The Dissertation Committee for Hongxiang Xie
certifies that this is the approved version of the following dissertation:

# Millimeter Wave Link Configuration in Practical Scenarios

Committee:

Nuria González-Prelcic, Co-Supervisor

Robert W. Heath, Jr., Co-Supervisor

Gustavo de Veciana

Antti Tölli

Haris Vikalo

# Millimeter Wave Link Configuration in Practical Scenarios

**by**

**Hongxiang Xie**

**DISSERTATION**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2021

Dedicated to my parents and sisters.

# Acknowledgments

I would like to first express the deepest appreciation to my supervisors Prof. Nuria González-Prelcic and Prof. Robert W. Heath, Jr, for their continuous strong support on me during the pursuit of my Ph.D. degree. I really appreciate their guidance, encouragement and persistent help for both my study, life, and professional career. Their passion and dedication to high-quality education and research have always been a great inspiration for me. I am also honored to be part of the WSIL group and have the opportunities to work with and learn from all the excellent colleagues. In addition, I would like to express my sincere gratitude for those people who helped me during my Ph.D. journey, including Dr. Amine Mezghani, Dr. Takayuki Shimizu and Dr. Joan Palacios. Discussions with them have been illuminating insightful ideas and directions on our projects and papers. Moreover, I would like to thank Prof. Gustavo de Veciana, Prof. Antti Tölli, and Prof. Haris Vikalo for serving as my committee members and providing precious comments and suggestions on my dissertation.

I feel grateful to have so many companions at UT. I thank all my friends in WSIL, Dalin Zhu, Yuyang Wang, Yi Zhang, Nitin Jonathan Myers, Anum Ali, Preeti Kumari, Khurram Usman Mazher, Travis Cuvelier, Karti Patel, Kevin Jinho Joe, etc., and I remember all the moments that we stayed

together and cherish them as the fortune of my life. I would also like thank all my other friends outside WSIL. It is my honor and luck to make acquaintance with them.

Finally, I would like to express my heartfelt gratitude to my family for their love and support on me. Without them, I will not be whom I am and cannot go this far.

# Millimeter Wave Link Configuration in Practical Scenarios

Publication No. _____

Hongxiang Xie, Ph.D.
The University of Texas at Austin, 2021

Supervisors: Nuria González-Prelcic
Robert W. Heath, Jr.

Acquiring channel state information (CSI) for link configuration in wideband millimeter wave (mmWave) massive multiple-input-multiple-output (MIMO) systems with hybrid architectures is challenging, due to the high dimensions of the channel matrices, the low signal-to-noise ratio (SNR) before beamforming, the various hardware constraints and the high mobility in the vehicular context. Previous work in this area exploits channel sparsity, statistical priors or side information to reduce the overhead associated to initial channel estimation or channel tracking. These works consider, however, a system model that neglects hardware imperfections. In addition, many of the proposed solutions are unable to operate in some realistic scenarios, such as vehicle-to-everything (V2X) communications. In this dissertation, we develop new signal processing solutions that can enable low-overhead mmWave link

configuration under various disturbances and practical limitations, e.g., hardware impairments, calibration errors, beam squint effect, channel blockage, high mobility, to name a few.

In the first part of this dissertation, we focus on the problem of wideband channel estimation for mmWave MIMO systems with different hardware imperfections. We first design a dictionary learning aided channel estimation strategy for wideband mmWave MIMO systems by explicitly considering the hardware uncertainties and calibration errors, and then derive algorithms that learn the optimal sparsifying dictionaries for channel representation and estimation. In a second contribution of this part, we further develop a dictionary learning aided compressive channel estimation scheme for mmWave MIMO systems by incorporating beam squint into the model of array responses. Numerical results show the proposed solutions can adapt to the practical scenarios and help reduce the overhead associated with channel estimation significantly.

In the second part of this dissertation, we deal with the problem of wideband channel tracking for mmWave MIMO systems with or without the impact of blockage. We first introduce statistical channel models that include the evolution models for channel gains and angles of arrival/departure, as well as the statistics of blockage events. Then, we design novel blockage detection schemes and efficient Bayesian channel tracking algorithms to facilitate the low-overhead tracking with or without blockage. Numerical results corroborate that the proposed solutions achieve better channel tracking performance even in mobile scenarios that suffer from highly dynamic blockage events.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Millimeter wave link configuration

Millimeter wave MIMO communication has become a key ingredient for the fifth generation (5G) of wireless communications [1], thanks to its potential to support Gigabits per second (Gbps) data rates with the aid of large available spectrum at mmWave band and the multi-antenna techniques. The design of MIMO communication at mmWave band is very different from that at lower frequency systems. To obtain enough link margin, large-scale antenna arrays comprised of hundreds or even thousands of antenna elements will be deployed at the transceivers, providing substantial beamforming gains. As a consequence, mmWave MIMO channels have a much higher dimension, which will induce overwhelming computational complexity and communication overhead associated to the link configuration procedures in practical scenarios.

The 5G New Radio (NR) considers a unified framework allowing an efficient use of mmWave bands, and thus a typical beam alignment mechanism for link configuration is adopted in 5G NR standards, i.e., the beam management protocol. This beam alignment scheme facilitates the directional communications at mmWave band by searching for the optimal beam pairs that provide

the largest link budget. Beam training can be performed either for initial access (IA) or beam tracking. The IA phase conducts the beam sweeping at both transmitter and receiver to determine their best beams. Once a beam pair link has been established between the access point (AP) and a user equipment (UE), the optimal beam pair will be tracked during beam tracking phases. To recover from a beam failure instance, several alternative beams neighboring the optimal one will also be monitored during the beam tracking procedure. Though the beam management protocol is robust to different channel conditions, it consumes a lot of time and frequency resources for the beam sweeping mechanism, especially when the number of beams in the codebook increases significantly and when the beam pair failure events (e.g., induced by channel blockage) happen frequently. In the case of high mobility V2X communications, the shorter channel coherence time interval makes it infeasible to do the beam alignment link configuration.

Channel estimation for IA is an effective alternative to beam training to configure mmWave antenna arrays. To reduce the high overhead associated to estimating the channel in mmWave MIMO systems, channel spatial sparsity has been exploited, e.g., [2, 3]. In most prior works, a narrow band channel model is considered, and the sparse channel matrices under a certain sparsifying dictionary are recovered from compressive channel measurements with few training resources. The dictionaries used in prior work are constructed from the transmit and receive array response vectors evaluated on a grid of quantized possible angles of arrival and departure (AoAs/AoDs) [4, 5]. The sparse

nature of frequency selective mmWave MIMO channels, both in the angular and delay domains, has also been considered to redefine the sparsifying dictionaries [6]. Unfortunately, in prior works, e.g., [2–6], ideal antenna arrays are considered for the channel model, without taking into consideration realistic models for their imperfections.

Though channel estimation has been studied in the literature for narrow band [3, 7, 8], and frequency-selective channels [5, 6, 9–12], little attention has yet been drawn towards the problem of wideband channel tracking. An interesting feature of the new 5G NR channel model [13] is that it incorporates spatial consistency. This way, a model for the temporal evolution of different channel parameters can be leveraged to design effective strategies for channel tracking. The problem of channel tracking under a narrow band communication model has been considered in recent work [14–21], although the mmWave channel is frequency-selective. Besides this unrealistic assumption, prior work on channel tracking considers different ad-hoc mathematical approaches to model the channel dynamics, that is, how the AoA, AoD, and channel gains evolve with time. Further, some prior work [21] considers the tracking problem only for a narrow band, multiple-input single-output (MISO) scenario, with single-antenna users, and considering channel vectors comprising of only two multipath components. Generally, tracking a frequency selective channel is significantly more challenging than tracking a narrowband channel. To operate in a frequency selective mmWave channel, a large number of channel matrices have to be tracked. In the literature, only a few works deal with the

problem of channel tracking assuming a hybrid mmWave MIMO system, and considering frequency-selective single-user scenarios [22–25]. The limitations of previous works include: 1) the considered channel model is not band-limited, thus leading to both unrealistic and sparser channel realizations; and 2) the channel gains are assumed to be time-invariant, which does not hold under realistic tracking periodicities.

## 1.2 Practical limitations in mmWave systems

In this section, we describe various common disturbances and uncertainties, including hardware imperfections, calibration errors, and beam squint effect, that will inevitably occur in practical wideband mmWave MIMO systems, but are usually ignored or erroneously handled.

Practically constructed antenna arrays deviate from the ideal case in many ways. Due to the manufacture and calibration errors, the antenna array will generate unexpected radiation patterns. The imperfect spacing between antenna elements also has to be considered. For example, the inter-element spacing in practical uniform linear arrays (ULA) is not the ideal half-wavelength, due to the limited manufacturing accuracy, which will result in irregular linear arrays rather than perfect ULAs. Therefore, the array response vectors do not longer follow the Vandermonde structure. The disturbance in the antenna spacing further increases the mutual coupling effect among antenna elements. There are also other hardware impairments in the radio frequency (RF) chains and the calibration errors that contribute to a general

mismatch between ideal and actual channel models. Given all the sources of mismatch, sparsifying dictionaries constructed from ideal array response vectors at quantized angels are not the best choice for exploiting channel sparsity.

Beam squint, namely, beams changing direction with frequency, is another significant effect. As shown in [26–28], the time delay of the same data symbol across the antenna array aperture is non-negligible in the large-scale MIMO configurations and/or the wideband systems. Due to the spatial delay difference of each data symbol at different antennas, the array steering vectors will have dissimilar responses at each frequency and then gives rise to the beam squint effect. Therefore, the adopted MIMO channel models in the existing literature, e.g., [4, 5, 29], that use the array response vectors evaluated at the carrier frequency for channel modeling at all subcarriers are no longer valid when operating with very large bandwidths. Moreover, the prior works on channel estimation under beam squint failed to consider the practical pulse shaping and filtering effects in the channel modeling, which is a non-trivial factor in practical wireless systems. We will show that the beam squint will not only induce the frequency-dependence on array steering vectors, but also yields to additional distortions at different antennas across all subcarriers, especially on side subcarriers in a multi-carrier system. This practical problem has not been considered in the previous literature.

Blockage is another feature that should be carefully considered for mmWave MIMO link configuration, especially for wideband channel tracking in high-mobility V2X scenarios. Due to the propagation properties of

mmWave signals, the directional mmWave links can be blocked by obstacles, like human body, neighboring vehicles, buildings, foliage and other infrastructures. Considering the practical scenarios like in urban areas, blockage events are expected to appear and disappear frequently, which will induce abrupt changes and fluctuations in the channel. Previous work, such as [14,15,17,30], derived different channel tracking strategies for narrowband channels, while only [22,31,32] consider a frequency selective channel model. These approaches consider an evolution model for the channel parameters and work well in the absence of blockage. Although [33–35] do consider blockage and propose algorithms to detect abrupt changes in the mmWave channel, only the narrowband case was considered. Moreover, [33–35] focus on fast re-access to the channel after link failure due to blockage, rather than on incorporating the statistics of blockage into the channel tracking strategy. Based on these, research on efficient wideband channel tracking strategies under blockage is essential for practical mmWave V2X scenarios.

## 1.3 Overview of contributions

This dissertation proposes new signal processing solutions that can enable low-overhead mmWave link configuration under various disturbances and practical limitations, including hardware impairments, calibration errors, beam squint effect, channel blockage, high mobility. The main contributions of this dissertation are summarized as follows.

1. We propose a novel dictionary learning based channel estimation solu-

tion for hybrid frequency-selective mmWave MIMO systems under the impact of hardware impairments. We develop the general model for such hybrid wideband mmWave MIMO systems by explicitly incorporating the hardware imperfections, like array manifold disturbances, antenna gain/phase errors, and array mutual coupling, and show why these hardware impairments should be embedded into the sparsifying dictionaries for channel representation. We also derive the Cramér-Rao Lower Bound (CRLB) for the estimation variance of frequency-domain channel matrices with unknown hardware impairments. We then design efficient dictionary learning approaches to optimize the best channel sparsifying dictionaries, which facilitate the compressive channel estimation in hybrid wideband mmWave MIMO systems and help reduce the training overhead significantly.

2. We propose another novel dictionary learning based channel estimation solution for hybrid frequency-selective mmWave MIMO systems under both the impact of hardware impairments and beam squint. We derive the new general channel model, which not only incorporates the hardware impairments, but also explicitly considers the impacts of combined pulse shaping and filtering on channels, showing the limitations of existing MIMO channel models and the associated channel estimation schemes. We then devise with beam squint new efficient dictionary learning algorithms by fully exploiting the channel properties at different subcarriers under beam squint.

3. We investigate the efficient wideband channel tracking and hybrid precoding design strategies for mmWave MIMO systems with or without the impact of blockage. We first introduce statistical channel models that include the evolution models for channel gains and AoAs/AoDs, and investigate efficient Bayesian channel tracking algorithms to facilitate the low-overhead tracking without channel blockage. We then extend the Bayesian tracking algorithm by generalizing the channel models with statistics of blockage events. To simultaneously obtain a high blockage detection probability and an accurate channel estimate, we also propose a double digital combiner architecture that uses different digital combiners for blockage detection and channel tracking.

The proposed solutions in this dissertation pay attentions to link configuration for hybrid wideband mmWave MIMO systems in different practical scenarios. The first and second contributions focus on compressive channel estimation with the aid of dictionary learning for various disturbances and practical limitations. The third contribution deals with the efficient channel tracking and blockage detection in the context of high mobility V2X scenarios. The proposed approaches in this dissertation have the advantages of lower complexity for online deployment. The first and second contributions make an assumption that enough computational power and capacity are available for offline dictionary training, while the third contributions assumes the statistical priors on the channel parameters are known or can be learned. Generally,

this dissertation has provided efficient and feasible solutions that are able to operate in realistic scenarios in wireless communication systems.

## 1.4 Thesis organization

Chapter 2 introduces the dictionary learning based channel estimation solution for hybrid wideband mmWave MIMO systems under the impact of hardware impairments. In Chapter 3, the beam squint effect is illustrated and a new dictionary learning aided wideband channel estimation approach, under both the impact of hardware impairments and beam squint, is proposed. Chapter 4 presents efficient wideband channel tracking and hybrid precoding design strategies for mmWave MIMO systems without the impact of blockage. The channel blockage modeling, detection and tracking schemes are illustrated in Chapter 5. Chapter 6 concludes the dissertation.

## 1.5 Notation

In this dissertation, vectors and matrices are denoted by boldface small and capital letters; the transpose, conjugate, Hermitian (conjugate transpose), inverse, and pseudo-inverse of the matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\overline{\mathbf{A}}$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$; $\mathbf{I}_M$ is an $M \times M$ identity matrix; $\mathbf{0}_{M\times N}$ is an $M \times N$ all-zero matrix and $\mathbf{1}_{M\times N}$ is an $M \times N$ all-one matrix; $[\mathbf{a}]_n$ denotes the $n$-th element of $\mathbf{a}$ and $[\mathbf{A}]_{:,j}$ denotes the $j$-th column vector of $\mathbf{A}$; $[\mathbf{A}]_{m,n}$ denotes the $(m,n)$-th element of $\mathbf{A}$; $\triangleq$ represents new definitions; $\mathcal{I}(N) \triangleq \{0, 1, \ldots, N-1\}$ denotes the index set of cardinality $N$; $\mathbb{C}$ and $\mathbb{R}$ denote the sets of complex and real numbers;

$\mathbb{E}\{\cdot\}$ returns expectation; $\mathrm{tr}\{\mathbf{A}\}$ is the trace of $\mathbf{A}$; $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with its diagonal elements given in $\mathbf{a}$ and $\mathrm{diag}\{\mathbf{A}\}$ formulates a vector by extracting the diagonal elements of $\mathbf{A}$; $\mathrm{j} = \sqrt{-1}$ is the imaginary unit; $|\mathcal{K}|$ denotes the cardinality of a set and $|\cdot|$ denotes the amplitude of a complex number; $\lfloor x \rfloor$ denotes the largest integer less than or equal to $x$ and $\lceil x \rceil$ denotes the smallest integer great than or equal to $x$; $\mathcal{F}(\cdot)[f]$ denotes the Fourier transform evaluated at frequency $f$; and $\|\mathbf{a}\|_0$ denotes the $\ell_0$ norm of vector, i.e., the number of nonzero elements of $\mathbf{a}$; $\otimes$, $\odot$ an $\star$ denote the Kronecker, Hadamard and Khatri-rao product; tensors are denoted by bold-faced calligraphic capital letters, e.g., $\boldsymbol{\mathcal{A}}$; For an $Q$-dimensional ($Q$-order) tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \times M_2 \times \ldots \times M_Q}$, the $q$-mode unfolding, denoted by $[\boldsymbol{\mathcal{A}}]_{(q)} \in \mathbb{C}^{M_q \times M_1 \cdots M_{q-1} M_{q+1} \cdots M_Q}$, represents a rearrangement of $\boldsymbol{\mathcal{A}}$ into a matrix, where the $q$-th index is used as a row index and all other indices are aligned along the columns (aligned in reverse cyclical ordering), and the columns of $[\boldsymbol{\mathcal{A}}]_{(q)}$ are referred to as mode-$q$ fibers (columns). The $q$-mode product between a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \times M_2 \times \ldots \times M_Q}$ and a matrix $\mathbf{B} \in \mathbb{C}^{N_q \times M_q}$ is denoted by $\boldsymbol{\mathcal{A}} \times_q \mathbf{B}$ and defined as $[\boldsymbol{\mathcal{A}} \times_q \mathbf{B}]_{(q)} = \mathbf{B}[\boldsymbol{\mathcal{A}}]_{(q)}$. Moreover, there is $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{A}} \times_1 \mathbf{B}^{(1)} \times_2 \mathbf{B}^{(2)} \cdots \times_Q \mathbf{B}^{(Q)} \Leftrightarrow [\boldsymbol{\mathcal{Y}}]_{(q)} = \mathbf{B}^{(q)}[\boldsymbol{\mathcal{A}}]_{(q)} \left(\mathbf{B}^{(Q)} \otimes \cdots \otimes \mathbf{B}^{(q+1)} \otimes \mathbf{B}^{(q-1)} \otimes \cdots \otimes \mathbf{B}^{(1)}\right)^T$. Furthermore, for a third-order tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$, its $i$-th horizontal, $j$-th lateral, and $c$-th frontal slides are denoted by $[\boldsymbol{\mathcal{A}}]_{i::} \in \mathbb{C}^{M_2 \times M_3}$, $[\boldsymbol{\mathcal{A}}]_{:j:} \in \mathbb{C}^{M_1 \times M_3}$, and $[\boldsymbol{\mathcal{A}}]_{::c} \in \mathbb{C}^{M_1 \times M_2}$. Moreover, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$ indicates that $\mathbf{x}$ is a circularly-symmetric complex Gaussian random variable vector with zero mean and covariance matrix $\mathbf{C}$. Time-domain vectors (matrices) are represented using $\mathbf{x}[n]$ ($\mathbf{X}[n]$), whilst

frequency-domain vectors (matrices) are represented using $\mathbf{x}[k]$ ($\mathbf{X}[k]$). We use $\boldsymbol{\mu}_{\mathbf{x}}$ to denote the mean of a vector $\mathbf{x}$, and $\mathbf{C}_{\mathbf{xy}}$ to denote the covariance matrix between two random vectors $\mathbf{x}$,$\mathbf{y}$.

## 1.6 Abbreviations

| | |
|---|---|
| **1D** | one-dimensional |
| **2D** | two-dimensional |
| **3D** | three-dimensional |
| **5G** | Fifth-generation |
| **6G** | Sixth-generation |
| **AoA** | Angle-of-arrival |
| **AoD** | Angle-of-departure |
| **AP** | Access point |
| **AWGN** | Additive white Gaussian noise |
| **CPD** | Change point detection |
| **CS** | Compressive sensing |
| **CSI** | Channel state information |
| **DFT** | Discrete Fourier transform |
| **GHz** | Gigahertz |
| **GMPF** | Generalized marginal particle filter |
| **IARM** | Ideal array response matrix |
| **IID** | Independent and identically distributed |
| **LOS** | Line-of-sight |

| | |
|---|---|
| **LTE** | Long Term Evolution |
| **MIMO** | Multiple-input multiple-output |
| **MISO** | Multi-input-single-output |
| **mmWave** | Millimeter wave |
| **MPF** | Marginal particle filter |
| **NLOS** | Non-line-of-sight |
| **NMSE** | Normalized mean square error |
| **NR** | New radio |
| **OFDM** | Orthogonal frequency division multiplexing |
| **OMP** | Orthogonal matching pursuit |
| **QPSK** | Quadrature phase shift keying |
| **RF** | Radio frequency |
| **SE** | Spectral efficiency |
| **SNR** | Signal-to-noise ratio |
| **SVD** | Singular value decomposition |
| **THz** | Terahertz |
| **UE** | User equipment |
| **ULA** | Uniform linear array |
| **V2X** | Vehicle-to-everything |

# Chapter 2

# Dictionary Learning for Channel Estimation in Hybrid Wideband mmWave MIMO Systems

In this chapter[1], we propose dictionary learning aided compressive channel estimation solutions for hybrid wideband mmWave MIMO systems under the impact of hardware impairments. Exploiting channel sparsity at mmWave frequencies can reduce the high training overhead associated with the channel estimation stage. Prior CS channel estimation techniques usually adopt the (overcomplete) Fourier transform matrix as sparsifying dictionary. This may not be the best choice when considering hardware impairments in practical arrays. We propose two dictionary learning algorithms to learn the best sparsifying dictionaries for channel matrices from observations obtained with practical hybrid frequency-selective mmWave MIMO systems. First, we optimize the combined dictionary, i.e., the Kronecker product of transmit and receive dictionaries, as it is used in practice to sparsify the channel matrix. This stage operates as a calibration phase, since all the hardware imperfections

[1]This chapter is based on our published work [36]: H. Xie and N. González-Prelcic, "Dictionary learning for channel estimation in hybrid frequency-selective mmWave MIMO systems," IEEE Trans. Wireless Commun., vol. 19, no. 11, pp. 7407–7422, Nov. 2020. This work was supervised by Prof. González-Prelcic. My contributions lie in designing and performing the research, doing simulations, analyzing data and writing the paper, etc.

are embedded into the learned dictionaries. Second, considering the different array structures at the transmitter and receiver, we exploit separable dictionary learning to find the best transmit and receive dictionaries. Once the channel is expressed in terms of the optimized dictionaries, various CS-based sparse recovery techniques can be applied for low overhead channel estimation. The effectiveness of the proposed dictionary learning algorithms under low SNR conditions has been corroborated via numerical simulations with different system configurations, array geometries and hardware impairments.

## 2.1 Introduction

To reduce the high overhead associated to estimating the channel in mmWave MIMO systems, channel spatial sparsity has been exploited, e.g., [2,3]. In most prior work, a narrowband channel model is considered, and the sparse channel matrices under a certain sparsifying dictionary are recovered from compressive channel measurements with few training resources. The dictionaries used in prior work are constructed from the transmit and receive array response vectors evaluated on a grid of quantized possible AoAs/AoDs [4,5]. The sparse nature of frequency selective mmWave MIMO channels, both in the angular and delay domains, has also been considered to redefine the sparsifying dictionaries [6]. Unfortunately, in prior work [2–6], ideal antenna arrays are considered in the channel model, without taking into consideration many practical effects, including hardware impairments and calibration errors.

Practically constructed antenna arrays deviate from the ideal case in

many ways. Due to the manufacture and calibration errors, the antenna array will generate unexpected radiation patterns. The imperfect spacing between antenna elements also has to be considered. For example, the inter-element spacing in practical ULA is not the ideal half-wavelength, due to the limited manufacturing accuracy, which will result in irregular linear arrays rather than perfect ULAs. Therefore, the array response vectors do not longer follow the Vandermonde structure. The disturbance in the antenna spacing further induces the mutual coupling effect among antenna elements. There are also other hardware impairments in the RF chains and the calibration errors that contribute to a general mismatch between ideal and actual channel models. Given all the sources of mismatch, sparsifying dictionaries constructed from ideal array response vectors at quantized angels are not the best choice for exploiting channel sparsity.

Learning a sparsifying dictionary using dictionary learning is one approach to capture the underlying practical structure in mmWave MIMO channels. This way, the compressive channel estimation stage will have the capability to adapt to all kinds of uncertainties and impairments. Dictionary learning for sparse signal representation has many applications in image processing including image denoising [37,38], component analysis [39], classification [40], or feature extraction [41] among others [42, 43]. Dictionary learning for wireless signal processing is not straightforward, however, given the different signal characteristics, the low SNR operation ranges and the different sparse structures to be exploited. The idea of dictionary learning based channel sparse

representation and estimation was proposed in [44] for massive MIMO systems operating at low frequencies. In that work, an overcomplete dictionary was learned from the channel measurement (training) data to substitute the predetermined discrete Fourier transform (DFT) dictionaries. A similar approach for dictionary learning based low-rank channel approximation was also considered in [45]. While [44] shows the power of leveraging dictionary learning for CSI acquisition, the formulation was limited to narrow band massive MIMO systems in high SNR regimes, without considering the aforementioned practical effects or the operating conditions at mmWave frequencies.

## 2.2 Contributions

In this chapter, we develop dictionary learning strategies for frequency selective mmWave MIMO systems with practical hybrid array architectures. The main contributions of this chapter are as follows:

- We propose a general model for frequency-selective mmWave MIMO systems that explicitly includes the array manifold disturbances, antenna gain/phase errors, and array mutual coupling. In the new model, the hardware impairments are embedded into the sparsifying dictionaries, and the channel parameters are represented by a sparse matrix. This general model will motivate and justify the formulation of our dictionary learning problems.

- We propose a channel estimation framework for mmWave channels based

on dictionary learning strategies. The dictionary learning process is implicitly calibrating the hardware at both ends, avoiding the need of conventional calibration and recalibration procedures based on extra transceiver hardware or an implicit feedback stage as the one described for example for the WiFi standard IEEE802.11n [46].

- We propose the *combined dictionary learning (CoDL)* algorithm to directly optimize a combined dictionary, i.e., the Kronecker product of transmit and receive dictionaries. By exploiting the common sparsity between subcarriers, CoDL is formulated as a non-convex optimization problem with two regularization terms to promote the common sparsity and combat high noise level. While common sparsity has been commonly considered for CS-based channel estimation, e.g., [5, 47, 48], it has never been exploited for learning a sparsifying dictionary for wireless channels.

- We propose the *separable dictionary learning (SeDL)* algorithm to optimize the transmit and receive dictionaries separately, which is more consistent with the practical system architecture, considering different array structures at transmitters and receivers. To exploit the separability of the Kronecker product between transmit and receive dictionaries, we formulate the SeDL problem in the tensor space, where the common sparsity among subcarriers is translated into the common sparsity support along the third-dimension of tensors. This is a typical property and constraint for our SeDL formulation that has not been considered in existing CS or dictionary learning problems. Though there is a performance

gap between SeDL and CoDL, SeDL has a much lower computational complexity due to the smaller sizes of transmit/receive dictionaries compared to the combined one used in CoDL. Therefore, SeDL achieves a good trade-off between performances and complexity.

- We derive the Cramér-Rao Lower Bound (CRLB) for the estimation variance of frequency-domain channel matrices with unknown hardware impairments. This helps to understand the performances of various compressive channel estimation techniques with different sparsifying dictionaries.

- We evaluate the proposed dictionary learning algorithms on different system configurations, array geometries and channel conditions. Numerical results show that the training overhead of channel estimation with learned dictionaries can be significantly reduced compared to that based on overcomplete dictionaries constructed from array response vectors. This corroborates the effectiveness of the proposed dictionary learning algorithms for hybrid wideband mmWave MIMO systems.

Compared with our prior work in [49], we develop a new DL algorithm (CoDL), we compute the CRLB for the original problem, we analyze the complexity of our approaches and we show the effectiveness of the proposed approaches via extensive simulations. The rest of the chapter is organized as follows. Section 2.3 describes the system, channel and signal models for the considered

wideband mmWave MIMO system based on fully connected hybrid architectures. The CRLB computation for the problem of estimating a mmWave channel and the hardware impairments is described in Section 2.4. Section 2.5 and Section 2.6 introduce the proposed CoDL and SeDL algorithms, respectively. Numerical simulations are provided in Section 2.7 to justify the effectiveness of the proposed dictionary learning algorithms and conclusions are drawn in Section 2.8.

## 2.3 System and channel models

### 2.3.1 System model

Consider a hybrid mmWave multi-user MIMO system with an AP of $N_{\mathrm{t}}$ antennas and $L_{\mathrm{t}}$ RF chains, as well as UEs using $N_{\mathrm{r}}$ antennas and $L_{\mathrm{r}}$ RF chains. The channel between the AP and the UE is assumed to be frequency-selective. An orthogonal frequency division multiplexing (OFDM)-based mmWave MIMO link employing $N_{\mathrm{c}}$ subcarriers is used to simultaneously transmit $N_{\mathrm{s}}$ ($\leq \min\{L_{\mathrm{t}}, L_{\mathrm{r}}\}$) data streams. The hybrid precoder and combiner adopted for such frequency-selective mmWave systems can be represented as $\mathbf{F}[c] = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[c]$ and $\mathbf{W}[c] = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}[c]$, for the $c$-th ($c \in \mathcal{I}(N_{\mathrm{c}})$) subcarrier, where $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{t}} \times L_{\mathrm{t}}}$ and $\mathbf{F}_{\mathrm{BB}}[c] \in \mathbb{C}^{L_{\mathrm{t}} \times N_{\mathrm{s}}}$ denote the analog and digital precoders, and $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{r}} \times L_{\mathrm{r}}}$ and $\mathbf{W}_{\mathrm{BB}}[c] \in \mathbb{C}^{L_{\mathrm{r}} \times N_{\mathrm{s}}}$ are the analog and digital combiners. The analog precoders/combiners are frequency-flat, while the baseband ones can be different for each subcarrier. In this chapter, we will consider a fully connected phase shifting network for the analog precoder and

combiner. During the channel estimation stage, prior knowledge of the training precoders and combiners is assumed at both the AP and the UE.

### 2.3.2 Channel model

We consider the frequency-selective channel model in [6, 50], consisting of $N_{\mathrm{p}}$ clusters with $N_{\mathrm{ray}}$ rays in each cluster and a delay tap length $N_{\mathrm{tap}}$. In the sequel, we will focus on dictionary learning and channel estimation for the downlink, although the analysis and proposed algorithms can be similarly developed for the uplink. The $d$-th delay tap of the downlink channel between the AP and a UE is denoted as $\boldsymbol{H}_d \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$, $d \in \mathcal{I}(N_{\mathrm{tap}})$, and can be expressed as

$$\boldsymbol{H}_d = \sqrt{\frac{N_{\mathrm{t}} N_{\mathrm{r}}}{N_{\mathrm{p}} N_{\mathrm{ray}}}} \sum_{\ell=1}^{N_{\mathrm{p}}} \sum_{k=1}^{N_{\mathrm{ray}}} \alpha_{\ell,k} p_{\mathrm{rc}}(dT_{\mathrm{s}} - \tau_\ell) \mathbf{C}_{\mathrm{R}} \boldsymbol{\Gamma}_{\mathrm{R}} \mathbf{a}_{\mathrm{R}}(\phi_{\ell,k}) (\mathbf{C}_{\mathrm{T}} \boldsymbol{\Gamma}_{\mathrm{T}} \mathbf{a}_{\mathrm{T}}(\theta_{\ell,k}))^*, \tag{2.3.1}$$

where $p_{\mathrm{rc}}(\tau)$ denotes a band-limited function including all filtering effects evaluated at $\tau$; $T_{\mathrm{s}}$ is the system sampling time; $\alpha_{\ell,k} \in \mathbb{C}$ is the complex gain; $\phi_{\ell,k} \in [-\pi, \pi)$ and $\theta_{\ell,k} \in [-\pi, \pi)$ are the AoA and AoD of the $k$-th ray in the $\ell$-th cluster; $\tau_\ell \in \mathbb{R}$ is the path delay of all rays in the $\ell$-th cluster. Moreover, $\mathbf{a}_{\mathrm{R}}(\phi_{\ell,k}) \in \mathbb{C}^{N_{\mathrm{r}} \times 1}$ and $\mathbf{a}_{\mathrm{T}}(\theta_{\ell,k}) \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$ denote the antenna array response vectors at UE and AP, which depend on the specific geometries of the antenna arrays and include any disturbance in the spacing between antenna elements due to manufacture errors. For instance, for a linear antenna array, instead of assuming a perfect ULA with an ideal uniform antenna spacing $d$, we denote

$\mathbf{a}_{\mathrm{R}}(\phi)$ as

$$\mathbf{a}_{\mathrm{R}}(\phi) \triangleq \frac{1}{\sqrt{N_{\mathrm{r}}}}\left[1, e^{-\mathrm{j}2\pi(d+\epsilon_{\mathrm{r},1})/\lambda \sin(\phi)}, \ldots, e^{-\mathrm{j}2\pi((N_{\mathrm{r}}-1)d+\epsilon_{\mathrm{r},N_{\mathrm{r}}-1})/\lambda \sin(\phi)}\right]^T, \quad (2.3.2)$$

where $\lambda$ is the carrier wavelength and $\epsilon_{\mathrm{r},1}, \ldots, \epsilon_{\mathrm{r},N_{\mathrm{r}}-1}$ denote the errors in the spacing between receive antenna elements. Furthermore, $\mathbf{C}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{r}}}$ and $\mathbf{C}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{t}}}$ in (2.3.1) are the mutual coupling matrices for the receive and transmit antenna arrays, representing the unwanted interchange of energy between elements in the arrays [51]. $\mathbf{\Gamma}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{r}}}$ and $\mathbf{\Gamma}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{t}}}$ are the antenna gain and phase error matrices, defined as $\mathbf{\Gamma}_{\mathrm{R}} \triangleq \mathrm{diag}\{g_{\mathrm{r},1}e^{\mathrm{j}\nu_{\mathrm{r},1}}, g_{\mathrm{r},2}e^{\mathrm{j}\nu_{\mathrm{r},2}}, \ldots, g_{\mathrm{r},N_{\mathrm{r}}}e^{\mathrm{j}\nu_{\mathrm{r},N_{\mathrm{r}}}}\}$, in which $\{g_{\mathrm{r},i}\}_{i=1}^{N_{\mathrm{r}}}$ are the receive gain error normalized to a reference amplitude and $\{\nu_{\mathrm{r},i}\}_{i=1}^{N_{\mathrm{r}}}$ are the additional receive phase errors. Note that these antenna gain and phase errors are due to the hardware impairments and calibration errors in production processes with respect to impedance matching networks, baluns, possible amplifiers, PCB materials, etc, [51]. Since our work focuses on the channel estimation problem with hardware impairments for a particular AP-UE mmWave link, during the channel coherence time both the physical channel parameters and the parameters modeling the impairments can be considered fixed and treated as deterministic. The deterministic consideration of the imperfections has also been assumed in many works on antenna array calibration such as [51–57].

We define an $N_{\mathrm{p}}N_{\mathrm{ray}} \times N_{\mathrm{p}}N_{\mathrm{ray}}$ diagonal matrix that contains the channel coeffients as $\mathbf{\Delta}_d \triangleq \sqrt{\frac{N_{\mathrm{t}}N_{\mathrm{r}}}{N_{\mathrm{p}}N_{\mathrm{ray}}}}\,\mathrm{diag}\{\alpha_{1,1}p_{\mathrm{rc}}(dT_{\mathrm{s}}-\tau_1), \ldots, \alpha_{N_{\mathrm{p}},N_{\mathrm{ray}}}p_{\mathrm{rc}}(dT_{\mathrm{s}}-\tau_{N_{\mathrm{p}}})\}$.

Then the compact expression for (2.3.1) is given as

$$\boldsymbol{H}_d = \mathbf{C}_{\mathrm{R}}\boldsymbol{\Gamma}_{\mathrm{R}}\mathbf{A}_{\mathrm{R}}\boldsymbol{\Delta}_d\mathbf{A}_{\mathrm{T}}^{*}\boldsymbol{\Gamma}_{\mathrm{T}}^{*}\mathbf{C}_{\mathrm{T}}^{*}, \tag{2.3.3}$$

where $\mathbf{A}_{\mathrm{R}} \triangleq \left[\mathbf{a}_{\mathrm{R}}(\phi_{1,1}), \ldots, \mathbf{a}_{\mathrm{R}}(\phi_{N_{\mathrm{p}},N_{\mathrm{ray}}})\right]$ and $\mathbf{A}_{\mathrm{T}} \triangleq \left[\mathbf{a}_{\mathrm{T}}(\theta_{1,1}), \ldots, \mathbf{a}_{\mathrm{T}}(\theta_{N_{\mathrm{p}},N_{\mathrm{ray}}})\right]$ collect the receive and transmit array response vectors evaluated at the actual AoAs and AoDs.

To exploit the sparsity within the channel matrix and enable the CS techniques, the exact expression of $\boldsymbol{H}_d$ in (2.3.3) can be approximated with the extended virtual channel model [4] as

$$\boldsymbol{H}_d \approx \mathbf{C}_{\mathrm{R}}\boldsymbol{\Gamma}_{\mathrm{R}}\mathbf{A}_{\mathrm{R}}^{\mathrm{v}}\boldsymbol{\Delta}_d^{\mathrm{v}}(\mathbf{A}_{\mathrm{T}}^{\mathrm{v}})^{*}\boldsymbol{\Gamma}_{\mathrm{T}}^{*}\mathbf{C}_{\mathrm{T}}^{*}, \tag{2.3.4}$$

where the dictionary matrices $\mathbf{A}_{\mathrm{R}}^{\mathrm{v}} \in \mathbb{C}^{N_{\mathrm{r}}\times G_{\mathrm{r}}}$ and $\mathbf{A}_{\mathrm{T}}^{\mathrm{v}} \in \mathbb{C}^{N_{\mathrm{t}}\times G_{\mathrm{t}}}$ generalize $\mathbf{A}_{\mathrm{R}}$ and $\mathbf{A}_{\mathrm{T}}$ in (2.3.3), while $\boldsymbol{\Delta}_d^{\mathrm{v}} \in \mathbb{C}^{G_{\mathrm{r}}\times G_{\mathrm{t}}}$ is the generalization of $\boldsymbol{\Delta}_d$ in (2.3.3). $\mathbf{A}_{\mathrm{R}}^{\mathrm{v}}$ and $\mathbf{A}_{\mathrm{T}}^{\mathrm{v}}$ collect the receive and transmit array response vectors evaluated on $G_{\mathrm{r}}$ quantized angles for AoAs and $G_{\mathrm{t}}$ quantized angles for AoDs, both sampled in $[-\pi, \pi)$, and $\boldsymbol{\Delta}_d^{\mathrm{v}}$ contains the path gains of these discrete quantized AoAs/AoDs at the non-zero elements. For the particular case of ULAs, the work in [58] proves that the virtual channel representation based on the array steering vectors evaluated on a grid of size equal to the number of antennas in the array is exact. In this case, the quantized dictionaries are in fact DFT and IDFT matrices. This proof can be extended to uniform planar arrays. This idea of building a basis for the representation of the channel based on virtual AoAs and AoDs, has been extended to general array geometries and grid sizes

larger than the number of antennas, using the extended virtual model defined in [4]. For this more general scenario, the extended model is only an approximation of the channel. Inspecting (2.3.4), if there is no prior knowledge on $\mathbf{C}_{\mathrm{R}}$, $\mathbf{C}_{\mathrm{T}}$, $\mathbf{\Gamma}_{\mathrm{R}}$, $\mathbf{\Gamma}_{\mathrm{T}}$, the existing channel estimation strategies based on CS techniques will not be applicable for this general model as the dictionary would be unknown. For this reason, in prior work like [5, 6], the mutual coupling matrices and gain/phase error matrices were all set as identity matrices and the antenna spacing disturbances were considered as zeros. Under this circumstance, a popular choice for $\mathbf{A}_{\mathrm{R}}^{\mathrm{v}}$ and $\mathbf{A}_{\mathrm{T}}^{\mathrm{v}}$ is the overcomplete DFT matrices if perfect ULAs are considered at the AP and the UE. Nevertheless, this is not an optimal choice for the general channel models that include hardware imperfections and callibration errors.

A natural solution is to substitute $\mathbf{C}_{\mathrm{R}}\mathbf{\Gamma}_{\mathrm{R}}\mathbf{A}_{\mathrm{R}}^{\mathrm{v}}$ and $\mathbf{C}_{\mathrm{T}}\mathbf{\Gamma}_{\mathrm{T}}\mathbf{A}_{\mathrm{T}}^{\mathrm{v}}$ in (2.3.4) with two general dictionaries $\mathbf{D}_{\mathrm{R}}$ and $\mathbf{D}_{\mathrm{T}}$, without any array structure related constraints, so that they can be applied to arbitrary antenna geometries and include all the hardware impairments. Under these assumptions, $\boldsymbol{H}_d$ in (2.3.3) can be generalized as

$$\boldsymbol{H}_d \approx \mathbf{D}_{\mathrm{R}}\mathbf{\Omega}[m,d]\mathbf{D}_{\mathrm{T}}^{*}, \tag{2.3.5}$$

where $\mathbf{D}_{\mathrm{R}} \triangleq [\mathbf{d}_{\mathrm{R},0}, \mathbf{d}_{\mathrm{R},1}, \ldots, \mathbf{d}_{\mathrm{R},K_{\mathrm{r}}-1}] \in \mathbb{C}^{N_{\mathrm{r}}\times K_{\mathrm{r}}}$ and $\mathbf{D}_{\mathrm{T}} \triangleq [\mathbf{d}_{\mathrm{T},0}, \mathbf{d}_{\mathrm{T},1}, \ldots, \mathbf{d}_{\mathrm{T},K_{\mathrm{t}}-1}] \in \mathbb{C}^{N_{\mathrm{t}}\times K_{\mathrm{t}}}$ denote the optimal receive and transmit dictionaries to be determined, and $\mathbf{\Omega}[m,d] \in \mathbb{C}^{K_{\mathrm{r}}\times K_{\mathrm{t}}}$ is a sparse channel matrix with few non-zero elements, similar to its counterpart $\mathbf{\Delta}_d^{\mathrm{v}}$ in (2.3.4).

The objective of the approximation from (2.3.4) to (2.3.5) is to operate with a more compact expression for the approximated channel that depends only on a dictionary for the transmit side $\mathbf{D}_{\mathrm{T}}$, a dictionary for the receive side $\mathbf{D}_{\mathrm{R}}$, both embedding the hardware impairments, and a sparse matrix representing the channel gains in the virtual domain $\boldsymbol{\Omega}[m,d]$. Notice that with this definition, the new virtual domain accounts for the hardware imperfections. To avoid the ambiguity between dictionaries and channel matrices, the dictionary atoms (columns) are normalized, i.e., $\|\mathbf{d}_{\mathrm{R},k_r}\|_2 = 1, \forall k_r \in \mathcal{I}(K_{\mathrm{r}})$ and $\|\mathbf{d}_{\mathrm{T},k_t}\|_2 = 1, \forall k_t \in \mathcal{I}(K_{\mathrm{t}})$, where $K_{\mathrm{r}}$ ($\geq N_{\mathrm{r}}$) and $K_{\mathrm{t}}$ ($\geq N_{\mathrm{t}}$) are the numbers of atoms of each dictionary. Note that with the optimized dictionaries accounting for practical antenna uncertainties and adapted to different channel effects, it is expected that the new channel matrix $\boldsymbol{\Omega}[m,d]$ will be sparser than $\boldsymbol{\Delta}_d^{\mathrm{v}}$.

For the geometric channel model in (2.3.5), the frequency-domain channel matrix at the $c$-th ($c \in \mathcal{I}(N_{\mathrm{c}})$) subcarrier can be written as

$$\begin{aligned}\mathbf{H}[c] &= \sum_{d=0}^{N_{\mathrm{tap}}-1} \boldsymbol{H}_d e^{-\mathrm{j}\frac{2\pi cd}{N_{\mathrm{c}}}} = \mathbf{C}_{\mathrm{R}}\boldsymbol{\Gamma}_{\mathrm{R}}\mathbf{A}_{\mathrm{R}} \underbrace{\left(\sum_{d=0}^{N_{\mathrm{tap}}-1} \boldsymbol{\Delta}_d e^{-\mathrm{j}\frac{2\pi cd}{N_{\mathrm{c}}}}\right)}_{\boldsymbol{\Delta}[c]} \mathbf{A}_{\mathrm{T}}^*\boldsymbol{\Gamma}_{\mathrm{T}}^*\mathbf{C}_{\mathrm{T}}^* \\ &\approx \mathbf{D}_{\mathrm{R}} \underbrace{\left(\sum_{d=0}^{N_{\mathrm{tap}}-1} \boldsymbol{\Omega}[m,d] e^{-\mathrm{j}\frac{2\pi cd}{N_{\mathrm{c}}}}\right)}_{\boldsymbol{\Omega}[c]} \mathbf{D}_{\mathrm{T}}^*, \end{aligned} \tag{2.3.6}$$

where $\boldsymbol{\Delta}[c]$ and $\boldsymbol{\Omega}[c]$ are defined accordingly and denote the channel gains in the frequency domain. In (2.3.6), we observe that $\boldsymbol{\Delta}[c]$ is computed as $\sum_{d=0}^{N_{\mathrm{tap}}-1} \boldsymbol{\Delta}_d e^{-\mathrm{j}\frac{2\pi cd}{N_{\mathrm{c}}}}$, $\forall c \in \mathcal{I}(N_{\mathrm{c}})$, which is the sum of the diagonal matrices

$\boldsymbol{\Delta}_d e^{-\mathrm{j}\frac{2\pi cd}{N_\mathrm{c}}}$. These matrices are diagonal because they are the product of a diagonal matrix and a scalar. We can conclude then that $\boldsymbol{\Delta}[c]$ is also diagonal. To be noted, we assume that $\mathbf{C}_\mathrm{R}, \mathbf{C}_\mathrm{T}, \boldsymbol{\Gamma}_\mathrm{R}, \boldsymbol{\Gamma}_\mathrm{T}$, as well as $\mathbf{A}_\mathrm{R}$ and $\mathbf{A}_\mathrm{T}$, are frequency-independent in this chapter, as we neglect the beam squint effect [26,27]. Therefore, the generalized dictionaries $\mathbf{D}_\mathrm{R}$ and $\mathbf{D}_\mathrm{T}$ are also frequency-independent in (2.3.6). Recalling that $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, the vectorization of (2.3.6) is given as

$$\begin{aligned}\mathrm{vec}\big(\mathbf{H}[c]\big) &= \big((\overline{\mathbf{C}}_\mathrm{T}\overline{\boldsymbol{\Gamma}}_\mathrm{T}\overline{\mathbf{A}}_\mathrm{T}) \otimes (\mathbf{C}_\mathrm{R}\boldsymbol{\Gamma}_\mathrm{R}\mathbf{A}_\mathrm{R})\big)\mathrm{vec}\big(\boldsymbol{\Delta}[c]\big) \\ &\approx \big(\overline{\mathbf{D}}_\mathrm{T} \otimes \mathbf{D}_\mathrm{R}\big)\mathrm{vec}\big(\boldsymbol{\Omega}[c]\big) = \boldsymbol{\Psi}\widetilde{\mathbf{h}}[c], \end{aligned} \tag{2.3.7}$$

where $\boldsymbol{\Psi} \triangleq \big(\overline{\mathbf{D}}_\mathrm{T} \otimes \mathbf{D}_\mathrm{R}\big) \in \mathbb{C}^{N_\mathrm{r}N_\mathrm{t} \times K_\mathrm{r}K_\mathrm{t}}$ is the combined dictionary and $\widetilde{\mathbf{h}}[c] \triangleq \mathrm{vec}(\boldsymbol{\Omega}[c]) \in \mathbb{C}^{K_\mathrm{r}K_\mathrm{t} \times 1}$ is the vectorized sparse channel matrix when this combined dictionary is used to build $\mathrm{vec}(\mathbf{H}[c])$.

### 2.3.3 Frequency domain signal model

According to [5,6], the received signal at the UE for the $c$-th subcarrier can be written as

$$\mathbf{y}[c] = \mathbf{W}^*_\mathrm{BB}[c]\mathbf{W}^*_\mathrm{RF}\mathbf{H}[c]\mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}[c]\mathbf{s}[c] + \mathbf{W}^*_\mathrm{BB}[c]\mathbf{W}^*_\mathrm{RF}\mathbf{z}[c], \tag{2.3.8}$$

where $\mathbf{s}[c] \in \mathbb{C}^{N_\mathrm{s} \times 1}$ is the transmitted signal vector at the $c$-th subcarrier and $\mathbf{z}[c] \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{N_\mathrm{r}})$ denotes the Gaussian noise vector with variance $\sigma^2$. During the DL and channel estimation phases, let the AP and the UE use the same frequency-flat precoder $\mathbf{F}_m \in \mathbb{C}^{N_\mathrm{t} \times L_\mathrm{t}}$ and combiner $\mathbf{W}_m \in \mathbb{C}^{N_\mathrm{r} \times L_\mathrm{r}}$ for all the subcarriers in the $m$-th OFDM symbol. Suppose that the transmitted

symbols satisfy $\mathbb{E}\{\mathbf{s}_m[c]\mathbf{s}_m^*[c]\} = \frac{P_{\rm tr}}{N_{\rm s}}\mathbf{I}_{N_{\rm s}}$, with $P_{\rm tr}$ the total power constraint and thus the SNR is defined as SNR $= \frac{P_{\rm tr}}{\sigma^2}$. To facilitate our proposed DL and channel estimation algorithms in the following, we decompose the transmitted symbol as $\mathbf{s}_m[c] = \mathbf{q}_m t_m[c]$ with $\mathbf{q}_m \in \mathbb{C}^{L_{\rm t}\times 1}$ a frequency-flat vector and $t_m[c]$ a scalar pilot symbol known at the receiver, as in [5]. To provide SNR gain for the DL algorithm to work properly, additional temporal spreading has to be considered. Therefore, during the learning stage, the training sequence is generated by repetition of the symbols in the auxiliary symbol sequence $r_n[c], n = 0, 1, 2, \ldots$ For a given spreading factor $N_{\rm rep}$, the training sequence is generated as $t_m[c] = r_{\lfloor \frac{m}{N_{\rm rep}} \rfloor}[c]$, i.e.,

$$t_m[c] = \underbrace{r_0[c], \cdots, r_0[c]}_{N_{\rm rep}\ \text{times}}, \underbrace{r_1[c], \cdots, r_1[c]}_{N_{\rm rep}\ \text{times}}, \cdots \tag{2.3.9}$$

Then the post-combining received training signals at $c$-th subcarrier, i.e., (2.3.8), in the $m$-th ($m = 0, 1, 2, \ldots$) training OFDM symbol is rewritten as

$$\mathbf{y}_m[c] = \mathbf{W}_m^*\mathbf{H}[c]\mathbf{F}_m\mathbf{q}_m t_m[c] + \mathbf{W}_m^*\mathbf{z}_m[c]. \tag{2.3.10}$$

To enable sparse reconstruction with a single subcarrier-independent measurement matrix, we multiply the received signal $\mathbf{y}_m[c]$ by $(t_m[c])^{-1}$ and vectorize it to get

$$\widetilde{\mathbf{y}}_m[c] \triangleq \mathrm{vec}\big((t_m[c])^{-1}\mathbf{y}_m[c]\big) = \big(\mathbf{q}_m^T\mathbf{F}_m^T \otimes \mathbf{W}_m^*\big)\,\mathrm{vec}\,(\mathbf{H}[c]) + \widetilde{\mathbf{z}}_m[c], \tag{2.3.11}$$

where $\widetilde{\mathbf{z}}_m[c] \triangleq \mathrm{vec}((t_m[c])^{-1}\mathbf{W}_m^*\mathbf{z}_m[c])$. Combining with (2.3.7), (2.3.11) can be approximated by

$$\begin{aligned}\widetilde{\mathbf{y}}_m[c] &\approx \left(\mathbf{q}_m^T\mathbf{F}_m^T \otimes \mathbf{W}_m^*\right)\left(\overline{\mathbf{D}}_{\mathrm{T}} \otimes \mathbf{D}_{\mathrm{R}}\right)\mathrm{vec}\big(\boldsymbol{\Omega}[c]\big) + \widetilde{\mathbf{z}}_m[c] \\ &= \boldsymbol{\Phi}_m\boldsymbol{\Psi}\widetilde{\mathbf{h}}[c] + \widetilde{\mathbf{z}}_m[c], \end{aligned} \tag{2.3.12}$$

where $\boldsymbol{\Phi}_m \triangleq \left(\mathbf{q}_m^T\mathbf{F}_m^T \otimes \mathbf{W}_m^*\right) \in \mathbb{C}^{L_{\mathrm{r}}\times N_{\mathrm{r}}N_{\mathrm{t}}}$ is defined accordingly. Note that $\boldsymbol{\Phi}_m$ is the sensing matrix for $m$-th OFDM symbol based on hybrid precoders and combiners, and $\boldsymbol{\Phi}_m\boldsymbol{\Psi}$ is the measurement matrix commonly defined in the literature of CS. To use the training spreading to average out the noise, we will keep the same $\mathbf{F}_m$ and $\mathbf{W}_m$ for the $N_{\mathrm{rep}}$ OFDM symbols during which the same training symbol $t_m[c]$ is transmitted, and then averaged the received signals as $\widetilde{\mathbf{y}}_{\mathrm{ave},i}[c] = \sum_{m=0}^{N_{\mathrm{rep}}-1}\widetilde{\mathbf{y}}_{iN_{\mathrm{rep}}+m}[c]/N_{\mathrm{rep}}$ for $i = 0, 1, 2, \ldots$. Next, we formulate a tall sensing matrix by stacking $M$ averaged measurements obtained above, and express the received signals at the $c$-th subcarrier as

$$\begin{aligned}\underbrace{\left[\widetilde{\mathbf{y}}_{\mathrm{ave},0}^T[c], \ldots, \widetilde{\mathbf{y}}_{\mathrm{ave},M-1}^T[c]\right]^T}_{\widetilde{\mathbf{y}}[c]} = \underbrace{\left[\boldsymbol{\Phi}_0^T, \ldots, \boldsymbol{\Phi}_{M-1}^T\right]^T}_{\boldsymbol{\Phi}}\boldsymbol{\Psi}\widetilde{\mathbf{h}}[c] + \\ \underbrace{\left[\widetilde{\mathbf{z}}_{\mathrm{ave},0}^T[c], \ldots, \widetilde{\mathbf{z}}_{\mathrm{ave},M-1}^T[c]\right]^T}_{\widetilde{\mathbf{z}}[c]}, \end{aligned} \tag{2.3.13}$$

where $\widetilde{\mathbf{y}}[c] \in \mathbb{C}^{ML_{\mathrm{r}}\times 1}$, $\boldsymbol{\Phi} \in \mathbb{C}^{ML_{\mathrm{r}}\times N_{\mathrm{r}}N_{\mathrm{t}}}$ and $\widetilde{\mathbf{z}}[c] \in \mathbb{C}^{ML_{\mathrm{r}}\times 1}$ are defined accordingly.

## 2.4 CRLB analysis for channel estimation with unknown dictionaries

The original problem we would like to solve is to obtain an estimation of $\mathbf{H}[c], c \in \mathcal{I}(N_{\mathrm{c}})$ in (2.3.6), which includes the physical channel parameters and the impairments. The derivations in this section and the Appendix A give us the CRLB [59], which characterizes the behavior of the best possible estimator of $\mathbf{H}[c]$ from the received signal defined in (2.3.11). Since the optimal estimator of this high number of parameters, some of them of high dimension, with a non linear dependence on the received signal and the channel $\boldsymbol{H}_d$, will likely involve a high complexity, we propose to consider an approximation of $\mathbf{H}[c]$. Then, in Section IV and Section V, we derive two estimators of this approximation of $\mathbf{H}[c]$ that operate in two stages. Our estimators are suboptimal, but with low complexity, and it is interesting to compare their performance with the CRLB that we derive in this section to understand how close to optimal our approaches perform.

Specifically, from (2.3.3) and (2.3.6), we first rewrite the vectorized channel matrix as $\mathrm{vec}(\mathbf{H}[c]) = \big((\overline{\mathbf{C}}_{\mathrm{T}}\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\overline{\mathbf{A}}_{\mathrm{T}}(\boldsymbol{\theta})) \star \mathbf{C}_{\mathrm{R}}\boldsymbol{\Gamma}_{\mathrm{R}}\mathbf{A}_{\mathrm{R}}(\boldsymbol{\phi})\big)\mathbf{g}[c]$ and $\mathbf{g}[c] \in \mathbb{C}^{N_{\mathrm{p}}N_{\mathrm{ray}}\times 1}$ contains the nonzero diagonal elements of $\boldsymbol{\Delta}[c]$ corresponding to the multipath gains at the $c$-th subcarrier. For the estimation of the channel matrices $\mathbf{H}[c]$, we collect the measurements at all $N_{\mathrm{c}}$ subcarriers based on

(2.3.13) as

$$\underbrace{\big[\widetilde{\mathbf{y}}[0],\ldots,\widetilde{\mathbf{y}}[N_{\rm c}-1]\big]}_{\widetilde{\mathbf{Y}}} = \mathbf{\Phi}\mathbf{\Psi}\,\underbrace{\big[\widetilde{\mathbf{h}}[0],\ldots,\widetilde{\mathbf{h}}[N_{\rm c}-1]\big]}_{\widetilde{\mathbf{H}}} + \underbrace{\big[\widetilde{\mathbf{z}}[0],\ldots,\widetilde{\mathbf{z}}[N_{\rm c}-1]\big]}_{\widetilde{\mathbf{Z}}}, \tag{2.4.1}$$

$$= \mathbf{\Phi}\big((\overline{\mathbf{C}}_{\rm T}\overline{\mathbf{\Gamma}}_{\rm T}\overline{\mathbf{A}}_{\rm T}(\boldsymbol{\theta}))\star\mathbf{C}_{\rm R}\mathbf{\Gamma}_{\rm R}\mathbf{A}_{\rm R}(\boldsymbol{\phi})\big)\,\underbrace{\big[\mathbf{g}[0],\ldots,\mathbf{g}[N_{\rm c}-1]\big]}_{\widetilde{\mathbf{G}}} + \widetilde{\mathbf{Z}}, \tag{2.4.2}$$

where $\widetilde{\mathbf{Y}} \in \mathbb{C}^{ML_{\rm r}\times N_{\rm c}}$, $\widetilde{\mathbf{H}} \in \mathbb{C}^{K_{\rm r}K_{\rm t}\times N_{\rm c}}$, $\widetilde{\mathbf{Z}} \in \mathbb{C}^{ML_{\rm r}\times N_{\rm c}}$ and $\widetilde{\mathbf{G}} \in \mathbb{C}^{N_{\rm p}N_{\rm ray}\times N_{\rm c}}$ are defined accordingly. Vectorizing (2.4.2), we have

$$\begin{aligned}\mathrm{vec}(\widetilde{\mathbf{Y}}) &= \Big(\mathbf{I}_{N_{\rm c}}\otimes\mathbf{\Phi}\big(\overline{\mathbf{C}}_{\rm T}\overline{\mathbf{\Gamma}}_{\rm T}\overline{\mathbf{A}}_{\rm T}(\boldsymbol{\theta})\star\mathbf{C}_{\rm R}\mathbf{\Gamma}_{\rm R}\mathbf{A}_{\rm R}(\boldsymbol{\phi})\big)\Big)\mathrm{vec}(\widetilde{\mathbf{G}}) + \mathrm{vec}(\widetilde{\mathbf{Z}}),\\ &= \Big(\mathbf{I}_{N_{\rm c}}\otimes\mathbf{\Phi}\big(\overline{\mathbf{C}}_{\rm T}\otimes\mathbf{C}_{\rm R}\big)\big(\overline{\mathbf{\Gamma}}_{\rm T}\otimes\mathbf{\Gamma}_{\rm R}\big)\big(\overline{\mathbf{A}}_{\rm T}(\boldsymbol{\theta})\star\mathbf{A}_{\rm R}(\boldsymbol{\phi})\big)\Big)\mathrm{vec}(\widetilde{\mathbf{G}}) + \mathrm{vec}(\widetilde{\mathbf{Z}}).\end{aligned} \tag{2.4.3}$$

Therefore, the received signal is Gaussian distributed as $\mathrm{vec}(\widetilde{\mathbf{Y}}) \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C}_{\mathrm{vec}(\widetilde{\mathbf{Z}})})$, where $\boldsymbol{\mu} \triangleq \big(\mathbf{I}_{N_{\rm c}}\otimes\mathbf{\Phi}\big(\overline{\mathbf{C}}_{\rm T}\otimes\mathbf{C}_{\rm R}\big)\big(\overline{\mathbf{\Gamma}}_{\rm T}\otimes\mathbf{\Gamma}_{\rm R}\big)\big(\overline{\mathbf{A}}_{\rm T}(\boldsymbol{\theta})\star\mathbf{A}_{\rm R}(\boldsymbol{\phi})\big)\big)\mathrm{vec}(\widetilde{\mathbf{G}})$ and $\mathbf{C}_{\mathrm{vec}(\widetilde{\mathbf{Z}})} \triangleq \sigma^2\mathbf{I}_{N_{\rm c}}\otimes\mathrm{blkdiag}\big\{\mathbf{W}_0^*\mathbf{W}_0, \mathbf{W}_1^*\mathbf{W}_1, \cdots, \mathbf{W}_{M-1}^*\mathbf{W}_{M-1}\big\}$. For simplicity, we can further whiten the received signal by $\sigma\mathbf{C}_{\mathrm{vec}(\widetilde{\mathbf{Z}})}^{-1/2}\mathrm{vec}(\widetilde{\mathbf{Y}})$, such that the whitened mean is $\boldsymbol{\mu}_{\rm w} \triangleq \big(\mathbf{I}_{N_{\rm c}}\otimes\mathbf{\Phi}_{\rm w}\big(\overline{\mathbf{C}}_{\rm T}\otimes\mathbf{C}_{\rm R}\big)\big(\overline{\mathbf{\Gamma}}_{\rm T}\otimes\mathbf{\Gamma}_{\rm R}\big)\big(\overline{\mathbf{A}}_{\rm T}(\boldsymbol{\theta})\star\mathbf{A}_{\rm R}(\boldsymbol{\phi})\big)\big)\mathrm{vec}(\widetilde{\mathbf{G}})$, with $\mathbf{\Phi}_{\rm w} \triangleq \mathrm{blkdiag}\big\{\mathbf{W}_0^*\mathbf{W}_0, \mathbf{W}_1^*\mathbf{W}_1, \cdots, \mathbf{W}_{M-1}^*\mathbf{W}_{M-1}\big\}^{-1/2}\mathbf{\Phi}$, and the whitened noise covariance matrix is $\sigma^2\mathbf{I}_{ML_{\rm r}N_{\rm c}}$.

Let us stack all the unknown parameters of interest into a vector $\boldsymbol{\xi}$, including $N_{\rm p}N_{\rm ray}$ pairs of AoAs/AoDs $\{\boldsymbol{\phi},\boldsymbol{\theta}\}$, the real and imaginary parts of the path gain vectors $\mathbf{g}^{\rm R}[c], \mathbf{g}^{\rm I}[c]$ (from $\mathbf{g}[c] \triangleq \mathbf{g}^{\rm R}[c] + \mathrm{j}\mathbf{g}^{\rm I}[c], \forall c \in \mathcal{I}(N_{\rm c})$), the antenna gain errors $\boldsymbol{g}_{\rm t} \triangleq \{g_{{\rm t},i}\}_{i=1}^{N_{\rm t}}$ and $\boldsymbol{g}_{\rm r} \triangleq \{g_{{\rm r},i}\}_{i=1}^{N_{\rm r}}$, the antenna phase errors

$\boldsymbol{\phi}_{\rm t} \triangleq \{\phi_{{\rm t},i}\}_{i=1}^{N_{\rm t}}$ and $\boldsymbol{\phi}_{\rm r} \triangleq \{\phi_{{\rm r},i}\}_{i=1}^{N_{\rm r}}$, the antenna spacing errors $\boldsymbol{\epsilon}_{\rm t} \triangleq \{\epsilon_{{\rm t},i}\}_{i=1}^{N_{\rm t}-1}$ and $\boldsymbol{\epsilon}_{\rm r} \triangleq \{\epsilon_{{\rm r},i}\}_{i=1}^{N_{\rm r}-1}$, and the coupling coefficients $\boldsymbol{c}_{\rm t} \triangleq \{c_{{\rm t},i,j}, \forall 1<i<j<N_{\rm t}\}$ and $\boldsymbol{c}_{\rm r} \triangleq \{c_{{\rm r},i,j}, \forall 1<i<j<N_{\rm r}\}$. Note that there are only $N_{\rm t}-1$ and $N_{\rm r}-1$ spacing error terms at the transmitter and receiver, respectively. Moreover, there are $\frac{N_{\rm t}(N_{\rm t}-1)}{2}$ and $\frac{N_{\rm r}(N_{\rm r}-1)}{2}$ unknown coupling coefficients in $\mathbf{C}_{\rm T}$ and $\mathbf{C}_{\rm R}$, respectively, for the coupling matrix for ULAs is Toeplitz matrices with diagonal elements being one (notice that the number of coupling coefficients is different if other types of coupling matrices are assumed). Therefore, $\boldsymbol{\xi} \triangleq [\boldsymbol{\theta}^T, \boldsymbol{\phi}^T, \mathbf{g}^{\rm R}[0]^T, \mathbf{g}^{\rm I}[0]^T, \ldots, \mathbf{g}^{\rm R}[N_{\rm c}-1]^T, \mathbf{g}^{\rm I}[N_{\rm c}-1]^T, \boldsymbol{g}_{\rm t}^T, \boldsymbol{g}_{\rm r}^T, \boldsymbol{\phi}_{\rm t}^T, \boldsymbol{\phi}_{\rm r}^T, \boldsymbol{\epsilon}_{\rm t}^T, \boldsymbol{\epsilon}_{\rm r}^T]^T \in \mathbb{C}^{2N_{\rm p}N_{\rm ray}+2N_{\rm p}N_{\rm ray}N_{\rm c}+2N_{\rm r}+2N_{\rm t}+(N_{\rm t}-1)+(N_{\rm r}-1)+N_{\rm r}(N_{\rm r}-1)/2+N_{\rm t}(N_{\rm t}-1)/2}$. Then the $(i,j)$-th entry in the overall Fisher Information Matrix (FIM) for $\boldsymbol{\xi}$, denoted by $[\mathbf{I}(\boldsymbol{\xi})]_{\xi_i,\xi_j}$, is given as [59]

$$[\mathbf{I}(\boldsymbol{\xi})]_{\xi_i,\xi_j} = 2\Re\left\{\frac{\partial \boldsymbol{\mu}^*(\boldsymbol{\xi})}{\partial \xi_i}\mathbf{C}^{-1}_{\rm vec(\widetilde{\mathbf{Z}})}(\boldsymbol{\xi})\frac{\partial \boldsymbol{\mu}(\boldsymbol{\xi})}{\partial \xi_j}\right\} = \frac{2}{\sigma^2}\Re\left\{\frac{\partial \boldsymbol{\mu}_{\rm w}{}^*(\boldsymbol{\xi})}{\partial \xi_i}\frac{\partial \boldsymbol{\mu}_{\rm w}(\boldsymbol{\xi})}{\partial \xi_j}\right\}. \tag{2.4.4}$$

Based on (2.4.4) , the detailed derivations of FIM $\mathbf{I}(\boldsymbol{\xi})$ are given in Appendix A. Once the FIM $\mathbf{I}(\boldsymbol{\xi})$ in (A.0.26) is available, considering that the channel $\text{vec}(\mathbf{H}[c]) = \big((\overline{\mathbf{C}}_{\rm T}\overline{\boldsymbol{\Gamma}}_{\rm T}\overline{\mathbf{A}}_{\rm T}(\boldsymbol{\theta})) \star \mathbf{C}_{\rm R}\boldsymbol{\Gamma}_{\rm R}\mathbf{A}_{\rm R}(\boldsymbol{\phi})\big)\mathbf{g}[c]$ is a function of the unknown variable $\boldsymbol{\xi}$, then the covariance matrix of any unbiased estimator for $\text{vec}(\mathbf{H}[c])$ is given as [59]

$$\mathbf{C}_{\text{vec}(\mathbf{H}[c])} \geq \mathbf{J}_{\boldsymbol{\xi}}(c)\mathbf{I}^{-1}(\boldsymbol{\xi})\mathbf{J}^*_{\boldsymbol{\xi}}(c), \quad \forall c \in \mathcal{I}(N_{\rm c}), \tag{2.4.5}$$

where $\mathbf{J}_{\boldsymbol{\xi}}(c) \triangleq \dfrac{\partial \text{vec}(\mathbf{H}[c])}{\partial \boldsymbol{\xi}^T}$ is the Jacobian matrix of the channel $\text{vec}(\mathbf{H}[c])$ w.r.t. $\boldsymbol{\xi}$. It can be expressed as a block column matrix as $\mathbf{J}_{\boldsymbol{\xi}}(c) =$

$[\mathbf{J}_{\boldsymbol{\theta}}(c)\ \mathbf{J}_{\boldsymbol{\phi}}(c)\ \mathbf{J}_{\mathbf{g}[c]}(c)\ \mathbf{J}_{\boldsymbol{g}_{\mathrm{t}}}(c)\ \mathbf{J}_{\boldsymbol{g}_{\mathrm{r}}}(c)\ \mathbf{J}_{\boldsymbol{\phi}_{\mathrm{t}}}(c)\ \mathbf{J}_{\boldsymbol{\phi}_{\mathrm{r}}}(c)\ \mathbf{J}_{\boldsymbol{\epsilon}_{\mathrm{t}}}(c)\ \mathbf{J}_{\boldsymbol{\epsilon}_{\mathrm{r}}}(c)\ \mathbf{J}_{\boldsymbol{c}_{\mathrm{t}}}(c)\ \mathbf{J}_{\boldsymbol{c}_{\mathrm{r}}}(c)]$, and the expressions of each column block can be derived similarly following the above procedure of derivation for the FIM $\mathbf{I}(\boldsymbol{\xi})$. Finally, the total CRLB for the channel matrices $\mathrm{vec}(\mathbf{H}[c]), \forall c \in \mathcal{I}(N_{\mathrm{c}})$ is

$$\mathrm{CRLB} = \sum_{c=0}^{N_{\mathrm{c}}-1} tr\{\mathbf{J}_{\boldsymbol{\xi}}(c)\mathbf{I}^{-1}(\boldsymbol{\xi})\mathbf{J}_{\boldsymbol{\xi}}^{*}(c)\}. \tag{2.4.6}$$

The value of the CRLB will be included in the simulations as a reference to evaluate the channel estimation error performance when using the learned dictionaries to model the channel.

## 2.5 Combined dictionary learning (CoDL)

In this section, we develop a solution to the problem of learning a combined dictionary for frequency-selective mmWave MIMO channels, i.e., the Kronecker product of transmit and receive dictionaries in (2.3.7).

### 2.5.1 Problem formulation of CoDL

Inspecting (2.3.13), there are two unknown parameters, i.e., the combined dictionary $\boldsymbol{\Psi}$ and the channel vector $\widetilde{\mathbf{h}}[c]$. The conventional DL problem for recovering the optimal $\boldsymbol{\Psi}$ as well as $\widetilde{\mathbf{h}}[c]$ can be formulated as

$$\min_{\boldsymbol{\Psi}\in\mathcal{D},\ \widetilde{\mathbf{h}}[c]} \quad \big\|\widetilde{\mathbf{y}}[c] - \boldsymbol{\Phi}\boldsymbol{\Psi}\widetilde{\mathbf{h}}[c]\big\|_2^2 + w\big\|\widetilde{\mathbf{h}}[c]\big\|_0, \tag{2.5.1}$$

where $\|\cdot\|_0$ is the $\ell_0$-norm of a vector, i.e., the number of non-zero elements and $w$ is a regularization factor that trades off the model fitting error and

sparsity level. Moreover, the unit-norm condition in (2.3.5) transfers to the unit-norm constraint for the combined dictionary as

$$\mathcal{D} = \Big\{ \mathbf{\Psi} \in \mathbb{C}^{N_{\mathrm{t}} N_{\mathrm{r}} \times K_{\mathrm{r}} K_{\mathrm{t}}} \big| \|[\mathbf{\Psi}]_{:,j}\|_2 = 1,\ \forall j \in \mathcal{I}(K_{\mathrm{r}} K_{\mathrm{t}}) \Big\}. \tag{2.5.2}$$

Note that the process of dictionary learning avoids the need for hardware calibration and recalibration procedures and human interaction, since the impairments are explicitly embedded into the sparsifying dictionaries.

Beyond the basic formulation in (2.5.1), the DL problem can be extended by considering the following aspects:

#### 2.5.1.1 Common support among subcarriers

The frequency domain channel vectors $\widetilde{\mathbf{h}}[c],\ c \in \mathcal{I}(N_{\mathrm{c}})$ exhibit an important property: common sparsity support among subcarriers [5]. Specifically, the rows of $\widetilde{\mathbf{H}}$ in (2.4.1) will be either all zeros or non-zeros. This property is commonly used when beam squint effect is not considered.

#### 2.5.1.2 Multiple measurements at different locations

In addition to increasing the number of training OFDM symbols and exploiting the common support between subcarriers, it is possible for the UE to collect multiple measurements at different positions. This way, a larger training set is created where different channel sparsity patterns are included, and then the sparsifying dictionary does not depend on the specific location, but on the specific impairments. Suppose there are $N_{\mathrm{sa}}$ measurement samples

collected at different positions, and each measurement $\widetilde{\mathbf{Y}}$ defined in (2.4.1) is then denoted by $\widetilde{\mathbf{Y}}^{(u)}, u \in \mathcal{I}(N_{\mathrm{sa}})$. Stacking them in a compact form gives

$$\underbrace{\left[\widetilde{\mathbf{Y}}^{(0)}, \widetilde{\mathbf{Y}}^{(1)}, \ldots, \widetilde{\mathbf{Y}}^{(N_{\mathrm{sa}}-1)}\right]}_{\mathbf{Y}} = \mathbf{\Phi}\mathbf{\Psi} \underbrace{\left[\widetilde{\mathbf{H}}^{(0)}, \widetilde{\mathbf{H}}^{(1)}, \ldots, \widetilde{\mathbf{H}}^{(N_{\mathrm{sa}}-1)}\right]}_{\mathbf{H}} + \underbrace{\left[\widetilde{\mathbf{Z}}^{(0)}, \widetilde{\mathbf{Z}}^{(1)}, \ldots, \widetilde{\mathbf{Z}}^{(N_{\mathrm{sa}}-1)}\right]}_{\mathbf{Z}}, \tag{2.5.3}$$

where $\mathbf{Y} \in \mathbb{C}^{ML_{\mathrm{r}} \times N_{\mathrm{c}} N_{\mathrm{sa}}}$, $\mathbf{H} \in \mathbb{C}^{K_{\mathrm{r}} K_{\mathrm{t}} \times N_{\mathrm{c}} N_{\mathrm{sa}}}$ and $\mathbf{Z} \in \mathbb{C}^{ML_{\mathrm{r}} \times N_{\mathrm{c}} N_{\mathrm{sa}}}$ are defined accordingly.

The overhead associated with the measurement collection can be neglected by operating in two different ways:

- Off-line operation: In many scenarios, an indoor WiFi system or an enterprise deployed cellular supported communication system, the learning process can be performed at the stage of network setup. In Section 2.7, we show that at mmWave frequencies the DL process can be performed in the order of hundred of microseconds, even at low SNR. This training time is negligible compared to the recalibration period, which will be very infrequent in these indoor environments.

- On-line operation: The mmWave link could be initially established without calibration, i.e., neglecting the hardware impairments. The achieved rate can be later improved by collecting enough measurements during different training periods for channel tracking. After a short time, when the number of measurements is sufficient, the sparsifying dictionaries

can be updated to account for hardware imperfections. The next channel estimate will be more accurate, and therefore, the achieved rate will increase, as we will show in Section 2.7. With this strategy, the training overhead associated to learning the dictionary can be distributed across several online training intervals and its overall effect can be neglected.

#### 2.5.1.3 Denoising option

The proposed DL algorithm has to be able to operate under the low SNR conditions in mmWave communications. Next, we denote a denoised (and unknown) version of $\mathbf{Y}$ in (2.5.3) by $\mathbf{X}$, i.e., $\mathbf{Y} = \mathbf{\Phi\Psi H} + \mathbf{Z} = \mathbf{X} + \mathbf{Z'}$, where $\mathbf{Z'}$ is comprised of $\mathbf{Z}$ and the mismatch error between $\mathbf{X}$ and $\mathbf{\Phi\Psi H}$. Then, an additional regularization term $\|\mathbf{Y} - \mathbf{X}\|_F^2$ can be added to the following DL problem formulation to alleviate the influence of noise.

Given these three aspects, the CoDL problem in (2.5.1) can be written as [37]

$$\begin{aligned} \min_{\substack{\mathbf{\Psi}\in\mathcal{D},\ \mathbf{X},\\ \widetilde{\mathbf{h}}^{(u)}[0],\ldots,\widetilde{\mathbf{h}}^{(u)}[N_{\mathrm{c}}-1]}} \quad & \left\|\mathbf{X} - \mathbf{\Phi\Psi H}\right\|_F^2 + w_1 \sum_{u=0}^{N_{\mathrm{sa}}-1} \sum_{c=0}^{N_{\mathrm{c}}-1} \left\|\widetilde{\mathbf{h}}^{(u)}[c]\right\|_0 + w_2 \|\mathbf{Y} - \mathbf{X}\|_F^2 \\ \text{s.t.} \quad & \operatorname{supp}\{\widetilde{\mathbf{h}}^{(u)}[0]\} = \cdots = \operatorname{supp}\{\widetilde{\mathbf{h}}^{(u)}[N_{\mathrm{c}}-1]\},\ \forall u \in \mathcal{I}(N_{\mathrm{sa}}). \end{aligned} \tag{2.5.4}$$

Note that an additional superscript $(u)$ is added as sample index for $\widetilde{\mathbf{h}}^{(u)}[c]$ to indicate which channel matrix $\widetilde{\mathbf{H}}^{(u)}$ it belongs to. Inspecting (2.5.4), the sparsity enhancement regularization term and the common sparsity support property can be considered jointly, i.e., using the joint sparsity regularization

$\ell_2/\ell_1$ norm of $\widetilde{\mathbf{H}}^{(u)}, \forall u \in \mathcal{I}(N_{\mathrm{sa}})$, rather than encoding each subcarrier channel vector $\widetilde{\mathbf{h}}^{(u)}[c]$ separately. In doing so, the sparsity enhancement regularization term together with the common sparsity support constraint in (2.5.4) can be integrated as $f(\mathbf{H})$, given by

$$f(\mathbf{H}) = \sum_{u=0}^{N_{\mathrm{sa}}-1} \left\|\widetilde{\mathbf{H}}^{(u)}\right\|_{2,1} = \sum_{u=0}^{N_{\mathrm{sa}}-1} \sum_{i=0}^{K_{\mathrm{r}}K_{\mathrm{t}}-1} \left\| \left[\widetilde{\mathbf{H}}^{(u)}\right]_{i,:} \right\|_2. \tag{2.5.5}$$

Then, we have the final formulation for the CoDL problem as

$$\min_{\mathbf{\Psi}\in\mathcal{D},\ \mathbf{H},\ \mathbf{X}} \left\|\mathbf{X} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{H}\right\|_F^2 + w_1 f(\mathbf{H}) + w_2 \|\mathbf{Y} - \mathbf{X}\|_F^2. \tag{2.5.6}$$

### 2.5.2 Problem optimization of CoDL

The objective function in (2.5.6) is not jointly convex w.r.t. $\mathbf{\Psi}$, $\mathbf{H}$ and $\mathbf{X}$, but it is convex in each of the variables when the others are kept fixed. Thereby, a possible approach to finding the solution of (2.5.6) involves solving three sub-problems: 1) updating the channel $\mathbf{H}$ with fixed $\mathbf{\Psi}$ and $\mathbf{X}$, 2) updating the dictionary $\mathbf{\Psi}$ with fixed $\mathbf{X}$ and newly updated $\mathbf{H}$, and 3) updating $\mathbf{X}$ with fixed newly updated $\mathbf{H}$ and $\mathbf{\Psi}$.

#### 2.5.2.1 Sparse coding stage

Let us first assume that $\mathbf{\Psi}$ and $\mathbf{X}$ are fixed, so that the optimization problem is reduced to a sparse coding problem for updating $\mathbf{H}$. Thanks to the separability of the objective function (2.5.6) w.r.t. each $\widetilde{\mathbf{H}}^{(u)}, u \in \mathcal{I}(N_{\mathrm{sa}})$, we can compute $\widetilde{\mathbf{H}}^{(u)}$ separately. Specifically, all $\widetilde{\mathbf{H}}^{(v)},\ v \neq u$ are fixed when computing $\widetilde{\mathbf{H}}^{(u)}$. Along these lines, the objective function in (2.5.6) can be

further simplified to

$$\min_{\widetilde{\mathbf{H}}^{(u)}} \ \left\|\widetilde{\mathbf{X}}^{(u)} - \mathbf{\Phi}\mathbf{\Psi}\widetilde{\mathbf{H}}^{(u)}\right\|_F^2 + w_1\left\|\widetilde{\mathbf{H}}^{(u)}\right\|_{2,1}. \tag{2.5.7}$$

Various sparse coding techniques can be used to solve this problem, such as the orthogonal matching pursuit (OMP) [60], the simultaneous-weighted (SW-OMP) [5], and the alternating direction method of multipliers (ADMM) [42, 61, 62].

#### 2.5.2.2 Dictionary update stage

Following the sparse coding step, we then update dictionary $\mathbf{\Psi}$ column by column with fixed $\mathbf{X}$ and newly estimated $\mathbf{H}$. When updating the $k$-th ($k \in \mathcal{I}(K_{\mathrm{r}}K_{\mathrm{t}})$) atom, all the other columns are fixed as well. Then the dictionary update is formulated as

$$\min_{\mathbf{\Psi}} \ \left\|\mathbf{X} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{H}\right\|_F^2, \quad \text{s.t.} \ \ \left\|[\mathbf{\Psi}]_{:,k}\right\|_2 = 1. \tag{2.5.8}$$

This problem can be solved by using the well-known K-SVD algorithm [63] (or approximate K-SVD [64] with reduced complexity), which updates the dictionary $\mathbf{\Psi}$ atom by atom, or by the canonical method of optimal direction (MOD) algorithm [65].

#### 2.5.2.3 Denoising stage

Following the sparse coding and dictionary update stages, the value of $\mathbf{X}$ is updated by computing

$$\min_{\mathbf{X}} \ \left\|\mathbf{X} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{H}\right\|_F^2 + w_2\|\mathbf{Y} - \mathbf{X}\|_F^2, \tag{2.5.9}$$

**Algorithm 1** : Combined dictionary learning (CoDL)

- **Input:** Training measurements $\mathbf{Y} \in \mathbb{C}^{ML_{\mathrm{r}} \times N_{\mathrm{c}} N_{\mathrm{sa}}}$, measurement matrix $\mathbf{\Phi} \in \mathbb{C}^{ML_{\mathrm{r}} \times N_{\mathrm{r}} N_{\mathrm{t}}}$, number of OFDM subcarriers $N_{\mathrm{c}}$, number of measurement samples $N_{\mathrm{sa}}$, regularization parameters $w_1, w_2$.
- **Initialization:** Set the dictionary matrix $\mathbf{\Psi} \in \mathbb{C}^{N_{\mathrm{t}} N_{\mathrm{r}} \times K_{\mathrm{r}} K_{\mathrm{t}}}$ using measurement data with normalized columns, and set $\mathbf{X} = \mathbf{Y}$.
- **While** *not converge* **do**

  1. *Sparse coding stage*: Solve (2.5.7) for each $u \in \mathcal{I}(N_{\mathrm{sa}})$ to get the channel matrix $\mathbf{H} = \left[\widetilde{\mathbf{H}}^{(0)}, \widetilde{\mathbf{H}}^{(1)}, \ldots, \widetilde{\mathbf{H}}^{(N_{\mathrm{sa}}-1)}\right]$.

  2. *Dictionary update stage*: If MOD is adopted, the update of $\mathbf{\Psi}$ is given by $\mathbf{\Psi} = \mathbf{\Phi}^{\dagger}\mathbf{X}\mathbf{H}^{*}(\mathbf{H}\mathbf{H}^{*})^{-1}$, followed by normalization on each column. If K-SVD is considered, $\mathbf{\Psi}$ is updated column by column as did in [63, 64].

  3. *Denoising stage*: Update $\mathbf{X}$ by solving (2.5.9), i.e., $[\mathbf{X}]^{\mathrm{new}} = (1 + w_2)^{-1}(w_2\mathbf{Y} + \mathbf{\Phi}\mathbf{\Psi}\mathbf{H})$.

  **end while**

- **Output:** The optimal dictionary $\mathbf{\Psi}$, the channel matrix $\mathbf{H}$ and the denoised version measurement $\mathbf{X}$.

which can be solved by least squares (LS) and given as $[\mathbf{X}]^{\mathrm{new}} = (1 + w_2)^{-1}(w_2\mathbf{Y} + \mathbf{\Phi}\mathbf{\Psi}\mathbf{H})$.

The whole procedure of the proposed CoDL algorithm is summarized in Algortihm 1. The proposed algorithm will stop either if the values of objective function (2.5.6) at adjacent iterations are sufficiently close, or if the maximum number of iterations is reached. In Section 2.7, we will show via numerical simulations that the objective function is to decrease quickly as the number of

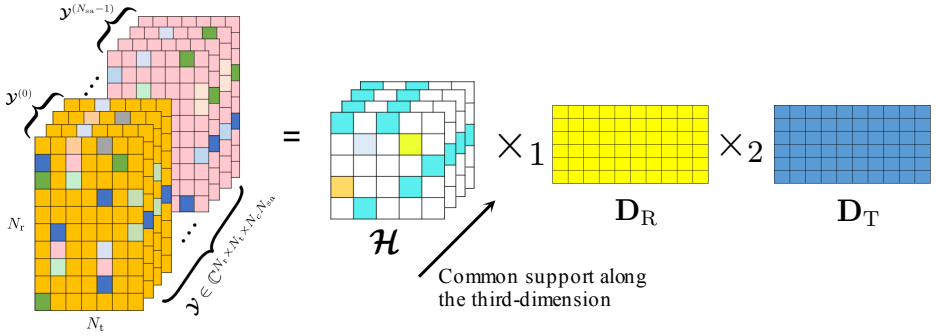



Figure 2.1: Illustration of tensor expression of $\mathcal{Y} \in \mathbb{C}^{N_r \times N_t \times N_c N_{sa}}$ and its tensor mode product expansion.

iterations increase, which helps to illustrate fast convergence of the proposed algorithms.

### 2.5.3 Convergence analysis and dictionary initialization

For sparse coding stage, the ADMM algorithm can compute the exact solution for each sub-problem, its convergence is guaranteed by the existing ADM theory [42,66]. In this stage, when the dictionaries are fixed, each sparse coding step decreases the value of the objective function. While for the dictionary update and denoising stages, as explained in [63], an additional reduction or no change in the mismatch error is guaranteed. Therefore, the alternating steps for optimizing the CoDL problem ensure a monotonic decrease in the objective function and then convergence to a local minimum is guaranteed.

As the alternating optimization can only guarantee to converge to a local minimum, the initialization is then of significant importance to avoid local minimizers and ensure the learned dictionaries to be closer to the actu-

ally dictionaries. The common initialization choices for current DL algorithms include random initial dictionary, an overcomplete wavelet/Fourier dictionary or a sample of data measurements. Some other initialization methods were also proposed for different DL problems in the literature, e.g., [67–69]. In this chapter, after thorough comparisons and investigations among different initialization methods, we finally adopt the dictionary initialization algorithm (DIA) proposed in [67]. The main idea of DIA is to use incoherent structures to create a very good initialization for a DL problem, which involves an iterative adaptation of the dictionary to the dataset with pruning of the less used atoms and constructions of new atoms that fit the data better. The detailed procedures of DIA can be found in the Algorithm 1 in [67].

## 2.6 Separable dictionary learning (SeDL)

Inspecting (2.5.6), the combined dictionary $\mathbf{\Psi}$ was learned without considering the specific structure constraints on $\mathbf{D}_{\mathrm{T}}$ and $\mathbf{D}_{\mathrm{R}}$, i.e., $\|\mathbf{d}_{\mathrm{T},k_t}\|_2 = 1$, $\forall k_t \in \mathcal{I}(K_{\mathrm{t}})$ and $\|\mathbf{d}_{\mathrm{R},k_r}\|_2 = 1$, $\forall k_r \in \mathcal{I}(K_{\mathrm{r}})$. In this section, we investigate the formulation and optimization for the SeDL problem. By separating the constraints on transmit and receive dictionaries, the SeDL problem will be more suited for the practical MIMO systems.

### 2.6.1 Problem formulation of SeDL

To facilitate the problem formulation for SeDL, we re-formulate (2.5.3) in a higher dimensional tensor space [70]. Specifically, the collected mea-

surements after removing the training sequences, i.e., $\mathbf{\Phi}^{\dagger}\mathbf{Y} \in \mathbb{C}^{N_{\rm r}N_{\rm t}\times N_{\rm c}N_{\rm sa}}$ in (2.5.3), is re-written as a three-order (three-dimensional) tensor $\mathcal{Y} \in \mathbb{C}^{N_{\rm r}\times N_{\rm t}\times N_{\rm c}N_{\rm sa}}$, which is stacked by $N_{\rm sa}$ sub-tensors $\mathcal{Y}^{(u)} \in \mathbb{C}^{N_{\rm r}\times N_{\rm t}\times N_{\rm c}}, u \in \mathcal{I}(N_{\rm sa})$ along the third dimension, as shown in Fig. 2.1. A similar definition goes for its unknown denoised version $\mathcal{X} \in \mathbb{C}^{N_{\rm r}\times N_{\rm t}\times N_{\rm c}N_{\rm sa}}$. We also replace $\mathbf{H} \in \mathbb{C}^{K_{\rm r}K_{\rm t}\times N_{\rm c}N_{\rm sa}}$ with a tensor $\mathcal{H} \in \mathbb{C}^{K_{\rm r}\times K_{\rm t}\times N_{\rm c}N_{\rm sa}}$, which is stacked by $N_{\rm sa}$ sub-tensors $\mathcal{H}^{(u)} \in \mathbb{C}^{K_{\rm r}\times K_{\rm t}\times N_{\rm c}}, u \in \mathcal{I}(N_{\rm sa})$ along the third dimension. Then we have the following formulation for the SeDL problem as

$$\begin{aligned}\min_{\mathcal{H},\mathcal{X},\mathbf{D}_{\rm R},\mathbf{D}_{\rm T}} \quad & \sum_{u=0}^{N_{\rm sa}-1} \left\|\mathcal{X}^{(u)} - \mathcal{H}^{(u)}{}_{\times 1}\, \mathbf{D}_{\rm R}\, {}_{\times 2}\, \overline{\mathbf{D}}_{\rm T}\right\|_F^2 \\ & \qquad + w_1 \sum_{u=0}^{N_{\rm sa}-1} \|\mathcal{H}^{(u)}\|_{1,1,2} + w_2\left\|\mathcal{Y} - \mathcal{X}\right\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{d}_{{\rm T},k_t}\|_2 = 1,\ \|\mathbf{d}_{{\rm R},k_r}\|_2 = 1,\ \forall k_t \in \mathcal{I}(K_{\rm t}), k_r \in \mathcal{I}(K_{\rm r}). \end{aligned} \tag{2.6.1}$$

where $\|\mathcal{H}^{(u)}\|_{1,1,2}$ denotes the summation of $\ell_2$ norm of all the mode-3 columns along the third dimension in $\mathcal{H}^{(u)}$ and is used to promote the common sparsity support between subcarriers.

### 2.6.2 Problem optimization of SeDL

Similar to Section 2.5.2, the optimization for SeDL in (2.6.1) is also divided into three parts, i.e., 1) sparse coding, 2) dictionary update, and 3) denosing stage.

**Sparse coding stage** Let us first assume that the transmit/receive dictionaries, $\mathbf{D}_{\rm T}$ and $\mathbf{D}_{\rm R}$, together with $\mathcal{X}$ are fixed, and then update the channel

tensor $\boldsymbol{\mathcal{H}}^{(u)}$ for each $u \in \mathcal{I}(N_{\text{sa}})$. Under these assumptions, the objective function is reduced to

$$\min_{\boldsymbol{\mathcal{H}}^{(u)}} \quad \left\|\boldsymbol{\mathcal{X}}^{(u)} - \boldsymbol{\mathcal{H}}^{(u)} \times_1 \mathbf{D}_{\text{R}} \times_2 \overline{\mathbf{D}}_{\text{T}}\right\|_F^2 + w_1 \left\|\boldsymbol{\mathcal{H}}^{(u)}\right\|_{1,1,2}. \tag{2.6.2}$$

After some mathematical manipulations, (2.6.2) is equivalent to

$$\min_{\boldsymbol{\mathcal{H}}^{(u)}} \quad \left\|[\boldsymbol{\mathcal{X}}^{(u)}]_{(3)}^T - \left(\overline{\mathbf{D}}_{\text{T}} \otimes \mathbf{D}_{\text{R}}\right)[\boldsymbol{\mathcal{H}}^{(u)}]_{(3)}^T\right\|_F^2 + w_1 \left\|[\boldsymbol{\mathcal{H}}^{(u)}]_{(3)}^T\right\|_{2,1}, \tag{2.6.3}$$

where the $\ell_{1,1,2}$ norm of tensor is simplified to the $\ell_2/\ell_1$ norm of matrix according to the definitions of these norms and the involved tensor operations. Inspecting (2.6.3), it is exactly in the form of (2.5.7). In fact, this is intuitive that when $\mathbf{D}_{\text{T}}$ and $\mathbf{D}_{\text{R}}$ are fixed, they can be combined to replace $\boldsymbol{\Psi}$ in (2.5.7) and thereby the update of $\boldsymbol{\mathcal{H}}^{(u)}$ is similar to that for $\widetilde{\mathbf{H}}^{(u)}$. Note that besides transforming back to the form of (2.5.7), the sparse coding algorithms, e.g., ADMM, can be re-derived for this SeDL problem, which leads to lower computational complexity for the smaller dimensions of $\overline{\mathbf{D}}_{\text{T}}$ and $\mathbf{D}_{\text{R}}$ compared to the Kronecker product $\left(\overline{\mathbf{D}}_{\text{T}} \otimes \mathbf{D}_{\text{R}}\right)$.

**Dictionary update stage** Following the sparse coding stage, the two dictionaries $\mathbf{D}_{\text{T}}$ and $\mathbf{D}_{\text{R}}$ are then updated column by column with fixed $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{X}}$. Note that rather than considering a combined dictionary $\boldsymbol{\Psi}$ as did in Section 2.5, the update of $\mathbf{D}_{\text{T}}$ and $\mathbf{D}_{\text{R}}$ here are separated. If MOD is considered, we first fix $\mathbf{D}_{\text{T}}$ and then the optimization of (2.6.1) w.r.t. $\mathbf{D}_{\text{R}}$ is equivalent to

$$\min_{\mathbf{D}_{\text{R}}} \quad \left\|[\boldsymbol{\mathcal{X}}]_{(1)} - \mathbf{D}_{\text{R}}[\boldsymbol{\mathcal{H}}]_{(1)}(\mathbf{I}_{N_{\text{sa}}} \otimes \mathbf{I}_{N_{\text{c}}} \otimes \overline{\mathbf{D}}_{\text{T}})^T\right\|_F^2, \tag{2.6.4}$$

where $[\mathcal{X}]_{(1)} \triangleq \left[[\mathcal{X}^{(0)}]_{(1)}, \cdots, [\mathcal{X}^{(N_{\rm sa}-1)}]_{(1)}\right] \in \mathbb{C}^{N_{\rm r} \times N_{\rm t} N_{\rm c} N_{\rm sa}}$ and its columns are the mode-1 fibers of $\mathcal{X}$. Similarly, $[\mathcal{H}]_{(1)} \triangleq \left[[\mathcal{H}^{(0)}]_{[(1)}, \cdots, [\mathcal{H}^{(N_{\rm sa}-1)}]_{(1)}\right] \in \mathbb{C}^{K_{\rm r} \times K_{\rm t} N_{\rm c} N_{\rm sa}}$ collects the mode-1 fibers of $\mathcal{H}$. Therefore, the updates for $\mathbf{D}_{\rm R}$ is readily available by LS as $\left[\mathbf{D}_{\rm R}\right]^{\rm new} = [\mathcal{X}]_{(1)} \Big( [\mathcal{H}]_{(1)} \big( \mathbf{I}_{N_{\rm sa}} \otimes \mathbf{I}_{N_{\rm c}} \otimes \left[\overline{\mathbf{D}}_{\rm T}\right]^{\rm old} \big)^T \Big)^{\dagger}$. Similarly, the update for $\overline{\mathbf{D}}_{\rm T}$ is given as $\left[\overline{\mathbf{D}}_{\rm T}\right]^{\rm new} = [\mathcal{X}]_{(2)} \Big( [\mathcal{H}]_{(2)} \big( \mathbf{I}_{N_{\rm sa}} \otimes \mathbf{I}_{N_{\rm c}} \otimes \left[\mathbf{D}_{\rm R}\right]^{\rm new} \big)^T \Big)^{\dagger}$. Besides, the basic idea of K-SVD for dictionary update could be extended accordingly in the tensor case by using the high-order SVD (HOSVD) [71].

**Denoising stage** Following the sparse coding and dictionary update stages, to update $\mathcal{X}^{(u)}, \forall u \in \mathcal{I}(N_{\rm sa})$, we can still transform it back to the denoising stage of CoDL, i.e.,

$$\min_{\mathcal{X}^{(u)}} \quad \left\| [\mathcal{X}^{(u)}]_{(3)}^T - \left(\overline{\mathbf{D}}_{\rm T} \otimes \mathbf{D}_{\rm R}\right) [\mathcal{H}^{(u)}]_{(3)}^T \right\|_F^2 + w_2 \left\| [\mathcal{Y}^{(u)}]_{(3)}^T - [\mathcal{X}^{(u)}]_{(3)}^T \right\|_F^2, \tag{2.6.5}$$

whose solution is also readily available by LS.

Based on these sparse coding, dictionary update and denoising steps, the whole procedure for SeDL can be summarized similar to Algortihm 1, which we omit for space limitation.

### 2.6.3 Complexity analysis

The convergence analysis and stopping rules of the SeDL problem are similar to those of CoDL in Section 2.5. The analysis of computational complexity in terms of complex multiplication operations per iteration for CoDL

Table 2.1: Complexity analysis for different sparse coding and dictionary update algorithms

<table>
<tr><td rowspan="2"></td><td colspan="2" rowspan="2">Sparse coding (online computation)</td><td colspan="2">Dictionary update (offline computation)</td></tr>
<tr><td>MOD</td><td>K-SVD/K-HOSVD</td></tr>
<tr><td>SW-OMP [5]</td><td colspan="2">$\mathcal{O}(N_{\rm sa}N_{\rm c}NK)$</td><td colspan="2">—</td></tr>
<tr><td>CoDL</td><td colspan="2">$\mathcal{O}\big(N_{\rm sa}(NK^2+K^3)+N_{\rm sa}N_{\rm c}(NK+K^2)\big)$</td><td>$\mathcal{O}\big(N_{\rm sa}N_{\rm c}NK+NK^2+N_{\rm sa}N_{\rm c}K^2+K^3\big)$</td><td>$\mathcal{O}\big(N^2(S_0N_{\rm sa}N_{\rm c}+NK)\big)$</td></tr>
<tr><td rowspan="2">SeDL</td><td>Reducing to CoDL</td><td>The same as CoDL</td><td rowspan="2">$\mathcal{O}\big(2N_{\rm sa}N_{\rm c}(N\sqrt{K}+K\sqrt{N})+6N_{\rm sa}N_{\rm c}K\sqrt{K}+2K\sqrt{K}\big)$</td><td rowspan="2">$\mathcal{O}\big(N^2(S_0N_{\rm sa}N_{\rm c}+NK)+2N_{\rm sa}N_{\rm c}N^3S_0\big)$</td></tr>
<tr><td>ADMM for SeDL</td><td>$\mathcal{O}\big(4N_{\rm sa}N_{\rm c}K\sqrt{K}\big)$</td></tr>
</table>

and SeDL is provided in Table 2.1, where we assume $N_{\rm r}=N_{\rm t}=\sqrt{N}$ and $K_{\rm r}=K_{\rm t}=\sqrt{K}$ for ease of comparison, and $S_0$ denotes the sparsity level of coefficient matrices. Note that the approximate K-SVD [64] can be used to alleviate the computational burden of K-SVD based algorithms. Moreover, the involved computations for dictionary update are implemented offline so that they will not affect the complexity of sparse coding once the dictionary is learned.

### 2.6.4 Channel estimation with learned dictionary

Once the optimal combined dictionary $\mathbf{\Psi}$ or separable dictionaries $\mathbf{D}_{\rm R}$ and $\mathbf{D}_{\rm T}$ are learned via CoDL and SeDL algorithms, they will be used for channel estimation in the following transmissions. Then, the problem of channel estimation at UE exactly boils down to a compressive sensing (sparse coding) problem, as formulated in (2.5.7) and (2.6.2) with fixed $\mathbf{\Psi}$ or $\mathbf{D}_{\rm R}$ and $\mathbf{D}_{\rm T}$, so that the various sparse coding algorithms for CoDL and SeDL can be directly applied, such as OMP, SW-OMP, ADMM, to name a few.

## 2.7 Numerical results

In this section, we provide numerical simulations to corroborate the effectiveness of the proposed dictionary learning and channel estimation algorithms for hybrid wideband mmWave MIMO systems. For comparison, compressive channel estimation (sparse coding) by SW-OMP [5] and the canonical OMP methods [60], using the overcomplete ideal array response matrix (IARM) uniformly sampled in the physical angle domain as the sparsifying dictionaries are also evaluated.

### 2.7.1 Simulation parameters

Unless otherwise specified, the default parameters in the simulations are summarized as follows. Both the AP and the UE deploy a ULA with (presumed) half-wavelength antenna spacing, and $N_{\mathrm{t}} = 32$, $N_{\mathrm{r}} = 8$, $N_{\mathrm{s}} = L_{\mathrm{t}} = L_{\mathrm{r}} = 2$ and $K_{\mathrm{t}} = 64$, $K_{\mathrm{r}} = 16$. The phase-shifters used at the AP and the UE are assumed to have $N_{\mathrm{Q}} = 2$ quantization bits, so that the phases of the entries of precoders $\mathbf{F}_m$ and combiners $\mathbf{W}_m$ are randomly chosen from $\left\{0, \frac{2\pi}{2^{N_{\mathrm{Q}}}}, \ldots, \frac{2\pi(2^{N_{\mathrm{Q}}}-1)}{2^{N_{\mathrm{Q}}}}\right\}$. The number of OFDM subcarriers is set as $N_{\mathrm{c}} = 128$ and $T_{\mathrm{s}} = \frac{1}{1760}\ \mu s$, as specified in the IEEE 802.11ad wireless standard. The channels are generated according to (2.3.1) with $N_{\mathrm{tap}} = 16$ delay taps and $N_{\mathrm{p}} = 6$ multipath components each of $N_{\mathrm{ray}} = 1$ ray (which is typical in indoor scenarios), whose delays are chosen uniformly from $[0, (N_{\mathrm{tap}}-1)T_{\mathrm{s}}]$. The band-limited filter $p_{\mathrm{rc}}(t)$ is assumed to be a raised-cosine filter with roll-off factor of 0.8. Moreover, all the AoAs and AoDs are supposed to be constrained

in a sector range of 120°, because we expect to have multiple antenna array panels covering different sectors. The number of measurement samples $N_{\text{sa}}$ is set as 100. The SNR is set to 0 dB for dictionary learning and channel estimation. During the dictionary learning phase, $M = 500$ and the spreading factor $N_{\text{rep}} = 10$ is used to increase the effective SNR in 10 dBs.

To evaluate the effects of array uncertainties on dictionary learning performances, we adopt the models and parameters in [51] to characterize the antenna gain and phase errors as well as their mutual coupling. Specifically, the gains and phase shifts in $\mathbf{\Gamma}_{\text{R}}$ and $\mathbf{\Gamma}_{\text{T}}$ are modeled as $g_i = 1.0 + 0.05 \cdot \sigma_g$ and $\nu_i = \frac{20^\circ \pi}{180^\circ} \cdot \sigma_\nu$, with $\sigma_g \sim \mathcal{N}(0, 1)$. This indicates that the gain and phase error variances for each antenna element are 5% and $20^\circ\pi/180^\circ$. The mutual coupling coefficients among antennas are in the range between 0.01 and 0.4 as assumed in [51]. Moreover, to characterize the manufacture error and evaluate irregular array geometries, we introduce the antenna spacing perturbation as in [72], and assume that the antenna spacing is not $0.5\lambda$, but is uniformly distributed between $0.4\lambda$ and $0.6\lambda$, with $\lambda$ the carrier wavelength.

For the optimization of CoDL and SeDL, unless otherwise specified, the ADMM-based algorithms are used for sparse coding, while approximate K-SVD based algorithms [64] are used for dictionary update. The aforementioned DIA algorithm is used for dictionary initialization. The regularization parameters $w_1, w_2$ can be tuned and their default values are set as $w_1 = 0.1, w_2 = 0.001$.

### 2.7.2 Performances of dictionary learning and channel estimation with learned dictionaries

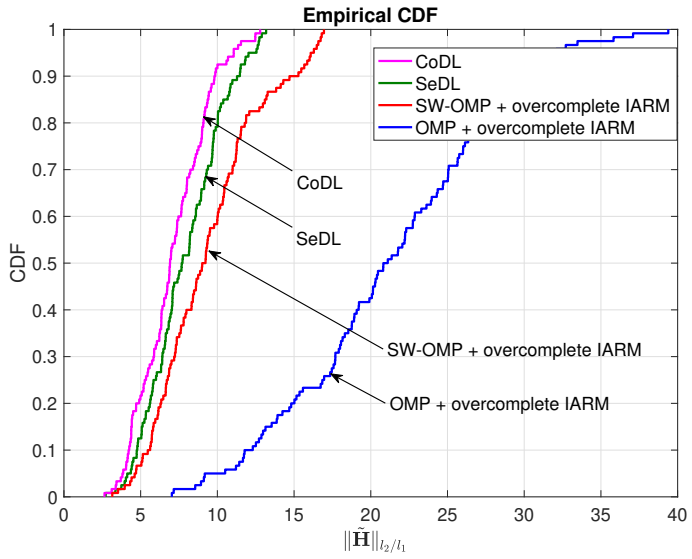



Figure 2.2: The cumulative density function of $\|\widetilde{\mathbf{H}}\|_{l_2/l_1}$ on different sparsifying dictionaries.

In Fig. 2.2, we compare the cumulative density functions (CDF) of $l_2/l_1$ norm of the estimated channel matrices, i.e., $\|\widetilde{\mathbf{H}}\|_{l_2/l_1}$, using different sparsifying dictionaries, including learned dictionaries from CoDL and SeDL, as well as the overcomplete IARM dictionaries. As explained before, since the $l_2/l_1$ norm of a matrix is a convex surrogate of its row sparsity, CDFs of $\|\widetilde{\mathbf{H}}\|_{l_2/l_1}$ on distinct dictionaries are then able to illustrate the capability of the corresponding dictionary for channel sparsity enhancement. It is clear from Fig. 2.2, the values of $\|\widetilde{\mathbf{H}}\|_{l_2/l_1}$ corresponding to CoDL and SeDL are more likely smaller than those of overcomplete IARM dictionary, which means that the learned dictionaries by CoDL and SeDL can enhance the channel common

sparsity compared to the overcomplete IARM dictionary. The performance gap between CoDL and SeDL lies in the fact that the unit-norm constraint on the combined dictionary in (2.5.6) is generally less strict than that on the separable transmit and receive dictionaries in (2.6.1). In other words, the predetermined structure constraints on separable dictionaries limit their feasible ranges.

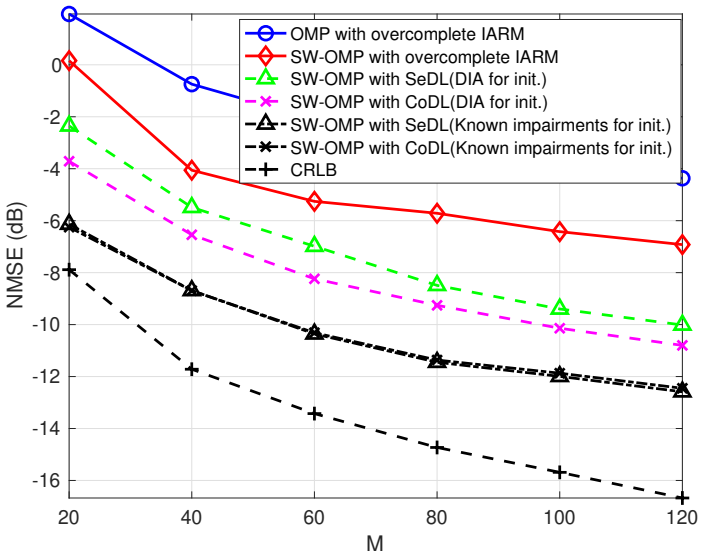
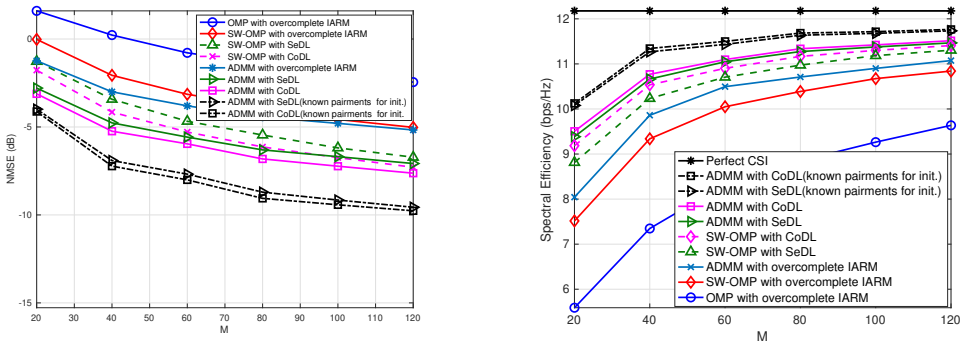



Figure 2.3: Example of NMSE versus $M$ performance comparison between the CRLB benchmark and various combinations of compressive channel estimation algorithms and dictionary learning methods with fixed channel and impairment parameters.

The performance metric for channel estimation is the normalized mean squared error (NMSE) in the estimation of $\hat{\mathbf{H}}[c]$. To understand the effectiveness of our approaches to learn the dictionaries and estimate the channel, we show in Fig. 2.3 the NMSE obtained by various combinations of SW-OMP and

different dictionary learning methods and initializations versus the number of training symbols when SNR= 0dB, the channel parameters and the impairments are fixed and set according to the description in 2.7.1. The CRLB for the joint estimation of all the parameters in the exact channel representation in (2.3.3) is also shown as benchmark. The NMSE obtained by SW-OMP when the hardware impairments are neglected is also shown for comparison. The gap between the performance of our approaches and the CRLB is due to three types of error: 1) approximation error in the channel model in (5); 2) error in the dictionary learning stage; and 3) error in the channel estimate based on the obtained dictionaries. Even if the performance gap could be further reduced, our approaches clearly outperform prior work that neglects hardware impairments [5]: around 4 dB reduction in NMSE for SW-OMP+CoDL, and 3dB for SW-OMP+SeDL with respect to SW-OMP when M=120.




(a) Comparison of NMSE performances for ULA.

(b) Comparison of SE performances for ULA.

Figure 2.4: Comparisons of NMSE and SE performances for ULA versus number of training OFDM symbols using various sparse coding algorithms and sparsifying dictionaries.

We now compare the channel estimation performances of various sparse coding algorithms and sparsifying dictionaries in Fig. 2.4(a) as a function of the number of training OFDM symbols averaging among 20 different channels and noise realizations. Since we have modeled the channel with a set of deterministic parameters, and the statistical CRLB is unknown, we include as benchmark for the averaged performance, the NMSE provided by our strategies when the impairments are assumed to be known and the dictionaries are initialized to $\mathbf{D}_{\mathrm{T}} = \mathbf{C}_{\mathrm{T}}\mathbf{\Gamma}_{\mathrm{T}}\mathbf{A}_{\mathrm{T}}^{\mathrm{v}}$, and $\mathbf{D}_{\mathrm{R}} = \mathbf{C}_{\mathrm{R}}\mathbf{\Gamma}_{\mathrm{R}}\mathbf{A}_{\mathrm{R}}^{\mathrm{v}}$. It can be seen that when the learned dictionaries from CoDL and SeDL are used, the NMSE performances of SW-OMP outperform all the cases based on an IARM dictionary. Furthermore, the ADMM-based sparse coding together with the optimized dictionaries is able to further reduce the training overhead, compared to all other sparse coding algorithms and dictionaries. The average reduction in NMSE when considering the impairments and learning the dictionaries is about 2.5 dB for ADMM+CoDL. Last but not least, though there is a performance gap between SeDL and CoDL, SeDL has lower computational complexity. In other words, SeDL provides a better trade-off between complexity and channel estimation performance.

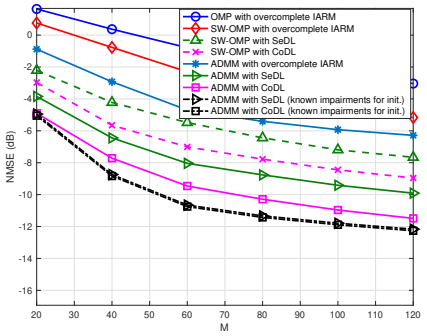
The comparison of spectral efficiency (SE) performances is then given in Fig. 2.4(b) as a function of the number of training OFDM symbols. The SE is computed by assuming fully-digital precoding and combining using estimates for the $N_{\mathrm{s}}$ dominant left and right singular vectors of the channel estimates [5]. Specifically, the effective channels are defined as $\mathbf{H}_{\mathrm{eff}}[c] \triangleq$

$\left[\hat{\mathbf{U}}_1[c]\right]^*_{:,1:N_\mathrm{s}}\mathbf{H}[c]\left[\hat{\mathbf{U}}_2[c]\right]_{:,1:N_\mathrm{s}}, \forall c \in \mathcal{I}(N_\mathrm{c})$, where $\hat{\mathbf{U}}_1[c]$ and $\hat{\mathbf{U}}_2[c]$ are the left and right singular vectors of the channel estimates $\hat{\mathbf{H}}[c]$. Then SE is defined as

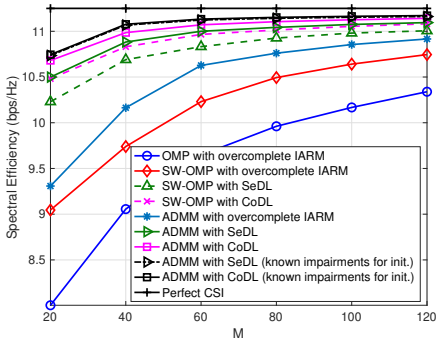
$$\mathrm{SE} = \frac{1}{N_\mathrm{c}} \sum_{c=0}^{N_\mathrm{c}-1} \sum_{n=1}^{N_\mathrm{s}} \log\left(1 + \frac{\mathrm{SNR}}{N_\mathrm{s}} \lambda_n(\mathbf{H}_{\mathrm{eff}}[c])^2\right), \tag{2.7.1}$$

where $\lambda_n(\mathbf{H}_{\mathrm{eff}}[c]), n = 1, \ldots, N_\mathrm{s}$ are the singular values of each effective channel $\mathbf{H}_{\mathrm{eff}}[c]$. For comparison, SE with perfect CSI is also provided as an upper bound. As can be seen from Fig. 2.4(b), SE is significantly increased by using SW-OMP with the learned dictionaries compared to those of SW-OMP with overcomplete IARM dictionary. Similar results can be found for ADMM-based sparse coding algorithm. It can be seen that a loss of only 0.25 bps appears when using ADMM+CoDL instead of ADMM+CoDL and impairment knowledge. These results illustrate the effectiveness of the proposed dictionary learning and channel estimation algorithms for the goal of designing the precoders and combiners that maximize the spectral efficiency.

To evaluate the effectiveness of the proposed dictionary learning algorithms on different array geometries, the compressive channel estimation algorithms based on the learned dictionaries have been evaluated for uniform circular arrays (UCAs), where the presumed spacing distances between adjacent antennas are half-wavelength (e.g., see [73]). From Fig. 2.5(a) and 2.5(b), it can be seen that the learned dictionaries from our proposed dictionary learning algorithms can result in significant performance gains for UCAs compared to the overcomplete IARM dictionary, both for SW-OMP or ADMM-based



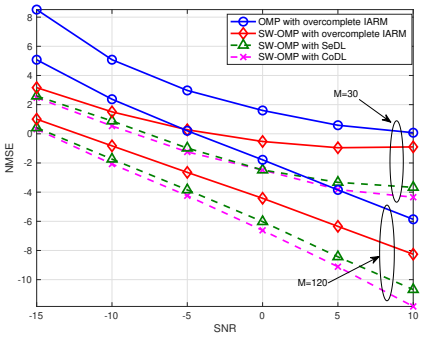

(a) Comparison of NMSE performances for UCA.



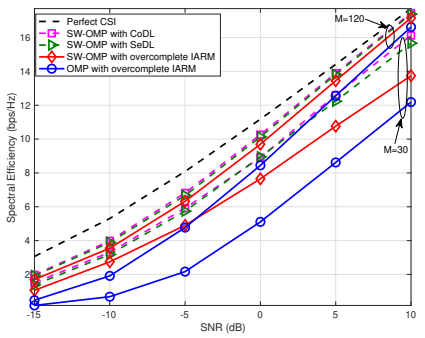

(b) Comparison of SE performances for UCA.

Figure 2.5: Comparisons of NMSE and SE performances for UCA versus number of training OFDM symbols using various sparse coding algorithms and sparsifying dictionaries.

sparse coding algorithms. This is consistent with the case of ULAs and thus corroborates the applicability of the proposed dictionary learning algorithms for various (irregular) array geometries. In Fig. 2.6, we compare the NMSE and




(a) Comparison of NMSE performances.




(b) Comparison of SE performances.

Figure 2.6: Comparisons of NMSE and SE performances of OMP and SW-OMP as a function of SNR using various sparsifying dictionaries.

SE performances of OMP and SW-OMP as a function of SNR, using overcomplete IARM dictionary or the learned dictionaries, and also include the impact of the number of training OFDM symbols. The parameters involved in this simulation are the same as Fig. 2.4(a). As can be seen, both the NMSE and SE performances of SW-OMP with learned dictionaries by CoDL and SeDL are superior to those using overcomplete IARM dictionary, even at a relatively low SNR. Of course, the performance gains of learned dictionaries at low SNR is relatively smaller, as the high noise level may mask the structural information of the channel so that the accuracy of learned dictionaries is deteriorated. Furthermore, the performance gains of the learned dictionaries are obvious when the number of training OFDM symbols is small, which means the dictionary learning methods can help to reduce the training overhead greatly.

## 2.8 Conclusions

In this chapter, we proposed two dictionary learning algorithms, i.e., CoDL and SeDL, to find the best sparsifying dictionaries for channel sparse representation in hybrid wideband mmWave MIMO systems with hardware impairments. The CoDL and SeDL problems were formulated by exploiting the common sparsity properties within the large bandwidth. The CoDL focused on the best combined dictionary, while SeDL incorporated the detailed constraints on separable transmit and receive dictionaries, and thus SeDL can achieve a better trade-off between performance gains and computational complexity. It has been shown that the learned dictionaries from both CoDL and

SeDL can result in a sparser channel representation, compared to commonly adopted overcomplete Fourier dictionary, especially when there exist irregular array geometries and array uncertainties. With the learned dictionaries, various compressive channel estimation techniques can be applied for CSI acquisition at low SNR with much reduced training overhead, which has been corroborated via numerical simulations with different system configurations, array geometries and channel environments.

# Chapter 3

# Dictionary Learning and Channel Estimation For Hybrid mmWave MIMO Systems Under Hardware Impairments and Beam Squint

In this chapter[1], we propose dictionary learning aided compressive channel estimation solutions for hybrid wideband mmWave MIMO systems under both the impact of hardware impairments and beam squint effect. Low overhead channel estimation based on CS has been widely investigated for hybrid wideband mmWave MIMO systems. The channel sparsifying dictionaries used in previous works typically are the (overcomplete) ideal array response vectors evaluated on discrete angles of arrivals/departures, and assumed to be the same for all subcarriers, without considering the impacts of either hardware impairments or beam squint effect. In this chapter, we propose a dictionary learning aided channel estimation strategy for hybrid wideband mmWave MIMO systems under both hardware impairments and beam squint. We first derive a general channel model by explicitly incorporating the impacts of hard-

[1]This chapter was based on our work [74]: H. Xie, J. Palacios, and Nuria González-Prelcic, "Dictionary learning and channel estimation for hybrid mmWave MIMO systems under hardware impairments and beam squint," In preparation for submission to IEEE Trans. Wireless Commun. This work was supervised by Prof. González-Prelcic. My contributions lie in designing and performing the research, doing simulations, analyzing data and writing the paper, etc.

ware impairments and practical pulse shaping functions, which will illustrate and compensate the limitations of existing MIMO channel modeling without considering beam squint. Then we propose a dictionary learning algorithm for learning hardware impairments related dictionaries, which exploits beam squint impacts but avoids learning beam squint itself. A novel CS channel estimation algorithm under beam squint is also designed, where the channel structures at different subcarriers are exploited to handle beam squint impacts and facilitate channel parameter estimation with lower complexity and higher accuracy. Numerical results have demonstrated the effectiveness and superiority of the proposed dictionary learning and channel estimation strategy.

## 3.1 Introduction

The acquisition of CSI is vital for link configuration in the mmWave massive MIMO systems with hybrid beamforming architectures. To reduce the training overhead associated with CSI acquisition, previous works have made full use of the sparsity properties of mmWave channels [2, 3] either in term of narrowband/wideband systems [4, 5] or from the perspective of angular/delay domains [6], which promote various CS based channel estimation strategies. These prior works have demonstrated the feasibility and effectiveness of compressive channel estimation for hybrid mmWave MIMO systems. Nevertheless, two significant practical factors have not been fully investigated in prior works, i.e., the beam squint effect and the hardware impairments. Specifically, the channel sparsifying dictionaries used in previous compressive

channel estimation works are typically assumed to be (overcomplete) DFT matrices or constructed from the ideal array response matrices (IARM) evaluated on discrete grids of quantized angles of AoAs/AoDs [4–6]. These assumptions of ideal antenna arrays are valid only when the beam squint effect is negligible and no hardware impairments or calibration errors exist on the antenna arrays. However, we will show in this chapter that the array response vectors will no long be the Vandermonde vectors under hardware impairments, and different array response vectors should be considered at every frequency for channel modeling under beam squint. In other words, the assumptions and modeling of wideband mmWave MIMO channels in most of prior works are not valid and thus their proposed CSI acquisition algorithms are not effective either under the circumstances of beam squint and hardware impairment impacts.

Many works have illustrated the beam squint effect, namely, beams changing direction with frequency. As shown in [26–28], the time delay of the same data symbol across the antenna array aperture is non-negligible in the large-scale MIMO configurations and/or the wideband systems. Due to the spatial delay difference of each data symbol at different antennas, the array steering vectors will have dissimilar responses at each frequency and then gives rise to the beam squint effect. Therefore, the adopted MIMO channel models in the existing literature, e.g., [4, 5, 29], that use the array response vectors evaluated at the carrier frequency for channel modeling at all subcarriers are no longer valid. The work of [26] derived a channel model for the large-scale MIMO system under beam squint and showed that the array re-

sponse vectors for channel modeling have to be frequency-dependent. Aware of this beam squint impact, [75] and [76] also considered the frequency domain channel models by using explicit frequency-dependent array response vectors for different subcarriers. Meanwhile, to avoid the impacts of beam squint on channel estimation at different subcarriers, [77–79] proposed to estimate the channel parameters from the perspective of angle and delay domains, together with user scheduling to alleviate inter-user interference. Generally, these prior works on channel estimation under beam squint are feasible but they failed to consider the practical pulse shaping and filtering effects in the channel modeling, which is a non-trivial factor in practical wireless systems. We will show in this chapter that the beam squint will not only induce the frequency-dependence on array steering vectors but also yield additional distortions at different antennas across all subcarriers, especially on side subcarriers. This has not been considered yet in the previous literature. Furthermore, these prior works either assumed fully-digital architectures for large-scale antenna systems at mmWave band, or neglected the impacts of hardware impairments.

The issue of hardware impairments on MIMO channels has been investigated in [36, 44, 51, 80]. There are many different hardware impairments in the practical RF chains and three main types of hardware impairments are typically considered in the channel modeling. Specifically, due to the manufacture and calibration errors, the antenna array will generate unexpected radiation patterns, including both gain and phase errors on each antenna element. Moreover, any perturbation on antenna locations or inter-element spac-

ing between antenna elements will result in irregular linear arrays rather than perfect half-wavelength ULAs. The antenna spacing disturbance also creates the mutual coupling effect between antenna elements. Taking into account all these hardware errors, it is apparent that the aforementioned sparsifying dictionaries constructed from IARM evaluated at quantized angles are no longer the best choice for exploiting channel sparsity. Previous works [36, 44, 45, 80] have shown that dictionary learning is an effective technique to capture the underlying practical structure of mmWave MIMO channels with various hardware impairments. Specifically, [44] proposed a joint uplink/downlink sparsifying Dictionary learning algorithm for narrow band massive MIMO systems operating at lower frequencies. Our previous work [36] further investigated the dictionary learning and channel estimation strategy for hybrid wideband mmWave MIMO systems under low SNR conditions. Nevertheless, the beam squint effect has not been incorporated in previous work, which is a significant phenomenon in wideband system and thus cannot be ignored.

## 3.2 Contributions

Motivated by the limitations discussed above in the literature, we in this chapter propose a dictionary learning based channel estimation strategy for hybrid mmWave MIMO systems under the impacts of both hardware impairments and beam squint. The main contributions of this chapter are summarized as follows:

- We derive a general model for wideband mmWave MIMO channels under

both hardware impairments and beam squint impacts. The new general model not only incorporates the hardware impairments of antenna spacing disturbances, gain/phase errors, and array mutual coupling, but also explicitly considers the impacts of combined pulse shaping and filtering on channels, showing the limitations of existing beam squint MIMO channel models and the associated channel estimation schemes. This general model has not been considered yet in the literature and will justify the motivation of our proposed dictionary learning strategies.

- We propose a dictionary learning algorithm for learning hardware impairments related dictionaries, which fully exploits the channel properties at different subcarriers under beam squint but does not learn the impacts of beam squint effect on the dictionaries. Comparing to existing dictionary learning strategies which has to learn a general dictionary for both hardware impairment and beam squint, i.e., a compromised dictionary, the newly proposed dictionary learning scheme enables better adaptation to the impacts of hardware impairments and then facilitates to address the beam squint impact at different subcarriers.

- We design a novel OMP based algorithm for CS channel estimation under beam squint impacts, which exploits the simple structures of the beam squint channel models at central subcarriers to obtain initial parameter estimates with low complexity, and then compensates the additional distortions at side subcarriers induced by beam squint, such that the

measurements at all subcarriers can be used to achieve higher parameter estimation accuracy with lower complexity.

- We evaluate the proposed dictionary learning and channel estimation algorithms via numerical simulations. Results show that the training overhead of channel estimation with learned dictionaries can be significantly reduced compared to traditional dictionaries without considering hardware impairment or beam squint. This corroborates the effectiveness of the proposed dictionary learning algorithms for hybrid wideband mmWave MIMO systems under both hardware impairments and beam squint effects.

## 3.3 System, channel and signal models

### 3.3.1 System model

As shown in Fig. 3.1, we in this chapter consider a hybrid mmWave MIMO system between a transmitter (TX) of $N_{\mathrm{t}}$ antennas and $L_{\mathrm{t}}$ RF chains, and a receiver (RX) of $N_{\mathrm{r}}$ antennas and $L_{\mathrm{r}}$ RF chains. The channel between TX and RX is assumed to be frequency-selective fading and the pulse shaping based OFDM modulation with $K$ subcarriers is considered to simultaneously transmit $N_{\mathrm{s}}$ ($\leq \min(L_{\mathrm{t}}, L_{\mathrm{r}})$) data streams. For ease of expression, the system sampling period is denoted by $T_{\mathrm{s}}$. If the pulse shaping function has a roll-off factor $\beta$, the overall system bandwidth will be $(1+\beta)/T_{\mathrm{s}}$. Moreover, the center carrier frequency and wavelength are denoted by $f_{\mathrm{c}}$ and $\lambda_{\mathrm{c}} = c/f_{\mathrm{c}}$ (with $c$ the speed of light). We will use the index $k$ to denote the frequency domain

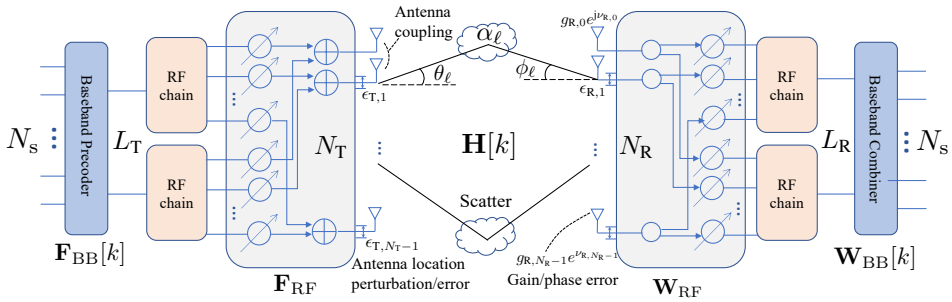



Figure 3.1: System model and architecture of a hybrid mmWave MIMO system with hardware impairments on antennas.

subcarriers and $f_k$ as the corresponding subcarrier frequency value. Therefore, we have [81]

$$f_k = f_c + \Delta f_k = f_c - \frac{1}{2T_\mathrm{s}} + \frac{k}{KT_\mathrm{s}}, \quad \text{for } k = 0, \ldots, K-1. \tag{3.3.1}$$

Moreover, we consider a fully connected phase shifting network for the hybrid architectures [4] as shown in Fig. 3.1.

### 3.3.2 Channel and signal models

In this section, we derive the new general time domain signal model and frequency domain channel model for the hybrid mmWave MIMO systems under the impact of both hardware impairments and beam squint. It will be seen that the assumed channel models for beam squint circumstances in previous works [26, 75–79] are not complete or valid in some cases. Let us start with the antenna array architectures. It is common in the literature to assume that all the antennas are arranged in the form of ULAs at both TX and RX, whose expected ideal antenna spacings are denoted by $d_\mathrm{T}$ and $d_\mathrm{R}$,

respectively. In this work, we also assume the linear architecture for TX and RX arrays but take into consideration some practical imperfections. As shown in [36, 44, 51, 82, 83] and illustrated in Fig 3.1, various hardware impairments on antennas exist in practical implementations, such as the gain/phase errors on each antenna element, the perturbation errors at each antenna element location, as well as the coupling effect between antenna elements, to name a few. For better exposition, let us first define some hardware impairments related variables. Specifically, we denote $\mathbf{C}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{r}}}$ as the symmetric mutual coupling matrix for the RX antenna array, representing the unwanted interchange of energy between elements in the arrays, and denote $\boldsymbol{\gamma}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}} \times 1}$ with $[\boldsymbol{\gamma}_{\mathrm{R}}]_{n_{\mathrm{R}}} = g_{\mathrm{R},n_{\mathrm{R}}} e^{\mathrm{j}\nu_{\mathrm{R},n_{\mathrm{R}}}}$ as the antenna gain and phase errors, in which $g_{\mathrm{R},n_{\mathrm{R}}}$ and $\nu_{\mathrm{R},n_{\mathrm{R}}}$ are respectively the receive gain error normalized to a reference amplitude and the additional receive phase error for the $n_{\mathrm{R}}$-th antenna element. Moreover, let $\boldsymbol{\epsilon}_{\mathrm{R}} \in \mathbb{R}^{N_{\mathrm{r}} \times 1}$ with $[\boldsymbol{\epsilon}_{\mathrm{R}}]_{n_{\mathrm{R}}} = \epsilon_{\mathrm{R},n_{\mathrm{R}}}$ be the vector of antenna location errors at all RX antenna elements, where the location error of the first antenna element is normalized to $\epsilon_{\mathrm{R},0} = 0$. Similar hardware impairment variables for TX antenna arrays are defined accordingly. Based on these definitions and notations, we next derive the new channel model including the impacts of both hardware impairments and beam squint effect.

We consider a measurement model in which the TX transmits an $L_{\mathrm{t}} \times 1$ signal vector $\boldsymbol{s}(t)$ at time instant $t$, where the $l_{\mathrm{T}}$-th ($l_{\mathrm{T}} = 1, \ldots, L_{\mathrm{t}}$) element

of $\boldsymbol{s}(t)$ is defined as

$$s_{l_{\mathrm{T}}}(t) \triangleq [\boldsymbol{s}(t)]_{l_{\mathrm{T}}} = \sum_{i=-\infty}^{+\infty} [\boldsymbol{S}]_{l_{\mathrm{T}},i}\delta(t - iT_{\mathrm{s}}), \tag{3.3.2}$$

where $\boldsymbol{S}$ collects all the time domain signal symbols to be transmitted. Moreover, $\boldsymbol{s}(t)$ satisfies $\mathbb{E}\{\boldsymbol{s}(t)\boldsymbol{s}(t)^*\} = \frac{P_{\mathrm{T}}}{L_{\mathrm{t}}}\mathbf{I}_{L_{\mathrm{t}}}$, with $P_{\mathrm{T}}$ the transmit power constraint. Using an analog precoder $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{t}} \times L_{\mathrm{t}}}$, the transmitted signal at each antenna element without considering any filtering effect can be computed as

$$f_{n_{\mathrm{T}}}^{(\mathrm{T})}(t) = \sum_{l_{\mathrm{T}}=1}^{L_{\mathrm{t}}} [\mathbf{F}]_{n_{\mathrm{T}},l_{\mathrm{T}}}[\boldsymbol{s}(t)]_{l_{\mathrm{T}}}. \tag{3.3.3}$$

Due to hardware imperfections, the signal at each antenna $n_{\mathrm{T}}$ receives a phase and amplitude perturbation $[\boldsymbol{\gamma}_{\mathrm{T}}]^*_{n_{\mathrm{T}}}$ and then leaks into the signal at each antenna $n'_{\mathrm{T}}$ through mutual coupling by a factor $[\mathbf{C}_{\mathrm{T}}]^*_{n_{\mathrm{T}},n'_{\mathrm{T}}}$, leading to the expression

$$f_{n'_{\mathrm{T}}}^{(\mathrm{T}')}(t) = \sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{t}}} [\mathbf{C}_{\mathrm{T}}]^*_{n_{\mathrm{T}},n'_{\mathrm{T}}}[\boldsymbol{\gamma}_{\mathrm{T}}]^*_{n_{\mathrm{T}}} f_{n_{\mathrm{T}}}^{(\mathrm{T})}(t) = \sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{t}}} [(\mathrm{diag}(\boldsymbol{\gamma}_{\mathrm{T}})\mathbf{C}_{\mathrm{T}})^*]_{n'_{\mathrm{T}},n_{\mathrm{T}}} f_{n_{\mathrm{T}}}^{(\mathrm{T})}(t). \tag{3.3.4}$$

Then the signal goes through the geometric channel and a series of filters, which is modeled as a convolution with a combined pulse shaping filter $p(t)$ with a time delay, due to the channel characteristics and antenna array geometry. An additive noise function $z_{n_{\mathrm{R}}}(t)$ is also added and thus the received signal at the $n_{\mathrm{R}}$-th receive antenna is expressed as

$$f_{n_{\mathrm{R}}}^{(\mathrm{R})}(t) = \sum_{\ell=1}^{L} \sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{t}}} \alpha_\ell \cdot (p * f_{n_{\mathrm{T}}}^{(\mathrm{T}')})(t - \tau_{\ell,n_{\mathrm{R}},n_{\mathrm{T}}}) + z_{n_{\mathrm{R}}}(t), \tag{3.3.5}$$

where $*$ denotes convolution operation, $L$ is the number of distinguishable paths of the mmWave channel, $\alpha_\ell \in \mathbb{C}$ is the path gain, and $\tau_{\ell,n_{\mathrm{R}},n_{\mathrm{T}}}$ is the delay of $\ell$-th path between the $n_{\mathrm{R}}$-th receive antenna and the $n_{\mathrm{T}}$-th transmit antenna. Similarly, owing to hardware imperfections, the signal at each antenna $n_{\mathrm{R}}$ leaks into the signal at each antenna $n'_{\mathrm{T}}$ through mutual coupling by a factor $[\mathbf{C}_{\mathrm{R}}]_{n'_{\mathrm{R}},n_{\mathrm{R}}}$ and then receives a gain/phase weight $[\boldsymbol{\gamma}_{\mathrm{R}}]_{n'_{\mathrm{R}}}$ as follows

$$f^{(\mathrm{R}')}_{n'_{\mathrm{R}}}(t) = \sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{r}}} [\boldsymbol{\gamma}_{\mathrm{R}}]_{n'_{\mathrm{T}}} [\mathbf{C}_{\mathrm{R}}]_{n'_{\mathrm{R}},n_{\mathrm{R}}} f^{(\mathrm{R})}_{n_{\mathrm{R}}}(t) = \sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{r}}} [\mathrm{diag}(\boldsymbol{\gamma}_{\mathrm{R}})\mathbf{C}_{\mathrm{R}}]_{n'_{\mathrm{R}},n_{\mathrm{R}}} f^{(\mathrm{R})}_{n_{\mathrm{R}}}(t). \tag{3.3.6}$$

Using an analog combiner $\mathbf{W} \in \mathbb{C}^{N_{\mathrm{R}} \times L_{\mathrm{R}}}$, the received signal vector is denoted by $\boldsymbol{y}(t) \in \mathbb{C}^{L_{\mathrm{r}} \times 1}$, with the $l_{\mathrm{R}}$-th element given as

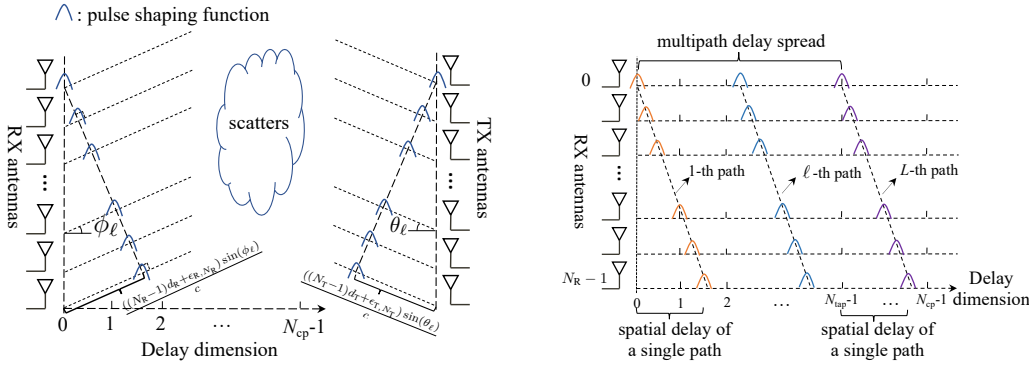
$$[\boldsymbol{y}(t)]_{l_{\mathrm{R}}} = \sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{r}}} [\mathbf{W}]^*_{n_{\mathrm{R}},l_{\mathrm{R}}} f^{(\mathrm{R}')}_{n_{\mathrm{R}}}(t). \tag{3.3.7}$$

Finally, RX measures the received signal samples with the system sampling period $T_{\mathrm{s}}$, yielding the final time domain discrete measurement expression at the $d$-th tap as

$$\big[\boldsymbol{y}[d]\big]_{l_{\mathrm{R}}} = [\boldsymbol{y}(dT_{\mathrm{s}})]_{l_{\mathrm{R}}}, \text{ for } d = 0, 1, \ldots \tag{3.3.8}$$

By developing the previous steps, we reach the final measurement expression as

$$\big[\boldsymbol{y}[d]\big]_{l_{\mathrm{R}}} = \sum_{\ell,n_{\mathrm{R}},n_{\mathrm{T}},l_{\mathrm{T}}} \alpha_\ell [\breve{\mathbf{W}}]^*_{n_{\mathrm{R}},l_{\mathrm{R}}} [\breve{\mathbf{F}}]_{n_{\mathrm{T}},l_{\mathrm{T}}} (p * s_{l_{\mathrm{T}}})(dT_{\mathrm{s}} - \tau_{\ell,n_{\mathrm{R}},n_{\mathrm{T}}}) + [\breve{\mathbf{W}}]^*_{n_{\mathrm{R}},l_{\mathrm{R}}} z_{n_{\mathrm{R}}}(dT_{\mathrm{s}}), \tag{3.3.9}$$




(a) Propagation of the $\ell$-th path between RX and TX with AoA $\phi_\ell$ and AoD $\theta_\ell$.

(b) Received signals at RX antennas, where the delay of each path varies across array geometry.

Figure 3.2: Illustration of the spatial wideband effect (from the perspective of time domain) or the beam squint effect (from the perspective of frequency domain) under large-scale antenna array regimes.

where $\breve{\mathbf{W}} \triangleq (\mathrm{diag}(\boldsymbol{\gamma}_\mathrm{R})\mathbf{C}_\mathrm{R})^* \mathbf{W}$ and $\breve{\mathbf{F}} \triangleq (\mathrm{diag}(\boldsymbol{\gamma}_\mathrm{T})\mathbf{C}_\mathrm{T})^* \mathbf{F}$ summarize the linear effects at RX and TX respectively.

Next, following the derivation in the Appendix B, the DFT of (3.3.9) over the time domain taps $d$, by considering the independent contributions of $s_{l_\mathrm{T}}(t)$, is given as

$$[\mathbf{y}[k]]_{l_\mathrm{R}} = \sum_{d=0}^{K-1} \left[\boldsymbol{y}[d]\right]_{l_\mathrm{R}} e^{-\mathrm{j}\frac{2\pi dk}{K}} = \sum_{\ell,n_\mathrm{R},n_\mathrm{T},l_\mathrm{T}} \alpha_\ell [\breve{\mathbf{W}}]^*_{n_\mathrm{R},l_\mathrm{R}} [\breve{\mathbf{F}}]_{n_\mathrm{T},l_\mathrm{T}} g(k,\tau_{\ell,n_\mathrm{R},n_\mathrm{T}}) \cdot e^{-\mathrm{j}2\pi f_k \tau_{\ell,n_\mathrm{R},n_\mathrm{T}}} [\mathbf{s}[k]]_{l_\mathrm{T}} + [\breve{\mathbf{W}}]^*_{n_\mathrm{R},l_\mathrm{R}} [\mathbf{z}[k]]_{n_\mathrm{R}}, \quad (3.3.10)$$

where $g(k,\tau_{\ell,n_\mathrm{R},n_\mathrm{T}})$ is defined in (B.0.8) in the Appendix B, representing the frequency domain pulse shaping function response at the $k$-th subcarrier for the $\ell$-th path between $n_\mathrm{R}$-th RX antenna and $n_\mathrm{T}$-th RX antenna. Moreover, $[\mathbf{s}[k]]_{l_\mathrm{T}}$ is the frequency domain transmitted signal and $[\mathbf{z}[k]]_{n_\mathrm{R}} \triangleq$

$\sum_{d=0}^{K-1} z_{n_\mathrm{R}}(dT_\mathrm{s})e^{-\mathrm{j}\frac{2\pi dk}{K}}$ is the frequency domain noise term. Regarding the delay $\tau_{\ell,n_\mathrm{R},n_\mathrm{T}}$, as shown in Fig. 3.2, it can be expressed through the different contributions on account of time of arrival and antenna geometry, i.e.,

$$\tau_{\ell,n_\mathrm{R},n_\mathrm{T}} = \tau_\ell + (n_\mathrm{R}d_\mathrm{R} + \epsilon_{\mathrm{R},n_\mathrm{R}})\sin(\phi_\ell)/c - (n_\mathrm{T}d_\mathrm{T} + \epsilon_{\mathrm{T},n_\mathrm{T}})\sin(\theta_\ell)/c. \quad (3.3.11)$$

Now by defining the frequency-dependent array steering vectors $\breve{\mathbf{a}}_{\mathrm{R},k}(\phi) \in \mathbb{C}^{N_\mathrm{R}\times 1}$ and $\breve{\mathbf{a}}_{\mathrm{T},k}(\theta) \in \mathbb{C}^{N_\mathrm{T}\times 1}$ as

$$\breve{\mathbf{a}}_{\mathrm{R},k}(\phi) = \mathrm{diag}(\boldsymbol{\gamma}_\mathrm{R})\mathbf{C}_\mathrm{R}\breve{\mathbf{a}}_{\mathrm{R},k}(\phi), \quad (3.3.12)$$

$$\breve{\mathbf{a}}_{\mathrm{T},k}(\theta) = \mathrm{diag}(\boldsymbol{\gamma}_\mathrm{T})\mathbf{C}_\mathrm{T}\breve{\mathbf{a}}_{\mathrm{T},k}(\theta), \quad (3.3.13)$$

with

$$[\breve{\mathbf{a}}_{\mathrm{R},k}(\phi)]_n = \frac{1}{\sqrt{N_\mathrm{R}}}e^{-\mathrm{j}2\pi f_k(nd_\mathrm{R}+\epsilon_{\mathrm{R},n})\sin(\phi)/c} \quad (3.3.14)$$

$$[\breve{\mathbf{a}}_{\mathrm{T},k}(\theta)]_n = \frac{1}{\sqrt{N_\mathrm{T}}}e^{-\mathrm{j}2\pi f_k(nd_\mathrm{T}+\epsilon_{\mathrm{T},n})\sin(\theta)/c}, \quad (3.3.15)$$

and denoting the distortion matrix by

$$[\mathbf{G}_k(\tau_\ell,\phi_\ell,\theta_\ell)]_{n_\mathrm{R},n_\mathrm{T}} = g(k,\tau_{\ell,n_\mathrm{R},n_\mathrm{T}}), \quad (3.3.16)$$

we can rewrite (3.3.10) in form of

$$\mathbf{y}[k] = \mathbf{W}^*\sum_{\ell=1}^{L}\left(\alpha_\ell e^{-\mathrm{j}2\pi f_k\tau_\ell}\mathbf{G}_k(\tau_\ell,\phi_\ell,\theta_\ell)\odot(\breve{\mathbf{a}}_{\mathrm{R},k}(\phi_\ell)\breve{\mathbf{a}}^*_{\mathrm{T},k}(\theta_\ell))\right)\mathbf{F}\hat{\mathbf{s}}[k] + \breve{\mathbf{W}}^*\mathbf{z}[k]. \quad (3.3.17)$$

Based on this expression, we can define the general frequency domain channel matrix per frequency, namely,

$$\mathbf{H}[k] = \sum_{\ell=1}^{L}\alpha_\ell e^{-\mathrm{j}2\pi f_k\tau_\ell}\mathbf{G}_k(\tau_\ell,\phi_\ell,\theta_\ell)\odot(\breve{\mathbf{a}}_{\mathrm{R},k}(\phi_\ell)\breve{\mathbf{a}}^*_{\mathrm{T},k}(\theta_\ell)), \quad (3.3.18)$$

and then the final measurement formula under these definitions becomes

$$\mathbf{y}[k] = \mathbf{W}^*\mathbf{H}[k]\mathbf{F}\mathbf{s}[k] + \breve{\mathbf{W}}^*\mathbf{z}[k]. \tag{3.3.19}$$

Note that the equivalent combiner for noise includes the hardware impairments as well and thereby the noise covariance matrix is $\mathbf{C}_{\mathbf{z}[k]} = \sigma^2\breve{\mathbf{W}}^*\breve{\mathbf{W}}$.

As per the definition of $g(k,\tau)$ in the Appendix B, i.e.,

$$g(k,\tau) = \begin{cases} \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} + \frac{\mathcal{F}(p)[f_k - 1/T_\mathrm{s}]}{T_\mathrm{s}} e^{\mathrm{j}2\pi\tau/T_\mathrm{s}} & \text{if } \ 2(f_k - f_\mathrm{c}) > (1-\beta)/T_\mathrm{s} \\ \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} & \text{if } \ 2|f_k - f_\mathrm{c}| \leq (1-\beta)/T_\mathrm{s} \\ \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} + \frac{\mathcal{F}(p)[f_k + 1/T_\mathrm{s}]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi\tau/T_\mathrm{s}} & \text{if } \ 2(f_k - f_\mathrm{c}) < -(1-\beta)/T_\mathrm{s} \end{cases} \tag{3.3.20}$$

where $\mathcal{F}(p)[f_k]$ is the Fourier transform of the pulse shaping function $p(t)$ evaluated at frequency $f_k$, the general channel matrix in (3.3.18) can be rewritten as

$$\mathbf{H}[k] = \begin{cases} \sum_{\ell=1}^{L} \alpha_\ell e^{-\mathrm{j}2\pi f_k \tau_\ell} \mathbf{G}_k(\tau_\ell, \phi_\ell, \theta_\ell) \odot (\breve{\mathbf{a}}_{\mathrm{R},k}(\phi_\ell)\breve{\mathbf{a}}^*_{\mathrm{T},k}(\theta_\ell)), \\ \qquad\qquad \forall |f_k - f_\mathrm{c}| > (1-\beta)/2T_\mathrm{s} \\ \sum_{\ell=1}^{L} \alpha_\ell \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi f_k \tau_\ell} \breve{\mathbf{a}}_{\mathrm{R},k}(\phi_\ell)\breve{\mathbf{a}}^*_{\mathrm{T},k}(\theta_\ell), \\ \qquad\qquad \forall |f_k - f_\mathrm{c}| \leq (1-\beta)/2T_\mathrm{s} \end{cases} \tag{3.3.21}$$

where the second case is due to the fact that when $|f_k - f_\mathrm{c}| \leq (1-\beta)/2T_\mathrm{s}$, the factor $g(k, \tau_{\ell,n_\mathrm{R},n_\mathrm{T}}) = \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}}$ is independent of delay $\tau_{\ell,n_\mathrm{R},n_\mathrm{T}}$ at different antenna indices, and thus enables the simplification. Note that (3.3.21) is valid for any random pulse shaping function (as shown in the derivation in Appendix B) and thus this is not a new assumption on the channel model.

*Remark* 1. We have derived the new general model for MIMO channels under both hardware impairments and beam squint impacts. This new model shows

that the beam squint will not only induce the frequency-dependence on array steering vectors but also yield additional distortions at different antennas across all subcarriers, especially on side subcarriers. Comparing (3.3.21) to the existing MIMO channel modeling for beam squint in [26, 75–79], it is obvious that current channel modeling only considered the second case in (3.3.21) and assumed the same expression for all subcarriers. In other words, the prior works only considered the frequency-dependence impact of beam squint but ignored the additional distortions. This is due to the fact that they derived the channel modeling in the continuous time domain and did not take into account the influences of extra bandwidth of pulse shaping functions (as shown in the detailed derivations in Appendix B). For ease of subsequent exposition, we will denote the set of central subcarriers by $\mathcal{K}_{\text{cen}} = \{k \big| |f_k - f_{\text{c}}| \leq (1-\beta)/2T_{\text{s}}\}$ and the set of side subcarriers by $\mathcal{K}_{\text{side}} = \{k \big| |f_k - f_{\text{c}}| > (1-\beta)/2T_{\text{s}}\}$. As shown in Fig. 3.3, there are approximately $|\mathcal{K}_{\text{cen}}| \approx \lfloor (1-\beta)K \rfloor$ central subcarriers inside $\mathcal{K}_{\text{cen}}$ and $|\mathcal{K}_{\text{side}}| \approx \lceil \beta K \rceil$ side subcarriers inside $\mathcal{K}_{\text{side}}$.

### 3.3.3 Frequency domain signal model for training

With the aforementioned system and channel models, the received signal model for training is described as follows. During the training phase, TX will send pilot signals over several OFDM frames for RX to collect measurements for initial channel estimation and dictionary learning. Specifically, during the $m$-th ($m = 1, 2, \ldots$) training OFDM symbol, TX will send an $L_{\text{t}} \times 1$ pilot signal $\mathbf{s}_m[k]$ at the $k$-th subcarriers using a frequency-flat precoder

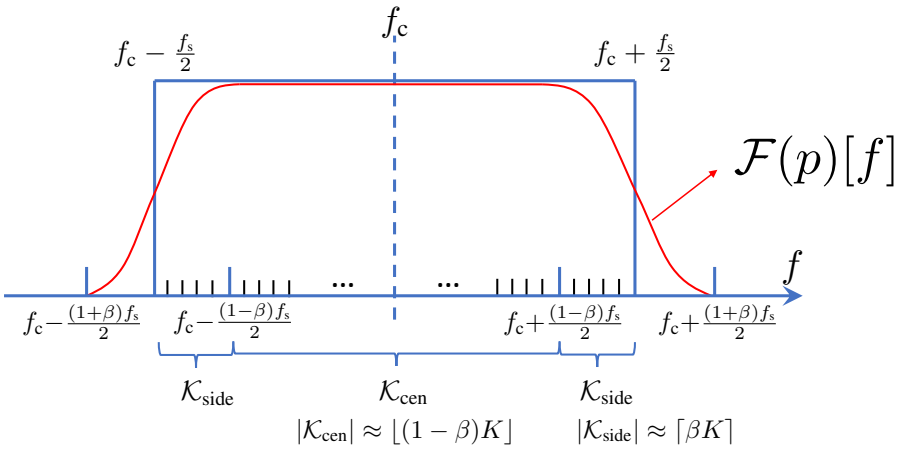



Figure 3.3: Frequency response of a random pulse shaping function with roll-off factor $\beta$, center frequency $f_c$ and $f_s = 1/T_s$. For OFDM systems with $K$ subcarriers, there are approximately $\lfloor (1-\beta)K \rfloor$ central subcarriers inside $\mathcal{K}_{\text{cen}}$ and $\lceil \beta K \rceil$ side subcarriers inside $\mathcal{K}_{\text{side}}$.

$\mathbf{F}_m \in \mathbb{C}^{N_t \times L_t}$, while RX receives with a frequency-flat combiner $\mathbf{W}_m \in \mathbb{C}^{N_r \times L_r}$, so that the received signal $\mathbf{y}_m[k]$ is given as

$$\mathbf{y}_m[k] = \mathbf{W}_m^* \mathbf{H}[k] \mathbf{F}_m \mathbf{s}_m[k] + \mathbf{n}_m[k], \tag{3.3.22}$$

where $\mathbf{n}_m[k] \in \mathbb{C}^{L_r \times 1}$ is the additive Gaussian noise vector, distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \breve{\mathbf{W}}_m^* \breve{\mathbf{W}}_m)$. Generally, we can decompose the transmitted signal $\mathbf{s}_m[k]$ as $\mathbf{s}_m[k] \triangleq \mathbf{q}_m s[k]$, with $\mathbf{q}_m \in \mathbb{C}^{L_t \times 1}$ a frequency-flat training vector and $s[k]$ a frequency-dependent training symbol. In doing so, we can multiply the received signal $\mathbf{y}_m[k]$ by $(s[k])^{-1}$ and get a frequency-flat measurement matrix at the RX as follows

$$\begin{aligned} \tilde{\mathbf{y}}_m[k] \triangleq \text{vec}\left(\mathbf{y}_m[k](s[k])^{-1}\right) &= (\mathbf{q}_m^T \mathbf{F}_m^T \otimes \mathbf{W}_m^*)\text{vec}(\mathbf{H}[k]) + \tilde{\mathbf{n}}_m[k] \\ &= \mathbf{\Phi}_m \text{vec}(\mathbf{H}[k]) + \tilde{\mathbf{n}}_m[k], \end{aligned} \tag{3.3.23}$$

where $\tilde{\mathbf{n}}_m[k] \triangleq \mathrm{vec}\big(\mathbf{n}_m[k](s[k])^{-1}\big)$ and $\mathbf{\Phi}_m$ is defined accordingly, representing the sensing matrix for the $m$-th OFDM symbol. To get a higher effective SNR for the received measurements, we will use training spreading to average out the noise [36]. Moreover, to enable the initial channel estimate, the measurements over in total $M$ OFDM symbols are stacked together such that we have

$$\underbrace{\big[\tilde{\mathbf{y}}_1[k]^T, \ldots, \tilde{\mathbf{y}}_M[k]^T\big]^T}_{\tilde{\mathbf{y}}[k]} = \underbrace{\big[\mathbf{\Phi}_1^T, \ldots, \mathbf{\Phi}_M^T\big]^T}_{\mathbf{\Phi}} \mathbf{h}[k] + \underbrace{\big[\tilde{\mathbf{n}}_1[k]^T, \ldots, \tilde{\mathbf{n}}_M[k]^T\big]^T}_{\tilde{\mathbf{n}}[k]}, \tag{3.3.24}$$

where $\tilde{\mathbf{y}}[k] \in \mathbb{C}^{ML_\mathrm{r} \times 1}$, $\mathbf{\Phi} \in \mathbb{C}^{ML_\mathrm{r} \times N_\mathrm{r} N_\mathrm{t}}$ and $\tilde{\mathbf{n}}[k] \in \mathbb{C}^{ML_\mathrm{r} \times 1}$ are defined accordingly. Therefore, we can have the initial channel estimate by least square (LS) as

$$\hat{\mathbf{h}}[k] = \mathbf{\Phi}^\dagger \tilde{\mathbf{y}}[k], \tag{3.3.25}$$

and the corresponding channel matrix $\hat{\mathbf{H}}[k] = \mathrm{unvec}(\hat{\mathbf{h}}[k])$.

During the training phase, it is also necessary to collect initial channel measurements at different locations across the coverage area as the RX moves around. This can not only create a large training data set for dictionary learning but also ensure the data set is diverse for the environment. In doing so, the learned sparsifying dictionary is not dedicated for a specific location, but adapted to the fixed hardware impairments. As the case in [36], this procedure of collecting measurements at multiple locations can be done at the stage of network setup, like the case of an indoor WiFi scenario. Therefore, we

assume the measurements are collected at $N_{\text{sa}}$ locations and the initial channel estimate at the $u$-th location is then represented as $\hat{\mathbf{h}}^{(u)}[k]$ for $u \in \mathcal{I}(N_{\text{sa}})$. In the following, these initial channel estimates obtained at the training phases will be used for learning the hardware impairments related dictionaries.

## 3.4 Dictionary learning for hardware impairments (DLHW)

In this section, we will discuss the problem formulation and optimization for learning the channel dictionaries adapted to hardware impairments. Generally, the overall channel representing dictionaries are partitioned into two parts, i.e., one frequency-flat part adapting to the hardware impairments and one frequency-dependent part accounting for the beam squint effect. The learning of hardware impairment related dictionaries will take into consideration of the channel properties under beam squint at different subcarriers.

### 3.4.1 Problem formulation of DLHW

As shown in (3.3.21), the frequency domain channel model can be partitioned into two sets. On one hand, for central subcarriers inside $\mathcal{K}_{\text{cen}}$, the channel model can be simplified so that it is more convenient and straightforward to separate the hardware impairments and beam squint impact on the channel matrix. Let us first rewrite the channel matrix for $k \in \mathcal{K}_{\text{cen}}$ as follows

$$\begin{aligned}\mathbf{H}[k] &= \sum_{\ell=1}^{L} \alpha_\ell \frac{\mathcal{F}(p)[f_k]}{T_{\text{s}}} e^{-\text{j}2\pi f_k \tau_\ell} \breve{\mathbf{a}}_{\text{R},k}(\phi_\ell) \breve{\mathbf{a}}^*_{\text{T},k}(\theta_\ell) = \breve{\mathbf{A}}_{\text{R},k}(\boldsymbol{\phi}) \boldsymbol{\Lambda}[k] \breve{\mathbf{A}}^*_{\text{T},k}(\boldsymbol{\theta}) \\ &\approx \breve{\mathbf{A}}^{\text{v}}_{\text{R},k} \boldsymbol{\Omega}[k] \breve{\mathbf{A}}^{\text{v}*}_{\text{T},k}, \end{aligned} \tag{3.4.1}$$

where $\mathbf{\Lambda}[k] \triangleq \mathrm{diag}\big\{[\alpha_1 \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi f_k \tau_1}, \ldots, \alpha_L \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi f_k \tau_L}]^T\big\}$ is a diagonal matrix containing the frequency domain path gains, and $\breve{\mathbf{A}}_{\mathrm{R},k}(\boldsymbol{\phi}) \triangleq [\breve{\mathbf{a}}_{\mathrm{R},k}(\phi_1), \ldots, \breve{\mathbf{a}}_{\mathrm{R},k}(\phi_L)] \in \mathbb{C}^{N_\mathrm{r} \times L}$ collects the receive array response vectors at all paths. Moreover, the approximation is obtained by discretizing the AoA/AoD spaces with on-grid angles, i.e., $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{R},k} \triangleq [\breve{\mathbf{a}}_{\mathrm{R},k}(\phi^{\mathrm{v}}_1), \ldots, \breve{\mathbf{a}}_{\mathrm{R},k}(\phi^{\mathrm{v}}_{K_\mathrm{r}})] \in \mathbb{C}^{N_\mathrm{r} \times K_\mathrm{r}}$ and $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{T},k} \triangleq [\breve{\mathbf{a}}_{\mathrm{T},k}(\theta^{\mathrm{v}}_1), \ldots, \breve{\mathbf{a}}_{\mathrm{T},k}(\theta^{\mathrm{v}}_{K_\mathrm{t}})] \in \mathbb{C}^{N_\mathrm{t} \times K_\mathrm{t}}$ collect the virtual receive and transmit array response vectors evaluated on $K_\mathrm{r}$ quantized angles $\boldsymbol{\phi}^{\mathrm{v}} = \{\phi^{\mathrm{v}}_i\}_{i=1}^{K_\mathrm{r}}$ for AoAs and $K_\mathrm{t}$ quantized angles $\boldsymbol{\theta}^{\mathrm{v}} = \{\theta^{\mathrm{v}}_i\}_{i=1}^{K_\mathrm{t}}$ for AoDs, and $\mathbf{\Omega}[k] \in \mathbb{C}^{K_\mathrm{r} \times K_\mathrm{t}}$ is a sparse matrix, containing the path gains of these discrete quantized AoAs/AoDs at its non-zero elements. To separate the impact of hardware impairments and beam squint on channel matrix, we can rewrite the array steering vector in (3.3.14) as

$$
\begin{aligned}
[\breve{\mathbf{a}}_{\mathrm{R},k}(\phi)]_{n_\mathrm{R}} &= \frac{1}{\sqrt{N_\mathrm{r}}} e^{-\mathrm{j}2\pi f_k (n_\mathrm{R} d_\mathrm{R} + \epsilon_{\mathrm{R},n_\mathrm{R}}) \sin(\phi)/c} \\
&= \frac{1}{\sqrt{N_\mathrm{r}}} e^{-\mathrm{j}2\pi (f_\mathrm{c} \epsilon_{\mathrm{R},n_\mathrm{R}} + \Delta f_k \epsilon_{\mathrm{R},n_\mathrm{R}}) \sin(\phi)/c} e^{-\mathrm{j}2\pi (f_k n_\mathrm{R} d_\mathrm{R}) \sin(\phi)/c} \\
&\approx e^{-\mathrm{j}2\pi (f_\mathrm{c} \epsilon_{\mathrm{R},n_\mathrm{R}}) \sin(\phi)/c} \cdot \frac{1}{\sqrt{N_\mathrm{r}}} e^{-\mathrm{j}2\pi (f_k n_\mathrm{R} d_\mathrm{R}) \sin(\phi)/c}, \qquad (3.4.2) \\
\breve{\mathbf{a}}_{\mathrm{R},k}(\phi) &\approx \mathbf{e}_\mathrm{R}(\phi) \odot \mathbf{a}_{\mathrm{R},k}(\phi), \qquad (3.4.3)
\end{aligned}
$$

where $[\mathbf{e}_\mathrm{R}(\phi)]_{n_\mathrm{R}} = e^{-\mathrm{j}2\pi (f_\mathrm{c} \epsilon_{\mathrm{R},n_\mathrm{R}}) \sin(\phi)/c}$ comprises the hardware antenna element location imperfection effects. Therefore, the expression of $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{R},k}$ in (3.4.1) can

be rewritten as

$$\begin{aligned}\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{R},k} &\triangleq [\breve{\mathbf{a}}_{\mathrm{R},k}(\phi^{\mathrm{v}}_1),\ldots,\breve{\mathbf{a}}_{\mathrm{R},k}(\phi^{\mathrm{v}}_{K_{\mathrm{r}}})] \\ &= \underbrace{\mathrm{diag}(\boldsymbol{\gamma}_{\mathrm{R}})\mathbf{C}_{\mathrm{R}}}_{\mathbf{D}_{\mathrm{R},1}} \Bigg( \underbrace{\left[\mathbf{e}_{\mathrm{R}}(\phi^{\mathrm{v}}_1),\ldots,\mathbf{e}_{\mathrm{R}}(\phi^{\mathrm{v}}_{K_{\mathrm{r}}})\right]}_{\mathbf{D}_{\mathrm{R},2}} \odot \underbrace{\left[\mathbf{a}_{\mathrm{R},k}(\phi^{\mathrm{v}}_1),\ldots,\mathbf{a}_{\mathrm{R},k}(\phi^{\mathrm{v}}_{K_{\mathrm{r}}})\right]}_{\mathbf{A}^{\mathrm{v}}_{\mathrm{R},k}} \Bigg) \\ &= \mathbf{D}_{\mathrm{R},1}\left(\mathbf{D}_{\mathrm{R},2} \odot \mathbf{A}^{\mathrm{v}}_{\mathrm{R},k}\right), \end{aligned} \tag{3.4.4}$$

where the general RX dictionary $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{R},k}$ is partitioned into three parts, i.e., $\mathbf{D}_{\mathrm{R},1}$ includes the impacts of antenna coupling and gain/phase errors, $\mathbf{D}_{\mathrm{R},2}$ accounts for the antenna location perturbations, and $\mathbf{A}^{\mathrm{v}}_{\mathrm{R},k} \in \mathbb{C}^{N_{\mathrm{r}}\times K_{\mathrm{r}}}$ handles the beam squint effect at each subcarriers. Note that $\mathbf{D}_{\mathrm{R},1}$ and $\mathbf{D}_{\mathrm{R},2}$ are hardware impairments related dictionary components to be learned while $\mathbf{A}^{\mathrm{v}}_{\mathrm{R},k}$ is known for all subcarriers. Moreover, recalling the notation of antenna location error vector $\boldsymbol{\epsilon}_{\mathrm{R}}$ and the definition of $\mathbf{e}_{\mathrm{R}}(\phi)$ in (3.4.2), we can further express $\mathbf{D}_{\mathrm{R},2}$ as a function of the antenna location error vector $\boldsymbol{\epsilon}_{\mathrm{R}}$, i.e., $\mathbf{D}_{\mathrm{R},2} = f(\boldsymbol{\epsilon}_{\mathrm{R}}) = e^{-\mathrm{j}2\pi f_{\mathrm{c}}\boldsymbol{\epsilon}_{\mathrm{R}}\cdot\sin(\boldsymbol{\phi}^{\mathrm{v}})^T/c}$. Similarly, the general TX dictionary $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{T},k}$ can be expressed as $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{T},k} = \mathbf{D}_{\mathrm{T},1}\left(\mathbf{D}_{\mathrm{T},2} \odot \mathbf{A}^{\mathrm{v}}_{\mathrm{T},k}\right)$ and $\mathbf{D}_{\mathrm{T},2}$ is a function of the antenna location error vector $\boldsymbol{\epsilon}_{\mathrm{T}}$, i.e., $\mathbf{D}_{\mathrm{T},2} = f(\boldsymbol{\epsilon}_{\mathrm{T}}) = e^{-\mathrm{j}2\pi f_{\mathrm{c}}\boldsymbol{\epsilon}_{\mathrm{T}}\cdot\sin(\boldsymbol{\theta}^{\mathrm{v}})^T/c}$. Then the vectorization of $\mathbf{H}[k], k \in \mathcal{K}_{\mathrm{cen}}$ in (3.4.1) can be expressed as

$$\begin{aligned}\mathbf{h}[k] = \mathrm{vec}(\mathbf{H}[k]) &\approx \left(\overline{\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{T},k}} \otimes \breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{R},k}\right)\mathrm{vec}(\boldsymbol{\Omega}[k]) \\ &= \left(\overline{\mathbf{D}_{\mathrm{T},1}} \otimes \mathbf{D}_{\mathrm{R},1}\right)\left[\left(\overline{\mathbf{D}_{\mathrm{T},2}} \otimes \mathbf{D}_{\mathrm{R},2}\right) \odot \left(\overline{\mathbf{A}^{\mathrm{v}}_{\mathrm{T},k}} \otimes \mathbf{A}^{\mathrm{v}}_{\mathrm{R},k}\right)\right]\mathrm{vec}(\boldsymbol{\Omega}[k]). \end{aligned} \tag{3.4.5}$$

On the other hand, the channel model for side subcarriers $k \in \mathcal{K}_{\mathrm{side}}$ in

(3.3.21) can be represented and approximated as follows

$$\begin{aligned}\mathbf{H}[k] &= \sum_{\ell=1}^{L} \alpha_\ell e^{-\mathrm{j}2\pi f_k \tau_\ell} \mathbf{G}_k(\tau_\ell, \phi_\ell, \theta_\ell) \odot (\breve{\mathbf{a}}_{\mathrm{R},k}(\phi_\ell)\breve{\mathbf{a}}^*_{\mathrm{T},k}(\theta_\ell)) \\ &\approx \sum_{l=1}^{L_\mathrm{v}} \sum_{i=1}^{K_\mathrm{r}} \sum_{j=1}^{K_\mathrm{t}} b_{l,i,j} e^{-\mathrm{j}2\pi f_k \tau_l^\mathrm{v}} \mathbf{G}_k(\tau_l^\mathrm{v}, \phi_i^\mathrm{v}, \theta_j^\mathrm{v}) \odot (\breve{\mathbf{a}}_{\mathrm{R},k}(\phi_i^\mathrm{v})\breve{\mathbf{a}}^*_{\mathrm{T},k}(\theta_j^\mathrm{v})) \end{aligned} \quad (3.4.6)$$

where the approximation is obtained by first discretizing the AoA/AoD spaces as did in (3.4.1) and then discretizing the delay space with $L_\mathrm{v}$ on-grid delays $\{\tau_l^\mathrm{v}\}_{l=1}^{L_\mathrm{v}}$. Moreover, $b_{l,i,j}$ is the corresponding path gain on each pair of discretized AoA/AoD/delay grids, which theoretically is nonzero only at the discrete AoA/AoD/delay bin $(\phi_i^\mathrm{v}, \theta_j^\mathrm{v}, \tau_l^\mathrm{v})$ corresponding to the true AoA/AoD/delay $(\phi_\ell, \theta_\ell, \tau_\ell)$. Then the vectorization of $\mathbf{H}[k], k \in \mathcal{K}_{\text{side}}$ can be given as

$$\begin{aligned}\mathbf{h}[k] = \mathrm{vec}(\mathbf{H}[k]) &= \sum_{\ell=1}^{L} \alpha_\ell e^{-\mathrm{j}2\pi f_k \tau_\ell} \mathrm{vec}(\mathbf{G}_k(\tau_\ell, \phi_\ell, \theta_\ell)) \odot (\overline{\breve{\mathbf{a}}_{\mathrm{T},k}(\theta_\ell)} \otimes \breve{\mathbf{a}}_{\mathrm{R},k}(\phi_\ell)) \\ &\approx \sum_{l=1}^{L_\mathrm{v}} \left[ \boldsymbol{\Psi}_k(\tau_l^\mathrm{v}, \boldsymbol{\phi}^\mathrm{v}, \boldsymbol{\theta}^\mathrm{v}) \odot \left( \overline{\breve{\mathbf{A}}^\mathrm{v}_{\mathrm{T},k}} \otimes \breve{\mathbf{A}}^\mathrm{v}_{\mathrm{R},k} \right) \right] \boldsymbol{b}_l, \end{aligned} \quad (3.4.7)$$

where $\boldsymbol{\Psi}_k(\tau_l^\mathrm{v}, \boldsymbol{\phi}^\mathrm{v}, \boldsymbol{\theta}^\mathrm{v}) \in \mathbb{C}^{N_\mathrm{r}N_\mathrm{t} \times K_\mathrm{r}K_\mathrm{t}}$ and its columns are defined as

$$\begin{aligned}\left[\boldsymbol{\Psi}_k(\tau_l^\mathrm{v}, \boldsymbol{\phi}^\mathrm{v}, \boldsymbol{\theta}^\mathrm{v})\right]_{:,(j-1)K_\mathrm{t}+i} &= e^{-\mathrm{j}2\pi f_k \tau_l^\mathrm{v}} \mathrm{vec}(\mathbf{G}_k(\tau_l^\mathrm{v}, \phi_i^\mathrm{v}, \theta_j^\mathrm{v})), \\ &\forall i = 1, \ldots, K_\mathrm{r}, j = 1, \ldots, K_\mathrm{t}, \end{aligned} \quad (3.4.8)$$

and $\boldsymbol{b}_l \in \mathbb{C}^{K_\mathrm{r}K_\mathrm{t} \times 1}$ collects $b_{l,i,j}$, satisfying $\|\boldsymbol{b}_l\|_0 \leq 1$ and $\sum_{l=1}^{L_\mathrm{v}} \|\boldsymbol{b}_l\|_0 \leq L$.

*Remark* 2. Comparing the channel approximation expressions for central subcarriers in (3.4.1) and for side subcarriers in (3.4.6), it is clear that the additional distortion matrix $\mathbf{G}_k(\tau_\ell, \phi_\ell, \theta_\ell)$ for side subcarriers entangles the three

parameters of delay, AoA and AoD, such that a combination of discretized delay, AoA and AoD grids is needed. This will increase the sparsifying dictionary dimension significantly and induce overwhelming computational complexity during the CS recovery of these parameters. Instead, for the central subcarriers, the discretization of delay, AoA and AoD is decoupled without the additional distortion matrix and thus even one-dimensional search can be done sequentially for each parameter, which will help reduce the computational complexity to a large extent.

Next, as per the approximate channel models in (3.4.1)-(3.4.5) as well as (3.4.6)-(3.4.7), we can formulate the final dictionary learning problem for hardware impairments. Specifically, stacking the initial channel estimates at all subcarriers from all locations $u \in \mathcal{I}(N_{\mathrm{sa}})$, the problem formulation of DLHW can be expressed as

$$\begin{aligned}
&\min_{\substack{\mathbf{D}_{\mathrm{T},1},\mathbf{D}_{\mathrm{T},2},\mathbf{D}_{\mathrm{R},1},\mathbf{D}_{\mathrm{R},2},\\ \boldsymbol{\Omega}^{(u)}[k],\boldsymbol{b}_l^{(u)}}} \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})}\sum_{k\in\mathcal{K}_{\mathrm{cen}}} \left\|\hat{\mathbf{h}}^{(u)}[k] - \left(\overline{\breve{\mathbf{A}}_{\mathrm{T},k}^{\mathrm{v}}}\otimes\breve{\mathbf{A}}_{\mathrm{R},k}^{\mathrm{v}}\right)\mathrm{vec}(\boldsymbol{\Omega}^{(u)}[k])\right\|_F^2 \\
&+\sum_{u\in\mathcal{I}(N_{\mathrm{sa}})}\sum_{k\in\mathcal{K}_{\mathrm{side}}} \left\|\hat{\mathbf{h}}^{(u)}[k] - \sum_{l=1}^{L_{\mathrm{v}}}\left[\boldsymbol{\Psi}_k(\tau_l^{\mathrm{v}},\boldsymbol{\phi}^{\mathrm{v}},\boldsymbol{\theta}^{\mathrm{v}})\odot\left(\overline{\breve{\mathbf{A}}_{\mathrm{T},k}^{\mathrm{v}}}\otimes\breve{\mathbf{A}}_{\mathrm{R},k}^{\mathrm{v}}\right)\right]\boldsymbol{b}_l^{(u)}\right\|_F^2, \\
&\quad \text{subject to} \qquad \left\|\mathrm{vec}(\boldsymbol{\Omega}^{(u)}[k])\right\|_0 \le L^{(u)}; \quad \left\|\boldsymbol{b}_l^{(u)}\right\|_0 \le 1, \quad \sum_{l=1}^{L_{\mathrm{v}}}\left\|\boldsymbol{b}_l^{(u)}\right\|_0 \le L^{(u)}.
\end{aligned} \tag{3.4.9}$$

Note that the two sum terms in (3.4.9) are format-consistent as $\left[\boldsymbol{\Psi}_k(\tau_l^{\mathrm{v}},\boldsymbol{\phi}^{\mathrm{v}},\boldsymbol{\theta}^{\mathrm{v}})\right]_{:,(j-1)K_{\mathrm{t}}+i} = \frac{\mathcal{F}(p)[f_k]}{T_{\mathrm{s}}}e^{-\mathrm{j}2\pi f_k\tau_l^{\mathrm{v}}}\mathbf{1}_{N_{\mathrm{r}}N_{\mathrm{t}}}$ for central subcarriers $k \in \mathcal{K}_{\mathrm{cen}}$. We will see that the simplified expression for central subcarriers can help facilitate the derivations of DL algorithms.

### 3.4.2 Problem optimization of DLHW

The optimization problem in (3.4.9) is not jointly convex with respect to the variables $\mathbf{D}_{\mathrm{T},1}$, $\mathbf{D}_{\mathrm{T},2}$, $\mathbf{D}_{\mathrm{R},1}$, $\mathbf{D}_{\mathrm{R},2}$, $\boldsymbol{\Omega}^{(u)}[k]$, and $\boldsymbol{b}_l^{(u)}$, but it can still be solved by the alternating optimization techniques. As in the typical dictionary learning problems [63], the optimization of (3.4.9) is split into two stages: sparse coding and dictionary update.

#### 3.4.2.1 Sparse coding stage

In this stage, we fix all dictionary parts and update the channel coefficients $\boldsymbol{\Omega}^{(u)}[k], \forall k \in \mathcal{K}_{\mathrm{cen}}, u \in \mathcal{I}(N_{\mathrm{sa}})$ and $\boldsymbol{b}_l^{(u)}, \forall k \in \mathcal{K}_{\mathrm{side}}, u \in \mathcal{I}(N_{\mathrm{sa}})$. Specifically, for each $u \in \mathcal{I}(N_{\mathrm{sa}})$, the optimization problem of (3.4.9) is reduced to (omitting the superscript $^{(u)}$ for simplicity)

$$\begin{aligned}
\min_{\boldsymbol{\Omega}[k],\boldsymbol{b}_l} \quad & \sum_{k\in\mathcal{K}_{\mathrm{cen}}} \left\| \hat{\mathbf{h}}[k] - \left(\overline{\breve{\mathbf{A}}_{\mathrm{T},k}^{\mathrm{v}}} \otimes \breve{\mathbf{A}}_{\mathrm{R},k}^{\mathrm{v}}\right) \mathrm{vec}(\boldsymbol{\Omega}[k]) \right\|_F^2 \\
& + \sum_{k\in\mathcal{K}_{\mathrm{side}}} \left\| \hat{\mathbf{h}}[k] - \sum_{l=1}^{L_{\mathrm{v}}} \left[ \boldsymbol{\Psi}_k(\tau_l^{\mathrm{v}}, \boldsymbol{\phi}^{\mathrm{v}}, \boldsymbol{\theta}^{\mathrm{v}}) \odot \left(\overline{\breve{\mathbf{A}}_{\mathrm{T},k}^{\mathrm{v}}} \otimes \breve{\mathbf{A}}_{\mathrm{R},k}^{\mathrm{v}}\right) \right] \boldsymbol{b}_l \right\|_F^2, \\
\text{subject to} \quad & \left\|\mathrm{vec}(\boldsymbol{\Omega}[k])\right\|_0 \le L; \quad \left\|\boldsymbol{b}_l\right\|_0 \le 1, \quad \sum_{l=1}^{L_{\mathrm{v}}} \left\|\boldsymbol{b}_l\right\|_0 \le L. \qquad (3.4.10)
\end{aligned}$$

Generally, this problem can be optimized by various CS techniques, such as OMP [60], simultaneous OMP (SOMP) [84], simultaneous weighted OMP (SWOMP) [5], to name a few. Compared to existing SOMP and SWOMP, we in this chapter propose a new sparse coding algorithm for (3.4.10), named *Dictionary Adaptive SWOMP under beam squint (DA-SWOMP-beam squint)* as shown in Algorithm 2. Crucially, the newly proposed DA-SWOMP-beam

squint exploits the following two important properties of the channel models under beam squint, which will enable low-complexity and high-accuracy recovery of the channel coefficients and parameters:

- First, the common sparsity support property of channel vectors across subcarriers will still be considered but with frequency-dependent dictionaries accounting for beam squint impacts. For SOMP and SWOMP in previous works, the common sparsity support of channel coefficients between subcarriers is assumed under the same dictionary for all subcarriers. This is an approximate result for the case without considering beam squint effect. When the steering vector at the center frequency is used for all subcarriers, there exist approximation errors no matter how significant beam squint effect is. Under beam squint circumstances, we can exploit the true property of common sparsity support between subcarriers but with different frequency-dependent dictionaries accounting for the beam squint impacts. The principle behind this argument is that the physical AoAs and AoDs associated with propagation paths are constant and independent of subcarriers, and thereby when the channel vector at each subcarrier is projected to the corresponding sparsifying dictionary, only a few bins out of the $K_{\mathrm{r}}$ or $K_{\mathrm{t}}$ virtual angular bins corresponding to the physical AoAs and AoDs are nonzero.

- Second, as mentioned above, the additional distortion matrix at side subcarriers tangles the three parameters of delay, AoA and AoD. Then if

traditional SOMP or SWOMP is directly applied for (3.4.10), the overall sparsifying dictionary will be a three-dimensional (3D) dictionary [6] and the dimension of this 3D dictionary $[\boldsymbol{\Psi}_k(\tau_1^{\mathrm{v}}, \boldsymbol{\phi}^{\mathrm{v}}, \boldsymbol{\theta}^{\mathrm{v}}) \odot \big(\overline{\breve{\mathbf{A}}_{\mathrm{T},k}^{\mathrm{v}}} \otimes \breve{\mathbf{A}}_{\mathrm{R},k}^{\mathrm{v}}\big), \ldots, \boldsymbol{\Psi}_k(\tau_{L_{\mathrm{v}}}^{\mathrm{v}}, \boldsymbol{\phi}^{\mathrm{v}}, \boldsymbol{\theta}^{\mathrm{v}}) \odot \big(\overline{\breve{\mathbf{A}}_{\mathrm{T},k}^{\mathrm{v}}} \otimes \breve{\mathbf{A}}_{\mathrm{R},k}^{\mathrm{v}}\big)]$ will be $N_{\mathrm{r}}N_{\mathrm{t}} \times L_{\mathrm{v}}K_{\mathrm{r}}K_{\mathrm{t}}$, proportional to the product of the numbers of delay/AoA/AoD grids. This will certainly induce overwhelming complexity. As for the newly proposed DA-SWOMP-BS, it will first exploit the channel models at central subcarriers to obtain initial estimates of the delay/AoA/AoD parameters. Without the impact of additional distortion matrix at central subcarriers, this can be done even by one-dimensional search over each parameter space iteratively. These initial estimates of parameters are then used to reduce the effective dictionary dimension at side subcarriers and compensate the distortions induced by beam squint, so that all the subcarriers can be collected to improve the estimates of delays/AoAs/AoDs again. In doing so, the sparse coding of (3.4.10) can be solved with much lower complexity and higher accuracy.

In line of these ideas, the DA-SWOMP-BS is expected to outperform SOMP and SWOMP without considering beam squint. We summarize the procedure of DA-SWOMP-BS in Algorithm 2. Once the sparse coding stage of (3.4.10) is done, the coefficients $\boldsymbol{\Omega}^{(u)}[k], k \in \mathcal{K}_{\mathrm{cen}}$ and $\boldsymbol{b}_l^{(u)}, k \in \mathcal{K}_{\mathrm{side}}$ as well as the estimates of path gains, delays, AoAs, and AoDs $\hat{\boldsymbol{\varphi}}_{\hat{l}}^{(u)} \triangleq \{\hat{\alpha}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}, \hat{\tau}_{\hat{l}}^{(u)}\}$ for $\hat{l} = 1, \ldots, \hat{L}^{(u)}, \forall u \in \mathcal{I}(N_{\mathrm{sa}})$ can be obtained, which will be used to update the hardware impairments related dictionaries in the next subsections.

#### 3.4.2.2 Dictionary update stage

Next, we fix the channel coefficients and path parameters in (3.4.9) in preparation of updating the hardware impairments related dictionaries $\mathbf{D}_{\mathrm{T},1}, \mathbf{D}_{\mathrm{T},2}, \mathbf{D}_{\mathrm{R},1}, \mathbf{D}_{\mathrm{R},2}$. Note that due to the special structure in this problem that the TX and RX dictionaries are entangled with the beam squint effect at different subcarriers, the typical dictionary update algorithms, like the method of optimal directions (MOD) [65] or K-SVD [63], cannot be directly used. Therefore, we apply alternating optimization in this sub-stage as well to subsequently update the four dictionaries.

For the update of $\mathbf{D}_{\mathrm{R},1}$, the problem of (3.4.9) can be reduced to

$$\begin{aligned}\min_{\mathbf{D}_{\mathrm{R},1}} \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})} \sum_{k\in\mathcal{K}_{\mathrm{cen}}} \Bigg\| \hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{\mathrm{R},1} \underbrace{\left(\mathbf{D}_{\mathrm{R},2} \odot \mathbf{A}^{\mathrm{v}}_{\mathrm{R},k}\right) \boldsymbol{\Omega}^{(u)}[k] \left(\mathbf{D}_{\mathrm{T},2} \odot \mathbf{A}^{\mathrm{v}}_{\mathrm{T},k}\right)^{*} \mathbf{D}^{*}_{\mathrm{T},1}}_{\mathbf{X}^{(u)}_{\mathrm{R},1}[k]} \Bigg\|^2_F \\ + \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})} \sum_{k\in\mathcal{K}_{\mathrm{side}}} \Bigg\| \hat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}^{(u)}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}^{(u)}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}^{(u)}_{\hat{l}}, \hat{\phi}^{(u)}_{\hat{l}}, \hat{\theta}^{(u)}_{\hat{l}}) \\ \odot \left(\breve{\mathbf{a}}_{\mathrm{R},k}(\hat{\phi}^{(u)}_{\hat{l}}) \breve{\mathbf{a}}^{*}_{\mathrm{T},k}(\hat{\theta}^{(u)}_{\hat{l}})\right) \Bigg\|^2_F . \end{aligned} \tag{3.4.11}$$

which is equivalent to

$$\begin{aligned}\min_{\mathbf{D}_{\mathrm{R},1}} \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})} \sum_{k\in\mathcal{K}_{\mathrm{cen}}} \Bigg\| \hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{\mathrm{R},1} \underbrace{\left(\mathbf{D}_{\mathrm{R},2} \odot \mathbf{A}^{\mathrm{v}}_{\mathrm{R},k}\right) \boldsymbol{\Omega}^{(u)}[k] \left(\mathbf{D}_{\mathrm{T},2} \odot \mathbf{A}^{\mathrm{v}}_{\mathrm{T},k}\right)^{*} \mathbf{D}^{*}_{\mathrm{T},1}}_{\mathbf{X}^{(u)}_{\mathrm{R},1}[k]} \Bigg\|^2_F \\ + \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})} \sum_{k\in\mathcal{K}_{\mathrm{side}}} \Bigg\| \hat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}^{(u)}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}^{(u)}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}^{(u)}_{\hat{l}}, \hat{\phi}^{(u)}_{\hat{l}}, \hat{\theta}^{(u)}_{\hat{l}})\end{aligned}$$

$$\odot \Big( \mathbf{D}_{\mathrm{R},1} \underbrace{\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}}^{(u)})\big)\big(\mathbf{e}_{\mathrm{T}}(\hat{\theta}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{\mathrm{T},k}(\hat{\theta}_{\hat{l}}^{(u)})\big)^{*} \mathbf{D}_{\mathrm{T},1}^{*}}_{\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]} \Big) \Big\|_{F}^{2}, \tag{3.4.12}$$

where $\mathbf{X}_{\mathrm{R},1}^{(u)}[k]$ and $\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]$ are defined accordingly for ease of expression. To update $\mathbf{D}_{\mathrm{R},1}$, we need to calculate the derivative of the objective function with respect to $\mathbf{D}_{\mathrm{R},1}$, i.e., $\dfrac{\partial J}{\partial \mathbf{D}_{\mathrm{R},1}}$, which is expressed as follows (A proof is provided in Appendix C.)

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{D}_{\mathrm{R},1}} = & - \sum_{u \in \mathcal{I}(N_{\mathrm{sa}})} \sum_{k \in \mathcal{K}_{\mathrm{cen}}} \overline{(\hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{\mathrm{R},1} \mathbf{X}_{\mathrm{R},1}^{(u)}[k])} (\mathbf{X}_{\mathrm{R},1}^{(u)}[k])^{T} \\
& - \sum_{u \in \mathcal{I}(N_{\mathrm{sa}})} \sum_{k \in \mathcal{K}_{\mathrm{side}}} \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}^{(u)}} \\
& \cdot \Bigg[ \Big( \overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \overline{\hat{\alpha}_{\hat{l}}^{(u)} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}^{(u)}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \odot \mathbf{D}_{\mathrm{R},1} \mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]} \Big) \\
& \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \Bigg] \cdot \big(\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]\big)^{T}.
\end{aligned} \tag{3.4.13}$$

Therefore, the update of $\mathbf{D}_{\mathrm{R},1}$ can be obtained by (stochastic) gradient decent as

$$[\mathbf{D}_{\mathrm{R},1}]^{\mathrm{new}} = [\mathbf{D}_{\mathrm{R},1}]^{\mathrm{old}} - \eta \frac{\partial J}{\partial \mathbf{D}_{\mathrm{R},1}}, \tag{3.4.14}$$

where $\eta$ is the step-size of gradient descent and can be determined by back-tracking line search. Similarly, the update of $\mathbf{D}_{\mathrm{T},1}$ can be obtained.

For the update of $\mathbf{D}_{\mathrm{R},2} = f(\boldsymbol{\epsilon}_{\mathrm{R}}) \triangleq e^{-\mathrm{j}2\pi f_{\mathrm{c}} \boldsymbol{\epsilon}_{\mathrm{R}} \cdot \sin(\boldsymbol{\phi}^{\mathrm{v}})^{T}/c}$, it is equivalent to updating $\boldsymbol{\epsilon}_{\mathrm{R}}$. By stacking all subcarriers, we have the objective of updating

$\boldsymbol{\epsilon}_{\mathrm{R}}$ as follows

$$\min_{\boldsymbol{\epsilon}_{\mathrm{R}}} \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})} \sum_{k\in\mathcal{K}_{\mathrm{cen}}} \Bigg\| \underbrace{\mathbf{D}_{\mathrm{R},1}^{\dagger}\hat{\mathbf{H}}^{(u)}[k]\mathbf{D}_{\mathrm{T},1}^{*\dagger}}_{\mathbf{Y}_{\mathrm{R},2}^{(u)}[k]} - \left(\mathbf{D}_{\mathrm{R},2}\odot\mathbf{A}_{\mathrm{R},k}^{\mathrm{v}}\right)\underbrace{\boldsymbol{\Omega}^{(u)}[k]\left(\mathbf{D}_{\mathrm{T},2}\odot\mathbf{A}_{\mathrm{T},k}^{\mathrm{v}}\right)^{*}}_{\mathbf{X}_{\mathrm{R},2}^{(u)}[k]} \Bigg\|_F^2$$
$$+ \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})} \sum_{k\in\mathcal{K}_{\mathrm{side}}} \Bigg\| \hat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-\mathrm{j}2\pi f_k\hat{\tau}_{\hat{l}}^{(u)}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)},\hat{\phi}_{\hat{l}}^{(u)},\hat{\theta}_{\hat{l}}^{(u)})$$
$$\odot\left(\mathbf{D}_{\mathrm{R},1}\left(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}^{(u)})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}}^{(u)})\right)\underbrace{\left(\mathbf{e}_{\mathrm{T}}(\hat{\theta}_{\hat{l}}^{(u)})\odot\mathbf{a}_{\mathrm{T},k}(\hat{\theta}_{\hat{l}}^{(u)})\right)^{*}\mathbf{D}_{\mathrm{T},1}^{*}}_{\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]}\right)\Bigg\|_F^2, \tag{3.4.15}$$

where $\mathbf{X}_{\mathrm{R},2}^{(u)}[k]$ and $\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]$ are defined accordingly for ease of expression. To update $\boldsymbol{\epsilon}_{\mathrm{R}}$, we need to calculate the derivative of the objective function with respect to $\boldsymbol{\epsilon}_{\mathrm{R}}$, which can be expressed as (A proof is given in Appendix D)

$$\frac{\partial J}{\partial\boldsymbol{\epsilon}_{\mathrm{R}}} = \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})}\sum_{k\in\mathcal{K}_{\mathrm{cen}}} 2\mathcal{R}\left\{\left\{\frac{\partial J_k^{(u)}}{\partial\mathbf{D}_{\mathrm{R},2}}\odot\mathbf{D}_{\mathrm{R},2}\right\}\cdot\frac{-\mathrm{j}2\pi f_{\mathrm{c}}\sin(\boldsymbol{\phi}^{\mathrm{v}})}{c}\right\}$$
$$- \sum_{u\in\mathcal{I}(N_{\mathrm{sa}})}\sum_{k\in\mathcal{K}_{\mathrm{side}}} 2\mathcal{R}\Bigg\{\sum_{\hat{l}=1}^{\hat{L}^{(u)}}\hat{\alpha}_{\hat{l}}^{(u)}e^{-\mathrm{j}2\pi f_k\hat{\tau}_{\hat{l}}^{(u)}}\cdot\frac{-\mathrm{j}2\pi f_{\mathrm{c}}\sin(\hat{\phi}_{\hat{l}}^{(u)})}{c}\cdot\Bigg[\mathbf{D}_{\mathrm{R},1}^{T}\Bigg(\Bigg(\overline{\hat{\mathbf{H}}^{(u)}[k]}$$
$$- \sum_{\hat{l}=1}^{\hat{L}^{(u)}}\hat{\alpha}_{\hat{l}}^{(u)}e^{-\mathrm{j}2\pi f_k\hat{\tau}_{\hat{l}}^{(u)}}\mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)},\hat{\phi}_{\hat{l}}^{(u)},\hat{\theta}_{\hat{l}}^{(u)})\odot\left(\mathbf{D}_{\mathrm{R},1}\left(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}^{(u)})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}}^{(u)})\right)\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]\right)\Bigg)$$
$$\odot\,\mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)},\hat{\phi}_{\hat{l}}^{(u)},\hat{\theta}_{\hat{l}}^{(u)})\Bigg)(\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k])^{T}\Bigg]\odot\left(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}^{(u)})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}}^{(u)})\right)\Bigg\}. \tag{3.4.16}$$

Therefore, the update of $\mathbf{D}_{\mathrm{R},2}$ is given as

$$[\mathbf{D}_{\mathrm{R},2}]^{\mathrm{new}} = f([\boldsymbol{\epsilon}_{\mathrm{R}}]^{\mathrm{new}}) = f\left([\boldsymbol{\epsilon}_{\mathrm{R}}]^{\mathrm{old}} - \eta\frac{\partial J}{\partial\boldsymbol{\epsilon}_{\mathrm{R}}}\right). \tag{3.4.17}$$

Similarly, the update of $\mathbf{D}_{\mathrm{T},2}$ can be obtained.

By iterative refinement between sparse coding and dictionary update stages, the learned hardware impairments related dictionaries can be obtained when the objective function in (3.4.9) converges. The overall procedure of the proposed dictionary learning scheme is summarized in Algorithm 3.

Once the hardware impairments related dictionaries $\mathbf{D}_{\mathrm{R},1}, \mathbf{D}_{\mathrm{R},2}, \mathbf{D}_{\mathrm{T},1}, \mathbf{D}_{\mathrm{T},2}$ are obtained, the overall RX and TX sparsifying dictionaries $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{R},k}$ and $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathrm{T},k}$ can be constructed at each subcarrier to incorporate the beam squint impact as in (3.4.4), and then used for subsequent online compressive channel estimation, which is expected to help reduce the training overhead significantly.

### 3.4.3 Convergence and complexity analysis

The alternating optimization involves two stages. For sparse coding stage, as explained in [63], when the sparsity level of signals is small related to its dimension, OMP based pursuit algorithms are known to perform well, i.e., obtaining the best approximation of signals and satisfying the sparsity constraint. Therefore, the proposed OMP based Algorithm 2 is able to decrease the value of the objective function during each sparse coding stage. While for the dictionary update, an additional reduction or no change in the mismatch error is guaranteed at each sub-stage by using gradient descent with backtracking line search. Therefore, the alternating steps for optimizing the dictionary learning problem can guarantee a monotonic decrease in the objective function and then ensure the convergence to a local minimum.

As for the complexity, since the dictionary learning phase can be implemented offline, the complexity involved in dictionary update stage does not increase the overall complexity of the online sparse coding stage. Therefore, we compare the computational overhead of the proposed DA-SWOMP-BS with those of SOMP [84] and SWOMP [5], in terms of complex multiplication operations for each iteration. For ease of comparison, we set the numbers of AoA/AoD/delay grids to be the same $K_{\mathrm{r}} = K_{\mathrm{t}} = L_{\mathrm{v}} = S$. For sparse coding with DA-SWOMP-BS, the parameters can be estimated by iterative one-dimensional search over each parameter space, and thus the complexity order per iteration is $\mathcal{O}(3N_{\mathrm{r}}N_{\mathrm{t}}SN_{\mathrm{c}}N_{\mathrm{sa}})$. While for SOMP or SWOMP, the joint search over the 3D parameter space induces a much higher complexity order of $\mathcal{O}\big(N_{\mathrm{r}}N_{\mathrm{t}}S^3N_{\mathrm{c}}N_{\mathrm{sa}}\big)$.

## 3.5 Numerical results

In this section, numerical results are presented to demonstrate the effectiveness of the proposed dictionary learning and channel estimation algorithms for hybrid wideband mmWave MIMO systems under both hardware impairments and beam squint.

### 3.5.1 Simulation parameters

The TX and RX are assumed to deploy ULAs with half-wavelength spacing, i.e., $d_{\mathrm{R}} = d_{\mathrm{T}} = \lambda_{\mathrm{c}}/2$, and $N_{\mathrm{t}} = 32$, $N_{\mathrm{r}} = 8$, $N_{\mathrm{s}} = L_{\mathrm{t}} = L_{\mathrm{r}} = 2$. The number of OFDM subcarriers is $N_{\mathrm{c}} = 64$ and $T_{\mathrm{s}} = 1\ ns$. Moreover,

for fair comparison, the numbers of discrete angle and delay grids are set as $K_{\mathrm{t}} = 2N_{\mathrm{t}}$, $K_{\mathrm{r}} = 2N_{\mathrm{r}}$ and $L_{\mathrm{v}} = 2N_{\mathrm{c}}$ for different algorithms in the simulations (To be noted that, the proposed Algorithm 2 is of much lower complexity so that finer angle and delay grids can be used to improve parameter estimation accuracy). The channels are generated based on (3.3.21) with $L = 6$ multipath components, as in the typical indoor scenarios. The pulse shaping function $p(t)$ is assumed to be a raised-cosine filter with roll-off factor of $\beta = 0.25$. The SNR is set to 0 dB for dictionary learning and channel estimation. During the dictionary learning phase, the number of OFDM symbols for training is set as 500 and a spreading factor 10 is used to increase the effective SNR in 10 dBs. For the parameters of hardware impairments, as in [36, 44, 51], the maximum gain and phase error variances for each antenna element are set as 5% and $20^{\circ}\pi/180^{\circ}$ respectively. Furthermore, the mutual coupling coefficients among antennas are within $[0.01, 0.4]$ and the antenna location errors are assumed to be uniformly distributed between $[-0.1\lambda_{\mathrm{c}}, 0.1\lambda_{\mathrm{c}}]$.

For the optimization of DLHW, a revised version of dictionary initialization algorithm (DIA) in [67] is used for initialization of the hardware impairments. Specifically, the original DIA algorithm [67] is first applied to initialize the combined dictionary $\mathbf{D}_{\mathrm{R}}^{\mathrm{init}} \triangleq \mathbf{D}_{\mathrm{R},1}(\mathbf{D}_{\mathrm{R},2} \odot \mathbf{A}_{\mathrm{R},k}^{\mathrm{v}})$. Then we set the initial values of the antenna spacing errors as zeros, i.e., $\boldsymbol{\epsilon}_{\mathrm{R}} = \mathbf{0}$ and $\mathbf{D}_{\mathrm{R},2} = \mathbf{1}_{N_{\mathrm{r}} \times K_{\mathrm{r}}}$, and thus the initialization of gain, phase and coupling matrix is obtained by $\mathbf{D}_{\mathrm{R},1} = \mathbf{D}_{\mathrm{R}}^{\mathrm{init}}(\mathbf{A}_{\mathrm{R},k}^{\mathrm{v}})^{\dagger}$.

For comparison, the sparsifying dictionaries constructed from overcom-

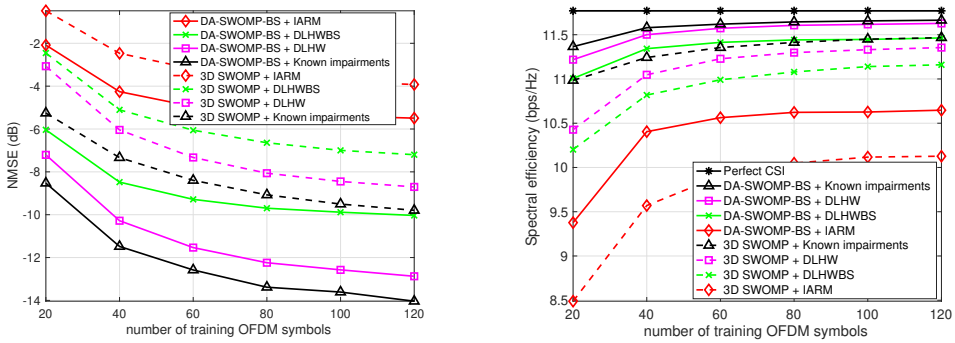
plete ideal array response matrix (IARM) $\mathbf{A}^{\mathrm{v}}_{\mathrm{R},k}$ and $\mathbf{A}^{\mathrm{v}}_{\mathrm{T},k}$ assuming no any hardware impairments, are also evaluated. Moreover, we take into account the general dictionary learning algorithm proposed in [36], which learns a frequency-flat combined dictionary for both hardware impairments and beam squint impact. For ease of exposition, we denote it by DLHWBS. Last but not least, we also assume full knowledge of the hardware impairments to gauge the performance gaps of DLHW from the true case. With these various sparsifying dictionaries, the online compressive channel estimation, which is exactly a sparse coding problem in (3.4.10), will be solved by the proposed DA-SWOMP-BS in Algorithm 2. The conventional SWOMP [5,6] described in section 3.4.2.1 will also be included for channel estimation for comparison.

### 3.5.2 Performance comparison between different dictionaries and sparse coding algorithms

In Fig. 3.4(a), we evaluate the normalized mean square error (NMSE) performances of online compressive channel estimation with different sparsifying dictionaries and sparse coding algorithms, as a function of the number of training OFDM symbols. The results are averaged over 100 realizations of channels. First, let us focus on channel estimation using the proposed DA-SWOMP-BS algorithm. Since the ideal IARM dictionaries do not have any prior information on the hardware impairments, the overall channel estimation performances are not satisfying. By contrast, the dictionary learning based approaches, i.e., DLHWBS and the newly proposed DLHW, can improve the channel estimation accuracy to a large extent, as they are able to capture the

intrinsic hardware impairments in the learned dictionaries. Comparing between DLHWBS and DLHW, the latter outperforms the former significantly. This is due to the fact that DLHWBS attempts to learn a frequency-flat dictionary for all subcarriers under both hardware impairments and beam squint effect, and assumes the channel models are all in the form of the second case of (3.3.21). In other words, the distortions on the channel models at side subcarriers are ignored and thus there already exist some modeling errors when DLHWBS is applied. Instead, the newly proposed DLHW is adapted only to the hardware impairments while the beam squint impacts are incorporated for different subcarriers, which enables it to achieve better accuracy. Moreover, the performances of DLHW are closer to the case of ideal knowledge of hardware impairments. As the ideal information of hardware impairments is not available in practice, the proposed DLHW algorithm has been a feasible and effective option. These results not only justify the motivation of DLHW in this chapter, but also demonstrate its effectiveness under both hardware impairments and beam squint effects.

Fig. 3.4(a) has also illustrated the channel estimation performances of the conventional SWOMP [5] with various dictionaries. As described in section 3.4.2.1, the additional channel distortion under beam squint will tangle the three parameters of delay/AoA/AoD, and thus the overall sparsifying dictionary for SWOMP in this case will be a 3D dictionary [6], with dimension proportional to the product of numbers of delay/AoA/AoD grids. This induces quite large computational complexity. Specifically, in this example, the aver-

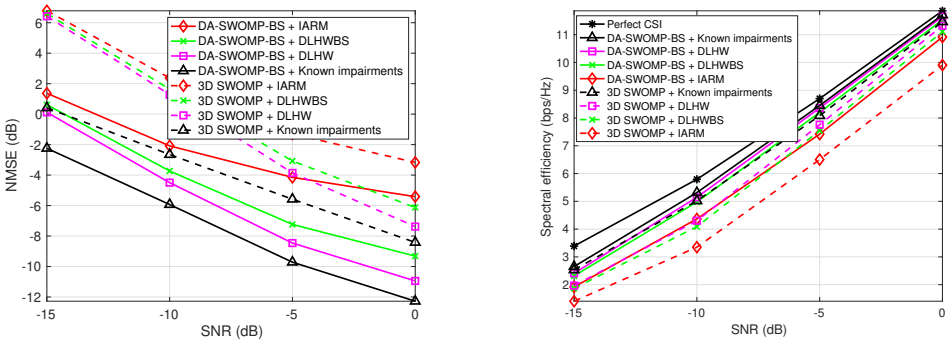



(a) Comparison of NMSE performances. (b) Comparison of SE performances.

Figure 3.4: Comparisons of NMSE and SE performances as a function of the number of training OFDM symbols for various sparsifying dictionaries and sparse coding algorithms, where SNR = 0 dB.

age simulation time is more than 10 times that of DA-SWOMP-BS in Matlab. Moreover, as can be seen from Fig. 3.4(a), the newly proposed DA-SWOMP-BS algorithm outperforms the conventional 3D SWOMP significantly with all various dictionaries. This is due to the fact that the newly proposed algorithm is able to conduct one-dimensional OMP for each parameter over the corresponding parameter space sequentially, and then iteratively update the estimated delay/AoA/AoD based on already computed ones, while the SWOMP performs only one projection over the 3D parameter space considering the limitation of overwhelming complexity. These results further demonstrate that the newly proposed DA-SWOMP-BS is able to estimate the channel parameters with much lower complexity and higher accuracy.

In Fig. 3.4(b), we also compare the spectral efficiency (SE) performances corresponding to the various sparsifying dictionaries and sparse cod-




(a) Comparison of NMSE performances. (b) Comparison of SE performances.

Figure 3.5: Comparisons of NMSE and SE performances as a function of SNR using the various sparsifying dictionaries and sparse coding algorithms, where the number of training OFDM symbols is set as $M = 60$.

ing algorithms in Fig. 3.4(a). As in previous work [5,36], the SE is computed by assuming fully-digital precoding and combining using estimates for the $N_{\mathrm{s}}$ dominant left and right singular vectors of the channel estimates. To be clear, SE is defined as

$$\mathrm{SE} = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{n=1}^{N_{\mathrm{s}}}\log\left(1+\frac{\mathrm{SNR}}{N_{\mathrm{s}}}\lambda_n(\mathbf{H}_{\mathrm{eff}}[k])^2\right), \tag{3.5.1}$$

where $\mathbf{H}_{\mathrm{eff}}[k]$ is the effective channel after precoding/combining and $\lambda_n(\mathbf{H}_{\mathrm{eff}}[k])$ takes the singular values of $\mathbf{H}_{\mathrm{eff}}[k]$. As can be seen that, for both DA-SWOMP-BS and conventional 3D SWOMP, SE can be significantly increased with the learned dictionaries compared to those of using IARM dictionaries. Moreover, the performance gap between the proposed DLHW algorithm and the case of ideal hardware impairment knowledge is small, while the performance gain of DLHW compared to DLHWBS is obviously larger. Moreover, when comparing the SE performances between DA-SWOMP-BS and the 3D SWOMP, the

former still outperforms the latter with various dictionaries, consistent with the channel estimation results in Fig. 3.4(a).

In Fig. 3.5, we further evaluate the channel estimation and SE performances for different sparsifying dictionaries and sparse coding algorithms, as a function of SNR. The parameters involved in this simulation are the same as those in Fig. 3.4. As can be seen from Fig. 3.5, both the NMSE and SE performances of DLHW are superior to those using DLHWBS or IARM dictionaries, even at a relatively low SNR. This means that the proposed DLHW algorithm is feasible and effective under different conditions.

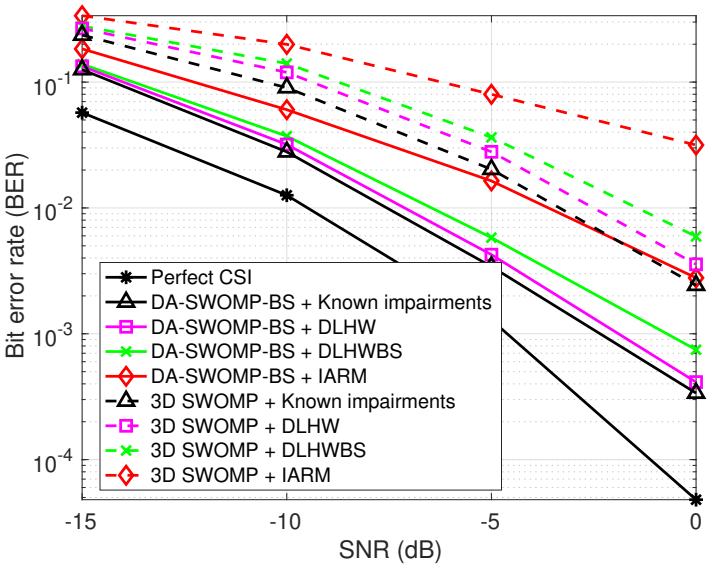
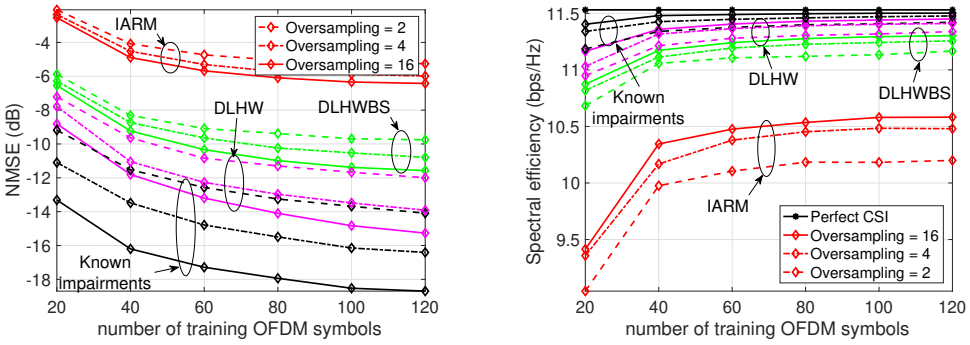



Figure 3.6: Comparison of BER performances as a function of SNR using QPSK modulation and MMSE equalizer, where the number of training OFDM symbols is set as $M = 60$.

We also evaluate the bit error rate (BER) performances in Fig. 3.6,

where the quadrature phase shift keying (QPSK) modulation and the minimum mean squared error (MMSE) detection are considered. As can be seen, the average BER performance gap between the proposed DLHW and the ideal case of known knowledge is less than 0.5 dB, while the performance gain of DLHW compared to DLHWBS is larger than 1 dB at the BER level of $10^{-3}$. This also corroborates the effectiveness of our proposed algorithm for hybrid wideband channel under hardware impairments and beam squint from a more practical perspective.

Last, as mentioned above, the advantage of lower complexity of the newly proposed DA-SWOMP-BS algorithm enables it to adopt finer grids for delay/AoA/AoD, so that parameter estimation accuracy can be further improved. To illustrate this, we evaluate the performances of DA-SWOMP-BS with various oversampling factors, i.e., the ratios of $K_{\mathrm{r}}/N_{\mathrm{r}}$, $K_{\mathrm{t}}/N_{\mathrm{t}}$ and $L_{\mathrm{v}}/N_{\mathrm{c}}$, in Fig. 3.7. Note that the conventional 3D SWOMP is not evaluated, for its overwhelming complexity with larger oversampling factors makes it unfeasible in a reasonable time. It is clear in Fig. 3.7 that when the oversampling factor is increased, the performances of DA-SWOMP-BS with various dictionaries can be improved accordingly due to the finer grids of parameters. This performance gain of larger oversampling factor is at the cost of increased computational complexity, which, however, is still much lower than that of conventional SWOMP with even a small oversampling factor.




(a) Comparison of NMSE performances. (b) Comparison of SE performances.

Figure 3.7: Comparison of NMSE and SE performances of the proposed DA-SWOMP-BS algorithm with various oversampling factors.

## 3.6 Conclusions

In this chapter, we investigated the dictionary learning and channel estimation problems for hybrid wideband mmWave MIMO systems under the impacts of both hardware impairments and beam squint. We first derived a general channel model for MIMO systems under beam squint, by explicitly considering the influences of hardware impairments and pulse shaping functions, which is a compensation and extension to existing MIMO channel models without considering beam squint. Based on this general channel modeling, we formulated a dictionary learning problem only for hardware impairments, which exploits the beam squint impacts but does not learn its impact on the dictionaries. Comparing to previous dictionary learning algorithm that learns the dictionary for both hardware impairments and beam squint, the newly proposed dictionary learning strategy ensures the learned dictionaries to be better adapted to hardware impairments. We also proposed a novel

compressive channel estimation algorithm under beam squint, which exploits the beam squint channel structures at different subcarriers to compensate the beam squint impacts and then facilitates the parameter estimations with much lower complexity and higher accuracy. Numerical results have demonstrated the effectiveness of the proposed dictionary learning and channel estimation strategies.

**Algorithm 2** : DA-SWOMP-BS algorithm for sparse coding

**procedure** DA-SWOMP-BS ($\mathbf{y}^{(u)}[k]$, $\mathbf{\Phi}$, $\mathbf{D}_{\mathrm{T},1}$, $\mathbf{D}_{\mathrm{T},2}$, $\mathbf{D}_{\mathrm{R},1}$, $\mathbf{D}_{\mathrm{R},2}, K_{\mathrm{r}}, K_{\mathrm{t}}, L_{\mathrm{v}}$)

**Construct** $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathbf{R},k}$ **and** $\breve{\mathbf{A}}^{\mathrm{v}}_{\mathbf{T},k}$ **at each subcarrier with input dictionaries**

**Compute the whitened equivalent sensing matrix and received signals**

$$\mathbf{L}_{\mathrm{w}} = \mathrm{blkdiag}\left\{\breve{\mathbf{W}}_1^*\breve{\mathbf{W}}_1, \ldots, \breve{\mathbf{W}}_M^*\breve{\mathbf{W}}_M\right\}^{-1/2} \quad \text{and} \quad \mathbf{\Phi}_{\mathrm{w}} = \mathbf{L}_{\mathrm{w}}\mathbf{\Phi}$$

$$\mathbf{y}_{\mathrm{w}}^{(u)}[k] = \mathbf{L}_{\mathrm{w}}\mathbf{y}^{(u)}[k], \text{ for } k = 0, \ldots, K-1, \ u \in \mathcal{I}(N_{\mathrm{sa}})$$

**for** $u = 0, \ldots, N_{\mathrm{sa}} - 1$ **do**

**Initialize the residual vectors and set of estimated parameters**

$$\mathbf{r}[k] = \mathbf{\Phi}_{\mathrm{w}}^*\mathbf{y}_{\mathrm{w}}^{(u)}[k], \ \forall k = 0, \ldots, K-1 \text{ and } \hat{L}^{(u)} = 0, \ \hat{\boldsymbol{\varphi}}_l^{(u)} = \emptyset$$

**while** MSE $> \epsilon$ **do:**

**Initial parameter estimation using central subcarriers**

$$\text{delay: } \hat{\tau} = \mathrm{argmax}_{\tau_l^{\mathrm{v}}} \ \Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} \mathbf{r}[k] e^{\mathrm{j}2\pi f_k \tau_l^{\mathrm{v}}} \Big\|_F^2$$

$$\text{AoA: } \hat{\phi} = \mathrm{argmax}_{\phi_i^{\mathrm{v}}} \ \frac{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\mathbf{I}_{N_{\mathrm{t}}} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\phi_i^{\mathrm{v}}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\mathbf{I}_{N_{\mathrm{t}}} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\phi_i^{\mathrm{v}}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}$$

$$\text{AoD: } \hat{\theta} = \mathrm{argmax}_{\theta_j^{\mathrm{v}}} \ \frac{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\theta_j^{\mathrm{v}})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\theta_j^{\mathrm{v}})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}$$

**Improve the initial parameter estimates**

$$\text{delay: } \hat{\tau} = \mathrm{argmax}_{\tau_l^{\mathrm{v}}} \ \frac{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\hat{\theta})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \tau_l^{\mathrm{v}}}) \Big\|_F^2}{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\hat{\theta})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \tau_l^{\mathrm{v}}}) \Big\|_F^2}$$

$$\text{AoA: } \hat{\phi} = \mathrm{argmax}_{\phi_i^{\mathrm{v}}} \ \frac{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\mathbf{I}_{N_{\mathrm{t}}} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\phi_i^{\mathrm{v}}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\mathbf{I}_{N_{\mathrm{t}}} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\phi_i^{\mathrm{v}}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}$$

$$\text{AoD: } \hat{\theta} = \mathrm{argmax}_{\theta_j^{\mathrm{v}}} \ \frac{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\theta_j^{\mathrm{v}})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}{\Big\| \sum_{k\in\mathcal{K}_{\mathrm{cen}}} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\theta_j^{\mathrm{v}})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}$$

**Distortion compensation at side subcarriers**

$$\mathbf{r}[k] = \mathrm{diag}\big\{\mathrm{vec}(\mathbf{G}_k(\hat{\tau}, \hat{\phi}, \hat{\theta}))\big\}^{-1}\mathbf{r}[k], \ \forall k \in \mathcal{K}_{\mathrm{side}}$$

**Update parameter estimates using all subcarriers**

$$\text{delay: } \hat{\tau} = \mathrm{argmax}_{\tau_l^{\mathrm{v}}} \ \frac{\Big\| \sum_{k=0}^{K-1} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\hat{\theta})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \tau_l^{\mathrm{v}}}) \Big\|_F^2}{\Big\| \sum_{k=0}^{K-1} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\hat{\theta})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \tau_l^{\mathrm{v}}}) \Big\|_F^2}$$

$$\text{AoA: } \hat{\phi} = \mathrm{argmax}_{\phi_i^{\mathrm{v}}} \ \frac{\Big\| \sum_{k=0}^{K-1} (\mathbf{I}_{N_{\mathrm{t}}} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\phi_i^{\mathrm{v}}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}{\Big\| \sum_{k=0}^{K-1} (\mathbf{I}_{N_{\mathrm{t}}} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\phi_i^{\mathrm{v}}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}$$

$$\text{AoD: } \hat{\theta} = \mathrm{argmax}_{\theta_j^{\mathrm{v}}} \ \frac{\Big\| \sum_{k=0}^{K-1} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\theta_j^{\mathrm{v}})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{r}[k] e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}{\Big\| \sum_{k=0}^{K-1} (\overline{\breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{T},k}(\theta_j^{\mathrm{v}})} \otimes \breve{\mathbf{a}}^{\mathrm{v}}_{\mathrm{R},k}(\hat{\phi}))^* (\mathbf{\Phi}_{\mathrm{w}}^* e^{\mathrm{j}2\pi f_k \hat{\tau}}) \Big\|_F^2}$$

**Update the set of estimated parameters**: $\hat{L}^{(u)} = \hat{L}^{(u)} + 1, \ \hat{\boldsymbol{\varphi}}_l^{(u)} = \{\hat{\tau}, \hat{\phi}, \hat{\theta}\}$

**Algorithm 2** : DA-SWOMP-BS algorithm for sparse coding (continued)

**Update the path gains $\hat{\alpha}_l$ and coefficient vectors by solving**

$\sum_{k=0}^{K-1} \left\| \mathbf{y}_{\mathrm{w}}^{(u)}[k] - \mathbf{\Phi}_{\mathrm{w}} \sum_{l=1}^{\hat{L}^{(u)}} \alpha_l e^{-\mathrm{j}2\pi f_k \hat{\tau}_l} \mathrm{vec}(\mathbf{G}_k(\hat{\tau}_l, \hat{\phi}_l, \hat{\theta}_l)) \odot \overline{(\breve{\mathbf{a}}_{\mathrm{T},k}(\hat{\theta}_l) \otimes \breve{\mathbf{a}}_{\mathrm{R},k}(\hat{\phi}_l))} \right\|_2^2$

**Update the residual for each subcarrier**

$\mathbf{r}[k] = \mathbf{\Phi}_{\mathrm{w}}^* \big( \mathbf{y}_{\mathrm{w}}^{(u)}[k] - \mathbf{\Phi}_{\mathrm{w}} \sum_{l=1}^{\hat{L}^{(u)}} \hat{\alpha}_l e^{-\mathrm{j}2\pi f_k \hat{\tau}_l} \mathrm{vec}(\mathbf{G}_k(\hat{\tau}_l, \hat{\phi}_l, \hat{\theta}_l)) \odot \overline{(\breve{\mathbf{a}}_{\mathrm{T},k}(\hat{\theta}_l) \otimes \breve{\mathbf{a}}_{\mathrm{R},k}(\hat{\phi}_l))} \big)$

**Update the current MSE:** $\mathrm{MSE} = \frac{1}{ML_{\mathrm{r}}K} \sum_{k=0}^{K-1} \| (\mathbf{\Phi}_{\mathrm{w}}^*)^{\dagger} \mathbf{r}[k] \|_2^2$

**end while**

**end for**

**Output:** The estimated parameters $\hat{\boldsymbol{\varphi}}_l^{(u)} = \{\hat{\alpha}_l, \hat{\tau}_l, \hat{\phi}_l, \hat{\theta}_l\}$, and the coefficient vectors $\mathrm{vec}(\hat{\mathbf{\Omega}}^{(u)}[k])$ and $\boldsymbol{b}_l^{(u)}$, for $u \in \mathcal{I}(N_{\mathrm{sa}})$

**Algorithm 3** : Dictionary learning for hardware impairment (DLHW)

- **Input:** Initial channel estimates $\hat{\mathbf{h}}^{(u)}[k], \forall k = 0, \ldots, K-1, u \in \mathcal{I}(N_{\mathrm{sa}})$.
- **Initialization:** Set the dictionary matrices $\mathbf{D}_{\mathrm{R},1} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{r}}}$ and $\mathbf{D}_{\mathrm{T},1} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{t}}}$ using measurement data based on DIA algorithm [67], and set the dictionary matrices $\mathbf{D}_{\mathrm{T},2}$ and $\mathbf{D}_{\mathrm{R},2}$ as all-one matrices.
- **While** *not converge* **do**

  1. *Sparse coding stage*: Fixing all dictionaries, solve (3.4.10) using Algorithm 2 to update channel coefficients $\boldsymbol{\Omega}^{(u)}[k]$ and $\boldsymbol{b}_l^{(u)}$, as well as the path parameters $\hat{\boldsymbol{\varphi}}_{\hat{l}}^{(u)}$.

  2. *Dictionary update stage*: Fixing coefficients, update dictionaries as follows

     **While** *not converge* **do**

       Update $\mathbf{D}_{\mathrm{R},1}$ using (3.4.14) and update $\mathbf{D}_{\mathrm{T},1}$ similarly,

       Update $\mathbf{D}_{\mathrm{R},2}$ using (3.4.17) and update $\mathbf{D}_{\mathrm{T},2}$ similarly.

     **end while**

  **end while**
- **Output:** The optimal dictionaries $\mathbf{D}_{\mathrm{T},1}, \mathbf{D}_{\mathrm{T},2}, \mathbf{D}_{\mathrm{R},1}, \mathbf{D}_{\mathrm{R},2}$.

# Chapter 4

# Channel Tracking for Wideband mmWave MIMO Systems Without Blockage

In this chapter[1], we investigate the efficient wideband channel tracking and hybrid precoding design strategies for mmWave MIMO systems in the context of high mobility V2X scenarios. A major source of difficulty when operating with large arrays at mmWave frequencies is to estimate the wideband channel, since the use of hybrid architectures acts as a compression stage for the received signal. Moreover, the channel has to be tracked and the antenna arrays regularly reconfigured to obtain appropriate beamforming gains when a mobile setting is considered. In this chapter, we focus on the problem of channel tracking for frequency-selective mmWave channels, and propose two novel channel tracking algorithms. One of them exploits the sparsity of the mmWave channel, while the other one leverages prior statistical information about the channel parameters. We also propose a hybrid precoder and combiner design method to increase the received signal-to-noise ratio (SNR)

[1]This chapter was based on our published work [32]: N. González-Prelcic, H. Xie, J. Palacios, and T. Shimizu, "Wideband channel tracking and hybrid precoding for mmWave MIMO systems," IEEE Trans. Wireless Commun., vol. 20, no. 4, pp. 2161-2174, Oct. 2020. This work was supervised by Prof. González-Prelcic. My contributions lie in doing simulations and analyzing data, etc.

during channel tracking, such that near-optimum data rates can be obtained with low-overhead. In our numerical results, we analyze the performance of our proposed algorithms for different system parameters. Simulation results show that our proposed channel tracking algorithms are able to achieve near-optimum data rates outperforming state-of-the-art methods.

## 4.1 Introduction

MmWave MIMO is a key ingredient for the fifth generation of wireless communications, which has been standardized as 5G NR [1]. 5G NR considers an unified framework allowing an efficient use of sub-6 GHz and millimeter wave bands. The current version of the standard lacks, however, of an efficient procedure for configuring the antenna arrays. The proposed beam training protocol incurs in a high overhead, and it is not appropriate for high mobility scenarios [85].

Channel estimation for initial access is an effective alternative to beam training to configure mmWave antenna arrays. It has been studied in the literature for narrowband [3, 7, 8], and frequency-selective channels [5, 6, 9–12], yet little attention has been drawn towards the problem of wideband channel tracking. An interesting feature of the new 5G NR channel model [13] is that it incorporates spatial consistency. This way, a model for the temporal evolution of different channel parameters can be leveraged to design effective strategies for channel tracking. The problem of channel tracking under a narrow band communication model has been considered in recent work

[14–21], although the mmWave channel is frequency-selective. Besides this unrealistic assumption, prior work on channel tracking considers different ad-hoc mathematical approaches to model the channel dynamics, that is, how the AoA, AoD, and channel gains evolve with time. Further, some prior work [21] considers the tracking problem only for a narrow band, multiple-input single-output (MISO) scenario, with single-antenna users, and considering channel vectors comprising of only two multipath components. For instance, prior work considering time correlation for the channel gains assumes a first-order Gauss-Markov process with very large correlation factor (i.e., 0.995), and makes no distinction between the correlation of line-of-sight (LOS) and non line-of-sight (NLOS) multipath components [15], [16], [86], [22], while the spatial correlation of channel gains has been shown to be different for LOS and NLOS components [87], the latter exhibiting lower correlation values. As to the AoA and AoD, prior work considers that the angles evolve according to a Gaussian distribution (which is a mathematical artifact to enable the use of a linearized Kalman filter (KF)) [86], [20], a uniform distribution [15], [16], or deterministically [22], while these parameters have been shown to follow a Laplacian distribution [88]. Finally, the spatial consistency model proposed in 5G NR differs in one or another sense from all these other models.

Tracking a frequency selective channel is significantly more challenging than tracking a narrowband channel. A frequency selective channel, in the context of a MIMO-OFDM millimeter wave system for example, is modeled in the frequency domain by a number of MIMO channel matrices equal to

the number of subcarriers, while a narrowband channel can be represented with a single matrix. This means that to operate in a frequency selective mmWave channel, a large number of channel matrices have to be tracked. A frequency-domain narrowband channel tracking algorithm could be applied in parallel to every subcarrier, but this is highly inefficient, because the spatial correlation of the channel matrices corresponding to different subcarriers would not be exploited. Alternatively, the problem can be analyzed considering a time domain representation of the received signal and the channel. In this case, the MIMO channel is represented by a number of MIMO channel matrices, each one associated to a given channel delay. Tracking these time domain channel matrices is equivalent to estimate the time evolution of the small scale parameters (gain, delay, AoA, AoD) associated to every multipath component in each channel matrix for all possible delays. The specific number of delays necessary to represent a given mmWave channel depends on the propagation scenario and the bandwidth, but it is also large. For example, the IEEE 802.11ad standard for Wi-Fi at 60 GHz considers a cyclic prefix length of 128, which means that channels composed from up to 128 delay taps are being considered.

To the best of our knowledge, only a few works deal with the problem of channel tracking assuming a *hybrid mmWave MIMO system*, and considering *frequency-selective single-user scenarios* [22], [23], [24], [25]. The limitations of [22] are: 1) the considered channel model is not band-limited, thus leading to both unrealistic and sparser channel realizations; 2) the channel gains are

assumed to be time-invariant, which does not hold under realistic tracking periodicities. In our prior work in [23], we developed a gradient algorithm for channel tracking that approximates the maximum likelihood (ML) estimator conditioned on a certain sparsity level, which was estimated during initial channel acquisition. Since the angular resolution of an $N$-element ideal antenna array is limited by $1/N$, two multipath components arriving with an angular separation smaller than this limit cannot be distinguished. Therefore, the ML estimator cannot be found in general owing to clustering and the use of finite-resolution antenna arrays. The main limitations of [23] are the high complexity, caused by the method used to design the hybrid precoders and combiners during tracking, and the still high tracking overhead. The drawbacks of [24] are: 1) that the strategy exploits position information at the user side, which is not always available; and 2) it does not specify a clear design of the precoder and combiners to be used during tracking. The design in [25] focuses on tracking the gains, assuming that the AoA/AoD are time invariant, which does not hold in practical scenarios when considering mobility.

## 4.2 Contributions

In this chapter, we focus on the problem of channel tracking for frequency-selective mmWave MIMO systems under the channel model used for 5G NR. We focus on the single-user scenario with transceivers being equipped with hybrid MIMO architectures [7], although the proposed channel tracking framework can be extended to the multi-user MIMO (MU-MIMO) sce-

nario similarly to [12]. We propose two novel algorithms to perform channel tracking at mmWave, considering both *compressive* and *Bayesian* solutions, suitable for different type of scenarios. The compressive solution is appropriate for general scenarios, where the only assumption that can be made about the mmWave communication channel is that it is sparse in the angular domain, as shown in different measurements campaigns [89] and exploited in previous work [17, 22, 23]. The Bayesian approach can be used when the statistical priors on the channel parameters are known or can be learnt. Statistical information has already been exploited in previous work on narrowband channel tracking [14–16, 18–21]. If not available, it could be easily obtained in some particular settings. For example, in vehicular communications, machine learning techniques could be applied to learn the distributions for the angular parameters, as in [90]. The proposed algorithms allow not only finding accurate estimates of the mmWave MIMO channel, but they also result in very low training overhead. Consequently, they enable us to obtain high spectral efficiency values, even when mobility is high and the distance between transmitter and receiver is small, thereby making the AoA and AoD harder to track. Further, we also focus on the problem of precoder and combiner design for channel tracking, whereby our interest is in achieving near-optimum data rates while keeping low overhead and low computational complexity as well. The detailed contributions of this chapter are listed hereafter:

- We formulate the channel tracking problem for a single-user frequency-selective mmWave MIMO system. We present a *general* formulation of

the time varying received signal, in which prior statistical information on the different small scale parameters may or may not be available.

- We propose a method to design hybrid precoders and combiners for channel tracking exploiting prior angular information from the channel estimates at the immediately preceding channel slot. Further, we prove that this method maximizes the received SNR.

- We derive a low complexity channel tracking algorithm building upon our prior work in [23]. We rely on an off-grid representation of the MIMO channel to obtain the *sparsity-constrained* ML estimator for the AoAs/AoDs. The channel gains for each channel slot are assumed uncorrelated, and are obtained by deriving the minimum variance unbiased estimator (MVUE).

- We develop a second algorithm that follows a Bayesian approach to finding the *sparsity-constrained* minimum mean squared error (MMSE) estimator of the mmWave MIMO channel matrices when prior information on the statistics of the small scale parameters is available. The priors related to the average number of clusters, average number of rays per cluster, and angular parameters are taken from the 5G channel model [13], while the statistical knowledge about the channel gains and their evolution has been proposed in previous work [16], [86], [22], [87].

- We obtain numerical results that show how our proposed strategies are capable to track the mmWave channel, outperforming state-of-the-art

techniques under different propagation conditions and system parameters.

## 4.3 System and channel models

In this section, we introduce the models and assumptions for the different blocks of the communication system considered in this chapter.

### 4.3.1 System model

We consider a single-user MIMO-OFDM communication link between a transmitter and a receiver equipped with $N_{\mathrm{t}}$ and $N_{\mathrm{r}}$ antennas, and $L_{\mathrm{t}}$ and $L_{\mathrm{r}}$ radio frequency (RF) chains. The system has to transmit $N_{\mathrm{s}}$ data streams. Both transceivers are assumed to use a fully-connected hybrid architecture [7]. We will use the super-index $n$ to denote the $n$-th channel slot, which is defined as the time window during which the channel may be considered to be constant.

Within a channel slot, there are two different transmission stages as shown in Fig. 4.1: i) a training phase to acquire CSI in order to configure the hybrid antenna arrays; and ii) a data phase to communicate data vectors from the transmitter to the receiver. The training/tracking phase consists of the transmission of several OFDM training (pilot) symbols, known at both the transmitter and the receiver. The received pilot symbols are used at the receiver side to estimate the channel (initial estimate) or the channel variations (channel tracking), which are then fed-back to the transmitter through the control channel. More specifically, we consider a tracking algorithm that

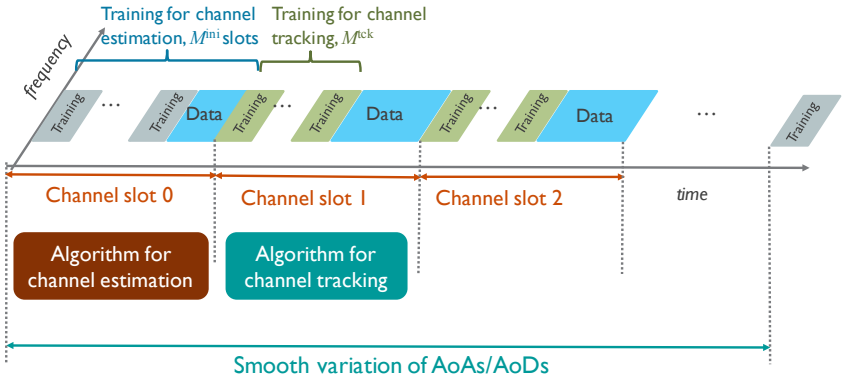



Figure 4.1: Illustration of how mmWave channel acquisition is obtained on a frame-by-frame basis, encompassing both a training phase and a data transmission phase.

provides a frequency domain estimation of the AoAs/AoDs and channel gains. During initial acquisition ($n = 0$), the training phase lasts longer than during tracking ($n \geq 1$), owing to the lack of prior information about the channel. In practical mmWave MIMO links, variations are expected between two consecutive slots, and each slot should be long enough to accommodate both the training and data phases.

During the data phase of the $n$-th channel slot, the transmitter uses a hybrid precoder $\mathbf{F}^{(n)}[k] \in \mathbb{C}^{N_{\text{t}} \times N_{\text{s}}}$ for the $k$-th subcarrier, $k = 0, \ldots, K-1$, to transmit a $N_{\text{s}} \times 1$ vector of data streams $\mathbf{s}^{(n)}[k]$, in which $N_{\text{s}}$ denotes the number of data streams to be transmitted. $\mathbf{F}^{(n)}[k] = \mathbf{F}_{\text{RF}}^{(n)} \mathbf{F}_{\text{BB}}^{(n)}[k]$, with $\mathbf{F}_{\text{RF}}^{(n)} \in \mathbb{C}^{N_{\text{t}} \times L_{\text{t}}}$ the analog precoder, common for all subcarriers, and $\mathbf{F}_{\text{BB}}^{(n)}[k] \in \mathbb{C}^{L_{\text{t}} \times N_{\text{s}}}$ the baseband precoder for the $k$-th subcarrier. We denote the transmit power as $P_{\text{tx}}$, and we assume that $\sum_{k=0}^{K-1} \|\mathbf{F}^{(n)}[k]\|_F^2 = P_{\text{tx}}$. Likewise, the receiver uses

a hybrid combiner $\mathbf{W}^{(n)}[k] \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{s}}$, $\mathbf{W}^{(n)}[k] = \mathbf{W}_{\mathrm{RF}}^{(n)} \mathbf{W}_{\mathrm{BB}}^{(n)}[k]$, with $\mathbf{W}_{\mathrm{RF}}^{(n)} \in \mathbb{C}^{N_\mathrm{r} \times L_\mathrm{r}}$ the analog combiner, and $\mathbf{W}_{\mathrm{BB}}^{(n)}[k] \in \mathbb{C}^{L_\mathrm{r} \times N_\mathrm{s}}$ the baseband combiner at the $k$-th subcarrier. If the mmWave MIMO channel for a given subcarrier $k$ and slot $n$ is denoted as $\mathbf{H}^{(n)}[k]$, the received signal during transmission of a data stream $\mathbf{s}^{(n)}[k] \in \mathbb{C}^{N_\mathrm{s} \times 1}$ is given by

$$\mathbf{y}^{(n)}[k] = \mathbf{W}_{\mathrm{BB}}^{(n)*}[k] \mathbf{W}_{\mathrm{RF}}^{(n)*} \mathbf{H}^{(n)}[k] \mathbf{F}_{\mathrm{RF}}^{(n)} \mathbf{F}_{\mathrm{BB}}^{(n)}[k] \mathbf{s}^{(n)}[k] + \mathbf{n}^{(n)}[k], \tag{4.3.1}$$

where the $N_\mathrm{s} \times 1$ vector $\mathbf{n}^{(n)}[k]$ is the complex-valued additive Gaussian-distributed received noise, $\mathbf{n}^{(n)}[k] \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{C}_\mathrm{w}^{(n)}[k])$, with $\mathbf{C}_\mathrm{w}^{(n)}[k]$ the noise covariance matrix given by $\mathbf{C}_\mathrm{w}^{(n)}[k] = \mathbf{W}_{\mathrm{BB}}^{(n)*}[k] \mathbf{W}_{\mathrm{RF}}^{(n)*} \mathbf{W}_{\mathrm{RF}}^{(n)} \mathbf{W}_{\mathrm{BB}}^{(n)}[k]$.

For the training phase, we need to sound the channel with a different combination of training precoder and combiner for every OFDM pilot. The training precoders and combiners for tracking are designed by the transmitter and receiver based on the CSI corresponding to the previous channel slot, which is assumed to be already available at both ends of the communication system. The specific design for training precoders and combiners, is derived in Section 4.4, and assumes frequency-flat hybrid filters. The reasons for this choice are the following: i) the analog precoder/combiner is frequency-flat due to hardware constraints, and is the one that captures the main angular characteristics of the channel; ii) to estimate the channel gains, matrix inverses (pseudoinverses) need to be computed at the receiver, so that the use of frequency-flat precoders and combiners allow computing these gains by using a single matrix inverse, thereby reducing computational complexity

during tracking, and iii) it reduces memory storage needs at both transmitter and receiver. Thus, during tracking, the transmitter uses a hybrid precoder $\mathbf{F}_{\mathrm{tr}}^{(m,n)} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{s}}}$ to transmit the $m$-th OFDM pilot. Such a precoder can be expressed as $\mathbf{F}_{\mathrm{tr}}^{(m,n)} = \mathbf{F}_{\mathrm{tr,RF}}^{(m,n)} \mathbf{F}_{\mathrm{tr,BB}}^{(m,n)}$, with $\mathbf{F}_{\mathrm{tr,RF}}^{(m,n)} \in \mathbb{C}^{N_{\mathrm{t}} \times L_{\mathrm{t}}}$ the training analog precoder, and $\mathbf{F}_{\mathrm{tr,BB}}^{(m,n)} \in \mathbb{C}^{L_{\mathrm{t}} \times N_{\mathrm{s}}}$ the training baseband precoder. We assume again that $\sum_{k=0}^{K-1} \|\mathbf{F}_{\mathrm{tr}}^{(m,n)}\|_F^2 = P_{\mathrm{tx}}$. Similarly to the transmitter, the training combiner can be expressed as $\mathbf{W}_{\mathrm{tr}}^{(m,n)} = \mathbf{W}_{\mathrm{tr,RF}}^{(m,n)} \mathbf{W}_{\mathrm{tr,BB}}^{(m,n)}$, with $\mathbf{W}_{\mathrm{tr,RF}}^{(m,n)} \in \mathbb{C}^{N_{\mathrm{r}} \times L_{\mathrm{r}}}$ the training analog combiner, and $\mathbf{W}_{\mathrm{tr,BB}}^{(m,n)} \in \mathbb{C}^{L_{\mathrm{r}} \times N_{\mathrm{s}}}$ the training baseband combiner at the $k$-th subcarrier. The $m$-th pilot symbol to be sent during training at the $n$-th channel slot is denoted as $\mathbf{s}_{\mathrm{tr}}^{(m,n)}[k] \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$. The received signal during training is then given by

$$\mathbf{y}_{\mathrm{tr}}^{(m,n)}[k] = \mathbf{W}_{\mathrm{tr}}^{(m,n)*} \mathbf{H}^{(n)}[k] \mathbf{F}_{\mathrm{tr}}^{(m,n)} \mathbf{s}_{\mathrm{tr}}^{(m,n)}[k] + \mathbf{n}^{(m,n)}[k], \tag{4.3.2}$$

where the $N_{\mathrm{s}} \times 1$ noise vector $\mathbf{n}^{(m,n)}[k]$ follows the same distribution as in the data phase, but with a covariance matrix depending on the training combiner.

### 4.3.2 Frequency-selective channel model

The $d$-th delay tap of the mmWave MIMO channel matrix corresponding to the $n$-th channel slot is defined using a clustered channel model [91] as follows. Let $D \in \mathbb{N}$ be the channel delay tap length, $\alpha_{c,r}^{(n)} \in \mathbb{C}$ be the complex gain of the $r$-th ray within the $c$-th cluster, $\phi_{c,r}^{n}$, $\theta_{c,r}^{n} \in \mathbb{R}$ be the AoA and AoD, $\tau_{c,r}^{n} \in \mathbb{R}$ be the time-delay, and $p(\tau)$ be the equivalent response of transmit and receive pulse-shapes and other analog filtering evaluated at $\tau$. The total number of clusters is denoted by $C \in \mathbb{N}$, and the $c$-th cluster consists

of $R_c \in \mathbb{N}$ rays. Last, let $\mathbf{a}_{\mathrm{T}}\left(\theta_{c,r}^{n}\right) \in \mathbb{C}^{N_{\mathrm{t}}\times 1}$ and $\mathbf{a}_{\mathrm{R}}\left(\phi_{c,r}^{n}\right) \in \mathbb{C}^{N_{\mathrm{r}}\times 1}$ be the transmit and receive array steering vectors evaluated on the AoD and AoA of each corresponding path. Then, the $N_{\mathrm{r}} \times N_{\mathrm{t}}$ channel matrix at delay tap $d = 0, \ldots, D-1$ is given by [91]

$$\mathbf{H}^{(n)}[d] = \sum_{c=1}^{C}\sum_{r=1}^{R_c} \alpha_{c,r}^{(n)} p\left(dT_{\mathrm{s}} - \tau_{c,r}^{(n)}\right) \times \mathbf{a}_{\mathrm{R}}\left(\phi_{c,r}^{(n)}\right) \mathbf{a}_{\mathrm{T}}^{*}\left(\theta_{c,r}^{(n)}\right). \tag{4.3.3}$$

From (4.3.3), we may express the MIMO channel in the frequency domain, whereby taking the $K$-point DFT of (4.3.3) yields

$$\mathbf{H}^{(n)}[k] = \sum_{d=0}^{D-1} \mathbf{H}^{(n)}[d] e^{-j\frac{2\pi kd}{K}} = \mathbf{A}_{\mathrm{R}}\left(\boldsymbol{\phi}^{(n)}\right) \mathbf{G}^{(n)}[k] \mathbf{A}_{\mathrm{T}}^{*}\left(\boldsymbol{\theta}^{(n)}\right), \tag{4.3.4}$$

where $\mathbf{A}_{\mathrm{T}}\left(\boldsymbol{\theta}^{(n)}\right) \in \mathbb{C}^{N_{\mathrm{t}}\times\sum_{c=1}^{C} R_c}$, $\mathbf{A}_{\mathrm{R}}\left(\boldsymbol{\phi}^{(n)}\right) \in \mathbb{C}^{N_{\mathrm{r}}\times\sum_{c=1}^{C} R_c}$ denote the transmit and receive antenna array responses, and $\mathbf{G}^{(n)}[k] \in \mathbb{C}^{\sum_{c=1}^{C} R_c \times \sum_{c=1}^{C} R_c}$ is the diagonal matrix containing the complex channel gains. Notice that we have neglected the beam squint effect in the channel model above. From the results in [92], [93], we can conclude that this effect is not critical when operating at frequencies in the order of 60 GHz and bandwidths of 1-3 GHz. For scenarios where the system parameters lead to a more noticeable beam squint effect, we have shown that it can be effectively handled by using a simple strategy, which consists of subcarrier grouping [92], [93]. The same idea could be applied to the channel tracking problem, so that any of the algorithms proposed in this chapter could run in parallel for several (few) subgroups of sucarriers.

## 4.4 Hybrid precoder and combiner design for channel tracking

In this section, we propose a method to design the training hybrid precoders and combiners for use during channel track for a single training step $m$, $1 \leq m \leq M_{\text{tck}}$ during the $n$-th channel slot. This is different than designing the training precoders and combiners for single shot channel estimation, as the entire angular space does not need to be explored. Channel tracking precoders are also different than those used for communication, which has the objective of maximizing spectral efficiency. Our design maximizes average received SNR, while allowing certain incoherence in the measurement process, i.e., making it possible to compressively estimate and update the channel parameters using a small number of measurements.

We make four assumptions about the precoders (and combiners). (A1) The precoder is frequency flat. The main reason, besides the complexity consideration discussed in Section 4.3, is that hybrid precoding in frequency selective channels involves a higher dimension frequency flat component common for all subcarriers and a lower dimension frequency selective component. The frequency flat component captures the main angular characteristics of the channel, which need to be tracked, and is generally the most important and challenging component to find. With this assumption, we write the precoder as $\mathbf{F}_{\text{tr}}^{(m,n)}$ as described in previous section, since the $k$ index is not needed. (A2) The number of training data streams is set as $N_{\text{s}} = L_{\text{t}}$ to exploit all the possible degrees of freedom provided by the hybrid architecture when sound-

ing the channnel. (A3) The precoder represents a subspace, which means that only the subspace spanned by the $L_{\mathrm{t}}$ columns of the $N_{\mathrm{t}} \times L_{\mathrm{t}}$ matrix $\mathbf{F}_{\mathrm{tr}}^{(m,n)}$ needs to be found. This is known as Grassmannian precoding. Expressed computationally, the columns of $\mathbf{F}_{\mathrm{tr}}^{(m,n)}$ are an orthonormal basis for the subspace. Then $\mathbf{F}_{\mathrm{tr}}^{(m,n)*}\mathbf{F}_{\mathrm{tr}}^{(m,n)} = \mathbf{I}$. Further, because the Grassmann manifold is quotient space of the Stiefel manifold, for any unitary $\mathbf{Q}$, $\mathbf{F}_{\mathrm{tr}}^{(m,n)}\mathbf{Q}$ is equivalent to $\mathbf{F}_{\mathrm{tr}}^{(m,n)}$ in the sense that both refer to the same subspace. Grassmannian precoding is widely used in MIMO communication systems, though normally for information transmission not training. (A4) The precoder is fully digital. At the end, that precoder is factorized into RF and BB components. The same assumptions also hold for the $N_{\mathrm{r}} \times L_{\mathrm{r}}$ combiner $\mathbf{W}_{\mathrm{tr}}^{(m,n)}$.

Let us consider the channel estimates obtained at the end of the $(n-1)$-th channel slot, denoted by $\hat{\mathbf{H}}^{(n-1)}[k]$, $0 \leq k \leq K-1$. We use this estimate to find the training precoder $\mathbf{F}_{\mathrm{tr}}^{(m,n)}$ and training combiner $\mathbf{W}_{\mathrm{tr}}^{(m,n)}$ for the next channel slot. As a performance measure, we consider an average received SNR as a function of a given precoder $\mathbf{F}$, $\mathbf{W}$, and $\hat{\mathbf{H}}^{(n-1)}[k]$

$$\begin{aligned}\mathrm{SNR}_{\mathrm{tr}}(\mathbf{F},\mathbf{W}|\hat{\mathbf{H}}^{(n-1)}[k]) &= \frac{1}{\sigma^2 K}\sum_{k=0}^{K-1}\frac{\|\mathbf{W}^*\hat{\mathbf{H}}^{(n-1)}[k]\mathbf{F}\|_F^2}{\|\mathbf{W}\mathbf{W}^*\|_F^2} \\ &\overset{(a)}{=} \frac{1}{\sigma^2 K L_{\mathrm{r}}}\sum_{k=0}^{K-1}\mathrm{tr}\left\{\mathbf{W}^*\hat{\mathbf{H}}^{(n-1)}[k]\mathbf{F}\mathbf{F}^*\hat{\mathbf{H}}^{(n-1)*}[k]\mathbf{W}\right\}\end{aligned} \tag{4.4.1}$$

where (a) results from using the definition of the Frobenius norm and the fact that the combiner represents an orthonormal basis. Pulling out the constant

combiner term gives

$$\mathrm{SNR}_{\mathrm{tr}}(\mathbf{F},\mathbf{W}|\hat{\mathbf{H}}^{(n-1)}[k]) = \frac{1}{\sigma^2 K L_{\mathrm{r}}}\mathrm{tr}\left\{\mathbf{W}^* \underbrace{\left(\sum_{k=0}^{K-1}\hat{\mathbf{H}}^{(n-1)}[k]\mathbf{F}\mathbf{F}^*\hat{\mathbf{H}}^{(n-1)*}[k]\right)}_{\mathbf{R}_{\mathrm{F}}(\mathbf{F})}\mathbf{W}\right\}, \tag{4.4.2}$$

while exploiting $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$ and pulling out the precoder gives another result

$$\mathrm{SNR}_{\mathrm{tr}}(\mathbf{F},\mathbf{W}|\hat{\mathbf{H}}^{(n-1)}[k]) = \frac{1}{\sigma^2 K L_{\mathrm{r}}}\mathrm{tr}\left\{\mathbf{F}^* \underbrace{\left(\sum_{k=0}^{K-1}\hat{\mathbf{H}}^{(n-1)*}[k]\mathbf{W}\mathbf{W}^*\hat{\mathbf{H}}^{(n-1)}[k]\right)}_{\mathbf{R}_{\mathrm{W}}(\mathbf{W})}\mathbf{F}\right\}. \tag{4.4.3}$$

We will use (4.4.2) to solve for $\mathbf{W}$ and (4.4.3) to solve for $\mathbf{F}$ in an iterative fashion.

Joint optimization of $\mathrm{SNR}_{\mathrm{tr}}(\mathbf{F},\mathbf{W})$ is challenging due to the frequency flatness of the precoders and combiners. An alternative is to exploit alternating maximization, which is in the same vein as alternating minimization [94], which is guaranteed to converge. Let $q$ denote the q-th step of the optimization. Initialize the precoders and combiners with random semi-unitary matrices to give $\mathbf{W}_0$ and $\mathbf{F}_0$. This may be done by generating for the precoder for example, an $N_{\mathrm{t}} \times N_{\mathrm{t}}$ matrix with IID $\mathcal{CN}(0,1)$ entries, taking the SVD, and choosing the first $L_{\mathrm{t}}$ left singular vectors. For $q > 0$, given $\mathbf{W}_{q-1}$ and $\mathbf{F}_{q-1}$, find the precoders and combiners as

$$\mathbf{F}_q = \underset{\mathbf{A}:\ \mathbf{A}^*\mathbf{A}=\mathbf{I}_{L_{\mathrm{t}}}}{\mathrm{argmax}}\ \mathrm{SNR}_{\mathrm{tr}}\left(\mathbf{A},\mathbf{W}_{q-1}|\hat{\mathbf{H}}^{(n-1)}[k]\right) \tag{4.4.4}$$

$$\mathbf{W}_q = \underset{\mathbf{A}:\ \mathbf{A}^*\mathbf{A}=\mathbf{I}_{L_{\mathrm{r}}}}{\mathrm{argmax}}\ \mathrm{SNR}_{\mathrm{tr}}\left(\mathbf{F}_q,\mathbf{A}|\hat{\mathbf{H}}^{(n-1)}[k]\right). \tag{4.4.5}$$

The form of the solution is similar in both cases and follows from linear algebra. The precoder $\mathbf{F}_q$ is a basis of the dominant $L_\mathrm{t}$ singular vectors of $\mathbf{R}_\mathrm{W}(\mathbf{W}_{q-1})$, while the combiner is a basis of the dominant $L_\mathrm{r}$ singular vectors of $\mathbf{R}_\mathrm{F}(\mathbf{F}_q)$. The algorithm may be stopped after a certain number of steps or once an error residual $\mathrm{SNR}_\mathrm{tr}(\mathbf{F}_q, \mathbf{W}_q|\hat{\mathbf{H}}^{(n-1)}[k]) - \mathrm{SNR}_\mathrm{tr}(\mathbf{F}_{q-1}, \mathbf{W}_{q-1}|\hat{\mathbf{H}}^{(n-1)}[k])$ falls below a target threshold. Similar alternating algorithms have been used in the past in the context of interference alignment, see e.g. [95], though with the objective minimized (taking the part corresponding to the smaller singular values). Assuming that the iteration stops at step $Q$, we define $\mathbf{F}_Q$ and $\mathbf{W}_Q$ as the solution provided by the algorithm. This solution is not unique. Another basis generated as a rotation of the initial one is also a basis for the same subspace. This fact let us to easily introduce some randomness in the final design for the precoder/combiner, which is interesting when estimating sparse vectors. Thus, the all-digital optimal precoders for training step $m$ and channel slot $n$ are defined as $\mathbf{F}_\mathrm{tr}^{(m,n)} = \mathbf{F}_Q\mathbf{\Omega}_\mathrm{F}^{(m)}$ and $\mathbf{W}_\mathrm{tr}^{(m,n)} = \mathbf{W}_Q\mathbf{\Omega}_\mathrm{W}^{(m)}$, where $\mathbf{\Omega}_\mathrm{F}^{(m)}$ and $\mathbf{\Omega}_\mathrm{W}^{(m)}$ are random rotation matrices. Note that we need to define a different pair of rotation matrices for every pilot symbol to be transmitted/received in a given channel slot to reduce the coherence of the measurement.

Finally, the fully digital assumption is dismissed and the hybrid constraint is imposed. Each matrix is decomposed into RF and baseband components using the hybrid factorization in [96], which has been shown to provide good performance and low complexity. Per-antenna power constraints could be included by including an extra term in the alternating maximization (along

the lines of [97, Algorithm 2]) or by using a suitable hybrid algorithm [98].

## 4.5 Channel tracking algorithms

In this section, we propose novel algorithms for tracking the variations of the mmWave MIMO channel. The main assumptions we make are summarized as follows:

- We assume that the number of paths is invariant along successive slots, i.e., there are no abrupt changes and only tracking is performed. If cluster blockage is present, the system needs to fully re-estimate the channel again. In this manuscript, we only focus on solving the problem of configuring the transmit and receive antenna arrays during channel slots that meet the spatial consistency model defined in [13].

- We assume that the AoAs and AoDs vary relatively smoothly between any two consecutively slots, such that estimates of these parameters obtained during the $(n-1)$-th slot can be used as prior information to refine those at the $n$-th slot. This assumption is reasonable if the mobile terminal is not moving extremely fast (i.e. less than 400 km/h), it has been verified in measurement campaigns [89], and it is part of the spatial consistency model defined in the 5G channel model [13, Sec. 7.6.3]. This assumption has been leveraged in all prior work on channel tracking [14–21], and it is a built-in feature of the QuaDRiGa [99] channel simulator, that we will use in our numerical experiments to generate test

channels.

### 4.5.1 Problem formulation

During the training phase of the $n$-th channel slot, $M_{\text{tck}}$ training frames are forwarded to the receiver. The received signal during training in (4.3.2) can be rewritten as

$$\mathbf{y}_{\text{tr}}^{(m,n)}[k] = \underbrace{\left(\mathbf{s}_{\text{tr}}^{(m,n)T}[k]\mathbf{F}_{\text{tr}}^{(m,n)T} \otimes \mathbf{W}_{\text{tr}}^{(m,n)*}\right)}_{\mathbf{\Phi}^{(m,n)}[k]} \text{vec}\{\mathbf{H}^{(n)}[k]\} + \mathbf{n}^{(m,n)}[k], \quad m = 1, \ldots, M_{\text{tck}}, \tag{4.5.1}$$

with $\mathbf{\Phi}^{(m,n)}[k] \in \mathbb{C}^{N_{\text{s}} \times N_{\text{t}} N_{\text{r}}}$. Now, owing to dimensionality reduction at the RF stage of both transmitter and receiver, it may be necessary to acquire several measurements to track the variations of the mmWave frequency-selective channel. Hence, we define

$$\mathbf{y}^{(n)}[k] \triangleq [\mathbf{y}_{\text{tr}}^{(1,n)T}[k], \ldots, \mathbf{y}_{\text{tr}}^{(M_{\text{tck}},n)T}[k]]^T \in \mathbb{C}^{N_{\text{s}} M_{\text{tck}} \times 1}, \tag{4.5.2}$$

$$\mathbf{\Phi}^{(n)}[k] \triangleq \begin{bmatrix} \mathbf{\Phi}^{(1,n)}[k] \\ \mathbf{\Phi}^{(2,n)}[k] \\ \vdots \\ \mathbf{\Phi}^{(M_{\text{tck}},n)}[k] \end{bmatrix} \in \mathbb{C}^{N_{\text{s}} M_{\text{tck}} \times N_{\text{t}} N_{\text{r}}}, \tag{4.5.3}$$

$$\mathbf{n}^{(n)}[k] \triangleq [\mathbf{n}^{(1,n)T}[k], \ldots, \mathbf{n}^{(M_{\text{tck}},n)T}[k]]^T \in \mathbb{C}^{N_{\text{s}} M_{\text{tck}} \times 1} \tag{4.5.4}$$

to extend (4.5.1) for $M_{\text{tck}}$ received training OFDM symbols as

$$\mathbf{y}^{(n)}[k] \approx \mathbf{\Phi}^{(n)}[k]\text{vec}\{\mathbf{H}^{(n)}[k]\} + \mathbf{n}^{(n)}[k], \tag{4.5.5}$$

The approximation comes from neglecting the beam squint effect and defining $\text{vec}\{\mathbf{H}^{(n)}[k]\}$ as

$$\text{vec}\{\mathbf{H}^{(n)}[k]\} \approx \underbrace{\left(\mathbf{A}_{\text{T}}^{\text{C}}\left(\boldsymbol{\theta}^{(n)}\right) \star \mathbf{A}_{\text{R}}\left(\boldsymbol{\phi}^{(n)}\right)\right)}_{\boldsymbol{\Psi}\left(\boldsymbol{\theta}^{(n)},\boldsymbol{\phi}^{(n)}\right)} \times \underbrace{\text{vec}\{\text{diag}\{\mathbf{G}^{(n)}[k]\}\}}_{\mathbf{g}^{(n)}[k]}, \tag{4.5.6}$$

where the AoAs/AoDs do not depend on the subcarrier index. Therefore, (4.5.5) can be expressed in terms of the unknown AoAs/AoDs, common for all the subcarriers, and the frequency selective channel gains as

$$\mathbf{y}^{(n)}[k] \approx \boldsymbol{\Phi}^{(n)}[k]\boldsymbol{\Psi}\left(\boldsymbol{\theta}^{(n)},\boldsymbol{\phi}^{(n)}\right)\mathbf{g}^{(n)}[k] + \mathbf{n}^{(n)}[k]. \tag{4.5.7}$$

The signal model in (4.5.7) can be used to derive different channel tracking algorithms. Nonetheless, using frequency-dependent training symbols $\mathsf{s}^{(m,n)}[k]$ leads to algorithms with high computational complexity. Then, to simplify subsequent calculations, we may express the training streams during tracking $\mathsf{s}^{(m,n)}[k]$ as $\mathsf{s}^{(m,n)}[k] = \mathbf{q}^{(m,n)}\mathsf{s}^{(m,n)}[k]$, where $\mathbf{q}^{(m,n)} \in \mathbb{C}^{N_{\text{s}}\times 1}$ is a frequency-flat spatial modulation vector consisting of independent and identically distributed energy-normalized QPSK constellation symbols, and $\mathsf{s}^{(m,n)}[k]$ is a frequency-dependent constellation symbol (i.e. QPSK) whose effect can be eliminated at the receiver by simply multiplying $\mathbf{y}^{(n)}[k]\mathsf{s}^{(m,n)-1}[k]$, without altering noise statistics. Further, as shown in [5], the use of $\mathbf{q}^{(m,n)}$ allows exploiting the $L_{\text{t}}$ degrees of freedom coming from the transmit hybrid MIMO architecture. Thereby, with this particular structure for the training sequence, $\boldsymbol{\Phi}^{(n)}$ is also frequency flat, and (4.5.7) can be simplified to

$$\mathbf{y}^{(n)}[k] = \boldsymbol{\Phi}^{(n)}\boldsymbol{\Psi}\left(\boldsymbol{\theta}^{(n)},\boldsymbol{\phi}^{(n)}\right)\mathbf{g}^{(n)}[k] + \mathbf{n}^{(n)}[k]. \tag{4.5.8}$$

### 4.5.2 Generalized marginalized particle filter (GMPF)

In this subsection, we present a new channel tracking algorithm. This strategy exploits prior statistical knowledge about the AoA, AoD, and channel gains so it is possible to obtain the minimum mean squared estimator for the channel matrix using a Bayesian approach. As long as the channel model can be accurately described using a clustered channel model, the proposed algorithm is directly applicable regardless of the particular statistics of the AoA/AoD. As for the channel gains, these must remain Gaussian, assumption that has been adopted in most of prior work. In summary, this Bayesian procedure can be applied to any channel model if we know the distribution of the channel parameters

To track off-grid variations of the AoA and AoD, some prior knowledge on the sparsity level (number of significant multipath components) is required. Channel models at mmWave are usually defined in terms of average number of clusters and average number of rays per cluster [13], for which there is prior statistical knowledge. Therefore, to devise our Bayesian tracking strategy, we consider that, on average, there are $\mu_L = C\mu_{R_c}$ multipath components in the channel, with $\mu_{R_c}$ the average number of rays per cluster and $C$ the average number of clusters. Notice that we do not consider that the total number of paths is known, but only the statistics on the number of paths is assumed to be known. This assumption is properly justified as these statistics are given in the 5G NR channel model.

We also consider that there is prior information about the distributions

of the AoAs and AoDs. We define $\boldsymbol{\xi}^{(n)} = [\boldsymbol{\theta}^{(n)T}, \boldsymbol{\phi}^{(n)T}]^T, \in \mathbb{R}^{2\mu_L \times 1}$ as the set of AoAs and AoDs for the $n$-th channel slot, and $\boldsymbol{\xi}^{(n-1)}$ the set of AoAs and AoDs for the $(n-1)$-th channel slot. In this chapter, we assume, as in the 5G NR channel model, that the conditional density function $p(\boldsymbol{\xi}^{(n)}|\boldsymbol{\xi}^{(n-1)})$ follows a Laplacian distribution for each multipath component. In our Bayesian channel tracking algorithm, for channel slot $n$, we will group the rays within clusters, and we will compute the mean of this Laplacian distribution for each multipath component as the sample mean of its corresponding cluster, as illustrated in Fig. 4.2. The mean computation is performed for each cluster and repeated for each channel slot. The variance can be computed in different ways depending on the practical scenario. For example, in a cellular system it could be set as the square of the RMS angular spread for AoA/AoD provided by the selected channel model (see for example Table 7.5-6, Part 1 in [13]). It could also be estimated at a particular mobile receiver if the maximum velocity is known and the trajectory is assumed to be linear during the channel slot (we will use this approach in our numerical simulations). The variance remains constant during the tracking stage, as specified in the 5G channel model.

Let $\mathbf{g}^{(n)} \in \mathbb{C}^{K\mu_L \times 1}$ denote the set of channel gains for the different subcarriers, $\mathbf{g}^{(n)} = [\mathbf{g}^{(n)T}[0], \ldots, \mathbf{g}^{(n)T}[K-1]]^T$. We assume that $\mathbf{g}^{(n)}$ follows a first-order Gauss-Markov process as in previous work [15]

$$\mathbf{g}^{(n)} = \mathbf{R}^{(n)}\mathbf{g}^{(n-1)} + (\mathbf{I} - \mathbf{R}^{(n)}\mathbf{R}^{(n)*})^{1/2}\boldsymbol{\Delta}\mathbf{g}^{(n)}, \tag{4.5.9}$$

for the correlation matrix $\mathbf{R}^{(n)} \in \mathbb{C}^{K\mu_L \times K\mu_L}$, which is assumed to be known, and a vector $\boldsymbol{\Delta}\mathbf{g}^{(n)} \in \mathbb{C}^{K\mu_L \times 1}$ with entries drawn from $\mathcal{CN}(\mathbf{0}, N_{\mathrm{t}}N_{\mathrm{r}}/\rho)$, being $\rho$

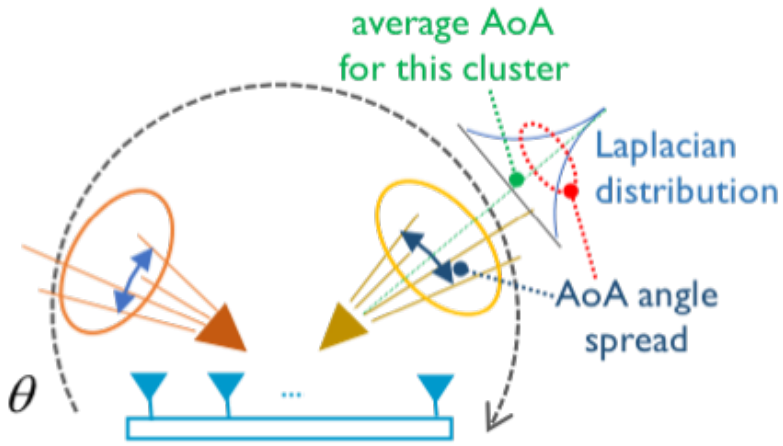



Figure 4.2: Illustration of the prior statistical knowledge about the he conditional density function $p(\boldsymbol{\xi}^{(n)}|\boldsymbol{\xi}^{(n-1)})$.

the path loss associated to the distance between the transmitter and receiver. $\mathbf{R}^{(n)}$ is the correlation matrix between channel gains in the frequency domain, as defined in $\mathbf{g}^{(n)}$. It has been analyzed and characterized in some measurements campaigns for narrowband systems [87]. The values in $\mathbf{R}^{(n)}$ correspond to the correlations between the gains corresponding to different paths $l, m$ and different subcarriers $j, k$ denoted as $R_{l,m}(j,k)$. The gains corresponding to different paths are assumed to be uncorrelated. This way $R_{l,m}(j,k) = 0, \forall l \neq m$. For our wideband system, we propose the following expression for the correlation between channel gains for the same path at two different subcarriers $i, j$, as a function of the frequency separation, for a given bandwidth $B$, tracking time $T_{\text{tck}}$, velocity of the receiver $v$ and number of subcarriers $K$

$$R_{l,l}(j,k) = \underbrace{(ae^{-bvT_{\text{tck}}} - c)}_{\text{Spatial autocorrelation}} \times \underbrace{e^{-|(j-k)|\frac{BT_{\text{tck}}}{K}}}_{\text{Frequency cross-correlation}} . \tag{4.5.10}$$

The first term captures the spatial autocorrelation and has been previously

proposed in [87] for outdoor environments. The second term in (4.5.10) captures the exponential decaying correlation between channel gains as their frequency separation increases. In the numerical results section, we will particularize and justify the choices of $a, b, c$ from the results in [87]. $B$ and the tracking time $T_{\text{tck}}$ are system parameters that will also be set in the simulations section. It also hods that the conditional pdf of $\mathbf{g}^{(n)}|\mathbf{g}^{(n-1)}$ follows $\mathbf{g}^{(n)}|\mathbf{g}^{(n-1)} \sim \mathcal{CN}\left(\mathbf{R}^{(n)}\mathbf{g}^{(n-1)}, (\mathbf{I} - \mathbf{R}^{(n)}\mathbf{R}^{(n)*})\right)$. Further, $\mathbf{g}^{(n)} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{\mathbf{gg}}^{(n)}\right)$, with $\mathbf{C}_{\mathbf{gg}}^{(n)} = \mathbf{R}^{(n)}\mathbf{C}_{\mathbf{gg}}^{(n-1)}\mathbf{R}^{(n)*} + (\mathbf{I} - \mathbf{R}^{(n)}\mathbf{R}^{(n)*})$.

In the following, we will derive our Bayesian filtering strategy based on these assumptions. In particular, we will focus on finding the *sparsity-constrained* MMSE estimator of $\{\mathbf{H}^{(n)}[k]\}_{k=0}^{K-1}$. Let $\mathbf{y}^{(n)} = [\mathbf{y}^{(n)T}[0], \ldots, \mathbf{y}^{(n)T}[K-1]]^T$ denote the received wideband signal, and $\mathbf{T}\left(\boldsymbol{\xi}^{(n)}\right) = \mathbf{I}_K \otimes \boldsymbol{\Phi}^{(n)}\boldsymbol{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)$ denote the transfer matrix of the system. Therefore, the wideband received signal yields

$$\mathbf{y}^{(n)} \approx \left(\mathbf{I}_K \otimes \boldsymbol{\Phi}^{(n)}\boldsymbol{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)\right)\mathbf{g}^{(n)} + \mathbf{n}^{(n)}. \tag{4.5.11}$$

Let us define $\mathbf{C}_{\text{w}}^{(n)} = \text{blkdiag}\{\mathbf{C}_{\text{w}}^{(1,n)}, \ldots, \mathbf{C}_{\text{w}}^{(M_{\text{tck}},n)}\}$. Then, $\mathbf{n}^{(n)}$ in (4.5.11) follows $\mathbf{n}^{(n)} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}_K \otimes \mathbf{C}_{\text{w}}^{(n)}\right)$. The measurement model in (4.5.11) is non-linear in $\boldsymbol{\xi}^{(n)}$, but the received signal $\mathbf{y}^{(n)}$ exhibits a linear sub-structure when conditioned on $\boldsymbol{\xi}^{(n)}$, i.e., $\mathbf{y}^{(n)}|\boldsymbol{\xi}^{(n)} \sim \mathcal{CN}\left(\mathbf{T}\left(\boldsymbol{\xi}^{(n)}\right)\mathbf{C}_{gg}^{(n)}\mathbf{T}^*\left(\boldsymbol{\xi}^{(n)}\right) + \mathbf{C}_{\text{w}}^{(n)}\right)$. Therefore, exploiting this signal structure can be used to obtain better estimates by analytically marginalizing out the linear state variables $\mathbf{g}^{(n)}$. The problem of finding the MMSE estimator of $\mathbf{h}^{(n)} = \text{vec}\{[\mathbf{H}^{(n)}[0], \ldots, \mathbf{H}^{(n)}[K -$

$1]]\} = \left(\mathbf{I}_K \otimes \mathbf{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)\right)\mathbf{g}^{(n)}$ reduces to finding its posterior conditional expectation as ($\mathbf{\Xi}^{(n)} = \{\boldsymbol{\xi}^{(i)}\}_{i=0}^{n}$, $\mathbf{Y}^{(n)} = \{\mathbf{y}^{(i)}\}_{i=0}^{n}$)

$$\hat{\mathbf{h}}_{\text{MMSE}}^{(n)} \triangleq \mathbb{E}_{\mathbf{h}^{(n)}|\mathbf{Y}^{(n)}}\{\mathbf{h}^{(n)}\}. \tag{4.5.12}$$

As shown in (4.5.13), an approximation of the MMSE estimator in (4.5.12) will be given by the combination of a KF to estimate the channel gains $\mathbf{g}^{(n)}$, a particle filter (PF) to estimate the angular parameters in $\boldsymbol{\xi}^{(n)}$, and a second PF to estimate the vectorized wideband channel $\mathbf{h}^{(n)}$. The joint posterior pdf of $\mathbf{g}^{(n)}$, $\mathbf{\Xi}^{(n)}$, and $\mathbf{h}^{(n)}$ is given by

$$p(\mathbf{g}^{(n)}, \mathbf{\Xi}^{(n)}, \mathbf{h}^{(n)}|\mathbf{Y}^{(n)}) = \underbrace{p(\mathbf{h}^{(n)}|\mathbf{Y}^{(n)}, \mathbf{g}^{(n)}, \mathbf{\Xi}^{(n)})}_{\text{PF \# 2}} \times \underbrace{p(\mathbf{\Xi}^{(n)}|\mathbf{Y}^{(n)})}_{\text{PF \# 1}} \underbrace{p(\mathbf{g}^{(n)}|\mathbf{\Xi}^{(n)}, \mathbf{Y}^{(n)})}_{\text{Optimum KF}}. \tag{4.5.13}$$

The different terms in the right hand side of (4.5.13) can be seen to yield

$$p(\mathbf{h}^{(n)}|\mathbf{Y}^{(n)}, \mathbf{g}^{(n)}, \mathbf{\Xi}^{(n)}) = \delta\left(\mathbf{h}^{(n)} - \left(\mathbf{I}_K \otimes \Psi\left(\boldsymbol{\xi}^{(n)}\right)\right)\mathbf{g}^{(n)}\right) \tag{4.5.14}$$

$$p(\mathbf{\Xi}^{(n)}|\mathbf{Y}^{(n)}) \propto p(\mathbf{y}^{(n)}|\mathbf{\Xi}^{(n)}, \mathbf{Y}^{(n-1)}) \times p(\boldsymbol{\xi}^{(n)}|\mathbf{\Xi}^{(n-1)}, \mathbf{Y}^{(n-1)}) \times p(\mathbf{\Xi}^{(n-1)}|\mathbf{Y}^{(n-1)}) \tag{4.5.15}$$

$$p(\mathbf{g}^{(n)}|\mathbf{\Xi}^{(n)}, \mathbf{Y}^{(n)}) = \mathcal{CN}\left(\hat{\mathbf{g}}^{(n|n)}, \mathbf{C}_{\mathbf{gg}}^{(n|n)}\right), \tag{4.5.16}$$

$$p(\mathbf{g}^{(n|n-1)}|\mathbf{\Xi}^{(n)}, \mathbf{Y}^{(n)}) = \mathcal{CN}\left(\hat{\mathbf{g}}^{(n|n-1)}, \mathbf{C}_{\mathbf{gg}}^{(n|n-1)}\right), \tag{4.5.17}$$

wherein $\delta(.)$ stands for the Dirac delta function and $p(\mathbf{x}) = \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ indicates that $p(\mathbf{x})$ takes the form of a circularly symmetric multivariate complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. The last two pdfs in (4.5.16) and (4.5.17) are of the form of the optimal KF, whereby exploiting

the linear sub-structure in (4.5.11) is clearly emphasized. Using straightforward application of the KF [100], the statistics in (4.5.16)-(4.5.17) yield the well-known KF prediction and update equations [101], which are omitted here for the sake of brevity.

The second pdf in (4.5.16) does not admit a closed-form for the different terms on its right hand side. Nonetheless, the terms $p\left(\mathbf{y}^{(n)}|\boldsymbol{\Xi}^{(n)}, \mathbf{Y}^{(n-1)}\right)$, $p\left(\boldsymbol{\xi}^{(n)}|\boldsymbol{\Xi}^{(n-1)}, \mathbf{Y}^{(n-1)}\right)$ can be expressed in closed form. Thus, we will approximate it using a standard PF, which corresponds to the PF #1. By direct inspection of $\mathbf{y}^{(n)}$ in (4.5.11), assuming that the noise realizations for two different channel slots are independent, it follows that the random vector $\mathbf{z}^{(n)} = \mathbf{y}^{(n)}|\boldsymbol{\Xi}^{(n)}, \mathbf{Y}^{(n-1)}$ is distributed as $\mathbf{z}^{(n)} \sim \mathcal{CN}\left(\boldsymbol{\mu}_{\mathbf{z}}^{(n)}, \mathbf{C}_{\mathbf{zz}}^{(n)}\right)$, with $\boldsymbol{\mu}_z^{(n)}$, $\mathbf{C}_z^{(n)}$ given by

$$\boldsymbol{\mu}_z^{(n)} = \mathbf{T}\left(\boldsymbol{\xi}^{(n)}\right)\hat{\mathbf{g}}^{(n|n-1)} \tag{4.5.18}$$

$$\mathbf{C}_{\mathbf{zz}}^{(n)} = \mathbf{T}\left(\boldsymbol{\xi}^{(n)}\right)\mathbf{C}_{\mathbf{gg}}^{(n|n-1)}\mathbf{T}^{*}\left(\boldsymbol{\xi}^{(n)}\right) + \sigma^2\mathbf{C}_{\mathrm{w}}. \tag{4.5.19}$$

The other random vector $\boldsymbol{\xi}^{(n)}|\boldsymbol{\Xi}^{(n-1)}, \mathbf{Y}^{(n-1)}$ is distributed with pdf $p(\boldsymbol{\xi}^{(n)}|\boldsymbol{\xi}^{(n-1)})$, which is assumed to be known, yet arbitrary. The pdf of $\boldsymbol{\Xi}^{(n-1)}|\mathbf{Y}^{(n-1)}$ corresponding to the last term in (4.5.15) can be approximated by the previous iteration of the first PF highlighted in (4.5.13). Now, the linear system in (4.5.9) and (4.5.11) can be formed for the $i$-th particle $\{\boldsymbol{\xi}^{(n,i)}\}_{i=1}^{N_{\mathrm{PF}}}$, $1 \leq i \leq N_{\mathrm{PF}}$, in which $N_{\mathrm{PF}}$ is the number of particles used by the PF. This requires one KF associated with each particle. Let the final sampled distribution of $\boldsymbol{\Xi}^{(n)}|\mathbf{Y}^{(n)}$ be denoted by the weights $\left\{w\left(\boldsymbol{\Xi}^{(n,i)}\right)\right\}_{i=1}^{N_{\mathrm{PF}}}$, which must satisfy

$\sum_{i=1}^{N_{\mathrm{PF}}} w\left(\boldsymbol{\Xi}^{(n,i)}\right) = 1$, since we are approximating a pdf by a probability mass function (pmf). Let us define $\mathbf{B}\left(\boldsymbol{\xi}^{(n)}\right) = \mathbf{I}_K \otimes \boldsymbol{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)$. Then, the vectorized channel can be estimated as

$$\hat{\mathbf{h}}^{(n)} = \int_{\mathbb{S}\left(\mathbf{g}^{(n)},\boldsymbol{\Xi}^{(n)}\right)} \mathbf{B}\left(\boldsymbol{\xi}^{(n)}\right)\mathbf{g}^{(n)} \times \underbrace{p(\mathbf{g}^{(n)}|\boldsymbol{\Xi}^{(n)},\mathbf{Y}^{(n)})w\left(\boldsymbol{\Xi}^{(n,i)}\right)}_{q^{(n)}} d\mathbf{g}^{(n)}d\boldsymbol{\Xi}^{(n)}. \tag{4.5.20}$$

The final estimator in (4.5.20) involves a complex integration over the joint support of $\mathbf{g}^{(n)}, \boldsymbol{\Xi}^{(n)}$, which cannot be solved in closed form. For this reason, we propose to approximate it using a second PF, which was highlighted in (4.5.13). Thus, normalizing the weights $q^{(n)}$ so that $\tilde{q}^{(n,i)} = q^{(n,i)} / \sum_{n=1}^{N_{\mathrm{PF}}} q^{(n,i)}\tilde{q}^{(n-1,i)}$, the MMSE estimator of the channel can be approximately expressed as a convex combination of $\mathbf{B}\left(\boldsymbol{\xi}^{(n,i)}\right)$ and $\hat{\mathbf{g}}^{(n|n,i)}$

$$\hat{\mathbf{h}}^{(n)} \approx \sum_{n=1}^{N_{\mathrm{PF}}} \tilde{q}^{(n,i)}\mathbf{B}\left(\boldsymbol{\xi}^{(n,i)}\right)\hat{\mathbf{g}}^{(n|n,j)}. \tag{4.5.21}$$

**Remark**: for the ML estimator, a function $f(\mathbf{z})$ of ML estimates $\hat{\mathbf{z}}_{\mathrm{ML}}$ will generally yield the ML estimator of a given vector $\mathbf{z}$. In the Bayesian framework, this property, called the *asymptotic invariance property* does not hold. We might be tempted to directly calculate the estimator of the channel by using the MMSE estimator of $\boldsymbol{\xi}^{(n)}$ and $\mathbf{g}^{(n)}$, but this would not result in the MMSE estimator of $\mathbf{h}^{(n)}$. To find the true MMSE estimator, the pdf of $\mathbf{h}^{(n)} = f\left(\boldsymbol{\xi}^{(n)}, \mathbf{g}^{(n)}\right)$ needs to be calculated to find the *posterior conditional mean*, as shown in (4.5.20).

Our proposed approach is a generalization of the marginalized particle filter (MPF) presented in [101], which has already been shown to perform well

**Algorithm 4** Generalized Marginalized Particle Filter (GMPF)

GMPF$\left(\mathbf{Y}^{(n)}, N_{\mathrm{PF}}, \mathbf{\Phi}^{(n)}, \hat{\theta}_{r,u}^{(0)}, \hat{\phi}_{r,u}^{(0)}\right)$

**Initialize angular particles based on prior information**

$\theta_{r,u}^{(1|0,i)} = \hat{\theta}_{r,u}^{(0)}$

$\phi_{r,u}^{(1|0,i)} = \hat{\phi}_{r,u}^{(0)}$

**Initialize linear particles**

$\hat{\mathbf{g}}^{(1|0,i)} = \mathbf{0}, \quad \hat{\mathbf{C}}_{\mathbf{gg}}^{(1|0,i)} = \frac{N_{\mathrm{t}} N_{\mathrm{r}}}{\mu_L} \mathbf{I}_{\mu_L} \otimes \mathbf{F}_1 \mathbf{F}_1^*$, with $\mathbf{F}_1 \in \mathbb{C}^{K \times Z_{\mathrm{p}}}$ the matrix comprising the first $Z_{\mathrm{p}}$ Fourier vectors.

**Evaluate the importance weights** $q^{(n,i)} = p(\mathbf{y}^{(n)}|\mathbf{\Xi}^{(n,i)}, \mathbf{Y}^{(n-1)}) p(\mathbf{g}^{(n)}|\mathbf{\Xi}^{(n)}, \mathbf{Y}^{(n)})$ **and normalize them**

$\tilde{q}^{(n,i)} = \frac{q^{(n,i)}}{\sum_{j=1}^{N_{\mathrm{PF}}} q^{(n,j)}} \tilde{q}^{(n-1,i)}$

**PF #1 measurement update (resampling $N_{\mathrm{PF}}$ particles with replacement)** $\mathbb{P}\left(\boldsymbol{\xi}^{(n|n,i)} = \boldsymbol{\xi}^{(n|n-1,j)}\right) = \tilde{q}^{(n,j)}$

**KF measurement update for** $1 \leq i \leq N_{\mathbf{PF}}$ $\mathbf{K}^{(n,i)} = \mathbf{C}_{\mathbf{gy}}^{(n|n-1,i)} \mathbf{C}_{\mathbf{yy}}^{(n|n-1,i)-1}$

$\hat{\mathbf{g}}^{(n|n,i)} = \hat{\mathbf{g}}^{(n|n-1,i)} + \mathbf{K}^{(n,i)} \left(\mathbf{y}^{(n,i)} - \boldsymbol{\mu}_y^{(n|n-1)}\right)$

$\hat{\mathbf{C}}_{\mathbf{gg}}^{(n|n,i)} = \mathbf{C}_{\mathbf{gg}}^{(n|n-1,i)} - \mathbf{K}^{(n,i)} \mathbf{C}_{\mathbf{gy}}^{(n|n-1,i)}$

**PF #1 prediction**

$\boldsymbol{\xi}^{(n+1|n,i)} \sim p\left(\boldsymbol{\xi}^{(n+1|n)}|\mathbf{\Xi}^{(n,i)}, \mathbf{Y}^{(n)}\right)$

**KF prediction for** $1 \leq i \leq N_{\mathbf{PF}}$

$\hat{\mathbf{g}}^{(n|n-1,i)} = \mathbf{R}^{(n)} \hat{\mathbf{g}}^{(n-1|n-1)}$

$\mathbf{C}_{\mathbf{gg}}^{(n|n-1,i)} = \mathbf{R}^{(n)} \mathbf{C}_{\mathbf{gg}}^{(n-1|n-1,i)} \mathbf{R}^{(n)*} + (\mathbf{I} - \mathbf{R}^{(n)} \mathbf{R}^{(n)*})$

**PF #2 to estimate the channel**

$\hat{\mathbf{h}}^{(n)} = \sum_{j=1}^{N_{\mathrm{PF}}} \tilde{q}^{(n,j)} \mathbf{B}\left(\boldsymbol{\xi}^{(n|n,j)}\right) \hat{\mathbf{g}}^{(n|n,j)}$

**Reshape vectorized channel estimate to find** $\hat{\mathbf{H}}^{(n)}[k]$

$\hat{\mathbf{H}}^{(n)}[k] = \mathrm{unvec}\{\hat{\mathbf{h}}^{(n)}\}$

in practice and has optimality guarantees as $N_{\mathrm{PF}}$ grows large. Owing to our proposed algorithm being based on the MPF in [101], we name the proposed algorithm as GMPF. The detailed steps the algorithm follows are given in Algorithm 4.

## 4.6 Numerical results

In this section, we include a performance evaluation of the channel tracking algorithms proposed in Section IV, GMPF, and a comparison with an extension of two state-of-the art narrowband tracking strategies. We also present numerical results to evaluate the effectiveness of the precoders and combiners especially designed for tracking in Section III.

### 4.6.1 Simulation parameters

The parameters used to configure the transmitter and the receiver in the simulations included in this section are summarized as follows. Both the transmitter and the receiver are assumed to be equipped with uniform linear arrays (ULAs) with half-wavelength separation. Such a ULA has steering vectors obeying the expressions $\{\mathbf{a}_{\mathrm{T}}(\theta_\ell)\}_n = \sqrt{\frac{1}{N_{\mathrm{t}}}} e^{\mathrm{j}n\pi\cos(\theta_\ell)}$, $n = 0, \ldots, N_{\mathrm{t}} - 1$, and $\{\mathbf{a}_{\mathrm{R}}(\phi_\ell)\}_m = \sqrt{\frac{1}{N_{\mathrm{r}}}} e^{\mathrm{j}m\pi\cos(\phi_\ell)}$, $m = 0, \ldots, N_{\mathrm{r}} - 1$. We take $N_{\mathrm{t}} = 64$, $N_{\mathrm{r}} = 32$ for illustration. The phase-shifters employed in both transmitter and receiver are assumed to have $N_{\mathrm{Q,Tx}}$ and $N_{\mathrm{Q,Rx}}$ quantization bits, so that the entries of the analog precoders and combiners are drawn from the sets $\mathcal{A}_{\mathrm{Tx}} = \left\{0, \frac{2\pi}{2^{N_{\mathrm{Q,Tx}}}}, \ldots, \frac{2\pi(2^{N_{\mathrm{Q,Tx}}}-1)}{2^{N_{\mathrm{Q,Tx}}}}\right\}$ and $\mathcal{A}_{\mathrm{Rx}} = \left\{0, \frac{2\pi}{2^{N_{\mathrm{Q,Rx}}}}, \ldots, \frac{2\pi(2^{N_{\mathrm{Q,Rx}}}-1)}{2^{N_{\mathrm{Q,Rx}}}}\right\}$. The number of quantization bits is set to $N_{\mathrm{Q,Tx}} = N_{\mathrm{Q,Rx}} = 4$ for illustration. The number of RF chains is set to $L_{\mathrm{t}} = L_{\mathrm{r}} = 4$. The number of OFDM subcarriers is set to $K = 32$, and a zero-prefix length of $Z_{\mathrm{P}} = K/4 = 8$ samples is assumed to remove intersymbol interference (ISI). The carrier frequency is set to $f_{\mathrm{c}} = 60$ GHz, the bandwidth is set to $B = 2.55$ GHz, with a roll-off

factor of 0.3 for the raised cosine pulse shaping factor. We use the sampling period defined in the 5G NR standard, $T_{\mathrm{s}} = 0.509$ ns. The OFDM symbol duration is $(K+Z_{\mathrm{p}})T_{\mathrm{s}} = 0.02$ $\mu$s, and the subcarrier spacing is $1/(KT_{\mathrm{s}}) = 60.8$ MHz. The transmitted power is set to $P_{\mathrm{tx}} = 35$ dBm for illustration. We also consider that the channel slot duration is $M = 240$ OFDM symbols.

The main performance metric we consider in this chapter is the ergodic spectral efficiency. To gain insight into how well the proposed channel tracking methods perform, we compute the ergodic spectral efficiency that could be achieved when operating with all-digital precoders and combiners obtained from the channel estimates provided by the different algorithms. If $\hat{\mathbf{H}}^{(n)}[k]$ is the channel estimate for subcarrier $k$ provided by a given algorithm, consider the SVD $\hat{\mathbf{H}}^{(n)}[k] = \hat{\mathbf{U}}_H^{(n)}[k]\hat{\mathbf{\Sigma}}_H^{(n)}[k]\hat{\mathbf{V}}_H^{(n)*}[k]$, and define $\hat{\mathbf{H}}_{\mathrm{eff}}^{(n)}[k] \in \mathbb{C}^{N_{\mathrm{s}}\times N_{\mathrm{s}}}$, $\hat{\mathbf{H}}_{\mathrm{eff}}^{(n)}[k] \triangleq \hat{\mathbf{U}}_H^{(n)*}[k]\mathbf{H}^{(n)}[k]\hat{\mathbf{V}}_H^{(n)}[k]$. Let $\eta$ quantify the performance loss owing to sending $M_{\mathrm{tck}}$ OFDM symbols to track the mmWave channel, so that $M_{\mathrm{data}} = M - M_{\mathrm{tck}}$ OFDM symbols are left to transmit data. In this case, $\eta = T_{\mathrm{s}}M_{\mathrm{data}}/(T_{\mathrm{s}}M_{\mathrm{data}} + T_{\mathrm{s}}M_{\mathrm{tck}})$. Then, we define the sample average ergodic spectral efficiency as

$$\mathcal{R}^{(n)} = \eta\frac{1}{K}\sum_{k=0}^{K-1}\log_2\left|\mathbf{I}_{N_{\mathrm{s}}} + \frac{\mathrm{SNR}}{N_{\mathrm{s}}}\hat{\mathbf{H}}_{\mathrm{eff}}^{(n)}[k]\hat{\mathbf{H}}_{\mathrm{eff}}^{(n)*}[k]\right|. \tag{4.6.1}$$

### 4.6.2 Performance evaluation of GMPF

We consider now the performance evaluation of GMPF, the Bayesian algorithm that we propose for channel tracking in scenarios where the channel

statistics are known or can be learnt. In these simulations, we consider that the channel has a LOS component with a single ray [13], and $C = 2$ NLOS clusters, each contributing with a number of rays $R_c \sim \mathcal{U}[6, 20]$ as in the 5G NR channel model. Therefore, the average number of multipath components $\mu_L = 27$ will be considered in GMPF. We choose the Rician K-factor for the generated test channels as -10dB, 0 dB, and 10 dB, to represent different propagation scenarios. We consider a relative velocity of $v = 30$ m/s for the receiver. We generate test channels that follow the statistical priors described in Section 4.5.2. To compute the correlation matrix in (4.5.10) we consider the parameters $a = 0.91$, $c = 0$ and $b = 0.1/\lambda_c$, with $\lambda_c$ the wavelength. Moreover, the tracking time is set as $T_{\text{tck}} = 0.1577$ ms. The mean of the Laplacian distribution that models the angle evolution for each multipath component is set as the estimated angle for the previous slot. To get a good trade off between performance and complexity, the proposed GMPF algorithm is configured to use $N_{\text{PF}} = 50$ particles to discretize the joint Laplacian pdf of the AoA/AoD. Considering that the transmitter and scatterers are fixed, and that the receiver follows a linear trajectory during every channel slot, we define the maximum displacement of the receiver in a channel slot as $d_{\max}^{\text{slot}} = v_{\max} T_{\text{tck}}$. Under the assumption $d >> d_{\max}^{\text{slot}}$, which always holds since $T_{\text{tck}}$ is very small, it can be easily shown that the variation in angle of arrival at the receiver for any multipath component satisfies $|\Delta \sin \phi_{c,r}^{(n)}| \approx |\Delta \phi_{c,r}^{(n)}| < \frac{d_{\max}^{\text{slot}}}{d}$. Considering a minimum distance $d$ between the transmitter and the receiver, and a maximum velocity, we can obtain an upper bound for $|\Delta \phi_{c,r}^{(n)}|$. In our simulation setting,

and considering a minimum distance between receiver and transmitter of 5 m and $v_{\max} = 30$ m/s, $|\Delta\phi_{c,r}^{(n)}| < 0.042°$. This way, we set the value for the standard deviation of the Laplacian distribution as 0.57°, ten orders the magnitude bigger than the maximum expected variation.

We choose two baselines strategies to be compared with GMPF. One of them is 5G NR and The second baseline algorithm is a modified version of the fast beam tracking strategy proposed in [14], which also exploits statistical priors. As explained before, due to the lack of wideband approaches to track the mmWave channel, we have enabled wideband operation of [14] by computing a narrowband version of the received signals for the different subcarriers and different pilot symbols. In particular, we collect the received signals in a tensor of size $L_{\mathrm{r}}\mu_L \times K \times M_{\mathrm{tck}}$. Then, after averaging over the $M_{\mathrm{tck}}$ measurements, we obtain a matrix of size $L_{\mathrm{r}}\mu_L \times K$. Finally, we obtain the dominant eigenvalue of this matrix, which provides an estimate of the centroid for the set of received signals as described in [102]. This narrowband version of the received signal is the input to the algorithm in [14], which provides the best analog precoder/combiners following a Bayesian approach. The digital precoders and combiners are computed from an estimate of the beamformed channel as in our 5G NR implementation.

Fig. 4.3 shows the additional benefit coming from exploiting prior statistical information. We present the spectral efficiency obtained by the different systems when the Rician K-factor is $-10$, 0 or 10 dB, and $N_{\mathrm{s}} = \{1, 2\}$ streams. GMPF outperforms the other Bayesian approach and 5G NR as

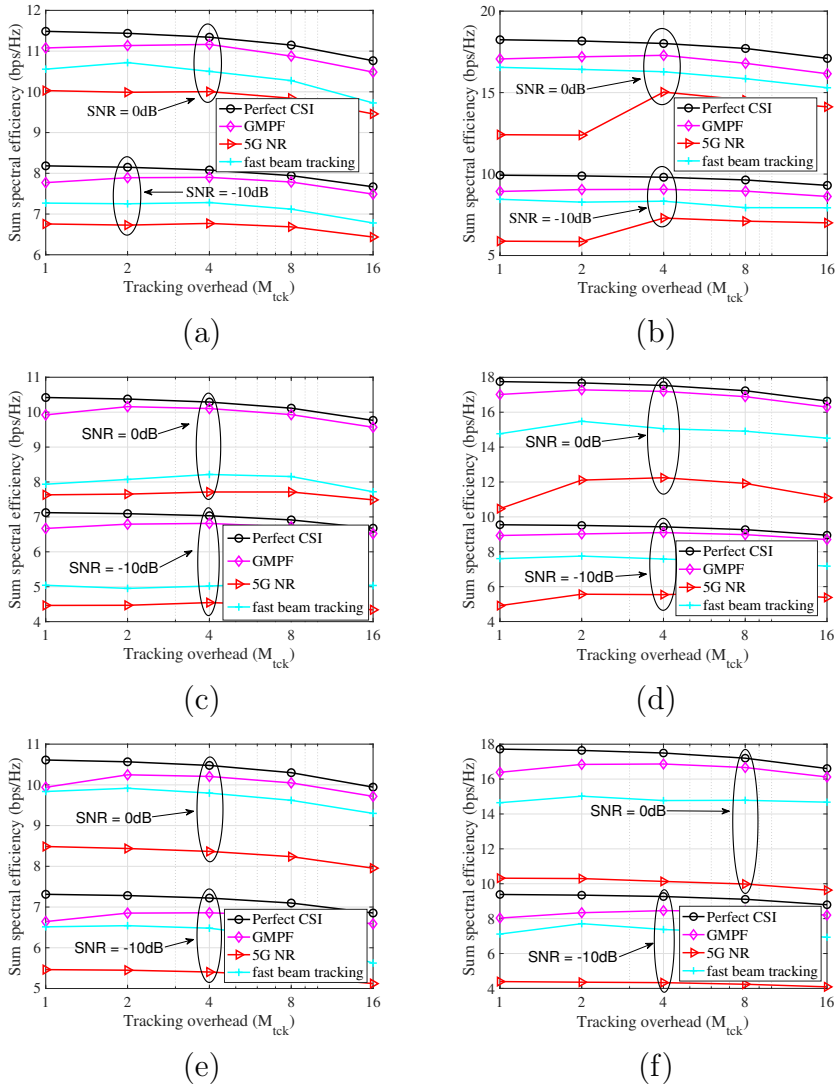



Figure 4.3: Comparison of evolution of the spectral efficiency versus number of tracking pilots for GMPF, 5G NR and fast beam training in for different SNRs. The Rician factor is $K_{\text{factor}} = 10\text{dB}$ for (a) and (b), $K_{\text{factor}} = 0\text{dB}$ for (c) and (d) and $K_{\text{factor}} = 0\text{dB}$ for (e) and (f). The number of data streams is set to $N_{\text{s}} = 1$ for (a), (c) and (e), and $N_{\text{s}} = 2$ for (b) and (d) and (f).

expected, for all considered scenarios. The performance gap increases with the number of transmit streams and the NLOS behavior of the channel. 5G NR only provides near optimal performance when the channel has a dominant LoS component.

## 4.7 Conclusions

In this chapter, we formulated the problem of channel tracking for an MIMO-OFDM system operating at mmWave, and proposed a non-linear Bayesian filter aiming at finding the global MMSE estimator. We also proposed an SNR-maximizing precoding and combining design method for channel tracking, leveraging prior information on the previously estimated AoA and AoD, such that low-overhead channel tracking can be performed efficiently. Simulation results showed the effectiveness of our proposed methods, which have been evaluated using realistic channel realizations extracted from the 5G NR channel model. We showed that, even under high mobility conditions, the proposed tracking framework is able to maintain near-optimum spectral efficiency values at low overhead, even if the distance between transmitter and receiver is small. Future work would conduct the extension of the proposed framework to deal with channel blockage, and consider a multi-user scenario, without assuming perfect synchronization, and accounting for the beam-squint effect that is present when the signal bandwidth is very large.

# Chapter 5

# Blockage Detection and Channel Tracking in Wideband mmWave MIMO Systems

In this chapter[1], we propose novel blockage detection and channel tracking strategies for wideband mmWave MIMO systems in the context of high mobility V2X scenarios under blockage. Tracking wideband mmWave MIMO systems based on a hybrid architecture is a challenging problem, especially in high mobility scenarios where links are likely to be blocked by obstacles, such as trees, pedestrians or vehicles. In this chapter, we propose a new strategy to track the frequency selective mmWave channel under blockage. We first introduce a statistical channel model that includes the evolution models for channel gains and angles of arrival and departure, as well as the statistics of blockage events. Then, we define a change point detection (CPD) test to identify the time instants where blockage appears/disappears, so the appropriate channel evolution models can be used for wideband channel tracking during the blockage events. To simultaneously achieve a high CPD success

[1]This chapter was based on our work [103]: H. Xie, N. González-Prelcic, and T. Shimizu, "Blockage detection and channel tracking in wideband mmWave MIMO systems,"' in Proc. of IEEE ICC'21, 2021. This work was supervised by Prof. González-Prelcic. My contributions lie in designing and performing the research, doing simulations, analyzing data and writing the paper, etc.

rate and a high accuracy in the channel estimate, we further propose a double digital combiner architecture that employs different digital combiners for CPD and channel tracking. Finally, we integrate into our framework a previously proposed Bayesian channel tracking algorithm. Simulation results show that the proposed approach achieves a good channel tracking performance even in mobile scenarios that suffer from highly dynamic blockage events.

## 5.1 Introduction

Configuring the wideband mmWave links for massive MIMO based on hybrid architectures is a challenging problem that has been addressed in prior works. After initial access, tracking the CSI with low overhead is necessary to maintain high-quality mmWave links, especially in high mobility use cases. The strategies proposed in prior work on channel tracking at mmWave assume a slow variation of channel parameters and/or a narrowband channel model. However, the mmWave channel with large bandwidth is typically frequency selective, and the directional mmWave links may be blocked by obstacles, like human body, neighboring vehicles, buildings, foliage and other infrastructures. Blockage may appear and disappear frequently, which induces abrupt changes in the channel.

## 5.2 Contributions

Previous work, such as [14, 17, 22, 30], derive different channel tracking strategies for narrowband channels, while only [22, 31, 32] consider a frequency

selective channel model. These approaches consider an evolution model for the channel parameters and work well in the absence of blockage. Although [33–35] do consider blockage and propose algorithms to detect abrupt changes in the mmWave channel, only the narrowband case was considered. Moreover, [33–35] focus on fast re-access to the channel after link failure due to blockage, rather than on incorporating the statistics of blockage into the channel tracking strategy.

Motivated by the limitations of previous work, we address, for the first time, the problem of tracking the frequency selective mmWave channel under blockage. Besides the statistical knowledge of the continuous-time evolution of the channel parameters, we further incorporate and exploit the statistics of blockage events described in prior work [104–107]. To derive a channel tracking algorithm robust to blockage, we adopt the CPD algorithm to identify the time instants where blockage appears/disappears. Then, this temporal information is used in the channel tracking stage to select the continuous-time evolution model of the channel parameters during smooth variation/blockage phases. This way, the tracking algorithm can use the appropriate evolution model at every time, providing a highly accurate channel estimate even during blockage. Therefore, we extend the Bayesian tracking algorithm in [32] by generalizing the channel models under blockage and integrating the blockage detection scheme. To simultaneously obtain a high blockage detection probability and an accurate channel estimate, we also propose in this chapter a double digital combiner architecture that uses different digital combiners for

CPD and channel tracking. Simulation results demonstrate the effectiveness of the proposed approach blockage even in high mobility scenarios.

## 5.3 System and signal models

We consider a single-user mmWave MIMO communication system, where the AP serves a moving UE using an OFDM link at mmWave. The AP and UE are equipped with $N_{\mathrm{t}}$ and $N_{\mathrm{r}}$ antennas, and $L_{\mathrm{t}}$ and $L_{\mathrm{r}}$ RF chains respectively. Moreover, both the AP and UE are assumed to use a fully-connected hybrid architecture.

We assume the channel is evolving between transmission intervals or time slots, but it keeps constant within each slot. We use the super-index $n$ to denote the $n$-th channel slot ($n = 1, 2, \ldots$). During the data transmission, the transmitter uses a hybrid precoder $\mathbf{F}^{(n)}[k] \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{s}}}$ for the $k$-th ($k = 0, \ldots, K-1$) subcarrier to transmit an $N_{\mathrm{s}} \times 1$ vector of data streams $\mathbf{s}^{(n)}[k]$, while the receiver uses a hybrid combiner $\mathbf{W}^{(n)}[k] \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{s}}}$ for reception. Let us denote the downlink channel between AP and UE as $\mathbf{H}^{(n)}[k]$ for the subcarrier $k$ and slot $n$. The received signal is given by

$$\mathbf{y}^{(n)}[k] = \mathbf{W}^{(n)*}[k]\mathbf{H}^{(n)}[k]\mathbf{F}^{(n)}[k]\mathbf{s}^{(n)}[k] + \mathbf{n}^{(n)}[k], \tag{5.3.1}$$

where the received additive Gaussian noise is denoted as $\mathbf{n}^{(n)}[k] \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{C}_{\mathrm{w}}^{(n)}[k])$, with $\mathbf{C}_{\mathrm{w}}^{(n)}[k] \triangleq \mathbf{W}^{(n)*}[k]\mathbf{W}^{(n)}[k] \in \mathbb{C}^{N_{\mathrm{s}} \times N_{\mathrm{s}}}$ the noise covariance matrix. Note that during the training phase of the $n$-th channel slot, frequency-flat precoders and combiners will be used. We also define the trans-

mitted signal $\mathbf{s}^{(n)}[k]$ as $\mathbf{s}^{(n)}[k]$=$\mathbf{q}^{(n)}s[k]$, with $\mathbf{q}^{(n)}\in\mathbb{C}^{N_\mathrm{s}\times 1}$ a frequency-flat training vector and $s[k]$ a frequency-dependent training symbol. Then, we can multiply the received signal $\mathbf{y}^{(n)}[k]$ by $(s[k])^{-1}$ and get a frequency-flat measurement matrix. Moreover, $M_\mathrm{tck}$ training OFDM symbols will be transmitted by AP to collect measurement at UE. Specifically, the received signal for the $m$-th training OFDM symbol can be expressed as

$$\begin{aligned}\mathbf{y}^{(m,n)}[k] = &\left(\mathbf{q}^{(m,n)T}\mathbf{F}^{(m,n)T}\otimes\mathbf{W}^{(m,n)*}\right)\mathrm{vec}\{\mathbf{H}^{(n)}[k]\}\\ &+\mathbf{n}^{(m,n)}[k],\quad m=1,\ldots,M_\mathrm{tck}.\end{aligned}\tag{5.3.2}$$

Stacking all the $M_\mathrm{tck}$ measurements as in [32], we have

$$\mathbf{y}^{(n)}[k]\approx\mathbf{\Phi}^{(n)}\mathrm{vec}\{\mathbf{H}^{(n)}[k]\}+\mathbf{n}^{(n)}[k],\tag{5.3.3}$$

where $\mathbf{y}^{(n)}[k]\triangleq[\mathbf{y}^{(1,n)T}[k],\ldots,\mathbf{y}^{(M_\mathrm{tck},n)T}[k]]^T\in\mathbb{C}^{N_\mathrm{s}M_\mathrm{tck}\times 1}$, $\mathbf{n}^{(n)}[k]\triangleq[\mathbf{n}^{(1,n)T}[k],\ldots,\mathbf{n}^{(M_\mathrm{tck},n)T}[k]]^T\in\mathbb{C}^{N_\mathrm{s}M_\mathrm{tck}\times 1}$, and $\mathbf{\Phi}^{(n)}\in\mathbb{C}^{N_\mathrm{s}M_\mathrm{tck}\times N_\mathrm{t}N_\mathrm{r}}$ stacks all the training OFDM symbols in the $n$-th slot accordingly.

## 5.4 Statistical channel model under blockage

In this section, we introduce a statistical model for the mmWave channel which includes the impact of blockage.

### 5.4.1 Geometrical channel model without blockage

First, a geometrical channel model [91] is adopted. Specifically, we denote the $d$-th delay tap of the mmWave MIMO channel at the $n$-th channel

slot as $\mathbf{H}^{(n)}[d] \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$, for $d = 0, \ldots, D-1$, with $D$ the channel delay tap length. Moreover, we assume that there are in total $R$ rays. Let $\alpha_r^{(n)} \in \mathbb{C}$ be the complex gain of the $r$-th ray, $\phi_r^n$, $\theta_r^n \in \mathbb{R}$ be the AoA and AoD, $\tau_r^n \in \mathbb{R}$ be the time-delay, and $p_{\mathrm{rc}}(\tau)$ be the equivalent response of transmit and receive pulse-shapes and other analog filtering evaluated at $\tau$. In addition, let $\mathbf{a}_\mathrm{T}\left(\theta_r^n\right) \in \mathbb{C}^{N_\mathrm{t} \times 1}$ and $\mathbf{a}_\mathrm{R}\left(\phi_r^n\right) \in \mathbb{C}^{N_\mathrm{r} \times 1}$ be the transmit and receive array steering vectors evaluated on the AoD and AoA of each corresponding path. Then, the $N_\mathrm{r} \times N_\mathrm{t}$ channel matrix at the $d$-th delay tap is given by [91]

$$\mathbf{H}^{(n)}[d] = \sqrt{\frac{N_\mathrm{t} N_\mathrm{r}}{R}} \sum_{r=1}^{R} \alpha_r^{(n)} p_{\mathrm{rc}}\left(dT_\mathrm{s} - \tau_r^{(n)}\right) \times \mathbf{a}_\mathrm{R}\left(\boldsymbol{\phi}_r^{(n)}\right) \mathbf{a}_\mathrm{T}^*\left(\boldsymbol{\theta}_r^{(n)}\right). \tag{5.4.1}$$

By taking the $K$-point DFT of (5.4.1), the frequency domain channel response can be expressed as

$$\mathbf{H}^{(n)}[k] = \mathbf{A}_\mathrm{R}\left(\boldsymbol{\phi}^{(n)}\right) \mathbf{G}^{(n)}[k] \mathbf{A}_\mathrm{T}^*\left(\boldsymbol{\theta}^{(n)}\right), \tag{5.4.2}$$

where $\mathbf{A}_\mathrm{T}\left(\boldsymbol{\theta}^{(n)}\right) \in \mathbb{C}^{N_\mathrm{t} \times R}$, $\mathbf{A}_\mathrm{R}\left(\boldsymbol{\phi}^{(n)}\right) \in \mathbb{C}^{N_\mathrm{r} \times R}$ contain the transmit and receive antenna array responses at all AoDs and AoAs, and $\mathbf{G}^{(n)}[k] \in \mathbb{C}^{R \times R}$ is the diagonal matrix containing the frequency responses of the channel gains filtered by the equivalent pulse-shape for all rays in (5.4.1). Vectorizing (5.4.2) gives

$$\begin{aligned} \mathrm{vec}\{\mathbf{H}^{(n)}[k]\} &= \left(\mathbf{A}_\mathrm{T}^\mathrm{C}(\boldsymbol{\theta}^{(n)}) \star \mathbf{A}_\mathrm{R}\left(\boldsymbol{\phi}^{(n)}\right)\right) \mathbf{g}^{(n)}[k] \\ &= \boldsymbol{\Psi}(\boldsymbol{\xi}^{(n)}) \mathbf{g}^{(n)}[k], \end{aligned} \tag{5.4.3}$$

where the superscript C of $\mathbf{A}_\mathrm{T}^\mathrm{C}\left(\boldsymbol{\theta}^{(n)}\right)$ denotes the conjugate of matrix $\mathbf{A}_\mathrm{T}\left(\boldsymbol{\theta}^{(n)}\right)$, and $\mathbf{g}^{(n)}[k]$ is an $R \times 1$ vector containing the diagonal elements of $\mathbf{G}^{(n)}[k]$. Moreover, $\boldsymbol{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)$ is defined from (5.4.3), with $\boldsymbol{\xi}^{(n)} = [\boldsymbol{\theta}^{(n)T}, \boldsymbol{\phi}^{(n)T}]^T$ being the set of AoDs and AoAs for the $n$-th channel slot.

### 5.4.2 Blockage modeling on geometrical channel model

Next, we introduce the statistical models of channel parameters and blockage effects. For better illustration, we first denote $g_r^{(n)}[d] = \sqrt{\frac{N_\mathrm{t}N_\mathrm{r}}{R}} p_\mathrm{rc}(dT_\mathrm{s} - \tau_r^{(n)})$, $r = 0, \ldots, R-1$, and then define the $D \times 1$ vector $\bar{\boldsymbol{g}}_r^{(n)} \triangleq [g_r^{(n)}[0], \ldots, g_r^{(n)}[D-1]]^T$ containing the complex channel gains in time domain for the $r$-th ray. Let $\mathbf{F}_1 \in \mathbb{C}^{K \times D}$ be the matrix containing the first $D$ column vectors of the unitary DFT matrix of size $K$ and we can define the frequency-domain channel gains $\boldsymbol{g}_r^{(n)} \in \mathbb{C}^{K \times 1}$ for the $r$-th ray as $\boldsymbol{g}_r^{(n)} \triangleq \mathbf{F}_1 \bar{\boldsymbol{g}}_r^{(n)}$ with

$$\boldsymbol{g}_r^{(n)} \sim \mathcal{CN}\left(\mathbf{0}, \frac{N_\mathrm{t}N_\mathrm{r}}{R}\mathbf{F}_1 \left(\mathrm{diag}\left\{p_\mathrm{rc}^2(dT_\mathrm{s} - \tau_r^{(n)})\right\}_{d=0}^{D-1}\right)\mathbf{F}_1^*\right), \tag{5.4.4}$$

where $\mathrm{diag}\{\ldots\}_{d=0}^{D-1}$ creates a $D \times D$ diagonal matrix with the given $D$ elements. As for the blockage modeling, the 5G NR standards [13] and some measurement works [104–107] show that blockage is an add-on feature. In other words, the modeling of channel under blockage can be obtained by adding the impacts of blockage to the original channel models in (5.4.4). From [104], the key parameters for blockage modeling during a long transmission time include the birth time of blockage events, the duration of each blockage event as well as the additional pathloss on blocked paths. Generally, the appearance of blockage can be viewed as random event and thus we assume its birth time is uniform across the communication intervals. Moreover, the duration of blockage events can be modeled as Weibull distribution [104]. As for the additional pathloss, it can be modeled as Gaussian distribution [104, 105, 107]. That is, if the $r$-th

path has been blocked during the $n$-th time slot, the (5.4.4) has to be modified as

$$\boldsymbol{g}_{r,b}^{(n)} \sim \mathcal{CN}\left(\mathbf{0}, \frac{N_{\mathrm{t}}N_{\mathrm{r}}}{R}\mathbf{F}_1\left(\mathrm{diag}\left\{\frac{p_{\mathrm{rc}}^2(dT_{\mathrm{s}}-\tau_r^{(n)})}{P_{blk}^2}\right\}_{d=0}^{D-1}\right)\mathbf{F}_1^*\right), \tag{5.4.5}$$

where $P_{blk}^2 \sim \mathcal{N}(\mu_{blk}, \sigma_{blk}^2)$ connotes the additional pathloss due to the blockage effect, with $\mu_{blk}$ and $\sigma_{blk}^2$ being the mean and variance of this pathloss variable. It indicates that the received power of the blocked path has been significantly degraded. If $P_{blk}^2$ is large, the $r$-th ray will completely disappear.

To perform channel tracking, the time evolution of the channel gains in the frequency domain, $\boldsymbol{g}_r^{(n)}$, is assumed to be a first-order Gauss-Markov process as in prior work [22, 22] without blockage. After taking into consideration of blockage effect, we will have different gain evolution models as follows. Specifically, when there is no abrupt change or blockage event emerging from $(n-1)$-th slot to the $n$-th slot, the gain evolution is given as

$$\begin{cases} \boldsymbol{g}_r^{(n)} = \mathbf{R}_r\boldsymbol{g}_r^{(n-1)} + (\mathbf{I}-\mathbf{R}_r\mathbf{R}_r^*)^{1/2}\,\boldsymbol{\Delta g}_r^{(n)}, \\ \text{with } \boldsymbol{\Delta g}_r^{(n)} \text{ defined as in (5.4.4).} \end{cases} \tag{5.4.6}$$

where, $\mathbf{R}_r \in \mathbb{C}^{K\times K}$ is the correlation matrix between channel gains in the frequency domain for any ray $r$ and its definition is given as in [31, 32]. Note that the gains corresponding to different paths are assumed to be uncorrelated. While if the $r$-th ray is blocked during the transition from $(n-1)$-th slot to the $n$-th slot, the gain evolution should be modified as

$$\begin{cases} \boldsymbol{g}_r^{(n)} = \mathbf{R}_r\boldsymbol{g}_r^{(n-1)}/P_{blk} + (\mathbf{I}-\mathbf{R}_r\mathbf{R}_r^*)^{1/2}\,\boldsymbol{\Delta g}_r^{(n)}, \\ \text{with } \boldsymbol{\Delta g}_r^{(n)} \text{ defined as in (5.4.5).} \end{cases} \tag{5.4.7}$$

Note that the differences of (5.4.7) compared to (5.4.6) exist not only on the covariance matrix $\boldsymbol{\Delta g}_r^{(n)}$, but also on the additional pathloss of the one-step prediction of the gains. Last but no least, the gain evolution model during the phase of a blockage interval should be adjusted as

$$\begin{cases} \boldsymbol{g}_r^{(n)} = \mathbf{R}_r \boldsymbol{g}_r^{(n-1)} + \left(\mathbf{I} - \mathbf{R}_r \mathbf{R}_r^*\right)^{1/2} \boldsymbol{\Delta g}_r^{(n)}, \\ \text{with } \boldsymbol{\Delta g}_r^{(n)} \text{ defined as in (5.4.5).} \end{cases} \tag{5.4.8}$$

The difference between (5.4.8) and (5.4.7) lies in that the additional pathloss for the one-step prediction of gains has already been added at the beginning of the blockage and thus should not be evaluated again during the blockage interval. Then, the joint definitions in (5.4.6), (5.4.7) and (5.4.8) provide the new evolution modeling for channel gains under blockage.

Regarding the time evolution of angle parameters, as defined in (5.4.3), let $\boldsymbol{\xi}^{(n)} = [\boldsymbol{\theta}^{(n)T}, \boldsymbol{\phi}^{(n)T}]^T$ be the set of AoDs and AoAs for the $n$-th channel slot, and $\boldsymbol{\xi}^{(n-1)}$ be the set of AoDs and AoAs for the $(n-1)$-th channel slot. In this chapter, we assume, as in the 5G NR channel model [13, 32], that the conditional density function $p(\boldsymbol{\xi}^{(n)}|\boldsymbol{\xi}^{(n-1)})$ follows a Laplacian distribution for each multipath component. That is, for the $r$-ray, we have

$$\begin{aligned} \theta_r^{(n)} &\sim \mathcal{L}(\theta_r^{(n-1)}, \sigma^2_{\Delta,AoD}), \\ \phi_r^{(n)} &\sim \mathcal{L}(\phi_r^{(n-1)}, \sigma^2_{\Delta,AoA}), \end{aligned} \tag{5.4.9}$$

where $\sigma^2_{\Delta,AoD}$ and $\sigma^2_{\Delta,AoA}$) can be set based on the environment, system and user mobility parameters. When new paths appear or the blockage disappears, the initial estimates of the new AoAs and AoDs will be denoted as $\phi_{\text{ini}}^{(n),\text{new}}$ and

$\theta_{\text{ini}}^{(n),\text{new}}$, which will be the initial points in the angle evolution model defined in (5.4.9).

## 5.5 Blockage detection and wideband channel tracking under blockage

In this section, we propose a channel estimation and tracking strategy suitable for mmWave MIMO communications under blockage. Our approach considers a blockage detection stage based on a CPD test that selects the appropriate channel evolution model for the channel tracking stage, as illustrated in Fig. 5.1. As channel tracking strategy we propose a modified version of the Generalized Marginal Partial Filter (GMPF) algorithm in [32]. To simultaneously obtain a good blockage detection performance and a high channel estimation accuracy we propose a double digital combiner architecture, as can be seen in Fig. 5.1. On one hand, one of the digital combiners aims to only capture the energy of the dominant multipath components, which in case of being blocked may cause a link failure. On the other hand, the second digital combiner operating in parallel is designed to maximize the SNR during the tracking stage that estimates the channel parameters. In the next subsection we explain in detail all the components of this blockage detection and tracking strategy.

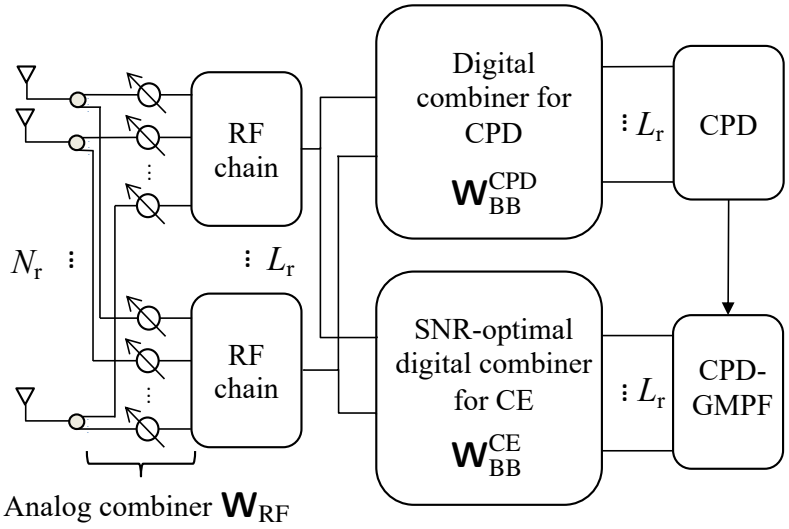



Figure 5.1: Block diagram of the proposed double digital combiner scheme and CPD-aided GMPF for wideband channel tracking under blockage.

### 5.5.1 Detection of blockage events and digital combiner design for CPD

From the blockage model in previous section, we know that when the blockage events appear or disappear, there will be significant changes in the channel responses due to the birth/death or additional attenuation of some propagation paths. In line of this idea, the detection of blockage events can be achieved with different change point detection or abrupt change detection algorithms, such as the ones described in [33–35]. In this chapter, we adopt the hypothesis test idea in [33, 35] for blockage detection, and introduce two modifications. First, the measurements at all $K$ subcarriers are collected for CPD, which can increase the effective number of measurements compared to the narrowband case. Second, a different digital combiner will be designed for

CPD rather than using the same combiner for channel tracking that maximizes the SNR.

Let us first consider a general precoder/combiner for transmission/reception. The CPD test defines two hypotheses regarding if there exist abrupt changes in the channel during transition from $(n-1)$-th channel slot to $n$ channel slot, i.e.,

$$\begin{cases} \mathcal{H}_0: \text{ No large changes exist in the channel,} \\ \mathcal{H}_1: \text{ Large changes exist in the channel.} \end{cases} \tag{5.5.1}$$

For the hypothesis test, we need to define the test statistic. To this end, we first combine (5.3.3) and (5.4.3), and stack the measurements at all subcarriers to get the following wideband signal model for all $M_{\text{tck}}$ training OFDM symbols as

$$\mathbf{y}^{(n)} \approx \left(\mathbf{I}_K \otimes \mathbf{\Phi}^{(n)} \mathbf{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)\right) \mathbf{g}^{(n)} + \mathbf{n}^{(n)}, \tag{5.5.2}$$

where $\mathbf{y}^{(n)} = [\mathbf{y}^{(n)T}[0], \ldots, \mathbf{y}^{(n)T}[K-1]]^T$ collects received wideband signals, $\mathbf{n}^{(n)} = [\mathbf{n}^{(n)T}[0], \ldots, \mathbf{n}^{(n)T}[K-1]]^T$ collects received noise, $\mathbf{g}^{(n)} = [\mathbf{g}^{(n)T}[0], \ldots, \mathbf{g}^{(n)T}[K-1]]^T$ collects the channel gains at all different subcarriers, and $\mathbf{T}\left(\boldsymbol{\xi}^{(n)}\right) = \mathbf{I}_K \otimes \mathbf{\Phi}^{(n)} \mathbf{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)$ denotes the transfer matrix of the system. Based on the measurements in (5.5.2), the test statistic can be defined as

$$\begin{aligned} T(\mathbf{y}^{(n)}) = &\left(\mathbf{y}^{(n)} - \mathbf{T}\big(\hat{\boldsymbol{\xi}}^{(n|n-1)}\big)\hat{\mathbf{g}}^{(n|n-1)}\right)^* \left(\sigma^2 \mathbf{C}_{\text{w}}^{(n)}\right)^{-1} \\ &\cdot \left(\mathbf{y}^{(n)} - \mathbf{T}\big(\hat{\boldsymbol{\xi}}^{(n|n-1)}\big)\hat{\mathbf{g}}^{(n|n-1)}\right), \end{aligned} \tag{5.5.3}$$

where the one-step prediction of angles and gains at the $n$-th slot from the previous $(n-1)$-th slot are $\hat{\boldsymbol{\xi}}^{(n|n-1)}$ and $\hat{\mathbf{g}}^{(n|n-1)}$. Moreover, $\mathbf{C}_{\text{w}}^{(n)}$ is

the covariance matrix of $\mathbf{n}^{(n)}$ and is defined as $\mathbf{C}_{\mathrm{w}}^{(n)} = \mathbf{I}_K \otimes \mathbf{D}_{\mathrm{w}}^{(n)}$, with $\mathbf{D}_{\mathrm{w}}^{(n)} \triangleq \mathrm{blkdiag}\{\mathbf{W}^{(1,n)}\mathbf{W}^{(1,n)*}, \ldots, \mathbf{W}^{(M_{\mathrm{tck}},n)}\mathbf{W}^{(M_{\mathrm{tck}},n)*}\} \in \mathbb{C}^{M_{\mathrm{tck}}L_{\mathrm{r}} \times M_{\mathrm{tck}}L_{\mathrm{r}}}$.

Inspecting (5.5.3), on one hand, if there are no abrupt changes in the channels, the difference between the received measurements and the predicted channels should be small, with only noise terms left. On the other hand, once there are abrupt changes, the difference between the received measurements and the predicted channels will contain the components of the blocked paths or the new paths. Thus, we can identify the appearance or disappearance of blockage when this difference is larger than a predetermined threshold $T_{\mathrm{th}}$, i.e., $T(\mathbf{y}^{(n)}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T_{\mathrm{th}}$. Note that this predetermined threshold can be determined based on a given false alarm probability, as shown in [33].

Note that different designs of precoders and combiners can be used in (5.5.2). In our previous work [32] we designed a frequency-flat SNR-optimal training precoder/combiner that can maximize the average received SNR based on the channel estimate of previous slot. In this chapter, we assume that this SNR-optimal precoder $\mathbf{F}^{(n)} = \mathbf{F}_{\mathrm{RF}}^{(n)}\mathbf{F}_{\mathrm{BB}}^{(n)}$ will be used at the transmitter and the SNR-optimal combiner $\mathbf{W}_Q^{(n)} = \mathbf{W}_{\mathrm{RF}}^{(n)}\mathbf{W}_{\mathrm{BB}}^{\mathrm{CE},(n)}$ will be used at the receiver for channel estimation/tracking.

For the change point detection stage, we design a different digital combiner $\mathbf{W}_{\mathrm{BB}}^{\mathrm{CPD},(n)}$ with the aim of maximizing the blockage detection probability. We exploit the fact that only a few multipath components in mmWave channels dominate the received energy, such as the LOS path and/or the strongest NLOS path. Therefore, we can focus on detecting the abrupt changes on the

main multipath components, rather than also in those marginal clusters. To this end, the new digital combiner $\mathbf{W}_{\mathrm{BB}}^{\mathrm{CPD},(n)}$ should be a spatial filter that remove the marginal clusters and emphasize the dominant ones. In line of this idea, we consider the following problem formulation

$$\mathbf{W}_{\mathrm{BB}}^{\mathrm{CPD},(n)} = \mathrm{argmin}_{\mathbf{W}} \ \|\mathbf{A}_{\mathrm{R}}\left(\hat{\boldsymbol{\phi}}^{(n|n-1)}\right) - \mathbf{W}_{\mathrm{RF}}^{(n)}\mathbf{W}\|_F^2, \tag{5.5.4}$$

where $\hat{\boldsymbol{\phi}}^{(n|n-1)}$ is the predicted mean AoAs of the main clusters from the previous channel estimate. Note that if the number of main clusters is smaller than the number of RF chains $L_{\mathrm{r}}$, the same digital combiner can be repeated to fully exploit all the RF chains. Once the new digital combiner is obtained, it can be used for CPD as shown in Fig. 5.1. We will show in Section V that this new digital combiner outperforms others in terms of CPD success rate. This justifies the proposed double digital combiner architecture.

### 5.5.2 CPD-aided GMPF strategy for wideband channel tracking under blockage

In this section, we extend the GMPF channel tracking algorithm proposed in [32] to take into consideration the impact of blockage. The difference lies in that the channel evolution models during the blockage phases should be adjusted accordingly during tracking, as indicated in (5.4.6)–(5.4.8).

Based on the measurement model in (5.5.2), the idea of GMPF is to exploit the conditional linear structures between measurements $\mathbf{y}^{(n)}$ and the gain vector $\mathbf{g}^{(n)}$, given the angle parameters $\boldsymbol{\xi}^{(n)}$. In other words, exploiting this conditional linear structure can help get better estimates for

**Algorithm 5** CPD-aided GMPF

1: CPD-aided GMPF $\left(\mathbf{Y}^{(n)},\, N_{\mathrm{PF}},\, \mathbf{\Phi}^{(n)}, \hat{\theta}_r^{(0)},\, \hat{\phi}_r^{(0)}\right)$
2: **Initialize angular particles based on prior information**
3: $\quad \theta_r^{(1|0,i)} = \hat{\theta}_r^{(0)},\, \phi_r^{(1|0,i)} = \hat{\phi}_r^{(0)}$
4: **Initialize linear particles**
5: $\quad \hat{\mathbf{g}}^{(1|0,i)} = \mathbf{0}, \quad \hat{\mathbf{C}}_{\mathbf{gg}}^{(1|0,i)} = \frac{N_{\mathrm{t}} N_{\mathrm{r}}}{\mu_L} \mathbf{I}_{\mu_L} \otimes \mathbf{F}_1 \mathbf{F}_1^*$, with $\mu_L$ the average number of multipath components in the channel, i.e., the approximation of $R$.
6: **for** $i = 1, \ldots, N_{\mathrm{PF}}$ **do**
7: $\quad$ **Evaluate the importance weights** $q^{(n,i)} = p(\mathbf{y}^{(n)}|\mathbf{\Xi}^{(n,i)}, \mathbf{Y}^{(n-1)}) p(\mathbf{g}^{(n)}|\mathbf{\Xi}^{(n)}, \mathbf{Y}^{(n)})$ **and normalize them**
8: $\quad \tilde{q}^{(n,i)} = \frac{q^{(n,i)}}{\sum_{j=1}^{N_{\mathrm{PF}}} q^{(n,j)}} \tilde{q}^{(n-1,i)}$
9: $\quad$ **PF #1 measurement update** $\quad \mathbb{P}\left(\boldsymbol{\xi}^{(n|n,i)} = \boldsymbol{\xi}^{(n|n-1,j)}\right) = \tilde{q}^{(n,j)}$
10: $\quad$ **KF measurement update for** $1 \leq i \leq N_{\mathbf{PF}}$
11: If blockage is detected:
12: $\quad$ Select corresponding gain evolution models and then update line 14–21.
13: If no blockage is detected:
14: $\quad \mathbf{K}^{(n,i)} = \mathbf{C}_{\mathbf{gy}}^{(n|n-1,i)} \mathbf{C}_{\mathbf{yy}}^{(n|n-1,i)-1}$
15: $\quad \hat{\mathbf{g}}^{(n|n,i)} = \hat{\mathbf{g}}^{(n|n-1,i)} + \mathbf{K}^{(n,i)} \left(\mathbf{y}^{(n,i)} - \boldsymbol{\mu}_y^{(n|n-1)}\right)$
16: $\quad \hat{\mathbf{C}}_{\mathbf{gg}}^{(n|n,i)} = \mathbf{C}_{\mathbf{gg}}^{(n|n-1,i)} - \mathbf{K}^{(n,i)} \mathbf{C}_{\mathbf{gy}}^{(n|n-1,i)}$
17: $\quad$ **PF #1 prediction**
18: $\quad \boldsymbol{\xi}^{(n+1|n,i)} \sim p\left(\boldsymbol{\xi}^{(n+1|n)}|\mathbf{\Xi}^{(n,i)}, \mathbf{Y}^{(n)}\right)$
19: $\quad$ **KF prediction for** $1 \leq i \leq N_{\mathbf{PF}}$
20: $\quad\quad \hat{\mathbf{g}}^{(n|n-1,i)} = \mathbf{R}^{(n)} \hat{\mathbf{g}}^{(n-1|n-1)}$
21: $\quad\quad \mathbf{C}_{\mathbf{gg}}^{(n|n-1,i)} = \mathbf{R}^{(n)} \mathbf{C}_{\mathbf{gg}}^{(n-1|n-1,i)} \mathbf{R}^{(n)*} + (\mathbf{I} - \mathbf{R}^{(n)} \mathbf{R}^{(n)*})$
22: $\quad$ **PF #2 to estimate the channel**
23: $\quad \hat{\mathbf{h}}^{(n)} = \sum_{j=1}^{N_{\mathrm{PF}}} \tilde{q}^{(n,j)} \mathbf{A}\left(\boldsymbol{\xi}^{(n|n,j)}\right) \hat{\mathbf{g}}^{(n|n,j)}$
24: $\quad$ **Reshape vectorized channel estimate to find** $\hat{\mathbf{H}}^{(n)}[k]$
25: $\quad \hat{\mathbf{H}}^{(n)}[k] = \mathrm{unvec}\{\hat{\mathbf{h}}^{(n)}\}$

the channel by analytically marginalizing out the linear state variables $\mathbf{g}^{(n)}$ when necessary. Then the problem of finding the MMSE estimator of $\mathbf{h}^{(n)} =$

$\text{vec}\{\mathbf{H}^{(n)}[0], \ldots, \mathbf{H}^{(n)}[K-1]\} = \left(\mathbf{I}_K \otimes \mathbf{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)\right)\mathbf{g}^{(n)}$ is reduced to finding its posterior conditional expectation as

$$\hat{\mathbf{h}}^{(n)}_{\text{MMSE}} \triangleq \mathbb{E}_{\mathbf{h}^{(n)}|\mathbf{Y}^{(n)}}\{\mathbf{h}^{(n)}\}, \tag{5.5.5}$$

where $\mathbf{Y}^{(n)} = \{\mathbf{y}^{(i)}\}_{i=0}^{n}$ accumulates the measurements from the very beginning until current slot. The detailed steps of the proposed CPD-aided GMPF algorithm for wideband channel tracking are summarized in Algorithm 5.

## 5.6 Numerical results

In this section, we provide some numerical results on the performance of the proposed wideband channel tracking strategy for mmWave MIMO communications under blockage.

### 5.6.1 Simulation parameters

We assume that both the AP and UE are equipped with ULAs with half-wavelength separation, with $N_{\text{t}} = 32, N_{\text{r}} = 16$ and $L_{\text{t}} = L_{\text{r}} = 4$. The phase-shifters have $N_{\text{Q,Tx}} = N_{\text{Q,Rx}} = 4$ quantization bits. The number of OFDM subcarriers is set to $K = 32$. The carrier frequency is set to $f_{\text{c}} = 60$ GHz, and the bandwidth is set to $B = 1.96$ GHz, obtained with a sampling rate $T_{\text{s}} = 0.509$ ns and a raised cosine with a roll-off factor of 0.3 to model the equivalent filtering effects.

We consider the channel tracking process along $N_{\text{slot}} = 20$ slots, while

each slot contains $M = 240$ OFDM symbols. The tracking time is set as $T_{\text{tck}} = 0.157$ ms. Moreover, we consider that the channel has a LOS component with a single ray [13], and $C = 2$ NLOS clusters, each contributing with a number of rays $R_c \sim \mathcal{U}[6, 10]$. We set $v = 30$ m/s the relative velocity of UE. To compute the correlation matrix $\mathbf{R}_r$, we consider the parameters $a = 0.91$, $c = 0$ and $b = 0.1/\lambda_c$ as in [32], with $\lambda_c$ the wavelength. We also assume that the correlation matrix is the same for all paths in the simulations. For the angle evolution, in our simulation setting, considering a minimum distance between receiver and transmitter of 5 m and $v_{\max} = 30$ m/s, we have $|\Delta\phi_r^{(n)}| < 0.042°$, and therefore we set the value for the standard deviation of the Laplacian distribution that model the angle spread as 0.57°, ten orders the magnitude bigger than the maximum expected variation. Last, $N_{\text{PF}} = 50$ is considered. For the blockage statistics, the additional pathloss is assumed to be Gaussian distributed as $\mathcal{N}(13.4, 2)$ dB.

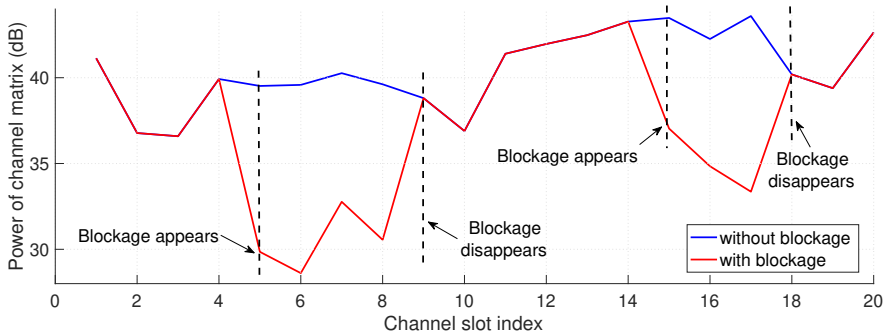



Figure 5.2: Comparison of channel power with and without blockage.

### 5.6.2 Illustration of channel blockage modeling

In Fig. 5.2 , we provide an example to illustrate the impact of blockage by comparing the power of channel matrices with and without blockage. It can be seen that the blockage appears from slot 5 to slot 8 and slot 15 to slot 17, as the significant decrease in channel power indicates additional attenuation by blockage.

### 5.6.3 Comparison of CPD performance with different combiners

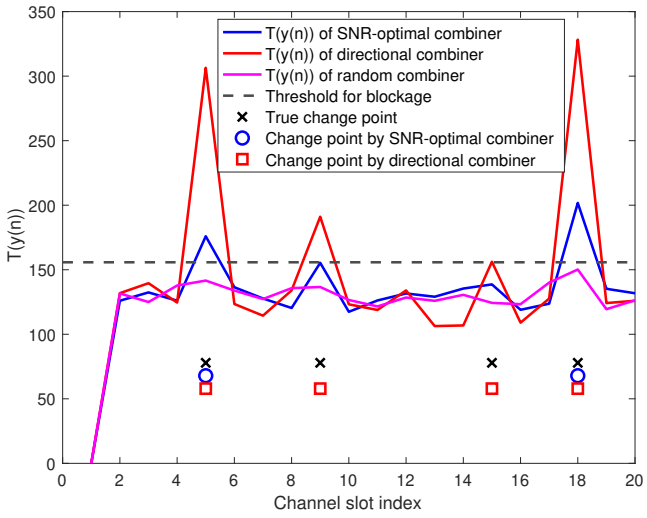



Figure 5.3: Comparison of CPD performance with different combiners, using $M_{\text{tck}} = 1$, SNR = -10dB, and with a threshold determined by a false positive rate of 0.01.

In Fig. 5.3, we illustrate one realization of the CPD test statistics with different combiners. It can be seen that the CPD test based on the new directional digital combiner can successfully identify all the abrupt changes. The

CPD test based on the SNR-optimal combiner may induce some missed alarm errors in this case. We further evaluate the receiver operating characteristics (ROCs), i.e., true positive rate vs. false positive rate, of the CPD test with different combiners in Fig. 5.4, which shows that the proposed directional digital combiner can improve the success rate at any given false alarm probability. These results demonstrate the effectiveness of our proposed new digital combiner design for blockage detection.

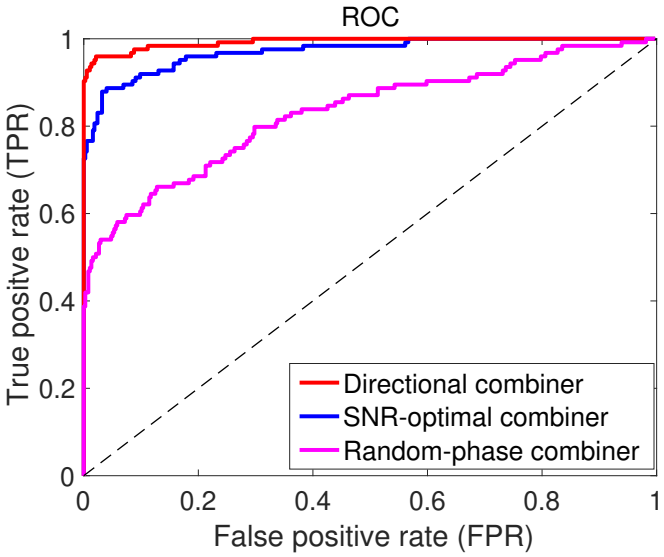



Figure 5.4: Receiver operating characteristic (ROC) curves of the CPD with different combiners, when $M_{\text{tck}} = 1$, SNR = -10dB.

### 5.6.4 Performances of channel tracking under blockage

In Fig. 5.5, we then compare the NMSE performances of the proposed wideband channel tracking strategy under blockage corresponding to the scenario in Fig. 5.2 . The NMSE of channel tracking along the $N_{\text{slot}}$ com-

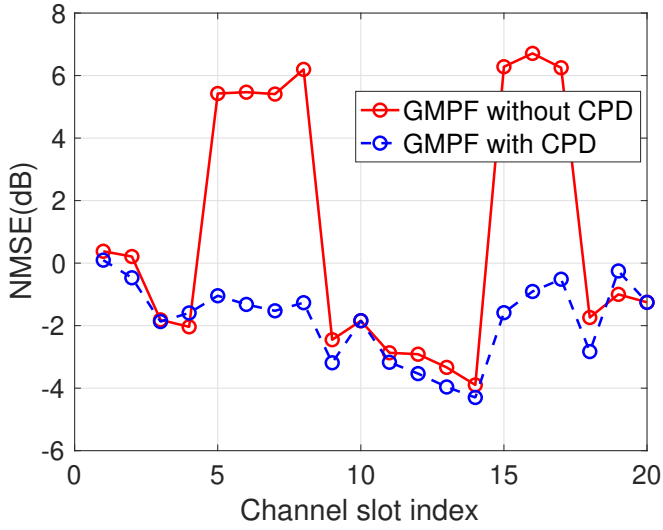



Figure 5.5: NMSE performances of wideband channel tracking under blockage, where $M_{\text{tck}} = 8$, SNR = -10dB, and Rician K-factor = 0 dB.

munication slots is defined as $\text{NMSE}(n) = 10\log_{10}\frac{1}{K}\sum_{k=0}^{K-1}\frac{\|\hat{\mathbf{H}}^{(n)}[k]-\mathbf{H}^{(n)}[k]\|_F^2}{\|\mathbf{H}^{(n)}[k]\|_F^2}$. For comparison, we also include the results of the original GMPF algorithm without CPD for the channel tracking under blockage. The number of OFDM symbols used for tracking is set to $M_{\text{tck}} = 8$ in this case. When the blockage events are detected with CPD algorithms, the evolution models under blockage can be adjusted for GMPF based channel tracking, as shown in line 14–21 in Algorithm 5. As can be seen, the proposed CPD aided GMPF channel tracking strategy can achieve a good channel tracking accuracy along the time, especially during the blockage phases. Nevertheless, without blockage detection, the original GMPF algorithm failed to track the channels when the blockage exists. This is due to the mismatch between the channel gain evolution models and the realistic channel measurements under blockage.

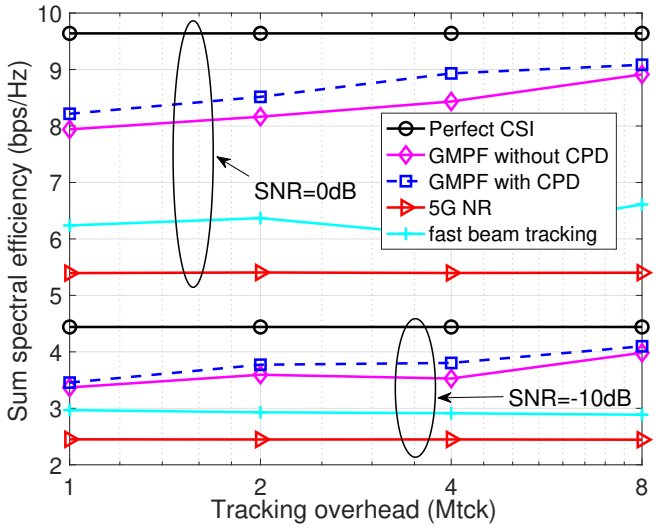



Figure 5.6: SE performances of wideband channel tracking under blockage as a function of tracking overhead $M_{\text{tck}}$, where $N_{\text{slot}} = 20$ and $N_{\text{s}} = 2$..

In Fig. 5.6, we further evaluate the ergodic SE performances as a function of the number of tracking overhead $M_{\text{tck}}$. The SE of the $n$-th slot is defined as $\mathcal{R}^{(n)} = \frac{M - M_{\text{tck}}}{M} \sum_{k=0}^{K-1} \log_2 \left| \mathbf{I}_{N_{\text{s}}} + \frac{\text{SNR}}{N_{\text{s}}} \hat{\mathbf{H}}_{\text{eff}}^{(n)}[k] \hat{\mathbf{H}}_{\text{eff}}^{(n)*}[k] \right|$, where $\hat{\mathbf{H}}_{\text{eff}}^{(n)}[k]$ is the effective channel estimates with eigen-beamforming. The sum SE is then averaged over the $N_{\text{slot}} = 20$ slots. For comparison, we also include the beam management procedure in 5G NR and the fast beam tracking scheme proposed in [14], which only exploits statistical priors of angles. As can be seen, the proposed CPD aided GMPF based wideband channel tracking scheme can achieve higher SE performances than the original GMPF algorithm, 5G NR beam management and the fast beam tracking scheme, for the better channel tracking accuracy during the blockage phases. Note that the SE has been averaged over the whole transmission interval, which means if the blockage

events happen frequently, the performance gains of the proposed strategy can be even larger. This further demonstrates the effectiveness of the proposed wideband channel tracking strategy under blockage.

## 5.7 Conclusions

In this chapter, we derived a channel tracking strategy for wideband mmWave MIMO systems robust to blockage. We first established a statistical wideband channel model for mmWave communications by exploiting the statistical knowledge of both the channel parameters and the blockage events. We then proposed a CPD algorithm to identify the appearance and disappearance of blockage events. Afterwards, we proposed a CPD-aided Bayesian algorithm for channel tracking under blockage. A double digital combiner structure was also designed to enhance both the CPD success rate and the channel tracking accuracy. Simulations show that the proposed CPD-aided wideband channel tracking scheme is robust to blockage, and can achieve a good channel tracking performance even in high mobility scenarios.

# Chapter 6

# Conclusion

Millimeter wave communication is a key technique for 5G cellular communication systems and wireless local area networks (WLAN). To realize the potential of mmWave communications, CSI acquisition is a vital stage for link configuration. Nevertheless, considering the high dimensions of the channel matrices, the low SNR conditions before beamforming, the inevitable hardware imperfections and the fast motion in the vehicular scenarios, efficient channel estimation or tracking are challenging at mmWave. Prior work on this topic either assumes ideal models or simplify the constraints, making previously proposed solutions infeasible in some realistic scenarios. Motivated by these limitations in previous work, in this dissertation we presented novel solutions to facilitate and improve mmWave MIMO link configuration under different practical conditions. We present a summary of the main contributions in this dissertation as follows.

## 6.1 Summary

In Chapter 2 of this dissertation, we investigated the channel estimation problem for hybrid wideband mmWave MIMO systems under the impact

of hardware impairments. We proposed novel dictionary learning based solutions to find the best sparsifying dictionaries for channel sparse representation. Two specific dictionary learning algorithms were formulated and optimized by exploiting the special channel properties of mmWave MIMO as well as the structures of hardware imperfections, achieving the advantages of both high performance gains and low computation burden. Numerical results show that with the proposed dictionary learning approaches, compressive channel estimation for mmWave MIMO systems can be done at low SNR with a low training overhead.

In Chapter 3 of this dissertation, we considered a more practical scenario for mmWave MIMO communications by incorporating the impact of both hardware impairments and the beam squint effect. We proposed a dictionary learning and channel estimation algorithm that exploit the channel response at different subcarriers in the presence of beam squint. The effectiveness and performance enhancement have also been demonstrated by numerical results under different system parameters.

In Chapter 4 and Chapter 5 of this dissertation, we proposed efficient wideband channel tracking solutions for mmWave MIMO systems when the variation of the channel is smooth and also when there are abrupt changes caused by blockage events. We first designed an efficient Bayesian channel tracking algorithm to facilitate the low-overhead tracking without channel blockage, and then extend the method by generalizing the channel models with statistics of blockage events. We also proposed a novel double digital combiner

architecture to enhance the channel blockage detection accuracy and the resultant Bayesian channel tracking performance under blockage. Simulations show that, even under high mobility conditions or frequent blockage disturbance, the proposed wideband channel tracking solutions are able to maintain satisfying CSI acquisition accuracy at low overhead.

## 6.2 Future work

This dissertation has drew attentions to the typical limitations or imperfections in practical mmWave MIMO communication systems and presented some solutions for specific problems. Nevertheless, the research on powerful mmWave MIMO link configuration solutions is far from over. On one hand, the work in this dissertation still considers some basic assumptions like single-user case, single polarization, or perfect time and frequency synchronization. On the other hand, new use cases and scenarios have their own problems to be solved. Dealing with these limitations or assumptions leads to new future research directions.

### 6.2.1 Link configuration for imperfect polarized antenna systems

This dissertation and most of prior works on the topic of mmWave MIMO link configuration have assumed the antenna geometry to be ULA or uniform rectangular array (URA), but in practical 5G LTE or NR standards, antenna arrays are typically comprised of single- or dual- polarized (DP) antennas. Polarization parameters are critical in successful operation on

mmWave MIMO systems [108]. It is widely used at the AP to account for the random orientations of the UE and considered for encrypting the classified information in secure mmWave communications. Therefore, accurate angle and polarization parameter estimations are particularly essential in mmWave MIMO systems with DP antennas.

The typical channel models for MIMO systems with DP antennas in the literature [108–112] assume same polarization angles for all antenna elements in the arrays, i.e., vertical/horizontal ($0°/90°$) polarization or ($\pm 45°$) cross polarization. Nevertheless, this kind of channel modeling is too ideal for practical implementation, where the polarization angles of each antenna may be different in a general case, due to calibration/manufacturing errors or deliberate design. In addition, non-ideal antenna radiation patterns should also be considered. Moreover, the difference in orientation angles between the transmit and receive antenna arrays, as well as the channel depolarization effect should be taken into consideration for the channel modeling and parameter estimation [113]. Based on these considerations, efficient parameter and channel estimation for mmWave MIMO systems with imperfect DP antennas should be carefully investigated. Inspired by the work in Chapter 2 and Chapter 3 of this dissertation, dictionary learning can be a good direction to deal with the aforementioned problems. Other potential solutions are also necessary and need more efforts.

### 6.2.2 Channel estimation for high mobility mmWave V2X under beam squint

In this dissertation, we investigated the impact of beam squint effect on mmWave MIMO link configuration in Chapter 3 for a time-flat frequency-selective channel. In practice, a more realistic mmWave V2X communication scenario will be both time-selective and frequency-selective. Therefore, it is necessary to consider the link configuration for practical high mobility wideband V2X communication systems under beam squint.

Specifically, the doubly-selective (DS) channel, i.e., both time- and frequency-selective channel, will be encountered in many wireless access links, such as high-speed trains, vehicle-to-vehicle (V2V), and mmWave communications. Considering the channel estimation for an MIMO-OFDM system over a DS channel, the number of channel coefficients to be estimated during each OFDM symbol is scaled with the numbers of subcarriers (sample-dimension), delay taps (delay-dimension) as well as the transmit/receive antennas (antenna-dimension) Inspecting the literature [114–116], their assumption on the common sparsity along the antenna-dimension is problematic for large-scale MIMO regimes or wideband systems when the beam squint effect is considered. Moreover, due to the delay differences of each data symbol at different antennas, the assumption that common sparsity support is shared among different transmit antennas does not longer hold true . More importantly, the commonly adopted antenna-dimension MIMO channel models in the existing literature that use the conventional array response vectors for

channel modeling are not valid either under beam squint circumstances. In other words, the previous works that uses Vandermonde structured array response vectors to form the non-orthogonal dictionary basis is questionable either. Considering these limitations, novel channel estimation solutions for high mobility mmWave V2X systems under beam squint are necessary and again the dictionary learning idea proposed in this dissertation is a promising direction.

### 6.2.3 Wideband channel tracking without statistical priors of channel parameter or blockage

In Chapter 4 and Chapter 5, we considered the problem of wideband channel tracking for mmWave MIMO systems without and with channel blockage, by introducing statistical channel models that include the evolution models for channel gains and AoAs/AoDs, as well as the statistics of blockage events. This has enabled the design of the proposed Bayesian channel tracking solutions. Though many measurement campaigns and field test works have provided some accurate statistical modeling of channel gains, AoAs/AoDs and blockage, these modeling is not enough for more general realistic scenarios. In other words, considering the complex practical communication scenario, the true statistical priors of channel parameters or blockage may not be known or accurate. In this sensing, the proposed Bayesian solutions may not be able to work well for a satisfying performance. Considering these, it is necessary to incorporate some schemes, which can learn the statistical information of the environment and adjust the models accordingly. The widely used artificial

intelligence (AI) or deep learning tools can be a potential candidate.

In the Chapter 5 of this dissertation, we adopted the blockage detection scheme to help correct the right channel statistical models and compensate the influence of blockage. To be noted that, this kind of solution still needs some training resources in each channel slot to help identify the status of channel blockage. Another alternative solution could be a general Bayesian algorithm that uses hierarchical channel modeling to incorporate all the channel statistics. For example, the variational Bayesian inference (VBI) algorithm [117] can be a promising direction, Specifically, VBI models the channel gain, angle and blockage priors with unknown hyperprior parameters. These hyperprior parameters can be updated and adjusted along the time as the algorithms operate, so that they can be better optimized and adapted to the realistic environments. In doing so, not only can the blockage detection step be omitted, but also can the channel tracking performances be improved. Last, other potential solutions to handle the wideband channel tracking without statistical priors of channel parameters and blockage are needed as well.

# Appendices

# Appendix A

# Calculation of FIM $\mathbf{I}(\boldsymbol{\xi})$

In this appendix, we provide the detailed derivations for the expressions of FIM $\mathbf{I}(\boldsymbol{\xi})$ based on (2.4.4). We first consider the diagonal blocks of the FIM $\mathbf{I}(\boldsymbol{\xi})$ with respect to (w.r.t.) each type of above unknown parameter. Specifically, the elements of the FIM $\mathbf{I}(\boldsymbol{\xi})$ w.r.t. the AODs/AOAs ($\forall i,j \in \mathcal{I}(N_{\mathrm{p}}N_{\mathrm{ray}})$) are given as

$$
\begin{aligned}
&[\mathbf{I}(\boldsymbol{\xi})]_{\theta_i,\phi_j} = \frac{2}{\sigma^2}\Re\left\{\frac{\partial \boldsymbol{\mu}_{\mathrm{w}}{}^*(\boldsymbol{\xi})}{\partial\theta_i}\frac{\partial \boldsymbol{\mu}_{\mathrm{w}}(\boldsymbol{\xi})}{\partial\phi_j}\right\} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Big\{\overline{g_i[c]}\frac{\partial \mathbf{a}_{\mathrm{T}}(\theta_i)^T\otimes\mathbf{a}_{\mathrm{R}}(\phi_i)^*}{\partial\theta_i}\\
&\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\frac{\partial\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)\otimes\mathbf{a}_{\mathrm{R}}(\phi_j)}{\partial\phi_j}g_j[c]\Big\}\\
&= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Big\{tr\Big\{\overline{g_i[c]}g_j[c]\big(\overline{\Gamma}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}\\
&\cdot\ \big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)\frac{\partial\mathbf{a}_{\mathrm{T}}(\theta_i)^T}{\partial\theta_i}\otimes\frac{\partial\mathbf{a}_{\mathrm{R}}(\phi_j)}{\partial\phi_j}\mathbf{a}_{\mathrm{R}}(\phi_i)^*\Big\}\Big\},
\end{aligned}
\tag{A.0.1}
$$

where, recalling the expression of $\mathbf{a}_{\mathrm{R}}(\phi_j)$ in (2.3.2), the derivative of $\mathbf{a}_{\mathrm{R}}(\phi_j)$ w.r.t. $\phi_j$ is

$$
\left[\frac{\partial\mathbf{a}_{\mathrm{R}}(\phi_j)}{\partial\phi_j}\right]_m = -\mathrm{j}\frac{2\pi(md+\epsilon_{\mathrm{r},m})}{\lambda}\cos(\phi_j)\left[\mathbf{a}_{\mathrm{R}}(\phi_j)\right]_m, \forall m = 0,\ldots,N_{\mathrm{r}}-1,
\tag{A.0.2}
$$

and the derivative of $\mathbf{a}_{\mathrm{T}}(\theta_i)$ w.r.t. $\theta_i$ can be obtained accordingly. Likewise, we have

$$[\mathbf{I}(\boldsymbol{\xi})]_{\phi_i,\theta_j} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1} \Re\Big\{ tr\Big\{\overline{g_i[c]}g_j[c]\big(\overline{\Gamma}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}} \cdot \big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\frac{\partial\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)}{\partial\theta_j}\mathbf{a}_{\mathrm{T}}(\theta_i)^T\otimes\mathbf{a}_{\mathrm{R}}(\phi_j)\frac{\partial\mathbf{a}_{\mathrm{R}}(\phi_i)^*}{\partial\phi_i}\Big\}\Big\}, \quad \text{(A.0.3)}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{\theta_i,\theta_j} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1} \Re\Big\{ tr\Big\{\overline{g_i[c]}g_j[c]\big(\overline{\Gamma}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}} \cdot\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\frac{\partial\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)}{\partial\theta_j}\frac{\partial\mathbf{a}_{\mathrm{T}}(\theta_i)^T}{\partial\theta_i}\otimes\mathbf{a}_{\mathrm{R}}(\phi_j)\mathbf{a}_{\mathrm{R}}(\phi_i)^*\Big\}\Big\}, \quad \text{(A.0.4)}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{\phi_i,\phi_j} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1} \Re\Big\{ tr\Big\{\overline{g_i[c]}g_j[c]\big(\overline{\Gamma}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}} \cdot \big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)\mathbf{a}_{\mathrm{T}}(\theta_i)^T\otimes\frac{\partial\mathbf{a}_{\mathrm{R}}(\phi_j)}{\partial\phi_j}\frac{\partial\mathbf{a}_{\mathrm{R}}(\phi_i)^*}{\partial\phi_i}\Big\}\Big\}. \quad \text{(A.0.5)}$$

Similar computations for the elements of the FIM $\mathbf{I}(\boldsymbol{\xi})$ w.r.t. the path gains $(\forall i,j\in\mathcal{I}(N_{\mathrm{p}}N_{\mathrm{ray}}),\forall c\in\mathcal{I}(N_{\mathrm{c}}))$ yield

$$[\mathbf{I}(\boldsymbol{\xi})]_{g_i^{\mathrm{R}}[c],g_j^{\mathrm{R}}[c]} = \frac{2}{\sigma^2}\Re\left\{\frac{\partial{\boldsymbol{\mu}_{\mathrm{w}}}^*(\boldsymbol{\xi})}{\partial g_i^{\mathrm{R}}[c]}\frac{\partial\boldsymbol{\mu}_{\mathrm{w}}(\boldsymbol{\xi})}{\partial g_j^{\mathrm{R}}[c]}\right\} = \frac{2}{\sigma^2}\Re\Big\{tr\Big\{\big(\overline{\Gamma}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^* \cdot{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}}\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\big(\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)\mathbf{a}_{\mathrm{T}}(\theta_i)^T\big)\otimes\big(\mathbf{a}_{\mathrm{R}}(\phi_j)\mathbf{a}_{\mathrm{R}}(\phi_i)^*\big)\Big\}\Big\}, \quad \text{(A.0.6)}$$

and $[\mathbf{I}(\boldsymbol{\xi})]_{g_i^{\mathrm{I}}[c],g_j^{\mathrm{I}}[c]} = [\mathbf{I}(\boldsymbol{\xi})]_{g_i^{\mathrm{R}}[c],g_j^{\mathrm{R}}[c]}$, and

$$[\mathbf{I}(\boldsymbol{\xi})]_{g_i^{\mathrm{R}}[c],g_j^{\mathrm{I}}[c]} = \frac{2}{\sigma^2}\Re\Big\{tr\Big\{\mathrm{j}\big(\overline{\Gamma}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}} \cdot\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)\mathbf{a}_{\mathrm{T}}(\theta_i)^T\otimes\mathbf{a}_{\mathrm{R}}(\phi_j)\mathbf{a}_{\mathrm{R}}(\phi_i)^*\Big\}\Big\}, \quad \text{(A.0.7)}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{g_i^{\mathrm{I}}[c],g_j^{\mathrm{R}}[c]} = \frac{2}{\sigma^2}\Re\left\{tr\left\{-\mathrm{j}\big(\overline{\Gamma}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)^*\big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}\right.\right.$$
$$\left.\left.\cdot\ \big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}\big)\big(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}\big)\overline{\mathbf{a}_{\mathrm{T}}}(\theta_j)\mathbf{a}_{\mathrm{T}}(\theta_i)^T\otimes\mathbf{a}_{\mathrm{R}}(\phi_j)\mathbf{a}_{\mathrm{R}}(\phi_i)^*\right\}\right\}. \quad \text{(A.0.8)}$$

Similar computations for the elements of FIM $\mathbf{I}(\boldsymbol{\xi})$ w.r.t. the antenna gain errors yield

$$[\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t},i},g_{\mathrm{r},j}} = \frac{2}{\sigma^2}\Re\left\{\frac{\partial\boldsymbol{\mu}_{\mathrm{w}}{}^*(\boldsymbol{\xi})}{\partial g_{\mathrm{t},i}}\frac{\partial\boldsymbol{\mu}_{\mathrm{w}}(\boldsymbol{\xi})}{\partial g_{\mathrm{r},j}}\right\}$$
$$= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\mathrm{diag}\{e^{\mathrm{j}\phi_{\mathrm{t},i}}\mathbf{e}_i\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}^*)(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}\right.$$
$$\left.\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\mathrm{diag}\{e^{\mathrm{j}\phi_{\mathrm{r},j}}\mathbf{e}_j\})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},\ \forall i\in\mathcal{I}(N_{\mathrm{t}}), j\in\mathcal{I}(N_{\mathrm{r}}), \quad \text{(A.0.9)}$$

where $\mathbf{e}_i$ is the $i$-th column of the identity matrix of appropriate dimension. Besides, we have

$$[\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r},i},g_{\mathrm{t},j}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\boldsymbol{\Gamma}_{\mathrm{T}}^T\otimes\mathrm{diag}\{e^{-\mathrm{j}\phi_{\mathrm{r},i}}\mathbf{e}_i\})(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\right.$$
$$\left.\cdot\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\mathrm{diag}\{e^{-\mathrm{j}\phi_{\mathrm{t},j}}\mathbf{e}_j\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},$$
$$\forall i\in\mathcal{I}(N_{\mathrm{r}}), j\in\mathcal{I}(N_{\mathrm{t}}), \quad \text{(A.0.10)}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t},i},g_{\mathrm{t},j}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\mathrm{diag}\{e^{\mathrm{j}\phi_{\mathrm{t},i}}\mathbf{e}_i\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}^*)(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\right.$$
$$\left.\cdot\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\mathrm{diag}\{e^{-\mathrm{j}\phi_{\mathrm{t},j}}\mathbf{e}_j\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},$$
$$\forall i,j\in\mathcal{I}(N_{\mathrm{t}}), \quad \text{(A.0.11)}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r},i},g_{\mathrm{r},j}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\boldsymbol{\Gamma}_{\mathrm{T}}^T\otimes\mathrm{diag}\{e^{-\mathrm{j}\phi_{\mathrm{r},i}}\mathbf{e}_i\})(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\right.$$
$$\left.\cdot\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\mathrm{diag}\{e^{\mathrm{j}\phi_{\mathrm{r},j}}\mathbf{e}_j\})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},$$
$$\forall i,j\in\mathcal{I}(N_{\mathrm{r}}). \quad \text{(A.0.12)}$$

Similar computations for the elements of FIM $\mathbf{I}(\boldsymbol{\xi})$ w.r.t. the antenna phase errors yield

$$[\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t},i},\phi_{\mathrm{t},j}} = \frac{2}{\sigma^2}\Re\left\{\frac{\partial \boldsymbol{\mu}_{\mathrm{w}}{}^*(\boldsymbol{\xi})}{\partial \phi_{\mathrm{t},i}}\frac{\partial \boldsymbol{\mu}_{\mathrm{w}}(\boldsymbol{\xi})}{\partial \phi_{\mathrm{t},j}}\right\}$$
$$= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\mathrm{diag}\{\mathrm{j}g_{\mathrm{t},i}e^{\mathrm{j}\phi_{\mathrm{t},i}}\mathbf{e}_i\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}^*)(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}\right.$$
$$\left.\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\mathrm{diag}\{-\mathrm{j}g_{\mathrm{t},j}e^{-\mathrm{j}\phi_{\mathrm{t},j}}\mathbf{e}_j\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},\quad \forall i,j\in\mathcal{I}(N_{\mathrm{t}}), \tag{A.0.13}$$

and

$$[\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r},i},\phi_{\mathrm{r},j}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\boldsymbol{\Gamma}_{\mathrm{T}}^T\otimes\mathrm{diag}\{-\mathrm{j}g_{\mathrm{r},i}e^{-\mathrm{j}\phi_{\mathrm{r},i}}\mathbf{e}_i\})\right.$$
$$\left.\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\mathrm{diag}\{\mathrm{j}g_{\mathrm{r},j}e^{\mathrm{j}\phi_{\mathrm{r},j}}\mathbf{e}_j\})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},$$
$$\forall i,j\in\mathcal{I}(N_{\mathrm{r}}), \tag{A.0.14}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t},i},\phi_{\mathrm{r},j}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\mathrm{diag}\{\mathrm{j}g_{\mathrm{t},i}e^{\mathrm{j}\phi_{\mathrm{t},i}}\mathbf{e}_i\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}}^*)\right.$$
$$\left.\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\mathrm{diag}\{\mathrm{j}g_{\mathrm{r},j}e^{\mathrm{j}\phi_{\mathrm{r},j}}\mathbf{e}_j\})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},$$
$$\forall i\in\mathcal{I}(N_{\mathrm{t}}),j\in\mathcal{I}(N_{\mathrm{r}}), \tag{A.0.15}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r},i},\phi_{\mathrm{t},j}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\left\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\boldsymbol{\Gamma}_{\mathrm{T}}^T\otimes\mathrm{diag}\{-\mathrm{j}g_{\mathrm{r},i}e^{-\mathrm{j}\phi_{\mathrm{r},i}}\mathbf{e}_i\})\right.$$
$$\left.\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*\boldsymbol{\Phi}_{\mathrm{w}}{}^*\boldsymbol{\Phi}_{\mathrm{w}}(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\mathrm{diag}\{-\mathrm{j}g_{\mathrm{t},j}e^{-\mathrm{j}\phi_{\mathrm{t},j}}\mathbf{e}_j\}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\right\},$$
$$\forall i\in\mathcal{I}(N_{\mathrm{r}}),j\in\mathcal{I}(N_{\mathrm{t}}). \tag{A.0.16}$$

Similar computations for the elements of FIM $\mathbf{I}(\boldsymbol{\xi})$ w.r.t. the antenna spacing errors yield

$$
\begin{aligned}
[\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t},i},\epsilon_{\mathrm{r},j}} &= \frac{2}{\sigma^2}\Re\left\{\frac{\partial {\boldsymbol{\mu}_{\mathrm{w}}}^*(\boldsymbol{\xi})}{\partial \epsilon_{\mathrm{t},i}}\frac{\partial \boldsymbol{\mu}_{\mathrm{w}}(\boldsymbol{\xi})}{\partial \epsilon_{\mathrm{r},j}}\right\} \\
&= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Big\{\mathbf{g}^*[c]\Big(\frac{\partial \overline{\mathbf{A}}_{\mathrm{T}}}{\partial \epsilon_{\mathrm{t},i}} \star \mathbf{A}_{\mathrm{R}}\Big)^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}}(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}}) \\
&\quad\cdot(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})\Big(\overline{\mathbf{A}}_{\mathrm{T}}\star\frac{\partial \mathbf{A}_{\mathrm{R}}}{\partial \epsilon_{\mathrm{r},j}}\Big)\mathbf{g}[c]\Big\}, \quad \forall i\in\mathcal{I}(N_{\mathrm{t}}), j\in\mathcal{I}(N_{\mathrm{r}}),
\end{aligned}
\tag{A.0.17}
$$

where the derivative of $\mathbf{A}_{\mathrm{R}}$ w.r.t. $\epsilon_{\mathrm{r},j}$ is expressed as

$$
\left[\frac{\partial \mathbf{A}_{\mathrm{R}}}{\partial \epsilon_{\mathrm{r},j}}\right]_{:,\ell} = \frac{\partial \mathbf{a}_{\mathrm{R}}(\phi_\ell)}{\partial \epsilon_{\mathrm{r},j}} = -\mathrm{j}\frac{2\pi}{\lambda}\sin(\phi_\ell)[\mathbf{a}_{\mathrm{R}}(\phi_\ell)]_j\cdot\mathbf{e}_j,\ \forall \ell\in\mathcal{I}(N_{\mathrm{p}}N_{\mathrm{ray}}), j\in\mathcal{I}(N_{\mathrm{r}}),
\tag{A.0.18}
$$

and the derivative of $\mathbf{A}_{\mathrm{T}}$ w.r.t. $\epsilon_{\mathrm{t},i}$ can be expressed accordingly. Besides, we have

$$
\begin{aligned}
[\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r},i},\epsilon_{\mathrm{t},j}} &= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Big\{\mathbf{g}^*[c]\Big(\overline{\mathbf{A}}_{\mathrm{T}}\star\frac{\partial \mathbf{A}_{\mathrm{R}}}{\partial \epsilon_{\mathrm{r},i}}\Big)^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}} \\
&\quