing gNBs with nearly identical portions of gNB 14 positions, such as gNB 8 and gNB 10 with 69.4% and 69.6%, respectively, and differing dataset sizes of 77.0% and 83.0% compared to the dataset size of gNB 14, respectively, we observe only minor differences in the accuracies, with $Ac_{t1} = 21.7\%$ and $Ac_{t5} = 58.5\%$ after fine-tuning on gNB 8 compared to $Ac_{t1} = 22.0\%$ and $Ac_{t5} = 55.6\%$ after fine-tuning on gNB 10. Comparing gNB 12 and 24 with dataset sizes of 18.4% and 19.6% compared to the dataset size of gNB 14, respectively, and portions of gNB 14 positions of 16.6% at gNB 12 and 9.6% at gNB 24, we observe that a higher portion of gNB 14 positions leads to lower accuracies, even though the dataset size of gNB 12 is lower. The resulting accuracies are $Ac_{t1} = 57.2\%$ and $Ac_{t5} = 89.0\%$ after fine-tuning on gNB 12, and to $Ac_{t1} = 66.6\%$ and $Ac_{t5} = 92.8\%$ after fine-tuning on gNB 24. We cannot observe this relationship between the dataset sizes or gNB 14 positions portions and the accuracies of the FCNN-S and FCNN-L, with gNB 18 showing one of the highest accuracies and gNB 13 one of the lowest. While FCNN-L show slightly higher accuracies compared to FCNN-S for most gNBs, there are several exceptions across gNBs with varying dataset properties, therefore we cannot assert a trend. Given the number of model parameters, as they are given in Table 5.1, is considerably higher in FCNN-L than in FCNN-S, an increase in model

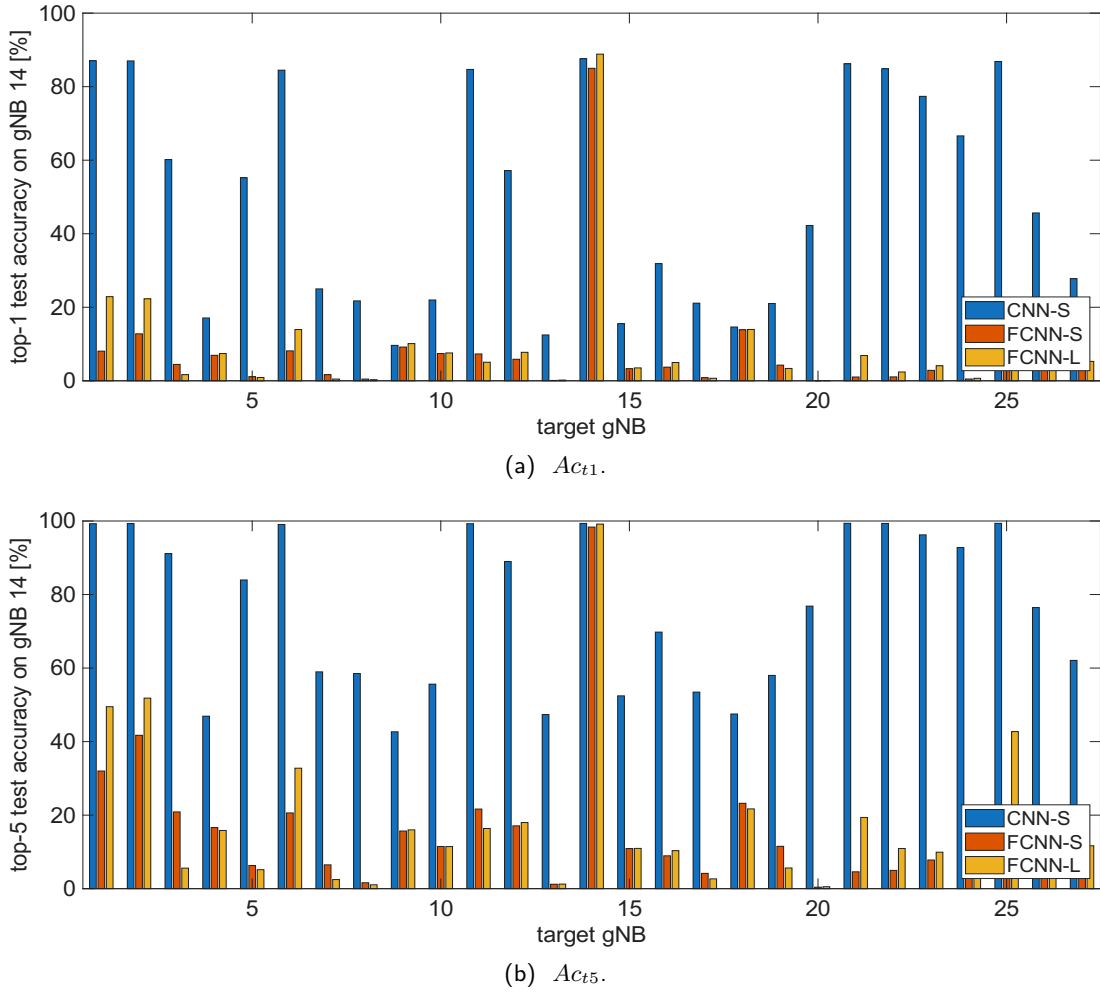


Figure 5.21: Beam prediction accuracy Ac_{t1} and Ac_{t5} on the reference gNB after the fine-tuning stage in B2 with $gNB_{ref} = 14$ for varying gNB_{tar} . The results at $gNB_{ref} = gNB_{tar} = 14$ represent B1 with $pd = 100\%$, thus the difference to the results at $gNB_{tar} = 14$ can be seen as accuracy lost on the reference gNB by fine-tuning on the respective gNB_{tar} . The results for FCNN-S are shown in blue, the orange and yellow bars show the results of FCNN-L and CNN-S, respectively.

size on its own does not increase the accuracy at the reference gNB after fine-tuning. Therefore, the better accuracies of the CNN-S are largely due to its structure.

The results open the question whether, using CNN-S, target gNBs that have disjunct UE positions in their datasets compared to the reference gNB profit from transfer learning, or whether the results for these gNBs in Figure 5.7 are a consequence of good accuracies in B1 with $pd = 5\%$. In Figure 5.6, we only investigated the pd dependent B1 prediction accuracies of gNB 14, which we use as gNB_{ref} . Therefore, we performed an additional test on gNB 1, which covers 0% of the UE positions included in the dataset of gNB 14, using B1 with $pd = 5\%$ on the CNN-S. The results confirm that the accuracy does improve even for gNBs with non-overlapping UE positions, as the mean of five simulations resulted

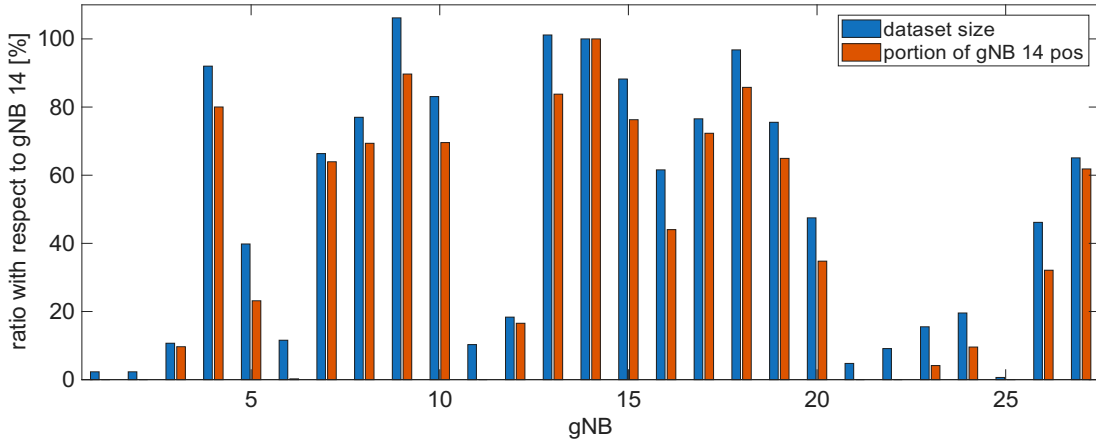


Figure 5.22: Dataset sizes and portion of UE positions included in the dataset in relation to gNB 14, in blue and orange, respectively.

in approximately $Ac_{t1} = 4\%$ and $Ac_{t5} = 19\%$, compared to $Ac_{t1} = 55\%$ and $Ac_{t5} = 86\%$ in the B2 results with $gNB_{ref} = 14$ shown in Figure 5.7.

5.2.2 Elastic Weight Consolidation

We now investigate EWC as an approach to handle the accuracy loss observed on the reference gNB after fine-tuning on the target gNB. Figure 5.23 shows the top-1 and top-5 accuracies on FCNN-SD after B2 with $gNB_{ref} = 14$ on the test data of the reference gNB and the target gNB in blue and orange, respectively. The corresponding accuracies for employing EWC in the fine-tuning step are shown in yellow and purple, respectively. The values shown in the bar-graph for gNB 14 represent B1 with $pd = 100\%$. We can observe in Figure 5.23a that the EWC is able to balance the $Ac_{t1}^{14,t-14}$ and $Ac_{t1}^{14,t-t}$ accuracies at a number of gNBs. Comparing with the dataset sizes in Figure 3.8 we observe that on gNBs with small datasets, the $Ac_{t1}^{14,t-t}$ values are very small compared to the $Ac_{t1}^{14,t-14}$, thus the EWC with the chosen parameters is too restrictive with very small datasets. While employing EWC achieves the goal of $Ac_{t1}^{14,t-14} \approx Ac_{t1}^{14,t-t}$ on target gNB with large datasets, the accuracies on the target gNB are significantly lower than without employing EWC, with the smallest difference on gNB 27 being 7 pp. The difference of $Ac_{t1}^{14,t-14}$ compared to the accuracy before the fine-tuning on other gNBs is even more significant, with over 50 pp for any gNB at which the difference between $Ac_{t1}^{14,t-14}$ and $Ac_{t1}^{14,t-t}$ is smaller than 10 pp. The top-5 accuracies in Figure 5.23b show a similar trend, with a higher accuracy loss between $Ac_{t5}^{14,t-t}$ on target gNBs with large dataset sizes, the lowest being 13 pp at gNB 27.

The loss in memory on old beam prediction tasks is high compared to our tests with FCNN-SD on the MNIST dataset, identifying even and uneven digits in tasks A and B, respectively. Using EWC, we achieved accuracies of 83% on task A and 85% on task B after retraining a model trained on data of task A with data of task B, compared to 37% on task A and 93% on task B without employing EWC. The accuracies after training

both tasks on their own training data separately amount to 95 %. One explanation for the better performance of EWC on the MNIST dataset can be found by observing the model parameters after training on the different tasks. Figure 5.24 shows the cumulative distribution functions (CDFs) of the biases of the second hidden layer after training on the gNBs 10, 14 and 19 following B1 with $pd = 100\%$ (a) in blue, orange and yellow, respectively, and after training separately on task A and task B on the MNIST dataset in (b), in blue and orange, respectively. We observe that the CDFs of the biases in 5.24b are nearly identical, while the CDFs of the biases on the beam prediction at the different gNBs show considerable deviations. Therefore, we can conclude that this lower difference between the optimal parameters of task A and task B is beneficial for retaining accuracy on task A while retraining the model on task B. The more complicated task of beam prediction on two different gNBs might result in a saturation of the NN. However, even with the larger difference between the optimal parameters, EWC on FCNN-SD is able to balance $Ac_{t1}^{14,t-14}$ and $Ac_{t1}^{14,t-t}$ for target gNBs with sufficiently large dataset sizes. A larger NN with more tunable parameters might be able to achieve higher accuracies with EWC.

5.2.3 Dataset Mixing

To conclude, we investigate the achievable accuracies of a model performing beam prediction on multiple gNBs simultaneously by mixing the training datasets of multiple gNBs to train the model. The top-1 and top-5 accuracy results of FCNN-S and FCNN-SG using the mixed dataset, including $pd = 60\%$ of all gNBs except gNB 10 and 13, are shown in Figure 5.25. We observe significantly higher top-1 and top-5 accuracies of FCNN-SG compared to FCNN-S over all gNBs in M1, with a mean top-1 accuracy across all gNBs in M1 of approximately 35 % and 66 % for FCNN-S and FCNN-SG, respectively, and a top-5 accuracy of approximately 67 % and 93 % for FCNN-S and FCNN-SG, respectively. The addition of the gNB position allows the FCNN to differentiate between the gNBs included in the mixed dataset, however the results for M2 show that this improvement does not hold for gNBs not included in the mixed dataset. The gNB position only functions as an identifier for the gNBs that are included in the training dataset, and the NN does not seem to be able to learn a relation between the position in the study area and optimal BPL otherwise. To test this hypothesis, we replaced the gNB position in the input data with the index of the gNB. The resulting accuracies were very similar to the accuracies of FCNN-SG, with deviations expected from the random selection of dataset portions, thus confirming that the gNB only functions as an index of the gNBs included in M1. The accuracies of FCNN-S on M2 follow a similar trend to some of the gNBs in M1 with the lower accuracies. Since FCNN-S has no way of differentiating between gNB, a possible factor for the differing accuracies inside M1 can be found in the dataset sizes, and therefore the portion of the mixed training dataset. By comparing with Figure 3.8, we can observe that gNBs with large datasets cannot be found amongst the gNBs with the lowest accuracies. Additionally, we observe that the portion of LOS has a significant impact on the top-1 accuracies. While the dataset sizes of gNB 1 and 2 are among the smallest, they achieve high top-1 accuracies as they consist only of LOS data. The top-1 accuracy of gNB 23 is higher than that of gNB 24, as it has a higher portion of LOS data, even though it has a smaller dataset size. The differing M2 results on gNB 10 and 13 might also be explained by the larger portion of LOS data in the dataset of gNB 13.

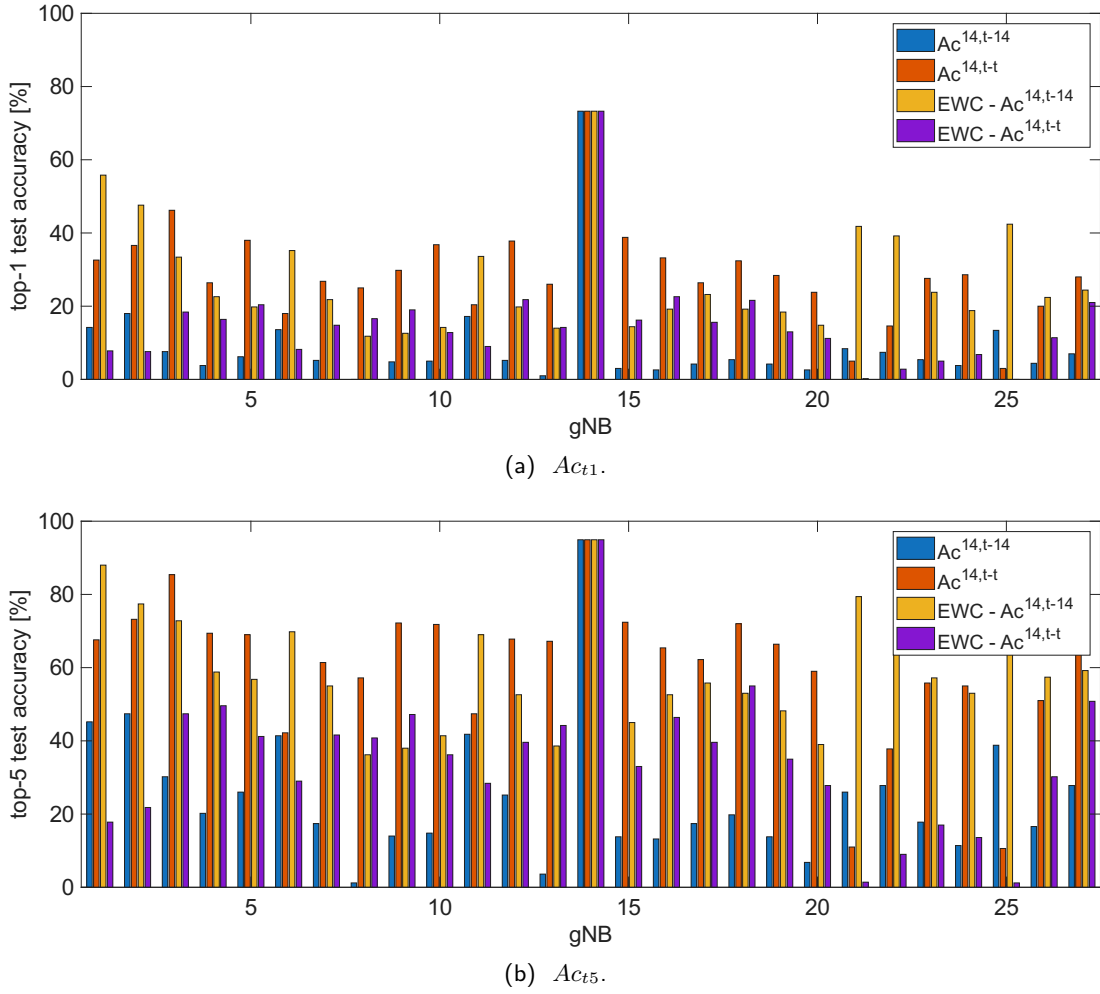


Figure 5.23: Top-1 and top-5 accuracies on FCNN-SD after B2 with $gNB_{ref} = 14$ on the test data of the reference gNB and the target gNB in blue and orange, respectively. The corresponding accuracies for employing EWC in the fine-tuning step are shown in yellow and purple, respectively. The values at gNB 14 represent B1 with $pd = 100\%$.

The corresponding results for FCNN-L and FCNN-LG shown in Figure 5.26 show similar results as observed in Figure 5.25, but with overall higher accuracies achieved in M1, with a mean top-1 accuracy across all gNB in M1 of approximately 39 % and 80 % for FCNN-L and FCNN-LG, respectively, and a top-5 accuracy of approximately 75 % and 98 % for FCNN-L and FCNN-LG, respectively. The difference in the M2 accuracies of FCNN-L and FCNN-LG, shown in Figure 5.26 in yellow and purple, respectively, with respect to the corresponding M1 and M2 accuracies for FCNN-S and FCNN-SG, shown in Figure 5.25, is smaller than 1 pp.

The results for CNN-S and CNN-SR in Figure 5.27 show no general advantage of either CNN in M1. Comparing with the dataset sizes in Figure 3.8, CNN-SR achieves higher accuracies on gNB with larger datasets. The good performance of CNN-S on gNB with large datasets can be explained by the fact that small datasets are more likely to include unique UE positions, thus the filter weights of the respective regions of the input data map

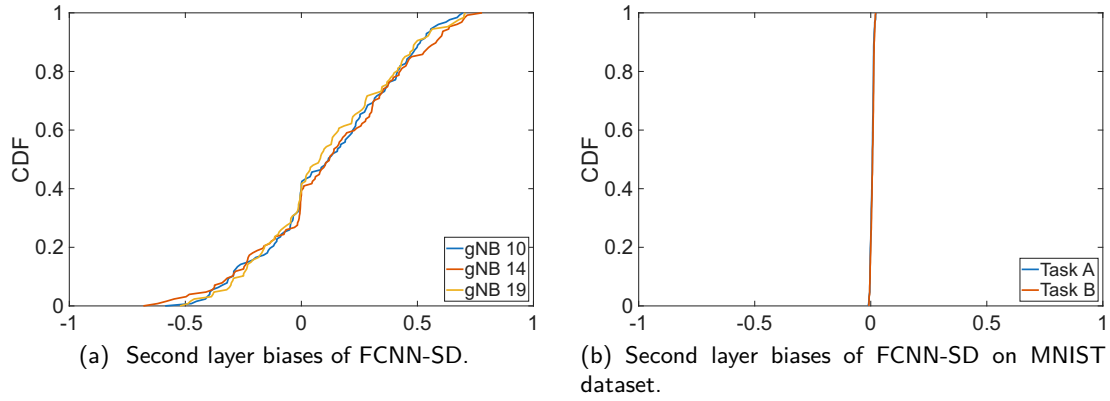


Figure 5.24: CDFs of the biases of the second hidden layer of FCNN-SD after B1 with $pd = 100\%$ on gNB 10, 14 and 19 are shown in (a) in blue, yellow and orange, respectively. The CDFs of the biases of the second layer of FCNN-SD after training on task A and task B of the MNIST dataset are shown in (b) in blue and orange, respectively.

are only influenced by data from their training datasets. The location of gNB 1 and 2 at the bottom left corner of the study area, partitioned from the rest of the area by the building of the main station, shown in Figure 3.3, leads to their covered UE positions only being included in their datasets. On the other hand, gNBs such as 8, 9, 13 and 18 have large datasets because they are located at the edge of the large open area and therefore share many UE positions, as can be seen by the portion of gNB positions shared with gNB 14 shown in Figure 5.22. As CNN-SR consists of a relative position centred on the centre of the input matrix, there is significantly more overlap between the input matrices. Therefore, gNBs with higher dataset sizes achieve higher accuracies than gNBs with lower dataset sizes, as they provide a larger portion of the total training data. Additionally, a larger LOS portion results in higher accuracies. The mean top-1 accuracy across all gNBs in M1 is approximately 21 % and 38 % for CNN-S and CNN-SR, respectively, and the mean top-5 accuracy is approximately 54 % and 73 % for CNN-S and CNN-SR, respectively.

In Figure 5.28, the M1 accuracies of FCNN-SG in orange and FCNN-LG in purple, using $pd = 5\%$ with all gNBs included in the mixed dataset, are compared with the B2 accuracies with $gNB_{ref} = 14$ of FCNN-S in blue and FCNN-L in yellow. Since gNB 14 is the reference gNB, the FCNN-S and FCNN-L shown in the bar-graph for gNB 14 are the B1 accuracies with $pd = 100\%$. We can observe significantly lower accuracies of the M1 of FCNN-SG compared to B2 of FCNN-S at several gNBs with small dataset sizes, e.g. gNB 1, 2, 21 and 25. At gNBs with high dataset sizes the accuracies are comparable. The location of gNB 7 at the edge of the study area, causing low accuracies of B2 with FCNN-S, does not negatively impact the M1 accuracies of FCNN-SG. The mean top-1 and top-5 accuracies of B2 of FCNN-S are approximately 6.0 pp and 5.1 pp higher than for M1 of FCNN-SG, respectively. The M1 accuracies of FCNN-LG closely match the B2 accuracies of FCNN-S, with notable outliers in the top-1 accuracies of gNB 3 and the accuracies of gNB 25. The mean differences between the top-1 and top-5 accuracies of FCNN-LG and FCNN-L, in scenarios M1 and B2, respectively, are below 1 pp, thus using multiple small datasets of different gNBs to construct a mixed training dataset can achieve similar accuracies as

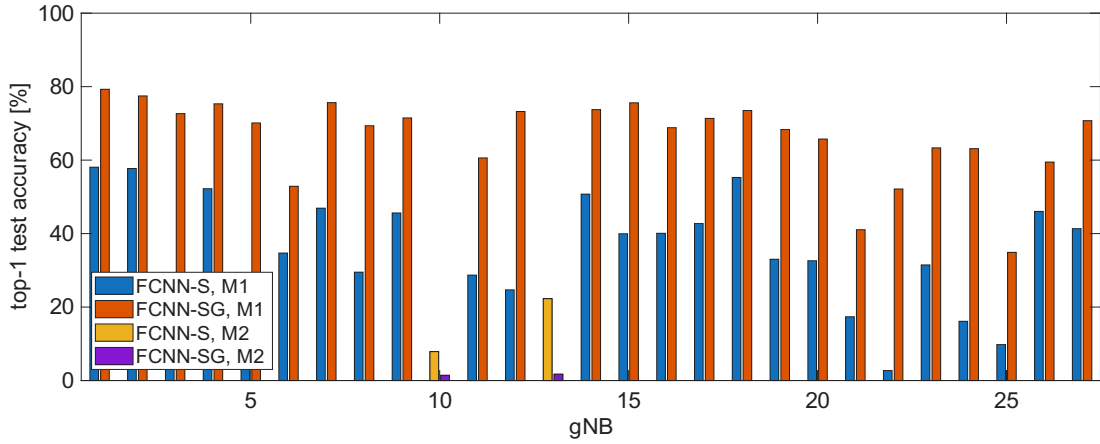
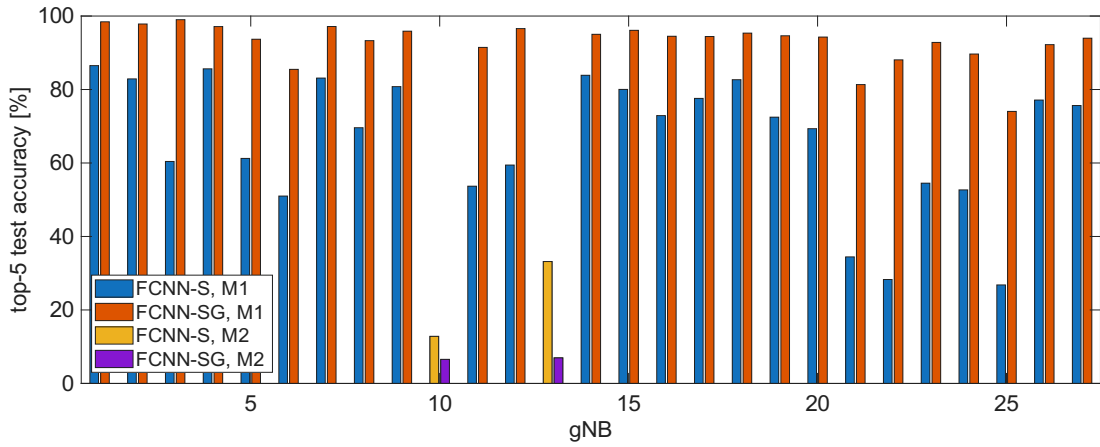
(a) Ac_{t1} .(b) Ac_{t5} .

Figure 5.25: M1 and M2 top-1 (a) and top-5 (b) accuracies of FCNN-S in blue and yellow, respectively. The corresponding M1 and M2 accuracies of FCNN-SG are shown in orange and purple, respectively. The mixed dataset contains every gNB except gNB 10 and 13, with $pd = 60\%$.

transfer learning with a large dataset as the reference, as long as we use a large enough NN. However, the accuracies of NN trained with mixed datasets might deteriorate at some point when more gNBs are added to the mixed training set.

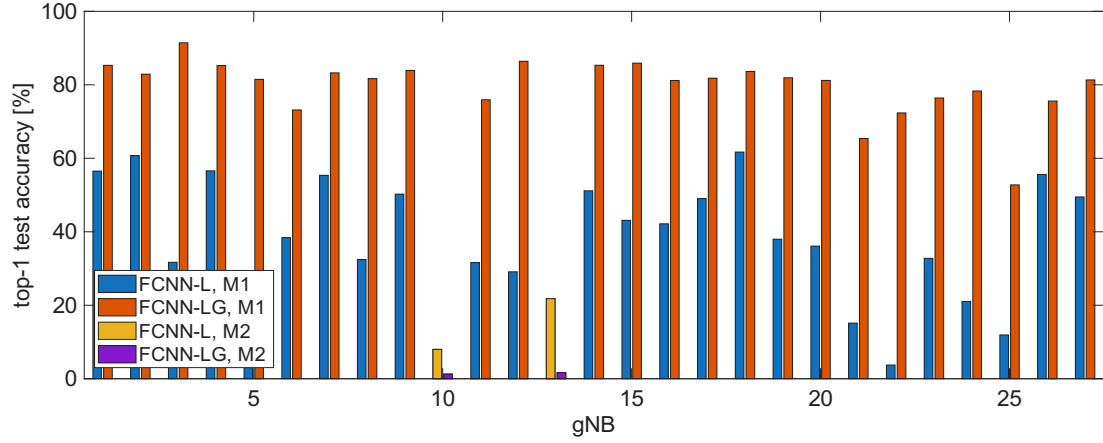
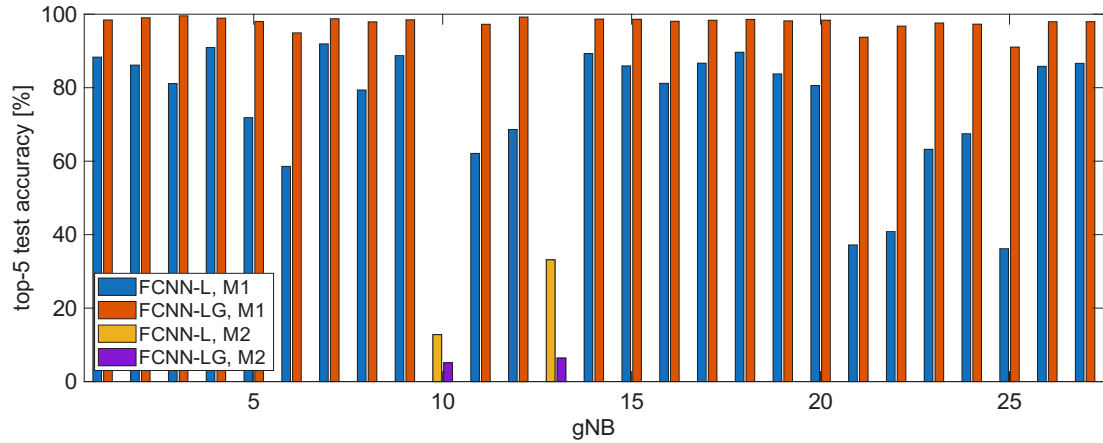
(a) Ac_{t1} .(b) Ac_{t5} .

Figure 5.26: M1 and M2 top-1 and top-5 accuracies of FCNN-L in blue and yellow, respectively. The corresponding M1 and M2 accuracies of FCNN-LG are shown in orange and purple, respectively. The mixed dataset contains every gNB except gNB 10 and 13, with $pd = 60\%$.

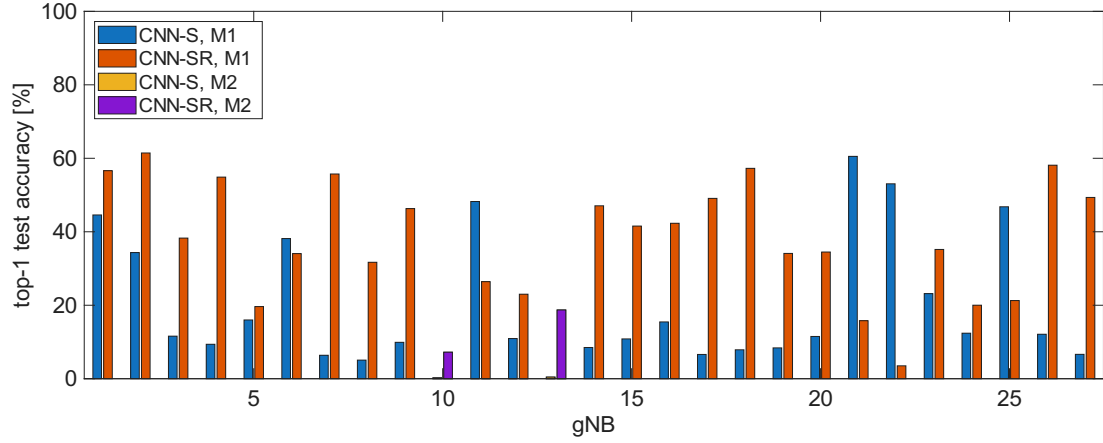
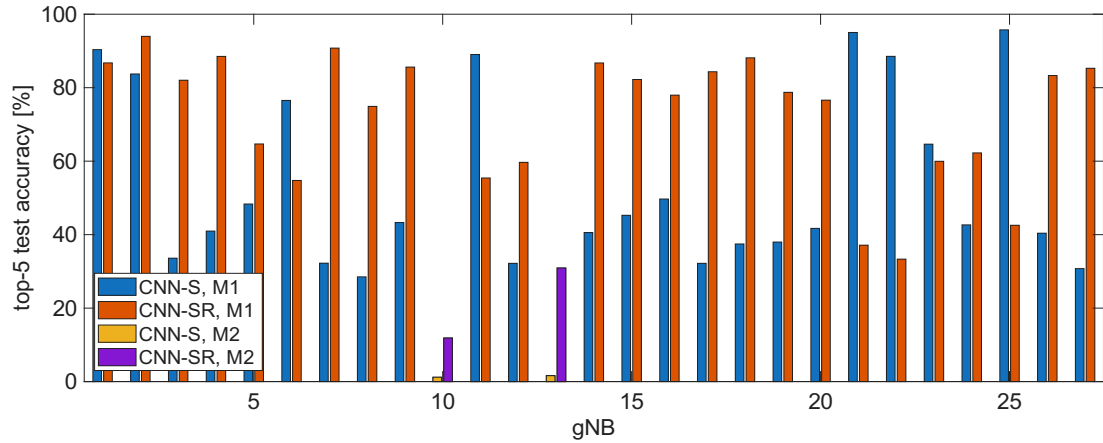
(a) Ac_{t1} .(b) Ac_{t5} .

Figure 5.27: M1 and M2 top-1 and top-5 accuracies of CNN-S in blue and yellow, respectively. The corresponding M1 and M2 accuracies of CNN-SR are shown in orange and purple, respectively. The mixed dataset contains every gNB except gNB 10 and 13, with $pd = 60\%$.

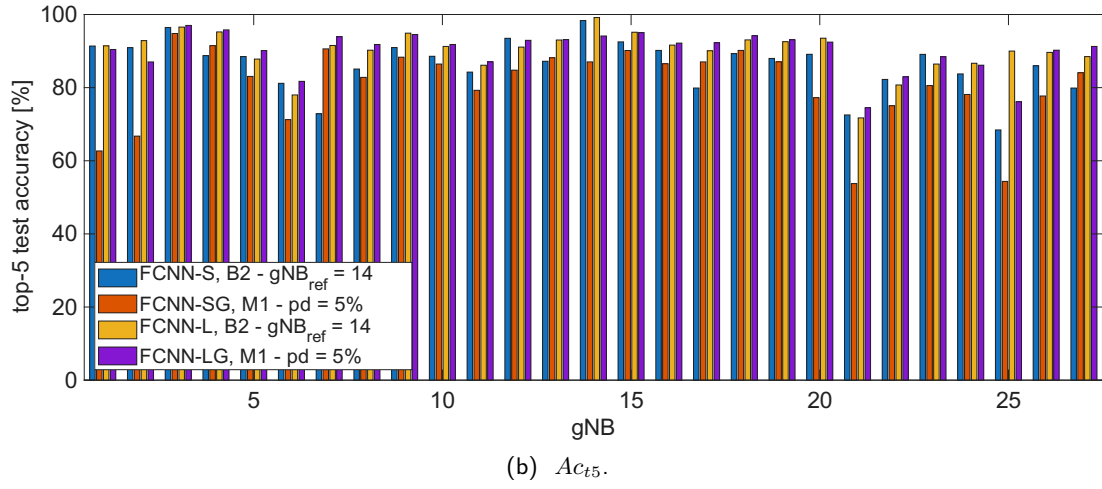
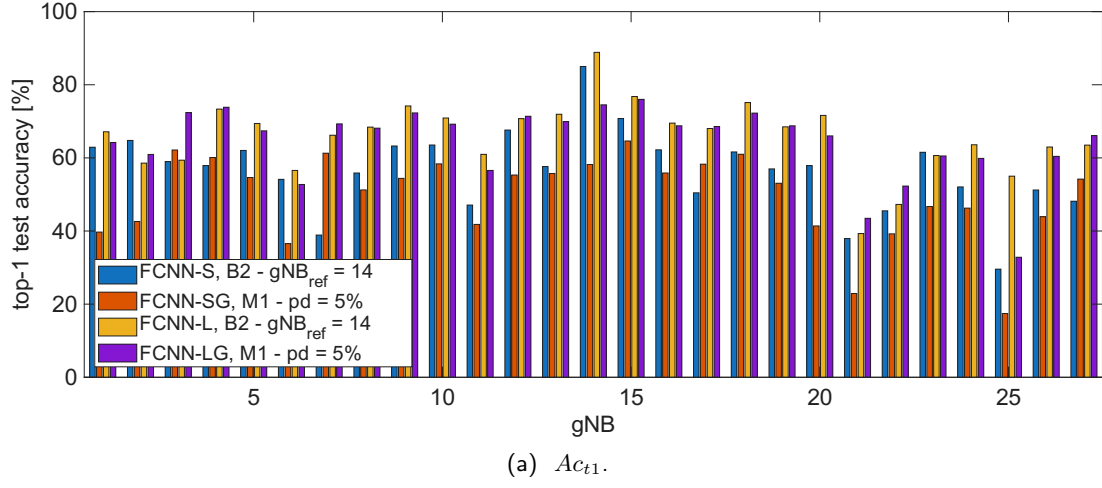


Figure 5.28: M1 accuracies of FCNN-SG in orange and FCNN-LG in purple. The B2 accuracies using $gNB_{ref} = 14$ of FCNN-S and FCNN-L are shown in blue and yellow, respectively, with gNB 14 being the B1 accuracy with $pd = 100\%$.

Conclusion and Future Work

This thesis presented an extensive investigation into the cross-environment performance of position-aided SL strategies for beam prediction in mmWave networks. The investigations were performed on data obtained by the commercial ray-tracing simulator Wireless InSite [57] for 27 gNBs in a 28 GHz downlink cellular MIMO network in a study area near the main station in Frankfurt, Germany, representing an urban-dense environment. Subsets of the datasets containing the BPLs, ordered by their RSS for each UE position, were used to train NN models, and their performance was evaluated based on the achieved accuracies of predicting the best BPL, as well as subsequent top BPLs. Transfer learning was shown in [5] to be a promising technique to reduce the required training data for new gNB locations. This thesis provides an in-depth analysis of the impact of the factors dataset size, NN size, noise in the position data, and the number of training labels on the performance of transfer learning. To achieve cross-environment generalizability of one NN model across multiple gNB locations simultaneously, the performance of EWC as a technique to preserve the performance on the reference gNB after fine-tuning in transfer learning was investigated. As a further approach to cross-environment generalizability, dataset mixing was proposed and analysed.

The results of the investigated FCNN, inspired by the FCNN proposed in [5], show a clear performance advantage of transfer learning compared to the training of a new model with the same limited dataset size as used for fine-tuning, with a mean improvement of the top-1 accuracy of approximately 11 pp. While a reference gNB with a sufficient dataset size is required to achieve good performances, the dataset size and LOS portion of the target gNB has the dominant impact on the achieved improvement in top-1 accuracies, resulting in a mean of up to 24 pp across all considered reference gNBs for target gNBs that feature small dataset sizes with high LOS portions. The FCNN achieves a mean top-1 accuracy across all considered reference and target gNB combinations of 53 % and a mean top-5 accuracy of 84 %. Employing the considered CNN results in lower variation between target gNBs, while achieving similar mean accuracies across all target gNBs. Increasing the size of the FCNN achieves a mean improvement of approximately 10 pp in the top-1 accuracy and 4 pp in the top-5 accuracies, however the larger FCNN shows signs of overfitting on target gNBs with small dataset sizes, achieving a lower improvement on these target gNBs. Noise in the UE position data affects the top-1 accuracies more than the top-5 accuracies, and the noise induced losses increase with the size of the NN, with the larger FCNN still achieving the highest accuracies. Considering noise in all training stages of the NNs shows a slight performance advantage across top-1 and top-5 accuracies in both FCNNs. Using multiple of the top BPLs as training labels has no negative impact on the prediction accuracy of the best BPL, while promoting the selection of subsequent top BPLs.

The investigations on cross-environment generalizability have shown severe catastrophic forgetting on the reference gNB after transfer learning for the considered FCNNs. The

effect on the CNN is significantly reduced for target gNBs that have low portions of the UE positions included in the dataset of the reference gNB. EWC is able to balance the beam prediction accuracies on reference and target gNBs after transfer learning on target gNBs with sufficient dataset sizes, however the loss in accuracy is significant, especially on the reference gNB, amounting to values over 50 pp. Dataset mixing shows promising results on producing a model that is able to perform beam prediction on multiple gNBs simultaneously. Employing dataset mixing, including the same limited portions of the datasets as used for the fine-tuning in transfer learning, the larger FCNN is able to achieve approximately the same mean accuracy across all considered gNBs with a single model, as transfer learning achieves with models separately fine-tuned for each target gNB.

Based on the presented results, there are multiple promising directions for future work. The results for small fine-tuning datasets have shown considerable variations between multiple repeated simulations. To achieve consistently high performance, the compositions of UE positions included in the training set that produce high accuracies have to be investigated. The performance of transfer learning on larger FCNNs may be further improved by employing techniques to limit overfitting, such as regularization and dropout. EWC might be more successful in retaining the performance on reference and target gNBs when employed in combination with larger NNs. Furthermore, it remains to be seen whether UE position noise induces the same total loss in beam prediction accuracy in models which employ EWC or dataset mixing. While dataset mixing achieves high accuracies on gNBs included in the training data, further research on approaches that achieve acceptable accuracies on unseen gNBs is needed. Finally, the achieved beam prediction accuracy results have to be assessed in the context of a more holistic view on the network, e.g. by examining the spectral efficiency achieved by the predicted beams, and assessing the overhead of beam prediction with respect to the achieved improvements in reactivity and reduced overhead from beam sweeping.

A

Abbreviations

AoA	angle of arrival
AoD	angle of departure
BIM	beam index map
BPL	Beam Pair Link
CDF	cumulative distribution function
CI	Context Information
CIR	channel impulse response
CNN	convolutional neural network
CS	compressive sensing
CSI	Channel State Information
CSI-RS	Channel State Information - Reference Signals
DL	downlink
eMBB	enhanced Mobile Broadband
EWC	elastic weight consolidation
FCNN	fully connected neural network
FSPL	Free Space Path Loss
gNB	Next Generation Node Base Station
GNSS	global navigation satellite system
GPS	Global Positioning System
IA	Initial Access
IoE	Internet of Everything
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union Radiocommunication Sector

LOS Line-of-Sight

MIMO Multiple-Input Multiple-Output

ML machine learning

mmWave millimeter-Wave

NLOS non-Line-of-Sight

NN neural network

NR New Radio

NSA non-standalone

pdf probability density function

PL Path Loss

PLE Path Loss exponent

pp percentage points

ReLU Rectified Linear Unit

RF radio frequency

RL reinforcement learning

RMS root-mean-squared

RSRP reference signal received power

RSS received signal strength

RX receiver

SA standalone

SINR Signal-to-Interference-plus-Noise Ratio

SL supervised learning

SM spatial multiplexing

SNR Signal-to-Noise Ratio

SRS Sounding Reference Signal

SS Synchronization Signal

TRP transmit-receive point

TX transmitter

UE User Equipment

UL uplink

URLLC Ultra-Reliable, Low Latency Communications

WLAN Wireless Local Area Network

List of Tables

3.1	Exemplary section of the dataset D_{gNB_z} .	26
4.1	NN architecture and hyperparameters for investigating UE position preprocessing.	28
4.2	CNN-S architecture and hyperparameters.	29
4.3	Parameters of the FCNN referred to as FCNN-S.	31
4.4	Parameters of FCNN-L, identical to the FCNN in Table 4.3 with the exception of the structure of hidden layers.	31
4.5	Weights employed in the weighted cross-entropy loss function given N_{label} , according to equation 3.6. The weights of $N_{label} = 5$ are equal to the weights used in other sections of this work.	34
5.1	Number of tunable parameters and median training- and inference-time of five simulations of FCNN-S, FCNN-L and CNN-S for case B1 at gNB 14 with $pd = 100\%$, resulting in 47,821 positions used in the training and 11956 positions used in the testing.	50

List of Figures

2.1	Attenuation due to atmospheric gases (in dB/km) as a function of radio frequency for an atmospheric pressure $P = 1013.25$ hPA, temperature $T = 15$ ÅC and a water vapour density $\rho = 7.5$ g/m ³ [11].	5
2.2	Simulated E-field plot for a phased-array antenna at 28 GHz demonstrating beam tilt [23].	8
2.3	3GPP beam management procedure for 5G NR [29].	10
2.4	Illustration of position based versus environment-aware beam management, adapted from [37].	12
2.5	Structure of a FCNN with T layers, S inputs and U outputs. Blue dots denote the neurons, black arrows denote the connections between the neurons.	14
2.6	Illustration of NLOS BPL failing in a changing environment while the relative positions between UE and gNB is unchanged.	16
3.1	Aerial image of the considered study area near the main station in Frankfurt, Germany. The red square represents the area border, the red arrow shows the location and direction of the pane in Figure 3.2.	19
3.2	Exemplary street level image of a street in the study area, as an illustration for the building density.	19
3.3	Building layout (grey) and gNB locations (blue dots) with their indices in the study area.	20
3.4	UE and gNB antenna patterns in azimuth cut for a selected set of some example ID entries in (a) UE and (b) gNB codebooks [56].	21
3.5	Beam direction of the codebook entries of (a) UE and (b) gNB.	22
3.6	Coverage map of gNB_{14} , showing the maximum RSS (in dBm), corresponding to the best BPL selection, at each UE position as a heat map.	23
3.7	Heat map representation of exemplary BPL matrix, with i denoting the UE beam ID and j denoting the gNB beam ID, at UE position (300,250), with gNB_{14} as TX.	24
3.8	Number of UE positions included in the datasets D_{gNB_z} , with the portion of LOS positions given in blue, and the portion of NLOS positions given in red.	25
4.1	Architecture of the FCNN used for beam prediction [5]. In Dense Layers, every neuron is connected to every neuron in the adjacent layers. The input consists of the two dimensional UE position, the dimensions of the input and each layer are given in brackets.	28
4.2	Architecture of the CNN referred to as CNN-S. The input consists of the UE position in form of the 500×500 one-hot map. In convolutional layers, the value F gives the number of filters, K gives the size of the filters, and S gives the stride. The flatten layer reduces the dimensions to fit the output layer.	30

4.3	Architecture of the FCNN-L. In Dense Layers, every neuron is connected to every neuron in the adjacent layers. The input consists of the two dimensional UE position, the dimensions of the input and each layer are given in brackets. .	32
5.1	Top-1 accuracy plots on initial learning rates and γ for the FCNN-S given in Table 4.1 on gNB_{11} and gNB_{14} without transfer learning. The l_{ri} are given on the x-axis, the multistep learning rate multipliers γ are given with solid, dashed and dotted lines, and the UE position preprocessing modalities S1-S5 are given in different colours.	38
5.2	Top-1 accuracies Ac_{t1} for all combinations of gNB_{ref} and gNB_{tar} , according to B2, with the diagonal $gNB_{tar} = gNB_{ref}$ corresponding to B1, using the UE position preprocessing modalities S1-S5, with $\gamma = 0.3$, $l_{ri} = 0.02$ for preprocessing modalities S2 and S4, and $l_{ri} = 0.002$ for preprocessing modalities S1, S3 and S5.	40
5.3	Coverage map of gNB_3 , representing the maximum RSS at each UE position as a colour map, showing a narrow propagation path, resulting in a low number of distinctive optimal BPLs in the dataset.	41
5.4	Colour map of prediction accuracies (a) Ac_{t1} and (b) Ac_{t5} for B1, using different percentages pd of the datasets D_{gNB}	42
5.5	Ac_{t1} accuracy difference colour map, giving the difference between B1 with $pd = 5\%$ and B2 in pp, with corresponding $gNB^{B1} = gNB_{tar}$ for each combination of gNB_{ref} and gNB_{tar} in B2. $Ac_{t1} > 0$ represents $Ac_{t1}^{B2} > Ac_{t1}^{B1}$. Values on the diagonal $gNB_{ref} = gNB_{tar}$ show the difference in Ac_{t1} between B1 with $pd = 5\%$ and B1 with $pd = 100\%$, with positive values representing higher values for the latter.	42
5.6	B1 prediction accuracies (a) Ac_{t1} and (b) Ac_{t5} at gNB 14 of the FCNN in blue and the CNN in red, over different percentages pd of the datasets $D_{gNB=14}$	44
5.7	B2 (a) Ac_{t1} and (b) Ac_{t5} with $gNB_{ref} = 14$ and gNB_{tar} spanning the x-axis. The blue bars are the accuracy results of the CNN, and the red bars depict the FCNN results. The yellow and purple bars at $gNB_{tar} = 14 = gNB_{ref}$ represent B1 with $pd = 100\%$ of the CNN and FCNN, respectively.	45
5.8	Top-5 accuracy (i.e. Ac_{t5}) of FCNN-S for all combinations of gNB_{ref} and gNB_{tar} , according to B2, with the diagonal $gNB_{tar} = gNB_{ref}$ corresponding to B1.	46
5.9	Top-5 accuracy difference colour map of FCNN-S between B1 with $pd = 5\%$ and B2, with corresponding $gNB^{B1} = gNB_{tar}$. $Ac_{t5} > 0$ represents $Ac_{t5}^{B2} > Ac_{t5}^{B1}$	47
5.10	FCNN-L top-1 accuracies Ac_{t1} (a) and top5 accuracies Ac_{t5} (b) for all combinations of gNB_{ref} and gNB_{tar} , according to B2, with the diagonal $gNB_{tar} = gNB_{ref}$ corresponding to B1.	48
5.11	Ac_{t1} and Ac_{t5} difference colour maps between the FCNN-S results given in Figure 5.2c and Figure 5.8, and the FCNN-L results in Figure 5.10. Positive values represent higher accuracies of FCNN-L.	48
5.12	Ac_{t1} and Ac_{t5} difference colour maps of FCNN-L between B1 with $pd = 5\%$ and B2, with corresponding $gNB^{B1} = gNB_{tar}$. Positive values represent $Ac_{tx}^{B2} > Ac_{tx}^{B1}$	49

5.13	Top-1 Ac_{t1} accuracy colour maps with noise N1-N3 of the FCNN-S on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-S, shown in Figure 5.2c, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3. .	51
5.14	Top-5 Ac_{t5} accuracy colour maps with noise N1-N3 of the FCNN-S on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-S, shown in Figure 5.8, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3. .	53
5.15	Top-1 Ac_{t1} accuracy colour maps with noise N1-N3 of the FCNN-L on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-L, shown in Figure 5.10a, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3. .	54
5.16	Top-5 Ac_{t5} accuracy colour maps with noise N1-N3 of the FCNN-L on the left, with corresponding accuracy difference colour maps with respect to the baseline of FCNN-L, shown in Figure 5.10b, on the right. Higher values in the accuracy difference maps correspond to a higher accuracy loss due to the noise N1-N3. .	55
5.17	CNN-S B2 Ac_{t1} and Ac_{t5} with $gNB_{ref} = 14$ and gNB_{tar} given on the x-axis. The results at $gNB_{ref} = gNB_{tar} = 14$ represent B1 with $pd = 100\%$. The blue and orange bars show the results without UE position noise, using a one-hot map and G1 as input, respectively. The yellow and purple bars show the corresponding results with N1 as UE position noise.	57
5.18	Top-x accuracies from the best BPL in (a) to the fifth best BPL in (e) using scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 10$ on FCNN-S. The values on the x-axis denote the x in the top-x accuracy Ac_{tx}^n , and the number of labels N_{label} used in the weighted cross-entropy loss function is given in different colours. .	59
5.19	Top-x accuracies from the best BPL in (a) to the fifth best BPL in (e) using scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 6$ on FCNN-S. The values on the x-axis denote the x in the top-x accuracy Ac_{tx}^n , and the number of labels N_{label} used in the weighted cross-entropy loss function is given in different colours. .	60
5.20	Top-x accuracies from the best BPL in (a) to the fifth best BPL in (e) using scenario B2 with $gNB_{ref} = 14$ and $gNB_{tar} = 10$, and noise according to N1 on FCNN-S. The values on the x-axis denote the x in the top-x accuracy Ac_{tx}^n , and the number of labels N_{label} used in the weighted cross-entropy loss function is given in different colours.	61
5.21	Beam prediction accuracy Ac_{t1} and Ac_{t5} on the reference gNB after the fine-tuning stage in B2 with $gNB_{ref} = 14$ for varying gNB_{tar} . The results at $gNB_{ref} = gNB_{tar} = 14$ represent B1 with $pd = 100\%$, thus the difference to the results at $gNB_{tar} = 14$ can be seen as accuracy lost on the reference gNB by fine-tuning on the respective gNB_{tar} . The results for FCNN-S are shown in blue, the orange and yellow bars show the results of FCNN-L and CNN-S, respectively.	63
5.22	Dataset sizes and portion of UE positions included in the dataset in relation to gNB 14, in blue and orange, respectively.	64

5.23	Top-1 and top-5 accuracies on FCNN-SD after B2 with $gNB_{ref} = 14$ on the test data of the reference gNB and the target gNB in blue and orange, respectively. The corresponding accuracies for employing EWC in the fine-tuning step are shown in yellow and purple, respectively. The values at gNB 14 represent B1 with $pd = 100\%$	66
5.24	CDFs of the biases of the second hidden layer of FCNN-SD after B1 with $pd = 100\%$ on gNB 10, 14 and 19 are shown in (a) in blue, yellow and orange, respectively. The CDFs of the biases of the second layer of FCNN-SD after training on task A and task B of the MNIST dataset are shown in (b) in blue and orange, respectively.	67
5.25	M1 and M2 top-1 (a) and top-5 (b) accuracies of FCNN-S in blue and yellow, respectively. The corresponding M1 and M2 accuracies of FCNN-SG are shown in orange and purple, respectively. The mixed dataset contains every gNB except gNB 10 and 13, with $pd = 60\%$	68
5.26	M1 and M2 top-1 and top-5 accuracies of FCNN-L in blue and yellow, respectively. The corresponding M1 and M2 accuracies of FCNN-LG are shown in orange and purple, respectively. The mixed dataset contains every gNB except gNB 10 and 13, with $pd = 60\%$	69
5.27	M1 and M2 top-1 and top-5 accuracies of CNN-S in blue and yellow, respectively. The corresponding M1 and M2 accuracies of CNN-SR are shown in orange and purple, respectively. The mixed dataset contains every gNB except gNB 10 and 13, with $pd = 60\%$	70
5.28	M1 accuracies of FCNN-SG in orange and FCNN-LG in purple. The B2 accuracies using $gNB_{ref} = 14$ of FCNN-S and FCNN-L are shown in blue and yellow, respectively, with gNB 14 being the B1 accuracy with $pd = 100\%$	71

Bibliography

- [1] P. Jonsson, P. Cerwall, A. Lundval, D. von Koch, S. Davis, R. R. B. Group), T. H. (Sony), and D. Y. (Sony), "Ericsson mobility report," Ericsson, Tech. Rep., 2025. [Online]. Available: <https://www.ericsson.com/49e9b6/assets/local/reports-papers/mobility-report/documents/2025/ericsson-mobility-report-june-2025.pdf>
- [2] W. Hong, Z. H. Jiang, C. Yu, D. Hou, H. Wang, C. Guo, Y. Hu, L. Kuai, Y. Yu, Z. Jiang, Z. Chen, J. Chen, Z. Yu, J. Zhai, N. Zhang, L. Tian, F. Wu, G. Yang, Z.-C. Hao, and J. Y. Zhou, "The role of millimeter-wave technologies in 5g/6g wireless communications," *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 101–122, Jan. 2021.
- [3] A. M. Nor, S. Halunga, and O. Fratu, "Survey on positioning information assisted mmwave beamforming training," *Ad Hoc Networks*, vol. 135, p. 102947, Oct. 2022.
- [4] M. Qurratulain Khan, A. Gaber, P. Schulz, and G. Fettweis, "Machine learning for millimeter wave and terahertz beam management: A survey and open challenges," *IEEE Access*, vol. 11, pp. 11 880–11 902, 2023.
- [5] E. Tosi, P. Hu, A. Ichkov, M. Petrova, and L. Simić, "Cross-environment transfer learning for location-aided beam prediction in 5g and beyond millimeter-wave networks," 2025.
- [6] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, May 2020.
- [7] I. T. Union, "Final acts," in *World Radiocommunication Conference 2023 (WRC-23)*. International Communication Union, 2023.
- [8] J. Skold and C. Taylor, "5g; nr; base station (bs) radio transmission and reception (3gpp ts 38.104 version 18.5.0 release 18)," Tech. Rep.
- [9] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.
- [10] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 2002.
- [11] G. A. Siles, J. M. Riera, and P. Garcia-del Pino, "Atmospheric attenuation in wireless communication systems at millimeter and thz frequencies [wireless corner]," *IEEE Antennas and Propagation Magazine*, vol. 57, no. 1, pp. 48–61, Feb. 2015.

- [12] I. T. U. (ITU), "Attenuation by atmospheric gases: Recommendation itu-r p.676-13," International Telecommunication Union (ITU), Tech. Rep., 2023. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.676-13-202208-!!PDF-E.pdf
- [13] G. R. MacCartney, J. Zhang, S. Nie, and T. S. Rappaport, "Path loss models for 5g millimeter wave propagation channels in urban microcells," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 3948–3953.
- [14] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1616–1653, 2018.
- [15] R. Medhurst, "Rainfall attenuation of centimeter waves: Comparison of theory and measurement," *IEEE Transactions on Antennas and Propagation*, vol. 13, no. 4, pp. 550–564, 1965.
- [16] I. A. 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 & Tutorials*, vol. 20, no. 2, pp. 870–913, 2018.
- [17] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 949–973, 2016.
- [18] 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.
- [19] I. Rodriguez, H. C. Nguyen, T. B. Sorensen, J. Elling, J. A. Holm, P. Mogensen, and B. Vejlgaard, "Analysis of 38 ghz mmwave propagation characteristics of urban scenarios," in *Proceedings of European Wireless 2015; 21th European Wireless Conference*, 2015, pp. 1–8.
- [20] 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 *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 5163–5167.
- [21] F. Khan and Z. Pi, "mmwave mobile broadband (mmb): Unleashing the 3-300ghz spectrum," in *34th IEEE Sarnoff Symposium*, 2011, pp. 1–6.
- [22] V. Raghavan, V. Podshivalov, J. Hulten, M. A. Tassoudji, A. Sampath, O. H. Koymen, and J. Li, "Spatio-temporal impact of hand and body blockage for millimeter-wave user equipment design at 28 ghz," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 46–52, 2018.
- [23] A. Bansal, C. J. Panagamuwa, and W. G. Whittow, "State-of-the-art millimeter-wave beam steering antennas for beyond 5g and 6g networks: A comprehensive review," *IEEE Antennas and Propagation Magazine*, vol. 66, no. 5, pp. 40–51, 2024.

- [24] H. Yang, M. H. A. J. Herben, I. J. A. G. Akkermans, and P. F. M. Smulders, "Impact analysis of directional antennas and multiantenna beamformers on radio transmission," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1695–1707, 2008.
- [25] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.
- [26] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3gpp nr at mmwave frequencies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 173–196, 2019.
- [27] S. Antipolis, "3gpp tr 21.915 version 15.0.0 release 15," 3GPP, techreport, 2019.
- [28] —, "Study on new radio access technology - physical layer aspects - release 14," 3GPP, techreport TR 38.802, 2017.
- [29] Q. Xue, C. Ji, S. Ma, J. Guo, Y. Xu, Q. Chen, and W. Zhang, "A survey of beam management for mmwave and thz communications towards 6g," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 3, pp. 1520–1559, 2024.
- [30] L. Wei, Q. Li, and G. Wu, "Exhaustive, iterative and hybrid initial access techniques in mmwave communications," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.
- [31] M. Giordani, M. Mezzavilla, C. N. Barati, S. Rangan, and M. Zorzi, "Comparative analysis of initial access techniques in 5g mmwave cellular networks," in *2016 Annual Conference on Information Science and Systems (CISS)*, 2016, pp. 268–273.
- [32] H. Hassanieh, O. Abari, M. Rodriguez, M. Abdelghany, D. Katabi, and P. Indyk, "Fast millimeter wave beam alignment," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 432–445. [Online]. Available: <https://doi.org/10.1145/3230543.3230581>
- [33] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [34] A. Zhou, T. Wei, X. Zhang, and H. Ma, "Fastnd: Accelerating directional neighbor discovery for 60-ghz millimeter-wave wireless networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 5, pp. 2282–2295, 2018.
- [35] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "BeamSpy: Enabling robust 60 GHz links under blockage," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*. Santa Clara, CA: USENIX Association, Mar. 2016, pp. 193–206. [Online]. Available: <https://www.usenix.org/conference/nsdi16/technical-sessions/presentation/sur>
- [36] A. Ichkov, S. Häger, P. Mähönen, and L. Simić, "Comparative evaluation of millimeter-wave beamsteering algorithms using outdoor phased antenna array measurements," in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2022, pp. 497–505.

- [37] Y. Zeng, J. Chen, J. Xu, D. Wu, X. Xu, S. Jin, X. Gao, D. Gesbert, S. Cui, and R. Zhang, "A tutorial on environment-aware communications via channel knowledge map for 6g," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 3, pp. 1478–1519, 2024.
- [38] M. Hashemi, C. E. Koksall, and N. B. Shroff, "Out-of-band millimeter wave beam-forming and communications to achieve low latency and high energy efficiency in 5g systems," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 875–888, 2018.
- [39] Y. Heng and J. G. Andrews, "Machine learning-assisted beam alignment for mmwave systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 4, pp. 1142–1155, 2021.
- [40] D. d. S. Brilhante, J. C. Manjarres, R. Moreira, L. de Oliveira Veiga, J. F. de Rezende, F. Müller, A. Klautau, L. Leonel Mendes, and F. A. P. de Figueiredo, "A literature survey on ai-aided beamforming and beam management for 5g and 6g systems," *Sensors*, vol. 23, no. 9, 2023. [Online]. Available: <https://www.mdpi.com/1424-8220/23/9/4359>
- [41] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, May 2014. [Online]. Available: <https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>
- [42] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," Nov. 2015.
- [43] J. Morais, A. Bchboodi, H. Pezeshki, and A. Alkhateeb, "Position-aided beam prediction in the real world: How useful gps locations actually are?" in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 1824–1829.
- [44] S. Rezaie, C. N. Manchón, and E. de Carvalho, "Location- and orientation-aided millimeter wave beam selection using deep learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.
- [45] M. Modsching, R. Kramer, and K. ten Hagen, "Field trial on gps accuracy in a medium size city: The influence of built-up," in *3rd workshop on positioning, navigation and communication*, vol. 2006, 2006, pp. 1–10.
- [46] D. Wu, Y. Zeng, S. Jin, and R. Zhang, "Environment-aware and training-free beam alignment for mmwave massive mimo via channel knowledge map," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, pp. 1–7.
- [47] A. Ichkov, P. Mäkelä, and L. Simić, "Is ray-tracing viable for millimeter-wave networking studies?" in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–7.
- [48] Z. Zhang, J. Ryu, S. Subramanian, and A. Sampath, "Coverage and channel characteristics of millimeter wave band using ray tracing," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 1380–1385.

- [49] M. Wang, Y. Lin, Q. Tian, and G. Si, "Transfer learning promotes 6g wireless communications: Recent advances and future challenges," *IEEE Transactions on Reliability*, vol. 70, no. 2, pp. 790–807, 2021.
- [50] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, Mar. 2017.
- [51] S. Madhekwana, M. A. Usman, A. Ayyub, and C. Politis, "Beam alignment for mmwave and thz: systematic review," *Telecommunication Systems*, vol. 88, no. 3, Jun. 2025.
- [52] Q. Xue, J. Guo, B. Zhou, Y. Xu, Z. Li, and S. Ma, "Ai/ml for beam management in 5g-advanced: A standardization perspective," *IEEE Vehicular Technology Magazine*, vol. 19, no. 4, pp. 64–72, 2024.
- [53] M. Akrouf, A. Feriani, F. Bellili, A. Mezghani, and E. Hossain, "Domain generalization in machine learning models for wireless communications: Concepts, state-of-the-art, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 3014–3037, 2023.
- [54] S. Rezaie, A. Amiri, E. de Carvalho, and C. N. Manchón, "Deep transfer learning for location-aware millimeter wave beam selection," *IEEE Communications Letters*, vol. 25, no. 9, pp. 2963–2967, 2021.
- [55] S. Rezaie, J. Morais, A. Alkhateeb, and C. N. Manchón, "Device-agnostic millimeter wave beam selection using machine learning," *arXiv preprint arXiv:2211.12125*, 2022.
- [56] E. Tosi, "Effect of localization error on the performance of location-aided beam-management in 5g and beyond," Master's thesis, ICT for internet and multimedia, Università degli Studi di Padova, 2023.
- [57] Remcom, "Wireless insite," <http://www.remcom.com/wireless-insite>, accessed: 2024-08-01.
- [58] P. Reisdorf, T. Pfeier, J. Breßler, S. Bauer, P. Weissig, S. Lange, G. Wanielik, and P. Protzel, "The problem of comparable gnss results - an approach for a uniform dataset with low-cost and reference data," 11 2016.
- [59] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," *IEEE Transactions on Nuclear Science*, vol. 44, no. 3, pp. 1464–1468, 1997.
- [60] Y. LeCun, C. Cortes, and C. J. Burges, *The mnist database of handwritten digits*, Std., 1998.
- [61] S. Moriarity, *Continual Learning with Elastic Weight Consolidation in TensorFlow 2*, Std., Oct. 2020. [Online]. Available: <https://seanmoriarity.com/2020/10/18/continual-learning-with-ewc/>

Eidesstattliche Versicherung

Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)

Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/
Masterarbeit* mit dem Titel

I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting)
erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.
Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich,
dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in
gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than
the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written
and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:

Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung
falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei
Jahren oder mit Geldstrafe bestraft.

Para. 156 StGB (German Criminal Code): False Statutory Declarations

Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely
testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so
tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Strafflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158
Abs. 2 und 3 gelten entsprechend.

Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence

(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not
exceeding one year or a fine.

(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2)
and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature