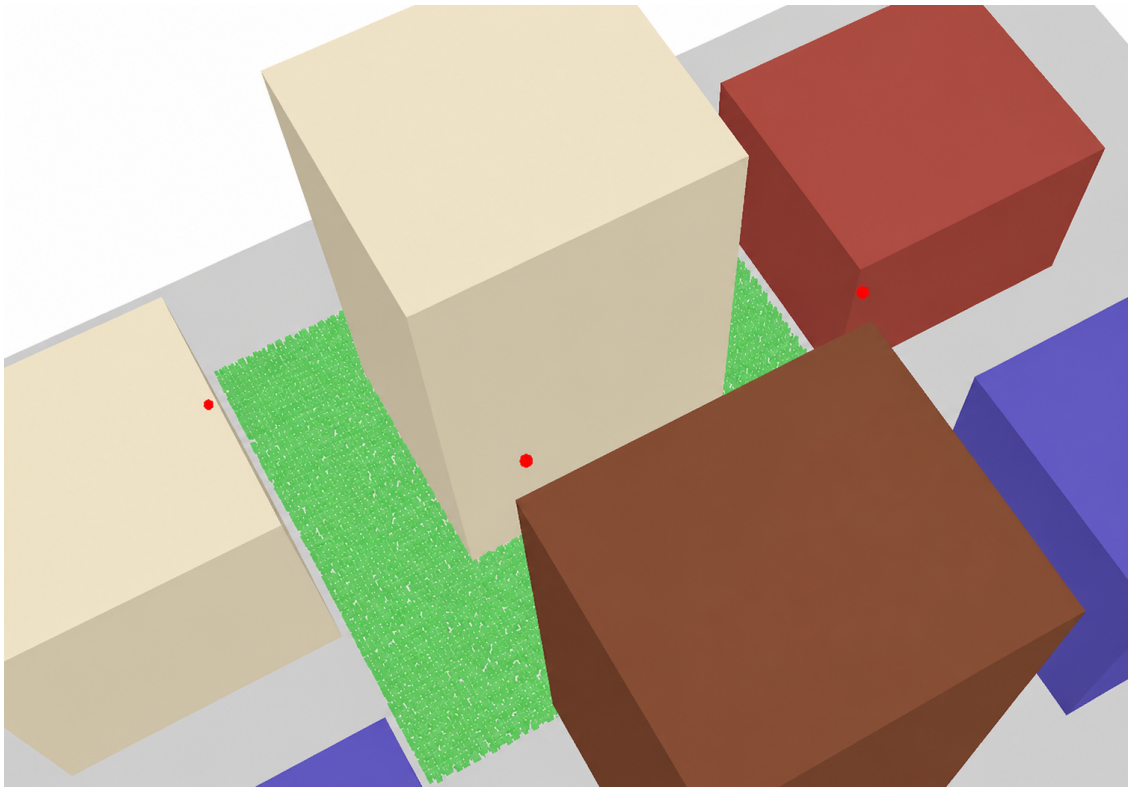
CHALMERS
UNIVERSITY OF TECHNOLOGY



AI for Improved Indoor Positioning via Multi-Band Channel Charting

Bachelor's thesis in Electrical Engineering

Philip Corsénsus
Viktor Jordansson
Max Monastyrski
Isac Nilsson

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Bachelor's thesis 2026

AI for Improved Indoor Positioning via Multi-Band Channel Charting

Philip Corsénsus
Viktor Jordansson
Max Monastyrski
Isac Nilsson

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

AI for Improved Indoor Positioning via Multi-Band Channel Charting
Philip Corsénsus
Viktor Jordansson
Max Monastyrski
Isac Nilsson

Supervisors: Yuhao Zhang & José Miguel Mateos Ramos, Department of Electrical Engineering
Examiner: Henk Wymeersch, Department of Electrical Engineering

Bachelor's thesis 2026
Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

Cover: Render of the simulated environment with transmitters (red) and receiver placement (green).

Typeset in LaTeX
Gothenburg, Sweden 2026

AI for Improved Indoor Positioning via Multi-Band Channel Charting
Philip Corsénsus
Viktor Jordansson
Max Monastyrski
Isac Nilsson
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Channel charting is an unsupervised positioning method that learns a low-dimensional spatial representation from channel state information (CSI), which describes how wireless signals propagate between a user and base stations. Unlike fingerprinting, it does not require ground-truth position labels during training. This thesis investigates whether using CSI from two frequency bands improves channel charting compared with conventional single-band CSI. Simulated CSI was generated in a street-canyon environment at 3.5 GHz and 12 GHz using Sionna RT. Three dual-band channel charting methods were evaluated: averaging dissimilarities between CSI samples, multiplying similarity scores from both bands, and aligning two separately trained networks. The methods were compared with single-band channel charting and with supervised fingerprinting baselines. The results show that dual-band fusion improves channel charting performance across both chart-quality metrics and positioning accuracy. The best channel charting result was obtained with similarity multiplication, reducing the mean absolute error from 6.67 m for the best single-band reference to 5.86 m. Dual-band fingerprinting also improved performance, reducing the mean absolute error from 1.09 m to 1.02 m, although the relative improvement was smaller than for channel charting. The gains were strongest in non-line-of-sight conditions, where the best channel charting error decreased from 7.64 m for the best single-band reference to 6.54 m with dual-band similarity multiplication. These results indicate that multi-band CSI provides complementary spatial information and is especially useful for unsupervised channel charting in challenging propagation environments.

Keywords: Channel Charting, Channel State Information, multi-band fusion, wireless localization, fingerprinting, unsupervised learning.

Acknowledgements

We would like to thank our supervisors, José Miguel Mateos Ramos and Yuhao Zhang, for their continuous support, insightful discussions and constructive feedback throughout this project. Their deep knowledge of the field and ability to clarify the current research landscape in Channel Charting have been essential in shaping our understanding and guiding the direction of this work. We would also like to thank Henk Wymeersch for his valuable feedback and suggestions during the project.

Philip Corsénsus, Gothenburg, May 2026
Viktor Jordansson, Gothenburg, May 2026
Max Monastyrski, Gothenburg, May 2026
Isac Nilsson, Gothenburg, May 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis in order of appearance:

ADP	Angle-Delay Profile
BS	Base Station
CC	Channel Charting
CDF	Cumulative Distribution Function
CFR	Channel Frequency Response
CIR	Channel Impulse Response
CSI	Channel State Information
CT	Continuity
FCF	Forward Charting Function
FFT	Fast Fourier Transform
FP	Fingerprinting
IFFT	Inverse Fast Fourier Transform
kNN	k-Nearest Neighbors
KS	Kruskal's Stress
LoS	Line-of-Sight
MAE	Mean Absolute Error
MSE	Mean Squared Error
NLoS	Non-Line-of-Sight
NN	Neural Network
OFDM	Orthogonal Frequency Division Multiplexing
TW	Trustworthiness
UE	User Equipment

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables that have been used throughout this thesis. Matrices and tensors are denoted with boldface capital letters, vectors are denoted with boldface lowercase letters.

Indices

l	Index for CSI datapoint
i, j	Indices for datapoint pairs
b	Index for base station array
m	Index for antenna
μ	Index for angular-domain antenna component after FFT
n	Index for OFDM subcarrier
τ	Index for delay tap
q	Index for carrier frequency

Sets and Dimensions

L	Total number of CSI datapoints
B	Number of base station arrays
M	Number of antennas per array
N_{sub}	Number of OFDM subcarriers
N	Training batch size
T	Number of time-domain taps
D	Spatial dimensionality of the position and channel chart space
K	Neighborhood size used for CT and TW evaluation

Parameters

$\tau_{\min}$, $\tau_{\max}$	Minimum and maximum delay tap indices
$t_{\text{threshold}}$	Time threshold for timestamp fusion
γ	Timestamp dissimilarity scaling factor
α	Alignment loss weighting factor
β	Stability constant in the Siamese loss function
f_1, f_2	Carrier frequencies of band 1 and band 2
k	Number of nearest neighbors in the kNN graph
$\mathbf{A}$	Affine transformation matrix
$\mathbf{b}$	Affine translation vector
$\hat{\mathbf{A}}$, $\hat{\mathbf{b}}$	Estimated affine transformation parameters
θ	Learnable channel charting model parameters
ϕ	Learnable fingerprinting model parameters

Variables

$\mathbf{H}_{f_q}^{(l)}$	Frequency-domain CSI for datapoint l at carrier frequency f_q
$\tilde{\mathbf{H}}_{f_q}^{(l)}$	Time-domain CSI for datapoint l at carrier frequency f_q
$\mathbf{F}^{(l)}$	Processed angle-delay feature tensor for datapoint l
$\mathbf{f}^{(l)}$	Feature vector for datapoint l
$\mathbf{f}_{f_1}^{(l)}$, $\mathbf{f}_{f_2}^{(l)}$	Feature vectors for datapoint l from carrier frequencies f_1 and f_2
$\mathbf{f}_{\text{dual}}^{(l)}$	Concatenated dual-band feature vector for datapoint l
$\mathbf{x}^{(l)}$	Ground-truth position for datapoint l
$\hat{\mathbf{x}}^{(l)}$	Predicted position from the fingerprinting model
$t^{(l)}$	Timestamp of datapoint l
$d_{\text{time},i,j}$	Timestamp dissimilarity between datapoints i and j
$d_{\text{ADP},i,j}$	ADP-based dissimilarity between datapoints i and j
$d_{\text{Fused},i,j}$	Timestamp-fused dissimilarity between datapoints i and j
$d_{\text{G-Fused},i,j}$	Geodesic dissimilarity between datapoints i and j
$d_{\text{SM-ADP},i,j}$	Similarity-multiplication ADP dissimilarity between datapoints i and j
$\mathbf{D}_{\text{ADP}}$	ADP-based dissimilarity matrix
$\mathbf{D}_{\text{time}}$	Timestamp dissimilarity matrix
$\mathbf{D}_{\text{Fused}}$	Timestamp-fused dissimilarity matrix
$\mathbf{D}_{\text{G-Fused}}$	Geodesic timestamp-fused dissimilarity matrix

$\mathbf{D}_{\text{Avg-ADP}}$	Dual-band ADP dissimilarity matrix obtained by averaging the two single-band ADP matrices
$\mathbf{D}_{\text{G-Avg-Fused}}$	Geodesic dissimilarity matrix after ADP averaging and timestamp fusion
$\mathbf{D}_{\text{G-SM-Fused}}$	Geodesic dissimilarity matrix after similarity multiplication and timestamp fusion
$\mathbf{G}_{k\text{-NN}}$	k-nearest-neighbor graph
$S_{b,i,j}^{f_q}$	Normalized cosine similarity score for frequency band q between datapoints i and j
$\mathcal{C}_\theta$	Forward charting function
$\mathcal{F}_\phi$	Fingerprinting regression model
$\mathbf{z}^{(l)} \in \mathbb{R}^D$	Channel chart embedding for datapoint l
$\hat{\mathbf{z}}^{(l)}$	Affine-aligned channel chart coordinate for datapoint l
$\mathbf{z}_{f_1}^{(l)}$, $\mathbf{z}_{f_2}^{(l)}$	Channel chart embeddings from the two single-band branches
$\mathbf{z}_{\text{Avg}}^{(l)}$	Averaged dual-network channel chart embedding
$r_{l,i}$	Rank of datapoint i as a neighbor of datapoint l in physical space
$\hat{r}_{l,i}$	Rank of datapoint i as a neighbor of datapoint l in representation space
$V_K^{(l)}$	Set of the K nearest neighbors of datapoint l in physical space
$u_K^{(l)}$	Set of the K nearest neighbors of datapoint l in representation space

Contents

List of Figures

List of Tables

1

Introduction

Global navigation satellite systems can locate a phone to within a few meters when used outdoors with a clear view of the sky [2]. In indoor and dense urban environments, however, buildings often block the line-of-sight (LoS) to satellites. As a result, positioning accuracy degrades sharply, signal acquisition becomes slower, and the estimated position may be incorrect [3].

Beyond satellite-based positioning, modern cellular networks are increasingly expected to provide geolocation services of their own [2]. The deployment of multi-antenna systems such as 5G has made rich channel state information (CSI) widely available [4]: from the CSI measured at each base-station antenna, properties such as angles of arrival, propagation delays, and the multipath structure of the channel can be inferred [1]. At present, however, this information is rarely retained or exploited for positioning [4].

Fingerprinting is an established CSI-based approach to localization [3]. Measurements collected at known locations are paired with their physical coordinates and used to train a model that maps subsequent measurements to position. The method is effective, but the upfront data collection – labeling each measurement with a ground-truth position – is laborious and costly [3]. If position could be inferred from CSI without any ground-truth labels, this burden would disappear. That is precisely the goal of "Channel Charting" (CC).

CC has emerged as an alternative to conventional fingerprinting-based localization [4,5]. CC is an unsupervised framework that learns a low-dimensional representation of wireless channel measurements while preserving local spatial relationships between users [4]. Since CC operates in an unsupervised manner and does not require ground-truth position labels during training [4], it avoids the costly data-labeling process associated with supervised methods, making it a more scalable solution for wireless networks.

Most existing CC approaches rely on measurements from a single frequency band. However, different frequency bands exhibit distinct propagation characteristics, such as penetration capability, path loss, and sensitivity to blockage [6]. These differences suggest that measurements collected across multiple frequency bands may provide complementary spatial information. Combining information from multiple bands could potentially improve localization performance.

The potential advantages of combining multiple frequency bands have already been demonstrated in supervised fingerprinting-based positioning systems, where multi-

band fusion has been shown to improve localization accuracy [7]. However, whether similar benefits can be achieved in unsupervised CC remains unexplored.

1.1 Purpose

The purpose of this thesis is to investigate the potential advantages of multi-band based CC, where CSI from multiple frequency bands is utilized instead of a single band. In addition, the performance of multi-band based CC is compared with conventional single-band CC as well as with both single-band and multi-band fingerprinting-based positioning approaches.

2

Background

In 2018, a research group published a scientific paper [4] in which they described a method to analyze signals in multi-antenna networks, such as 5G and future 6G networks [5], using machine learning to determine relative positions of receiver devices. By exploiting various signal properties and propagation characteristics together with unsupervised machine learning, they created a framework for dimensionality reduction, which they chose to call Channel Charting.

This chapter provides the necessary background for the work. It begins with an introduction to CSI and its role in wireless propagation modeling. The concept of CC is then presented, along with its connection to fingerprinting-based positioning and neural network-based mapping methods. Finally, the project's limitations and demarcations are outlined.

2.1 Channel State Information

When a signal is transmitted over a wireless channel, it is affected by path loss, multipath propagation, scattering, and fading [8]. The combined effect of these phenomena on the transmitted signal can be described by the Channel Impulse Response (CIR) [8], which characterizes how the channel responds to an impulse at a given point in time. In the time domain, the received signal $y(t)$ is related to the transmitted signal $x(t)$ through the convolution

$$y(t) = h(t) * x(t) + n(t) \tag{2.1}$$

where $h(t)$ is the CIR and $n(t)$ is additive noise. In practice, the channel is estimated in the frequency domain, where convolution becomes elementwise multiplication

$$Y(f) = H(f) \cdot X(f) + N(f). \tag{2.2}$$

Here, $H(f)$ is the Channel Frequency Response (CFR), which is the Fourier transform of $h(t)$ and describes the complex-valued gain of the channel at each frequency. In an orthogonal frequency-division multiplexing (OFDM) system [8], with N_{sub} OFDM subcarriers, M antennas and B base stations, the CFR is estimated at the base station as part of the normal communication protocol [1] [8], yielding the CSI tensor

$$\mathbf{H}^{(l)} \in \mathbb{C}^{B \times M \times N_{\text{sub}}} \tag{2.3}$$

where each entry $\mathbf{H}^{(l)}$ encodes the complex-valued amplitude and phase response of the channel at a particular antenna array, antenna, and subcarrier. The variable l indexes samples points, this is described further in Section 3.1. This tensor $\mathbf{H}^{(l)}$ is what is referred to as CSI [8]. When needed, the time-domain representation $\tilde{\mathbf{H}}^{(l)} \in \mathbb{C}^{B\times M\times T}$ is recovered by applying an inverse fast Fourier transform (IFFT) along the subcarrier axis

$$\tilde{\mathbf{H}}^{(l)}_{b,m,\tau} = \mathrm{IFFT}_n \left\{ \mathbf{H}^{(l)}_{b,m,n} \right\} \tag{2.4}$$

where τ indexes the time-domain taps, and each tap corresponds to a discrete propagation delay that aggregates the multipath components arriving with similar delays.

In realistic propagation environments, a transmitted signal reaches the receiver via multiple paths, each caused by reflections, diffractions and scattering from surrounding objects [8]. These multipath components arrive at the receiver with different delays, attenuations, and angles of arrival, which manifests in the CIR $\tilde{\mathbf{H}}^{(l)}$ [8] as energy spread across multiple taps τ. The structure of the CIR is highly dependent on the physical surroundings of the transmitter and receiver, which is precisely what makes CSI a useful input for position-related inference. The dominant path arrives at the earliest tap and typically carries the most reliable geometric information. These taps are predominantly LoS or short NLoS components, whose delays are almost directly proportional to the UE-antenna distance. Later taps are largely composed of reflected components with longer paths that depend on the surrounding environment rather than on the direct UE-antenna distance [1].

2.1.1 Frequency Dependence

The propagation characteristics of a wireless channel are inherently frequency-dependent [8]. At lower frequencies, signals experience less path loss and are better able to penetrate obstacles such as walls and buildings, making them suitable for coverage in challenging environments. At higher frequencies, the shorter wavelength results in finer resolution in both the delay and angular domains, enabling more precise estimation of multipath components [8]. As a consequence, channels at substantially different carrier frequencies interact with the same physical environment in different ways: wavelength-dependent reflection, scattering, and diffraction lead each band to emphasise different aspects of the propagation scene.

Most CC systems operate on CSI collected from a single frequency band, which captures only a partial view of the radio environment. Using multiple frequency bands simultaneously diversifies the spatial content of the CSI, providing richer and more complementary information about the propagation scene [7].

The waveform used in this project is OFDM since it is the standard signaling format in modern wireless systems from 4G onwards [9]. It naturally provides frequency-domain CSI across all subcarriers within its operating bandwidth, although the propagation characteristics within a single band vary only modestly, given that the bandwidth is small relative to the carrier frequency. The motivation for using multiple frequency bands simultaneously in CC, therefore, does not come from within-band

variation, but from the fact that channels at substantially different carrier frequencies interact with the propagation environment differently. Wavelength-dependent reflection, scattering, and diffraction cause two bands to provide complementary views of the same physical scene, which is the property exploited in this work [7].

The benefits of multi-band information have already been demonstrated for fingerprinting-based positioning [7], where combining features from multiple bands consistently reduces localization error. Whether equivalent gains can be achieved in the unsupervised CC setting – where no ground-truth labels are available and spatial information must be extracted purely from CSI structure – remains an open question and constitutes one of the primary motivations for this work.

2.2 Positioning with CSI

CSI, introduced in Section 2.1, depends on the physical propagation environment and therefore implicitly reflects the spatial locations of the transmitter and receiver. CSI typically varies smoothly over space, so that measurements taken at nearby locations tend to be similar [4, 5]. As a result, relative spatial relationships between users can be inferred from the structure of the CSI data alone. This makes CSI well-suited for positioning applications, where location is inferred directly from measured channel characteristics rather than explicit geometric models.

One common approach that leverages this idea is fingerprinting, where previously observed CSI measurements are associated with known positions and later used for localization [3, 7, 10]. Another approach that leverages this idea is CC, where a low-dimensional representation — a chart — is learned directly from CSI measurements such that nearby measurements in physical space remain nearby in the chart, without requiring any ground-truth positions during training [4].

2.2.1 Fingerprinting

Fingerprinting is a widely used approach for positioning in wireless systems, particularly in indoor environments where satellite-based positioning systems fail [10]. The method relies on collecting CSI at known locations and associating these measurements with physical coordinates. Typical features used in fingerprinting include Received Signal Strength (RSS), Time-of-Arrival (ToA) or Angle-of-Arrival (AoA).

Fingerprinting consists of two phases [10], as illustrated by Figure 2.1. The offline phase includes constructing a database of fingerprints by measuring radio features at a set of reference locations and labeling them with ground-truth coordinates. Machine learning models are then trained to learn a mapping from radio features to spatial positions. In the online phase, features from unseen real-time samples are fed to the machine learning algorithm to infer the estimated position.

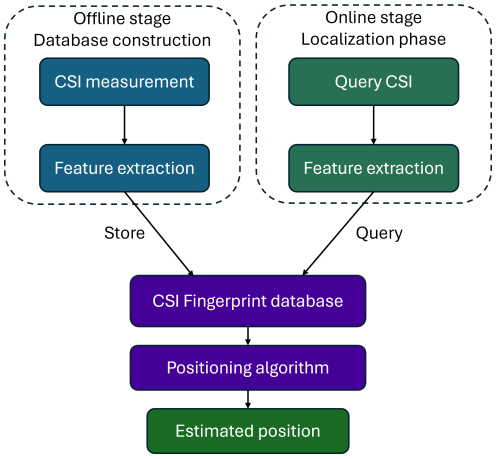


Figure 2.1: Offline and online phases in CSI fingerprinting.

Despite its effectiveness, fingerprinting suffers heavily from several limitations [11]. The requirement for extensive position-labeling campaigns makes the offline phase costly, time-consuming, and inefficient, especially in larger environments. Furthermore, fingerprinting methods are highly sensitive to changes in the radio environment, such as moving objects, human presence, or modifications to infrastructure, which significantly degrade positioning accuracy. This lack of robustness and scalability limits the long-term applicability of the method.

Nevertheless, fingerprinting remains an important reference method and will be used as a performance baseline for evaluating more advanced positioning techniques. It is included as a baseline due to its widespread use in indoor positioning literature [3, 10], enabling meaningful comparison between supervised and unsupervised approaches. Its limitations motivate the development of alternative approaches, such as CC, which aim to extract spatial information from wireless measurements without requiring labeled position data.

2.2.2 Dissimilarity-Metric based Channel Charting

CC is an approach to directly extract the spatial geometry of the radio environment from wireless channel measurements, without requiring ground-truth location information [4]. Instead of learning an explicit mapping from channel features to absolute coordinates, CC aims to construct a low-dimensional representation, referred to as the channel chart, that preserves the relative spatial relationships among users.

The assumption that enables CC is that CSI often varies gradually with user position [8], so that CSI measurements collected from nearby locations tend to be more similar than those collected farther apart [4]. This relationship enables the direct inference of spatial information from the similarity structure of the CSI data.

In practice, a common approach is to compute pairwise dissimilarities between all collected samples using a dissimilarity metric [1, 12]. The metric takes two CSI measurements as input and returns a single number that grows with the physical distance between the two corresponding UE positions, so that two samples taken

close to each other yield a small dissimilarity, while two samples taken at opposite ends of the room yield a large one.

However, the gradual-variation assumption only holds locally [12]. Over short distances, CSI changes smoothly with position; over longer distances, the relationship between CSI similarity and physical distance breaks down, and direct dissimilarities between distant samples become unreliable proxies for true distance. The geodesic transformation [12] addresses this by making the dissimilarities better reflect the underlying data manifold. Instead of using the direct dissimilarity between two distant samples, the distance between them is computed as the sum of small, locally reliable dissimilarities along the shortest path through neighboring samples in the dataset, as shown in Figure 2.2.

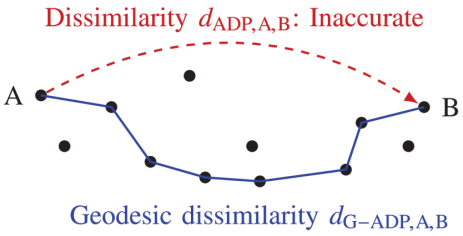


Figure 2.2: A figure describing a geodesic dissimilarity from [1] © 2024 IEEE

CC does not use CSI strictly, it also benefits from side information such as timestamps, which were suggested in early work of CC [13] and are now commonly used. It is based on the fact that two CSI measurements between a BS and the same UE, measured close in time, are also spatially close to each other [1]. This is particularly useful for the UE transition between LoS and NLoS, which often leads to a high CSI dissimilarity despite the samples being spatially close. In such cases, a timestamp can often provide a more accurate distance measurement than CSI [1]. Timestamps can be viewed as a dissimilarity metric and are defined as the absolute time difference between two measurements. However, timestamps as a dissimilarity metric fail to provide meaningful information in cases with large absolute time differences and may give misleading information when the UE does not move. Therefore, timestamps have to be fused with a CSI-based dissimilarity metric to provide more meaningful information.

The pairwise dissimilarities described above define the local geometric structure of the dataset, but do not directly provide a low-dimensional representation of the radio environment. To obtain a channel chart, a mapping must therefore be learned from the high-dimensional CSI features to a low-dimensional embedding space that preserves these dissimilarity relationships [4].

$$\mathcal{C}_\theta : \mathbf{H}^{(l)} \mapsto \mathbf{z}^{(l)} \in \mathbb{R}^D \tag{2.5}$$

such that the distance in the channel chart preserves local neighborhood structure between UE position, i.e.,

$$\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|_2 \propto \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2, \quad \forall i, j \in \{1, \ldots, L\}. \tag{2.6}$$

In practice, the mapping $\mathcal{C}_\theta$ is commonly implemented using a neural network trained in an unsupervised manner [14]. Rather than using ground-truth position labels, the network is trained to preserve the pairwise dissimilarity relationships derived from the CSI measurements. The resulting mapping from CSI to low-dimensional channel chart coordinates is referred to as a Forward Charting Function (FCF) [4].

The resulting CC preserves neighborhood relations rather than absolute positions [15], which is sufficient for many applications, such as user clustering, mobility prediction, beam management, and relative positioning [5]. However, the obtained channel chart can be rotated, sheared, or scaled relative to the true positions in physical space [15]. To relate the channel chart to real-world coordinates, some form of reference information is therefore required.

This reference information can be obtained in different ways. One possibility is to use anchor points, where both the CSI measurement and the corresponding physical position are known. For example, through a model-based approach, where LoS conditions are identified, and physical propagation models are used to estimate ground-truth positions from the measured channel [16, 17]. Once such reference points are available, the CC can be transformed into real-world coordinates via an affine transformation, see Section 3.2.6 for specific implementation details.

In contrast to fingerprinting, CC does not require a costly offline data collection phase with labeled reference positions [4]. This makes CC inherently more scalable and robust in response to environmental changes, as CC leverages the fact that CSI data is already being collected, used, and then discarded. This presents the opportunity to recycle CSI data into a usable CC.

Overall, CC provides a data-driven alternative to traditional positioning techniques, particularly suited for dense wireless networks and future communication systems, where labeled location data may be unavailable or impractical to obtain [5].

2.3 Research Objectives

The objective of this project is to develop and evaluate CC methods that exploit CSI from multiple frequency bands. By combining the complementary information provided by different bands, the aim is to improve positioning performance in challenging environments, such as the indoor NLoS scenarios commonly encountered in shopping centers and industrial facilities, where conventional single-band approaches struggle. Improved indoor localization could also support other services, such as emergency response, where rapid and accurate positioning of individuals can be life-saving.

2.4 Limitations

This project is limited to simulated CSI data generated using the Sionna RT [18] ray-tracing framework. Real-world measurements were not included due to time and resource constraints. The evaluation was further restricted to a single simulated urban street canyon environment. Consequently, environments with different dimensions and conditions, such as varying distributions between LoS and NLoS, were not considered. In terms of multi-band fusion, only three multi-band fusion strategies were implemented and evaluated. Other possible fusion approaches were outside the scope of this work. Finally, the affine transform relied on ground truth

positions. Such information would not be available in a real-world scenario, since CC is based on unsupervised learning and is intended to operate without labeled position data. This was done to restrict the scope to lateral aspects.

3

Methods

This chapter describes the methodology used in this work. It begins with the dataset and pre-processing of the CSI for two frequencies. It further outlines the general framework for single-band CC and the three different approaches considered for dual-band CC fusion. A fingerprinting-based method is also included as a baseline for comparison.

3.1 CSI Dataset and Processing

For this work, CSI was generated using the ray-tracing framework provided by Sionna RT [18]. See Section 4.1 for our full experimental setup.

For each receiver position l, frequency-domain CSI was generated at two carrier frequencies, denoted f_1 and f_2. The CSI corresponding to carrier frequency f_q, where $q \in \{1, 2\}$, is represented by

$$\mathbf{H}_{f_q}^{(l)} \in \mathbb{C}^{B \times M \times N_{\text{sub}}} \tag{3.1}$$

where each entry encodes the complex-valued coefficient at a particular base station, antenna, and subcarrier. Timestamps $t^{(l)}$ are recorded alongside each CSI sample and denoted in seconds. From a total of L samples, a dataset is obtained as

$$\mathcal{S} = \left\{ \mathbf{H}_{f_1}^{(l)}, \mathbf{H}_{f_2}^{(l)}, \mathbf{x}^{(l)}, t^{(l)} \right\}_{l=1,\ldots,L} \tag{3.2}$$

where $\mathbf{x}^{(l)} \in \mathbb{R}^D$ denotes the ground-truth position of sample l and D denotes the number of spatial dimensions. These positions were used solely to evaluate the learned channel chart representations. The CSI at the two carrier frequencies was generated sequentially using the same transmitter and receiver configuration. This ensures that the resulting CSI samples correspond to the same physical geometry and propagation conditions across both frequency bands. The bandwidth for each frequency was set to scale proportionally with the frequency.

As the dissimilarity metrics in Sections 3.2 and 3.3 are computed using the time-domain CSI, the simulated frequency-domain CSI is transformed using an inverse fast Fourier transform (IFFT) over the subcarrier axis. This transforms the CSI from

$$\mathbf{H}_{f_q}^{(l)} \in \mathbb{C}^{B \times M \times N_{\text{sub}}} \tag{3.3}$$

to a time-domain representation

$$\tilde{\mathbf{H}}_{f_q}^{(l)} \in \mathbb{C}^{B \times M \times T} \tag{3.4}$$

where the subcarrier dimension is replaced by time-domain taps with a total amount of T time-domain taps . Most of the information is concentrated in a select few time taps τ and can thus be further truncated to

$$\tilde{\mathbf{H}}_{f_q}^{(l)} \in \mathbb{C}^{B \times M \times (\tau_{\max}^{f_q} - \tau_{\min}^{f_q})} \tag{3.5}$$

where the truncation bounds $\tau_{\max}^{f_q}$ and $\tau_{\min}^{f_q}$ depend on both the propagation environment and carrier bandwidth. Since the time resolution increases with bandwidth, the two carrier frequencies require different truncation ranges to represent the same physical time window and contain all relevant LoS and multipath components.

3.2 Single-Band Channel Charting

The implemented CC pipeline is based on the CSI dissimilarity metric and the timestamp fusion framework proposed in [1]. This method largely follows the described dissimilarity-metric-based CC in Section 2.2.2. Some modifications were performed during the timestamp fusion stage and in the feature engineering. It consists of learning an FCF that maps high-dimensional CSI to a low-dimensional channel chart coordinate using pairwise dissimilarities as a proxy for physical distance. Sections 3.2.1 – 3.2.3 describe the computation of these dissimilarities used to capture the spatial relationships between CSI samples. Section 3.2.4 describes the feature extraction process and resulting input to the FCF. Section 3.2.5 describes the Siamese neural network used to learn the FCF from the pairwise dissimilarities. Finally, Section 3.2.6 describes the final transformation to the real space coordinates, assuming a few known ground-truth positions.

3.2.1 Angle-Delay Profile-based Dissimilarity

To compute pairwise dissimilarities between CSI samples, an Angle-Delay Profile (ADP) dissimilarity metric from [1] was used. The ADP dissimilarity compares CSI samples by measuring the similarity of their angle-of-arrival power distributions across multiple time taps. The purpose of the metric is to compute a dissimilarity, $d_{i,j}$, which fulfills

$$d_{i,j} \propto \left\| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \right\|_2 \tag{3.6}$$

The ADP dissimilarity is computed as:

$$d_{\mathrm{ADP},i,j} = \sum_{b=1}^{B} \sum_{\tau=1}^{\tau_{\max}-\tau_{\min}} \left(1 - \frac{\left| \sum_{m=1}^{M} \left(\tilde{\mathbf{H}}_{b,m,\tau}^{(i)} \right)^* \tilde{\mathbf{H}}_{b,m,\tau}^{(j)} \right|^2}{\left(\sum_{m=1}^{M} \left| \tilde{\mathbf{H}}_{b,m,\tau}^{(i)} \right|^2 \right) \left(\sum_{m=1}^{M} \left| \tilde{\mathbf{H}}_{b,m,\tau}^{(j)} \right|^2 \right)} \right) \tag{3.7}$$

where the inner product is a normalized cosine similarity of the channel vectors between samples i and j. This is then converted into a dissimilarity measure and

summed over all time taps and base stations. These dissimilarities are computed for all possible pairs and turn a dataset of L samples of CSI belonging to $\tilde{\mathbf{H}}^{(l)} \in \mathbb{C}^{B\times M\times(\tau_{\max}-\tau_{\min})}$ into a dissimilarity matrix described as

$$\mathbf{D}_{\text{ADP}} = \begin{pmatrix} d_{1,1} & \cdots & d_{1,L} \\ \vdots & \ddots & \vdots \\ d_{L,1} & \cdots & d_{L,L} \end{pmatrix}, \tag{3.8}$$

where $\mathbf{D}_{\text{ADP}}$ is of size $L \times L$ and symmetric along the diagonal. The individual scalar values of the matrix will be denoted by $d_{\text{ADP},i,j}$.

3.2.2 Timestamp Fusion

Timestamp-based dissimilarities are fused with the preceding ADP-based dissimilarity, based on the application from [1]. The timestamp dissimilarity between two samples i and j is defined as

$$d_{\text{time},i,j} = \left| t^{(i)} - t^{(j)} \right|. \tag{3.9}$$

Using equation (3.9), the raw timestamps of L samples are used to construct a timestamp dissimilarity matrix $\mathbf{D}_{\text{time}}$ of size $L \times L$. To fuse $\mathbf{D}_{\text{time}}$ with the ADP dissimilarity matrix $\mathbf{D}_{\text{ADP}}$, the two matrices must first be brought to a comparable scale [1]. The ratio between the metrics is computed for sample pairs whose time difference is below a threshold $t_{\text{threshold}}$,

$$R = \left\{ \frac{d_{\text{ADP},i,j}}{d_{\text{time},i,j}} \;\middle|\; 0 < d_{\text{time},i,j} < t_{\text{threshold}};\ i,j \in 1,\ldots,L \right\}, \tag{3.10}$$

which restricts the comparison to sample pairs for which the timestamp metric is reliable. A scaling factor γ is determined from a statistic of these ratios to ensure that the two dissimilarity metrics are on a comparable scale for the relevant samples. Finally, the matrices are fused element-wise by taking the minimum

$$d_{\text{Fused},i,j} = \min\{d_{\text{ADP},i,j},\ \gamma \cdot d_{\text{time},i,j}\}. \tag{3.11}$$

Sample pairs for which the ADP metric overestimates the separation are thereby replaced by the scaled timestamp dissimilarity whenever they are close in time. The fused dissimilarity matrix is denoted $\mathbf{D}_{\text{Fused}}$.

3.2.3 Geodesic Dissimilarity

CSI-based dissimilarity matrices are most reliable for physically close samples and therefore better describe local structure than global structure [12]. Consequently, a geodesic dissimilarity matrix $\mathbf{D}_{\text{G-Fused}}$ is computed from the $\mathbf{D}_{\text{Fused}}$ matrix to predict the true distances more accurately. The geodesic dissimilarity between two samples corresponds to the sum of local dissimilarities along the shortest path between them.

The construction proceeds in two steps [12]. First, a k-nearest neighbor graph $\mathbf{G}_{k-NN}$ is built from $\mathbf{D}_{\text{Fused}}$, where each sample is connected to its k closest neighbors and the corresponding edge weights are taken from $\mathbf{D}_{\text{Fused}}$. Second, Dijkstra's

shortest-path algorithm [19] is applied to $\mathbf{G}_{k-NN}$ to compute pairwise geodesic distances between all samples, which are stored in a new dissimilarity matrix. The resulting matrix $\mathbf{D}_{\text{G-Fused}}$ approximates distances along the underlying data manifold rather than through the ambient signal space, which more faithfully reflects the true spatial separation between samples.

3.2.4 Feature Engineering Stage

The time-domain CSI is processed to extract features that more effectively capture large-scale fading in the wireless channel while reducing sensitivity to small-scale fading effects [4]. The feature extraction procedure is based on [15, 16]. First, an FFT is applied over the antenna dimension [16]

$$\tilde{\mathbf{H}}^{(l)}_{b,\mu,\tau} = \text{FFT}_m \left\{ \tilde{\mathbf{H}}^{(l)}_{b,m,\tau} \right\} \tag{3.12}$$

to transform the CSI to the angular domain denoted by μ. This representation emphasizes the angle of arrival of the wireless channel. Next, the magnitude of the transformed CSI is computed to reduce the sensitivity to small-scale phase shifts [15]. The tensor is then normalized by its Frobenius norm as

$$\mathbf{F}^{(l)}_{b,\mu,\tau} = \frac{|\tilde{\mathbf{H}}^{(l)}_{b,\mu,\tau}|}{||\tilde{\mathbf{H}}^{(l)}_{b,\mu,\tau}||_F} \tag{3.13}$$

which suppresses variations in overall received power and emphasizes the relative structure instead of the absolute received signal strength. Finally, the tensor $\mathbf{F}^{(l)}_{b,\mu,\tau}$ is flattened into a vector $\mathbf{f}^{(l)} \in \mathbb{R}^{BM(\tau_{\max}-\tau_{\min})}$, which serves as the input to the neural network.

3.2.5 Neural Network

Since no ground-truth positions are available, the FCF is trained in an unsupervised manner using a Siamese neural network architecture, illustrated in Figure 3.1. In this architecture, two inputs $\mathbf{f}^{(i)}$ and $\mathbf{f}^{(j)}$ are processed simultaneously through two weight-sharing neural networks, and each sub-network maps its input to their corresponding position $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$ in the channel chart [14]. The loss function is computed as the distance between the two predicted positions and the corresponding dissimilarity $d_{i,j}$. The network is then tuned through backpropagation. The full loss function used is from [1] and can be described as

$$\mathcal{L}_{\text{Siamese}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{d_{i,j} + \beta} \left(d_{i,j} - \left\| \mathbf{z}^{(i)} - \mathbf{z}^{(j)} \right\|_2 \right)^2 \tag{3.14}$$

where N refers to the batch size for which the loss is averaged over, $d_{i,j}$ is the dissimilarity between the sample i and j, and the two mappings produced by the FCF from samples i and j is referred to as $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$. The parameter β is included for numerical stability in case of small values of $d_{i,j}$. Minimizing this loss function encourages the network to place measurements that are similar in CSI space close

together in the channel chart, and dissimilar measurements far apart [1]. Since $d_{i,j}$ is derived from CSI alone – without any ground-truth position labels – this constitutes an unsupervised learning objective.

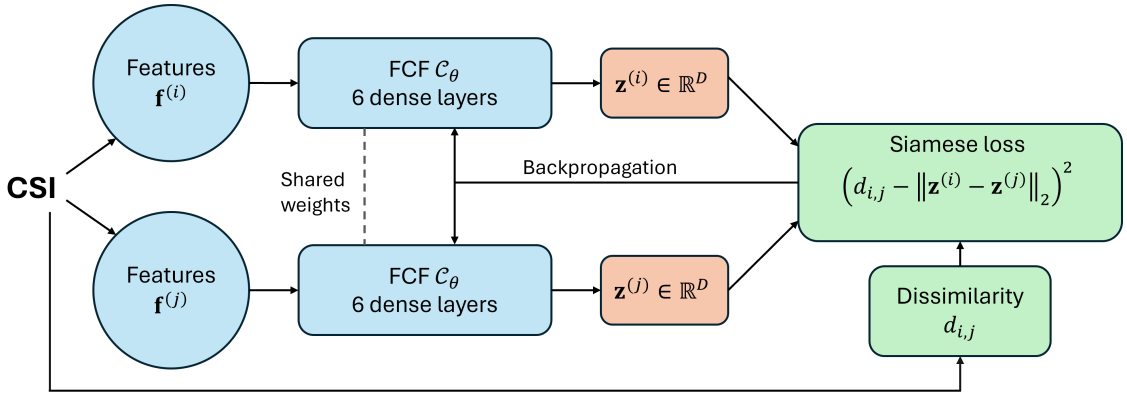


Figure 3.1: Siamese network structure used to train the FCF

The Siamese training scheme operates on CSI sample pairs rather than on individual measurements. At each training step, two indices are drawn at random from the dataset and their corresponding CSI features $\boldsymbol{f}^{(i)}$ and $\boldsymbol{f}^{(j)}$ are passed through the shared-weight FCF, yielding their respective channel chart embeddings $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$. The Euclidean distance $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|_2$ is compared against the corresponding entry $d_{i,j}$ of the precomputed dissimilarity matrix, and the squared deviation contributes to the loss in equation (3.14). Backpropagation through this loss updates the shared weights so that the embedding distances progressively align with the dissimilarity structure of the data. It works by passing the prediction error backward through the network and computing how much each weight contributed to it. The weights are then adjusted to reduce the error in the next iteration. This pairwise construction yields $L(L-1)/2$ unique training pairs, providing a substantially larger effective training set than the number of original CSI samples.

3.2.6 Affine Alignment of Channel Charts

Since CC aims to preserve local geometric relationships, the resulting embedding may still be rotated, sheared, or scaled relative to the true positions in physical space [15]. To address these problems, an affine transform is applied to align the channel chart with the true coordinates in physical space by assuming a few ground-truth positions are known.

Recall that $\mathbf{f}^{(l)}$ denotes the CSI feature vector for sample l, and $\mathbf{z}^{(l)}$ denotes its corresponding channel chart embedding learned by the model. The transformation aims to align these embeddings with their respective ground-truth positions $\mathbf{x}^{(l)}$. The affine transformation is parameterized by a matrix $\mathbf{A} \in \mathbb{R}^{D \times D}$ and a translation vector $\mathbf{b} \in \mathbb{R}^D$ [15]. The optimal transformation parameters are obtained by solving the following least-squares problem [1, 15]:

$$(\hat{\mathbf{A}}, \hat{\mathbf{b}}) = \arg\min_{\mathbf{A},\mathbf{b}} \sum_{l=1}^{L} \left\| (\mathbf{A}\,\mathbf{z}^{(l)} + \mathbf{b}) - \mathbf{x}^{(l)} \right\|_2^2 \tag{3.15}$$

Furthermore, the transformed channel chart coordinates are then given by

$$\hat{\mathbf{z}}^{(l)} = \hat{\mathbf{A}}\,\mathbf{z}^{(l)} + \hat{\mathbf{b}} \tag{3.16}$$

where $\hat{\mathbf{z}}^{(l)}$ denotes the affine-aligned chart coordinates [15], i.e., the final position proxy.

3.3 Dual-Band Channel Charting

To exploit complementary frequency information, three fusion methods were evaluated for CC, each applied at different stages of the CC pipeline. Firstly, averaging the ADP dissimilarity matrices from the two different frequencies. Secondly, multiplying the similarity scores during the ADP calculation. Lastly, performing a late-stage fusion in which each frequency has its own neural network trained using an additional loss function. This loss function aligns the outputs corresponding to the same location, and the final embedding is averaged across the two networks.

3.3.1 ADP Averaging

For each frequency, the ADP dissimilarity matrix is computed using equation (3.7), as in the single-band case. This operation is then followed by simple element-wise averaging as

$$\mathbf{D}_{\text{Avg-ADP}} = \frac{\mathbf{D}^{f_1}_{\text{ADP}} + \mathbf{D}^{f_2}_{\text{ADP}}}{2} \tag{3.17}$$

after which the result is fused with timestamps and converted into a geodesic dissimilarity $\mathbf{D}_{\text{G-Avg-Fused},i,j}$. Additionally, during feature engineering, feature vectors are generated separately for each frequency and concatenated into a single input vector for the FCF where $\mathbf{f}^{(l)} \in \mathbb{R}^{BM((\tau^{f_1}_{\max}-\tau^{f_1}_{\min})+(\tau^{f_2}_{\max}-\tau^{f_2}_{\min}))}$.

3.3.2 Similarity Multiplication

Since the geodesic conversion estimates global structure from local kNN relations, accurate nearest-neighbor relationships are particularly important. Therefore, fusion performed prior to the geodesic computation should primarily aim to improve the local structure of the pairwise dissimilarities. A potential method for implementing this is to only regard two samples as similar if both bands agree that they are similar. Practically, this can be done by computing the cosine similarity for each band and then multiplying the two similarities. If one band views two samples as dissimilar, the multiplicative action is sufficiently harsh to force the fused product to have a low similarity and will be excluded during the geodesic conversion. Whereas, if both bands assign high similarity, the fused product remains high and relatively undistorted, since both values are close to one. The ADP formula can be modified as

$$S^{f_q}_{b,i,j} = \left(\frac{1}{(\tau^{f_q}_{\max} - \tau^{f_q}_{\min})} \sum_{\tau=1}^{\tau^{f_q}_{\max}-\tau^{f_q}_{\min}} \frac{\left| \sum_{m=1}^{M} \left(\tilde{\mathbf{H}}^{(i)}_{b,m,\tau}\right)^{*} \tilde{\mathbf{H}}^{(j)}_{b,m,\tau} \right|^2}{\left(\sum_{m=1}^{M} \left| \tilde{\mathbf{H}}^{(i)}_{b,m,\tau} \right|^2 \right) \left(\sum_{m=1}^{M} \left| \tilde{\mathbf{H}}^{(j)}_{b,m,\tau} \right|^2 \right)} \right) \tag{3.18}$$

where $S_{b,i,j}^{f_q}$ is a similarity score for a single band, which is calculated by summing the cosine similarity over the time taps and then normalized to between $[0, 1]$. These similarity scores are calculated for each band and then fused by

$$d_{\text{SM-ADP},i,j} = \sum_{b=1}^{B} \left(1 - S_{b,i,j}^{f_1} \cdot S_{b,i,j}^{f_2}\right) \tag{3.19}$$

where the similarities are multiplied and converted to a dissimilarity, and finally summed over the base stations. This new dissimilarity matrix is then fused with timestamps and converted into a geodesic dissimilarity matrix, $\mathbf{D}_{\text{G-SM-Fused}}$. As in Section 3.3.1, feature vectors are generated separately for each frequency and concatenated into a single input vector for the FCF where $\mathbf{f}^{(l)} \in \mathbb{R}^{BM((\tau_{\max}^{f_1} - \tau_{\min}^{f_1}) + (\tau_{\max}^{f_2} - \tau_{\min}^{f_2}))}$.

3.3.3 Dual-Network Alignment

A late-stage fusion method where each frequency has its own time-fused geodesic dissimilarity matrix and neural network, as was described in section 3.2. Prior to training, each dissimilarity matrix is normalized by its 95th percentile so that the dissimilarities from the two frequencies are brought to a comparable scale before fusion. This helps the networks learn embeddings with comparable geometric scale, which is necessary for the subsequent alignment of channel charts. The 95th percentile was chosen instead of the maximum because it yields a more robust normalization in the presence of outliers. The networks are then fused during training using an additional alignment loss

$$\mathcal{L}_{\text{Alignment}} = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{z}_{f_1}^{(i)} - \mathbf{z}_{f_2}^{(i)} \right\|_2 \tag{3.20}$$

where the corresponding embeddings for each frequency are trained to map to nearby locations in the channel chart. The final loss for the dual-network becomes

$$\mathcal{L}_{\text{SSA}} = \mathcal{L}_{\text{Siamese}}^{f_1} + \mathcal{L}_{\text{Siamese}}^{f_2} + \alpha \mathcal{L}_{\text{Alignment}} \tag{3.21}$$

where each Siamese loss only affects its corresponding network. As each network creates its own embedding, the final embedding is defined as the average of the two embeddings

$$\mathbf{z}_{\text{Avg}}^{(l)} = \frac{\mathbf{z}_{f_1}^{(l)} + \mathbf{z}_{f_2}^{(l)}}{2} \tag{3.22}$$

To enable a more meaningful comparison of the effects of adding frequencies, two new single-band references were created that work as described above, except that both inputs use the same frequency. One reference model uses f_1 for both inputs, while the other uses f_2 for both inputs. This isolates the impact of adding additional frequencies and eliminates any effects from the architectural changes.

3.4 Fingerprinting Baseline

Fingerprinting is used as a supervised baseline for comparison with the unsupervised CC methods. The same simulated dataset, delay-domain preprocessing, and feature

engineering described for CC are reused here. The key difference from CC is the learning objective. Instead of learning a low-dimensional embedding from pairwise dissimilarities using a Siamese loss, fingerprinting is trained in a supervised manner to predict the receiver position directly from the extracted CSI features.

For each sample l, let the preprocessed CSI after delay-domain conversion and tap cropping be denoted by

$$\tilde{\mathbf{H}}_{f_q}^{(l)} \in \mathbb{C}^{B\times M\times(\tau_{\max}^{f_q}-\tau_{\min}^{f_q})}. \tag{3.23}$$

Applying the same feature engineering as in the CC forward charting function yields a feature vector

$$\mathbf{f}^{(l)} \in \mathbb{R}^{BM(\tau_{\max}^{f_q}-\tau_{\min}^{f_q})}, \tag{3.24}$$

which is used as input to a neural network regressor

$$\mathcal{F}_\phi : \mathbf{f}^{(l)} \mapsto \hat{\mathbf{x}}^{(l)} \in \mathbb{R}^2, \tag{3.25}$$

where $\hat{\mathbf{x}}^{(l)}$ is the predicted receiver position.

To ensure a fair architectural comparison, the regressor uses the same dense-layer backbone as the FCF used in CC, but is trained directly on ground-truth positions instead of pairwise dissimilarities. The model parameters ϕ are optimized using Adam by minimizing the mean squared error

$$\mathcal{L}_{\mathrm{FP}}(\phi) = \frac{1}{|\mathcal{S}_{\mathrm{train}}|}\sum_{l\in\mathcal{S}_{\mathrm{train}}} \left\|\mathcal{F}_\phi\left(\mathbf{f}^{(l)}\right) - \mathbf{x}^{(l)}\right\|_2^2. \tag{3.26}$$

For dual-band fingerprinting, the same preprocessing and feature extraction are applied separately to the CSI from each carrier frequency. Let $\mathbf{f}_{f_1}^{(l)}$ and $\mathbf{f}_{f_2}^{(l)}$ denote the resulting feature vectors for the two bands. These are concatenated into a single fused input representation

$$\mathbf{f}_{\mathrm{dual}}^{(l)} = \left[\mathbf{f}_{f_1}^{(l)}, \mathbf{f}_{f_2}^{(l)}\right]. \tag{3.27}$$

A supervised neural network with the same dense regression backbone is then trained to map the fused feature vector to the receiver position.

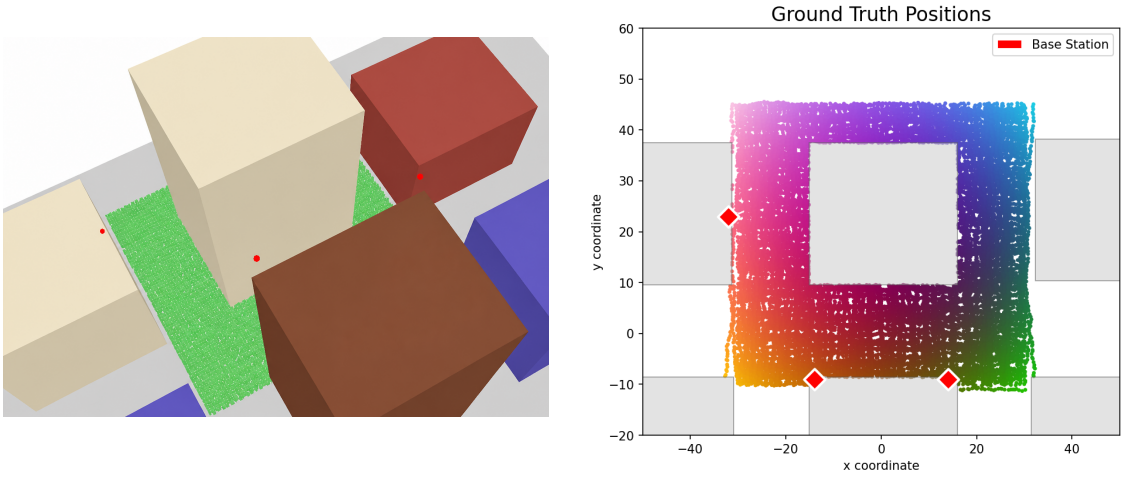
4

Results

This chapter presents the results. It begins with a description of the scene configuration, data generation, and preprocessing. Parameter selection for the dissimilarity metrics and training setup is also presented. Subsequently, the resulting channel charts are then presented for the different methods, covering both single-band and dual-band approaches. In addition, results for the fingerprinting-based methods are included as a baseline for comparison.

4.1 Scene configuration

CSI was generated using the ray-tracing framework provided by Sionna RT [18]. The simulations were performed in the predefined simple street-canyon scenario, which contains multiple buildings to represent an urban environment. Figures 4.1a and 4.1b depict the simulated radio environment.



(a) 3D view showing the street canyon geometry and BS positions.

(b) 2D view of the BS and UE placement.

Figure 4.1: Simulated radio environment alongside BS & UE placement

Simulated radio environment is based on the `simple_street_canyon` scene in Sionna RT. The scene contains three rooftop BS (red), each equipped with an eight-element antenna array with two rows and four columns, while the user equipment is modeled

with a single antenna at a height of 1.5 m. The specific propagation and material settings used in Sionna RT are presented in Appendix A.1. This configuration yields a compact yet spatially informative multi-BS setup, in which the channel observations vary with both the receiver position and the BS visibility. Figure 4.2 illustrates the spatial distribution of LoS connectivity, where each point is colored according to the number of BS with LoS visibility to the UE.

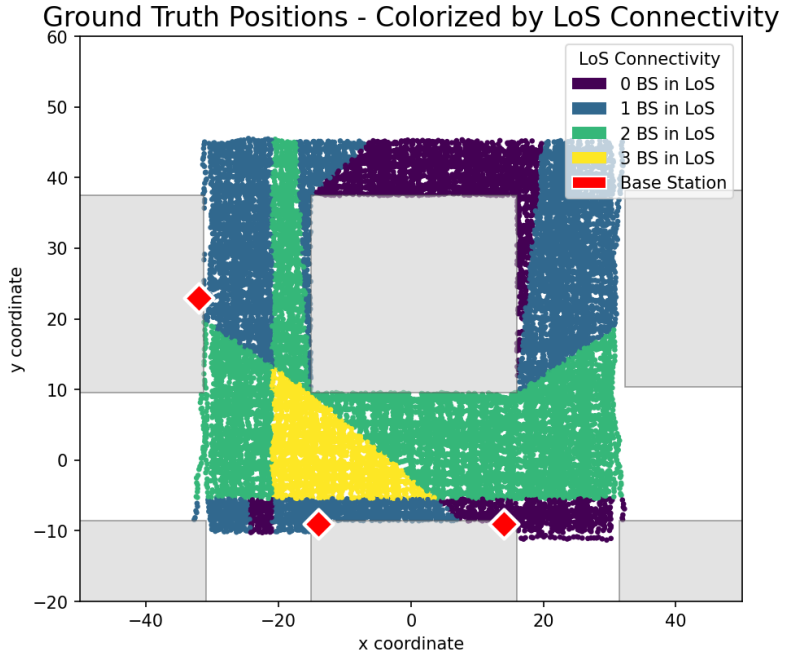


Figure 4.2: LoS connectivity of samples. 84% of the samples are within LoS with at least one base station

Due to the street-canyon geometry and rooftop-mounted BS placement, the coverage conditions vary significantly throughout the environment.

Simulations were performed for the 3.5 and 12 GHz bands; both bands use the same number of subcarriers of $N_{\text{sub}} = 1024$, with the subcarrier spacing scaled proportionally to the carrier frequency. This gives a total bandwidth of 30.69 MHz at 3.5 GHz and 105.369 MHz at 12 GHz. The shared geometry and comparable OFDM structure make the two bands suitable for evaluating whether multi-band CSI provides additional spatial information for CC.

Receiver positions were generated using a boustrophedon sweep, simulating a single UE traversing the scene's accessible outdoor area at a fixed height of 1.5 m, corresponding to a typical handheld UE. Placements overlapping with building footprints were excluded. This traversal strategy is important for the timestamp-based dissimilarity described in Section 3.2.2, since it ensures that samples collected close in simulated time are also close in physical space.

A total of 19,586 UE positions were simulated, providing a sufficiently dense spatial sampling of the environment for CC evaluation. This dataset was split into separate

train and test sets with a 50/50 split. The simulation-based approach provides full access to ground-truth positions, enabling quantitative evaluation, as described in Section 4.3.

4.1.1 Timetap Truncation

During CSI preprocessing, the timetap truncation for the 3.5 GHz band was set from $\tau_{min}^{f_1} = 507$ to $\tau_{max}^{f_1} = 537$. Since the 12 GHz band has a higher bandwidth than the 3.5 GHz band, it will have a finer time-domain resolution and will require a different timetap truncation if both bands are to represent the same time window. For the 12 GHz band, this was calculated as $\tau_{min}^{f_2} = 495$ to $\tau_{max}^{f_2} = 598$.

Figure 4.3 shows the CSI magnitude over the selected time-domain tap range in the 3.5 GHz band, plotted every ten samples across all antennas for a single base station.

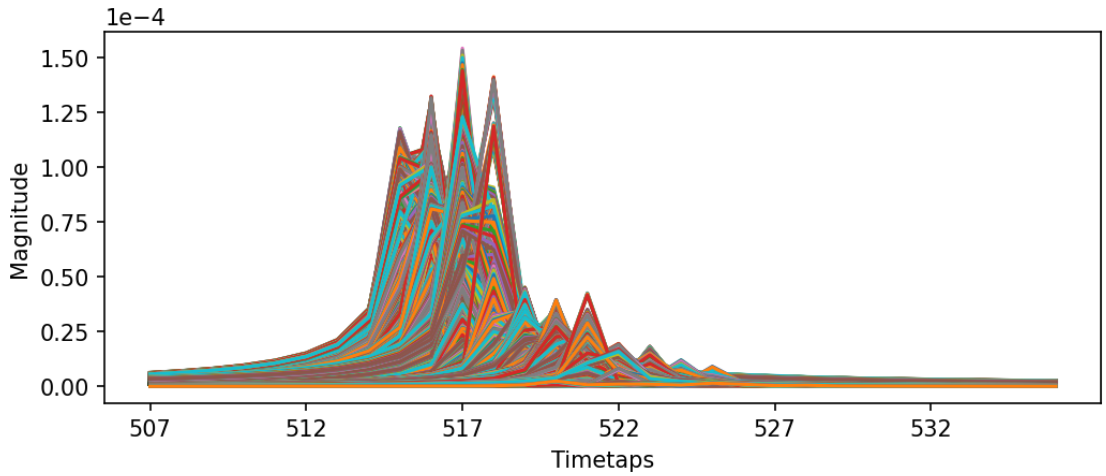


Figure 4.3: CSI magnitude over the used time-domain truncation for the 3.5 GHz band.

Although most of the channel energy is concentrated in a smaller subset of taps than the selected truncation range, experiments with more aggressive truncation resulted in degraded performance. Therefore, a wider tap range was retained.

4.1.2 Dissimilarity Matrix Parameters

Figure 4.4 shows a histogram of the ADP and time dissimilarity ratios from equation (3.10), along with the determined scaling factor γ, which was selected as the median of the ratios.

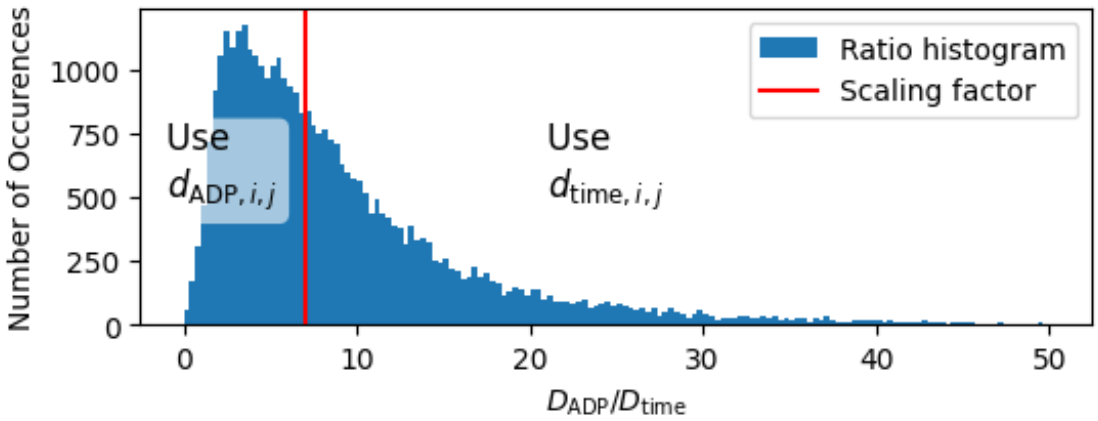


Figure 4.4: Histogram of the dissimilarity ratios and the scaling factor

From equation (3.11), the selection of which dissimilarity metric to retain for a sample is determined by comparing the ratio to the scaling factor γ. If the ratio exceeds γ, the timestamp dissimilarity $d_{\text{time},i,j}$ is used; otherwise, the ADP dissimilarity $d_{\text{ADP},i,j}$ is used. All timestamp fusions were performed using a time threshold of $t_{\text{threshold}} = 5\,\text{s}$. Timestamp fusion does not provide a significant improvement to the dissimilarity matrix on its own, as it only replaces a few of the closest samples. Instead, the benefits arise during the geodesic transformation described in section 3.2.3 where only the k nearest neighbors are preserved. For the geodesic transformation, $k = 20$ was selected to construct the kNN graph for all dissimilarity matrices.

4.1.3 Training Parameters

All CC models were implemented in TensorFlow. The single-band and early-fusion models use a neural network following the same architecture as in [1] and comprise 6 dense layers with 1024, 512, 256, 128, 64, and 2 nodes, respectively. The first 5 layers use ReLU activations followed by batch normalization, and the last layer has a linear activation. For the dual-network alignment-loss architecture, fusion was performed using an additional alignment term with weight $\alpha = 1$.

Training was performed using Adam Optimizer with a Cosine-Decay learning-rate scheduler with warmup. The learning rate increased during the first 5% of the training steps up to a target learning rate of 5×10^{-2}, after which cosine decay was applied with a final learning rate of 2.5×10^{-4}. All models were trained for 100 epochs, with a batch size of $N = 4096$ sample pairs and 150000 randomly generated pairs per epoch. The β term in the loss function (3.14) was set to the 1st-percentile of the dissimilarity matrix trained against. To ensure reproducibility, both sample pair selection and weight initialization were seeded between models.

The learned channel charts were aligned with the physical ground-truth positions using an affine transformation estimated from 40 calibration points. These points were selected from the test set via Poisson disk sampling [20] with a minimum spacing of 5 m.

4.2 Channel Chart Visualization

The spatial domain is chosen as $D = 2$ as elevation is not considered in this type of positioning. The ground-truth positions of the unseen test set used to evaluate the CC methods are shown in Figure 4.5 and serve as a visual reference for the learned charts presented in the following figures. Since CC is only expected to preserve local neighborhood structure, the learned charts are evaluated in terms of how faithfully the local geometry of the ground-truth layout is recovered, rather than by direct pointwise agreement.

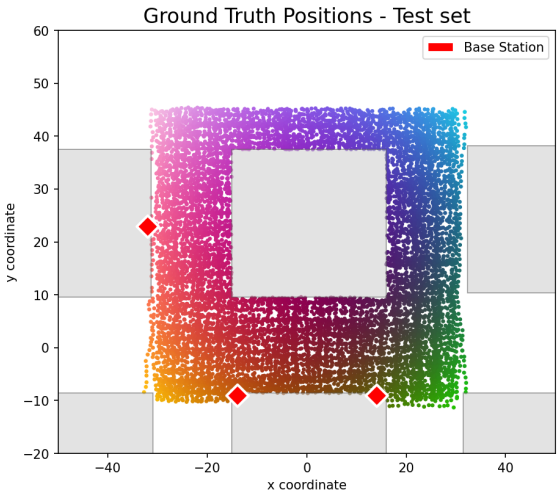


Figure 4.5: Ground-truth positions for the test set

Figure 4.6 shows the learned charts for the two single-band references, trained on 3.5 GHz and 12 GHz, respectively. Each model follows the single-band procedure described in Section 3.2, using a Siamese network with a time-fused geodesic dissimilarity matrix that combines ADP and timestamp dissimilarities.

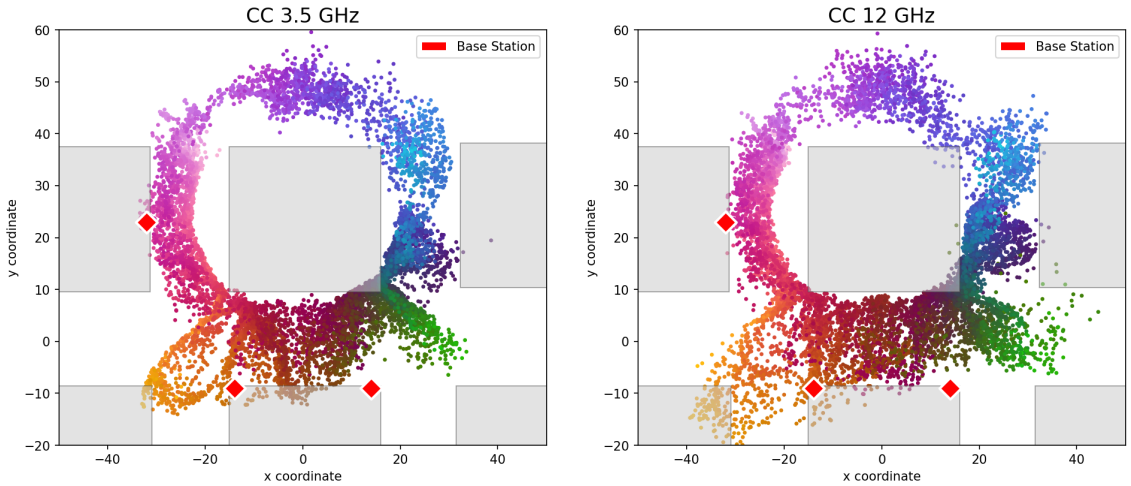


Figure 4.6: Transformed channel charts from single-band 3.5 GHz (left) and 12 GHz (right).

Figure 4.7 shows the same single-band setup as in Figure 4.6, but trained using the dual-network architecture from Section 3.3.3 with both inputs set to the same frequency band. This isolates the effect of the alignment loss while keeping the input information identical to the single-band case.

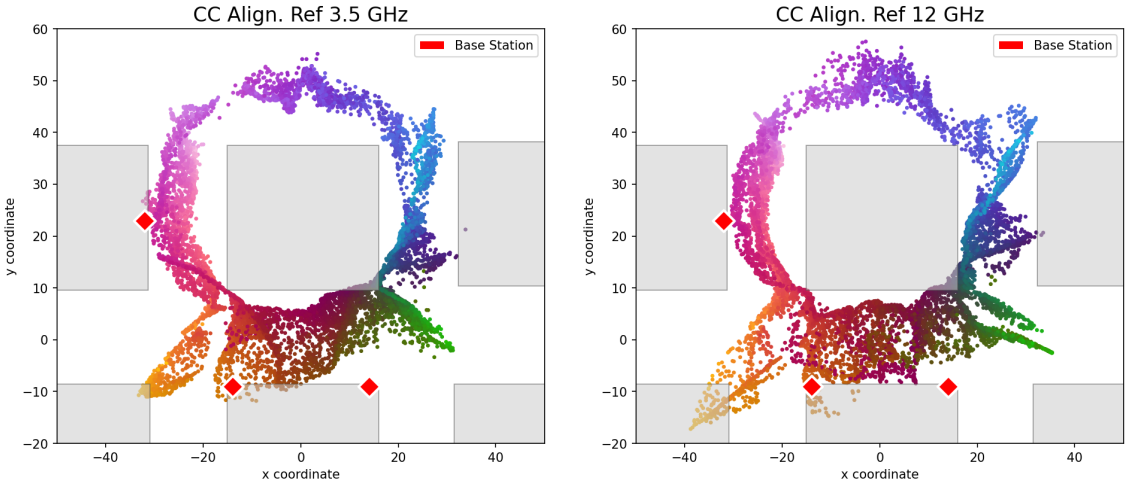


Figure 4.7: Single-band charts at 3.5 GHz (left) and 12 GHz (right), trained with the dual-network architecture and alignment loss using identical inputs in both branches.

Compared to Figure 4.6, the resulting charts appear visibly less scattered, which may indicate better preservation of local neighborhood structure. This improvement does not stem from the alignment loss intrinsically penalizing scattering or local structure — neither of which it directly affects — but rather from its regularising effect, which guides the networks toward a more stable solution.

Figures in 4.8 display the separate effects of three methods for dual-band fusion, namely: ADP-average, Similarity Multiplication, and Alignment Loss.

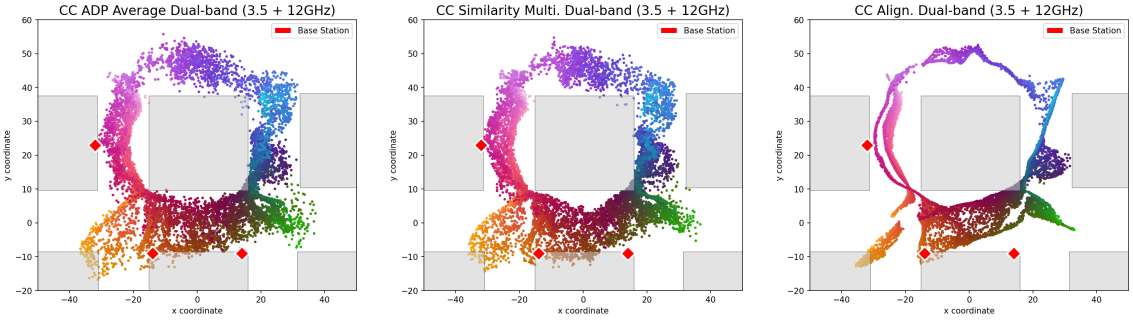


Figure 4.8: Channel charts from dual-band fusion

Across the evaluation metrics mentioned later in Section 4.3, the early-stage fusion methods — ADP averaging and similarity multiplication — and the alignment-loss-based dual-network model show mixed performance, with no single method

completely dominating. Similarity multiplication performs slightly better than ADP averaging on all metrics, while the dual-network model produces a visibly more compact chart. This tighter grouping does not necessarily indicate better pointwise accuracy, as visual compactness and faithfulness to the ground-truth layout are distinct properties.

4.3 Evaluation Metrics

The generated channel charts are evaluated using Continuity (CT), Trustworthiness (TW), Kruskal's Stress (KS), and Mean Absolute Error (MAE). CT and TW evaluate local neighborhood preservation [4], whereas KS evaluates preservation of the global geometry of the channel chart [1]. MAE is used to quantify the positioning accuracy after the affine transformation.

CC aims to preserve relative positions such that neighborhood structures are maintained during dimensionality reduction. This is evaluated by comparing the K-nearest neighbors (kNN) in the physical space with those in the low-dimensional representation [4]. Let L denote the number of data points and K the neighborhood size. Furthermore, let $r_{l,i}$ denote the rank of point i as a neighbor of point l in the physical space, and $\hat{r}_{l,i}$ denote the corresponding rank in the representation space [1]. Define

$$V_K^{(l)} = \{i \mid r_{l,i} \leq K\}$$

as the set of the K nearest neighbors of point l in the physical space, and

$$U_K^{(l)} = \{i \mid \hat{r}_{l,i} \leq K\}$$

as the corresponding set in the representation space [1]. For channel charts, the representation-space ranks are computed from the Euclidean distances between channel chart coordinates, i.e., $\|\mathbf{z}^{(l)} - \mathbf{z}^{(i)}\|_2$ for $i = 1, \ldots, L$ [1]. The CT metric measures whether points that are neighbors in the physical space remain neighbors in the representation space [4], thereby penalizing missing neighborhood relationships. CT is defined as [1]:

$$CT(K) = 1 - \frac{2}{LK(2L - 3K - 1)} \sum_{l=1}^{L} \sum_{i \in V_K^{(l)}} \max\{0, \hat{r}_{l,i} - K\} \tag{4.1}$$

Conversely, TW evaluates whether points that are neighbors in the representation space are also neighbors in the physical space, thereby penalizing false neighborhood relationships. TW is defined as [1]:

$$TW(K) = 1 - \frac{2}{LK(2L - 3K - 1)} \sum_{l=1}^{L} \sum_{i \in U_K^{(l)}} \max(0, r_{l,i} - K) \tag{4.2}$$

Both CT and TW take values in the interval $[0, 1]$, where higher values indicate better preservation of neighborhood relationships [1, 4].

CT and TW are computed following standard practice in CC literature, using a neighborhood size of $K = 0.05L$ [1, 4], where L denotes the number of samples in the test subset. The neighborhood relationships are determined based on pairwise Euclidean distances between samples in the ground truth position space and the corresponding pairwise Euclidean distances in the channel chart space [1].

KS quantifies how well the global structure is preserved in the channel chart [1]. This is achieved by comparing pairwise distances between points in the physical space, $\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|$, with the corresponding Euclidean distances in the representation space, $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|$. A value of $KS = 0$ indicates perfect preservation of pairwise distances between the physical space and the representation space. Since KS depends only on Euclidean distances, rigid transformations and uniform scaling do not affect the metric. KS is defined as

$$KS = \sqrt{\frac{\sum_{i=1}^{L-1}\sum_{j=i+1}^{L}\left(\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2 - \beta\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|_2\right)^2}{\sum_{i=1}^{L-1}\sum_{j=i+1}^{L}\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2^2}} \tag{4.3}$$

with the scaling factor

$$\beta = \frac{\sum_{i=1}^{L-1}\sum_{j=i+1}^{L}\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2\,\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|_2}{\sum_{i=1}^{L-1}\sum_{j=i+1}^{L}\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|_2^2} \tag{4.4}$$

that ensures that KS is bounded to the interval $[0, 1]$ [1]. Table 4.1 summarizes the CT, TW, and KS for the evaluated channel charts.

Table 4.1: CT, TW, and KS of the final channel charts

	Dissimilarity	↑ CT	↑ TW	↓ KS
3.5 GHz	$d_{\text{G-Fused}}$	0.9825	0.9830	0.2262
12 GHz	$d_{\text{G-Fused}}$	0.9821	0.9806	0.2108
Dual-band	$d_{\text{G-Avg-Fused}}$	0.9864	0.9858	0.1919
Dual-band	$d_{\text{G-SM-Fused}}$	0.9883	0.9870	**0.1849**
Align. ref 3.5 GHz ($\alpha = 1$)	$d_{\text{G-Fused}}$	0.9867	0.9893	0.2189
Align. ref 12 GHz ($\alpha = 1$)	$d_{\text{G-Fused}}$	0.9859	0.9849	0.2057
Align. Dual-band ($\alpha = 1$)	$d_{\text{G-Fused}}$	**0.9896**	**0.9911**	0.2007

All three fusion methods improved CT, TW, and KS compared with their single-band counterparts. When comparing the two single-band methods, the new dual-network architecture yielded higher CT and TW values and lower KS values than the single-network method at both frequencies. The dual-band alignment loss yielded the highest CT and TW values, whereas both ADP averaging and similarity multiplication yielded a lower KS value. The largest improvement in KS was observed with similarity multiplication, which also had the highest CT and TW among the single-network methods.

4.3.1 Localization Performance

The MAE quantifies the mean Euclidean distance between the ground-truth positions $\mathbf{x}^{(l)}$ and the corresponding affine-transformed channel chart coordinates $\hat{\mathbf{z}}^{(l)}$

over the test subset. It is computed as

$$\mathrm{MAE} = \frac{1}{L}\sum_{l=1}^{L} \|\mathbf{x}^{(l)} - \hat{\mathbf{z}}^{(l)}\|_2 \tag{4.5}$$

In addition to the total MAE, error statistics were computed separately for LoS and NLoS samples to assess the impact of propagation conditions on localization performance. Table 4.2 contains the positioning performance for the CC and fingerprinting methods used. It reports the MAE and the 90th-percentile error for the positioning of the test set.

Table 4.2: Positioning performance for CC and fingerprinting

		MAE (m)			90th
		Total	LoS	NLoS	percentile (m)
Channel Charting					
3.5 GHz	$d_{\text{G-Fused}}$	7.30	7.21	7.82	12.21
12 GHz	$d_{\text{G-Fused}}$	6.67	6.35	8.38	11.80
Dual-band	$d_{\text{G-Avg-Fused}}$	6.03	5.84	7.05	11.23
Dual-band	$d_{\text{G-SM-Fused}}$	**5.86**	**5.74**	**6.54**	10.99
Align. ref 3.5 GHz ($\alpha = 1$)	$d_{\text{G-Fused}}$	6.92	6.79	7.64	11.75
Align. ref 12 GHz ($\alpha = 1$)	$d_{\text{G-Fused}}$	6.52	6.21	8.18	11.33
Align. Dual-band ($\alpha = 1$)	$d_{\text{G-Fused}}$	6.30	5.99	7.95	**10.94**
Fingerprinting					
3.5 GHz		1.15	1.13	1.25	1.97
12 GHz		1.09	1.07	1.19	1.83
Dual-band		**1.02**	**1.01**	**1.08**	**1.72**

All three fusion methods achieved higher positioning accuracy than their single-band counterparts. with similarity multiplication providing the lowest MAE. Comparing the two single-band architectures, the dual-network alignment loss yielded a lower MAE than the single-network method. The dual-band alignment loss also yielded the lowest 90th-percentile error, although it did not have the lowest MAE. Both early-stage fusion methods, ADP averaging and similarity multiplication, achieved better performance than the late-stage alignment loss fusion. Fingerprinting showed the best localization performance, both for its single- and dual-band versions.

The localization performance of the different fingerprinting and CC methods is evaluated using the empirical CDF of the positioning error. This makes it possible to compare not only the average performance, but also how the errors are distributed across the full test set. The results are first shown for all samples, followed by a separate evaluation of NLoS conditions.

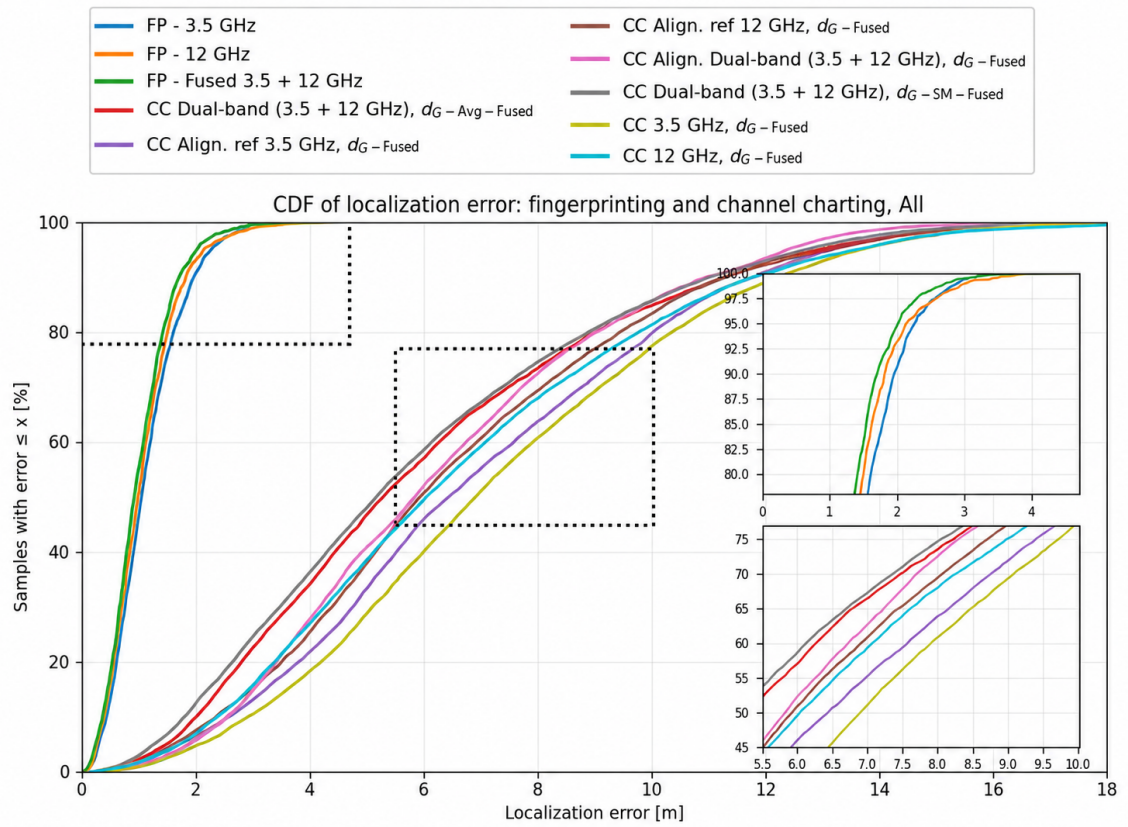


Figure 4.9: CDF of the localization error and zoomed results for all samples.

Figure 4.9 shows the CDF of the localization error for all samples. For fingerprinting, the fused 3.5 GHz and 12 GHz model (green) consistently achieves lower localization error than the two single-band models, with the green curve shifted slightly to the left of the other fingerprinting curves. For CC, the differences between the methods are most visible across the middle part of the CDF. In this region, the dual-band $d_{\text{G-Avg-Fused}}$ method (red) and the dual-band $d_{\text{G-SM-Fused}}$ method (grey) generally achieve the lowest errors, with their curves lying furthest to the left for much of the distribution.

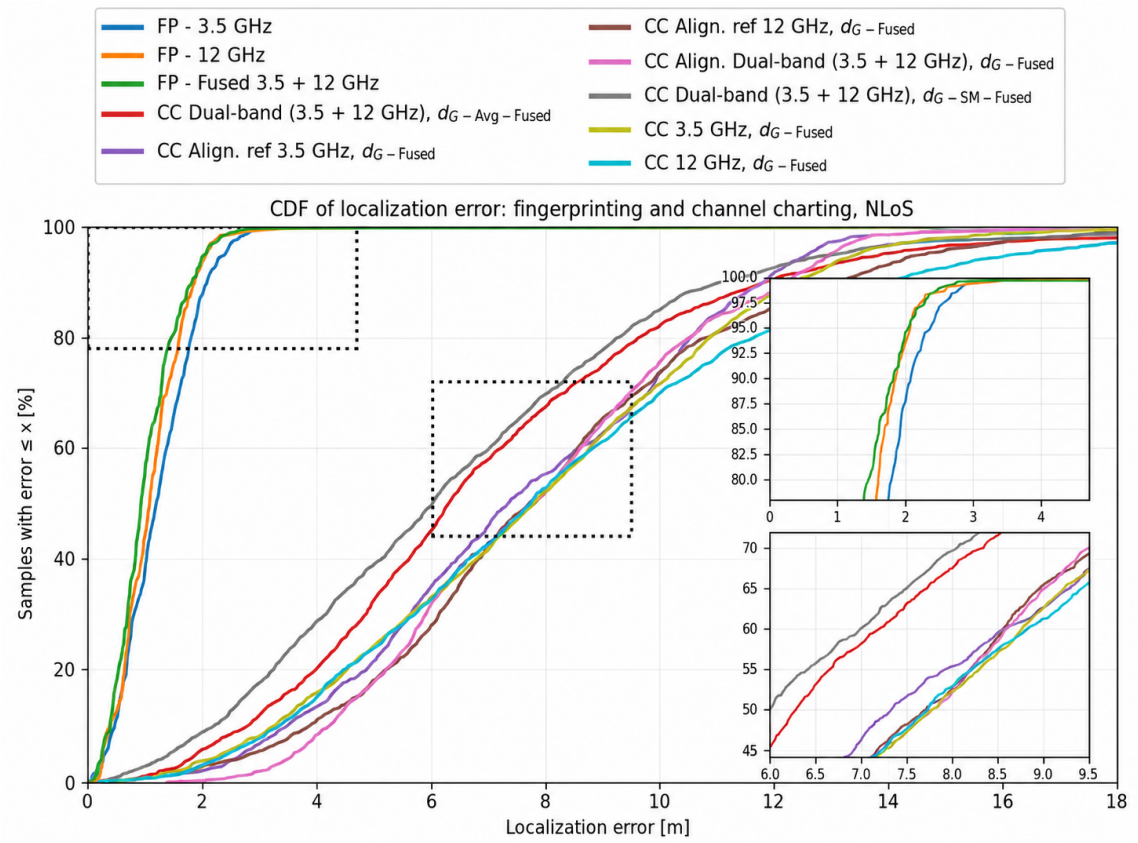


Figure 4.10: CDF of the localization error and zoomed results for NLoS samples.

Figure 4.10 shows the CDF of the localization error for NLoS samples. For fingerprinting, the overall pattern remains essentially the same as for all samples, with the fused 3.5 GHz and 12 GHz model (green) again achieving the lowest localization error among the fingerprinting methods.

For CC, the separation between the methods is even more pronounced, especially across the middle part of the CDF. In this region, the dual-band $d_{\text{G-Avg-Fused}}$ method (red) and the dual-band $d_{\text{G-SM-Fused}}$ method (grey) perform best, with a clearer gap to the other CC methods than in the CDF for all samples. This indicates that a large part of the difference observed previously for all samples is driven by the stronger performance differences in NLoS conditions.

4.3.2 Multi-band Performance Comparison

To compare the selected methods more directly, Figure 4.11 summarizes the MAE for the selected single-band and dual-band variants, separated into All, LoS, and NLoS samples. The Figure highlights the best-performing variant within each method group and shows how the dual-band approaches behave relative to their own single-band references.

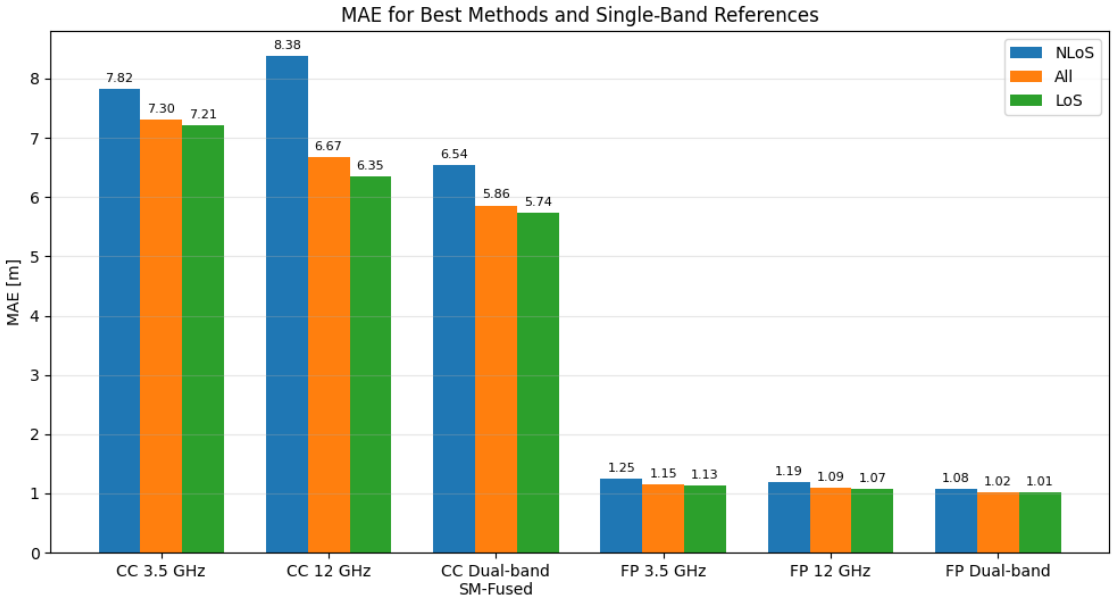


Figure 4.11: MAE for selected dual-band methods and single-band references.

Figure 4.11 shows that dual-band fusion improves performance for both CC and fingerprinting, with similarity multiplication giving the lowest MAE among the selected CC methods and the dual-band model performing best among the fingerprinting methods. More importantly, the LoS and NLoS split shows that the improvement is not evenly distributed. While the gain is visible in both cases, it is larger for NLoS samples, where the relationship between CSI and physical position is less direct. This indicates that the second frequency band provides especially useful complementary information in the parts of the radio environment where single-band observations are harder to interpret.

To make the relative effect of dual-band fusion clearer, Figure 4.12 presents the same comparison as a percentage MAE reduction relative to the best single-band reference within each method group.

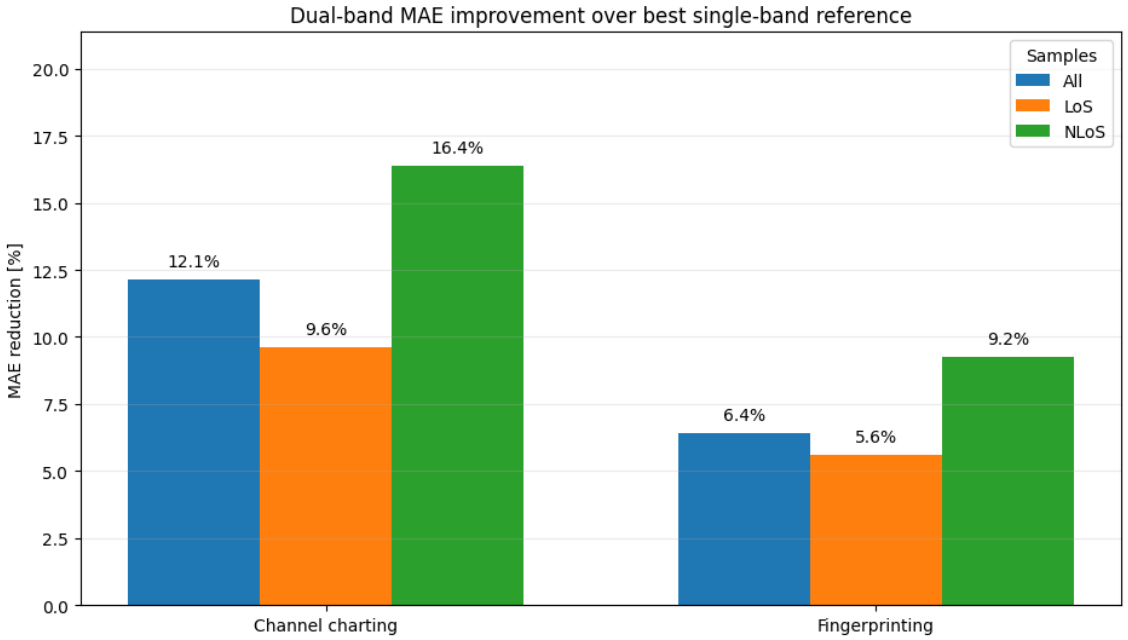


Figure 4.12: Percentage MAE reduction of the dual-band methods compared to the best single-band reference within each method group.

Figure 4.12 confirms that the relative gain from dual-band fusion is larger for CC than for fingerprinting. This suggests that the supervised fingerprinting model already extracts strong position information from a single band, while the unsupervised CC methods benefit more from the additional frequency information. In both method groups, the largest relative improvement occurs for NLoS samples, which supports the interpretation that multi-band fusion is most useful in the more challenging parts of the radio environment.

Overall, these results show that dual-band fusion improves both method groups, but with a stronger relative effect for CC and for NLoS conditions. This aligns with the main motivation of the multi-band experiments: to investigate whether additional frequency information can improve positioning when single-band CSI is less reliable.

5

Discussion

This chapter discusses the results in relation to the purpose of the thesis. Since the main numerical comparisons are presented in Chapter 4, the focus here is on interpreting why the observed improvements occur, what they suggest about multi-band CC, and which limitations should be considered when drawing conclusions from the results.

5.1 Interpretation of Multi-band Fusion

The results suggest that the benefit of multi-band CC is not only caused by adding more input data, but by improving the structure of the pairwise dissimilarities used to train the channel chart. This follows from the CC pipeline described in Section 3.2, where pairwise dissimilarities are used as the training signal for the learned chart. Since the Siamese network is trained to preserve distances derived from CSI [14], the quality of the dissimilarity matrix has a direct influence on the learned representation. If the dissimilarity metric better reflects physical proximity, the resulting chart is also more likely to preserve the underlying spatial structure.

This helps explain why the early-stage fusion methods performed well. Both ADP averaging and similarity multiplication combine information before the geodesic transformation, as described in Sections 3.3.1 and 3.3.2, meaning that the local neighborhood graph is constructed from already-fused frequency information. This is important because the geodesic step relies heavily on local neighbor relations, as described in Section 3.2.3. If incorrect neighbors are introduced at this stage, later parts of the pipeline cannot easily correct them. Similarity multiplication appears particularly suitable in this setting because it is more conservative than simple averaging. Two samples are only treated as strongly similar if both frequency bands indicate similarity. This reduces the chance that an unreliable similarity estimate from one band alone causes two physically distant samples to be placed close in the neighborhood graph.

The fingerprinting results provide a useful reference point for this interpretation. Since fingerprinting is trained directly on ground-truth positions, unlike the unsupervised CC methods described in Section 3.2.5, it can extract position information in a more direct way than CC. The smaller relative improvement from dual-band fingerprinting suggests that the supervised model already captures much of the available position information from a single band. For CC, the additional band has a

larger relative effect because it improves the pairwise dissimilarities that the unsupervised model is trained to preserve.

5.2 LoS and NLoS Conditions

The greater improvement in NLoS conditions is consistent with the motivation for using multi-band CSI. In LoS conditions, the direct path provides relatively clear geometric information, so a single frequency band can already contain useful position-related structure. In NLoS conditions, this relationship becomes less direct because the received signal is dominated by reflected, scattered, or diffracted paths, as discussed in Section 2.1. The second frequency band is therefore most useful where the single-band channel observation is more ambiguous.

Since different frequencies interact with the same environment differently, as described in Section 2.1.1, the additional band can provide complementary information about the propagation scene. This does not remove the difficulty of NLoS positioning, but it can make the dissimilarity structure more reliable in cases where one band alone gives an incomplete or misleading view of the environment. This also explains why the chosen radio environment affected the interpretation of the results. Because the environment included a significant amount of NLoS and partial NLoS samples, it created a setting where multi-band fusion had room to show a benefit. In a mostly LoS environment, the relative improvement from adding a second band may have been smaller.

The fingerprinting baseline also showed improvements from dual-band CSI in NLoS conditions, although the relative gain was smaller than for CC. This supports the interpretation that the second band contains useful complementary spatial information, but that its impact is stronger when the method depends on unsupervised CSI similarity rather than direct supervised position labels.

5.3 Limitations of the Interpretation

The results should be interpreted in the context of the simulated setup. The use of ray-tracing made it possible to evaluate the methods under controlled conditions and with available ground-truth positions, but it also means that real-world effects are only approximated. Movement patterns, scattering behavior, antenna deployment, and environmental variation may differ from practical systems. This is consistent with the scope described in the project limitations.

The timestamp fusion is another factor that affects interpretation. Since the receiver positions were generated using a structured movement pattern, samples close in time were also generally close in space. This makes timestamps especially useful in the present setup. As described in Section 3.2.2, timestamp fusion relies on the assumption that samples close in time are also likely to be close in space. In real deployments, user motion may include stops, irregular movement, or repeated visits to nearby locations at very different times, which could reduce the reliability of timestamp information.

The multi-band setup also assumes that both carrier frequencies share the same base-station geometry, as described in Chapter 3. This isolates the effect of frequency-dependent propagation, but it is more idealized than real deployments where different bands may use different antenna arrays, positions, orientations, or downtilts. The results therefore show the potential benefit of multi-band CSI under strong spatial alignment between bands, rather than fully realistic deployment conditions.

5.4 Implications

Overall, the results indicate that multi-band CSI is most useful for CC when it improves the local dissimilarity structure before the chart is learned. This relates to the ADP-based dissimilarity, timestamp fusion, and geodesic dissimilarity stages described in Sections 3.2.1–3.2.3. This suggests that future improvements should focus not only on neural network architecture, but also on how multi-band information is converted into pairwise dissimilarities.

The results also suggest that the value of multi-band CC depends on the propagation conditions. The gain is most relevant in difficult NLoS cases, where single-band CSI is less directly tied to physical position. This makes multi-band fusion especially interesting for environments where conventional positioning methods struggle, but less clearly necessary in simpler LoS-dominated settings.

The comparison with fingerprinting further suggests that multi-band fusion has different roles depending on the learning setup. In supervised fingerprinting, the second band improves an already direct mapping from CSI to position. In CC, it improves the structure from which the position-related representation is learned. This distinction follows from the methodological difference between the fingerprinting baseline and the unsupervised CC pipeline described in Chapter 3. This makes the impact of multi-band CSI especially relevant for unsupervised methods, where the training signal is derived entirely from the channel measurements themselves.

6

Conclusion

This work investigated whether multi-band CSI can improve CC and positioning performance compared to single-band CSI. Three dual-band CC fusion methods were developed and evaluated: ADP averaging, similarity multiplication, and dual-network alignment. In addition, a dual-band fingerprinting method was developed as a supervised baseline for comparison.

The results showed that multi-band fusion improved the CC performance compared to the corresponding single-band references. The improvements were visible both in the chart evaluation metrics and in the positioning accuracy after affine alignment. Among the CC methods, early-stage fusion was generally the most effective, with similarity multiplication giving the strongest overall positioning performance.

The fingerprinting baseline also benefited from using both frequency bands, although the relative improvement was smaller than for CC. This suggests that supervised fingerprinting already extracts strong position information from single-band CSI, while unsupervised CC gains more from the additional frequency information.

The improvements were most apparent in NLoS conditions, which were one of the main motivations for using a more challenging radio environment. Overall, the results indicate that multi-band CSI can provide complementary information that improves both CC and positioning, particularly in scenarios where single-band CSI is less reliable.

Future work should investigate additional frequency combinations, more realistic multi-band deployments with non-identical antenna configurations, and more diverse environments to determine how well the observed improvements generalize beyond the simulated street-canyon scenario.

Bibliography

[1] P. Stephan, F. Euchner, and S. t. Brink, "Angle-delay profile-based and timestamp-aided dissimilarity metrics for channel charting," *IEEE Transactions on Communications*, vol. 72, no. 9, pp. 5611–5625, 2024.

[2] K. Radnosrati, C. Fritsche, F. Gunnarsson, F. Gustafsson, and G. Hendeby, "Localization in 3gpp lte based on one rtt and one tdoa observation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3399–3411, 2020.

[3] X. Wang, L. Gao, S. Mao, and S. Pandey, "Csi-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2017.

[4] C. Studer, S. Medjkouh, E. Gonultaş, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[5] P. Ferrand, M. Guillaud, C. Studer, and O. Tirkkonen, "Wireless channel charting: Theory, practice, and applications," *IEEE Communications Magazine*, vol. 61, no. 6, pp. 124–130, 2023.

[6] H. Al-Tous, T. Ponnada, C. Studer, and O. Tirkkonen, "Multipoint channel charting-based radio resource management for v2v communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, p. 132, 2020.

[7] Z. Hu, X. Chen, Z. Zhou, and S. Mumtaz, "Localization with cellular signal rsrp fingerprint of multiband and multicell," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 9, pp. 2380–2394, 2024.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[9] K.-L. Du and M. N. S. Swamy, *Orthogonal frequency division multiplexing*. Cambridge University Press, 2010, p. 290–336.

[10] N. Rafie and B. Wang, "Simultaneous indoor localization based on wi-fi rss fingerprints," in *Advances of Science and Technology*, M. L. Berihun, Ed. Springer International Publishing, 2022, pp. 382–400. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-93709-6_25

[11] Y. Yang, M. Chen, Y. Blankenship, J. Lee, Z. Ghassemlooy, J. Cheng, and

S. Mao, "Positioning using wireless networks: Applications, recent progress, and future challenges," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 9, pp. 2149–2178, 2024.

[12] M. Stahlke, G. Yammine, T. Feigl, B. M. Eskofier, and C. Mutschler, "Indoor localization with robust global channel charting: A time-distance-based approach," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 1, pp. 3–17, 2023.

[13] P. Ferrand, A. Decurninge, L. G. Ordoñez, and M. Guillaud, "Triplet-based wireless channel charting: Architecture and experiments," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2361–2373, 2021.

[14] E. Lei, O. Castañeda, O. Tirkkonen, T. Goldstein, and C. Studer, "Siamese neural networks for wireless positioning and channel charting," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2019, pp. 200–207.

[15] S. Taner, V. Palhares, and C. Studer, "Channel charting in real-world coordinates with distributed mimo," *IEEE Transactions on Wireless Communications*, vol. 24, no. 9, pp. 7286–7300, 2025.

[16] Y. Zhang, G. Pan, M. F. Keskin, O. Kaltiokallio, M. Valkama, and H. Wymeersch, "Unilocpro: Unified localization integrating model-based geometry and channel charting," 2025. [Online]. Available: https://arxiv.org/abs/2510.27394

[17] F. Euchner, P. Stephan, and S. t. Brink, "Augmenting channel charting with classical wireless source localization techniques," in *2023 57th Asilomar Conference on Signals, Systems, and Computers*, 2023, pp. 1641–1647.

[18] J. Hoydis, S. Cammerer, F. Ait Aoudia, M. Nimier-David, L. Maggi, G. Marcus, A. Vem, and A. Keller, "Sionna," 2022, https://nvlabs.github.io/sionna/.

[19] E. DIJKSTRA, "A note on two problems in connexion with graphs." *Numerische Mathematik*, vol. 1, pp. 269–271, 1959. [Online]. Available: http://eudml.org/doc/131436

[20] R. Bridson, "Fast poisson disk sampling in arbitrary dimensions." *SIGGRAPH sketches*, vol. 10, no. 1, p. 1, 2007.

A

Appendix 1

A.1 Sionna RT Propagation Settings

Table A.1 summarizes the propagation variables used in Sionna RT [18] to generate the simulated CSI dataset.

Table A.1: Sionna RT propagation settings

Variable	**Value**
max_depth	3
max_num_paths_per_src	300
samples_per_src	300 000
synthetic_array	True
los	True
specular_reflection	True
diffuse_reflection	True
refraction	False
diffraction	True
edge_diffraction	False

The Sionna RT parameters describe the ray-tracing configuration used for CSI generation. They define the main propagation settings of the simulated environment and were kept fixed across all datasets, so that performance differences can be attributed to the channel charting method or the frequency band rather than to changes in the simulation setup. All materials in the simulated scene were set with a scattering coefficient of 0.3.

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY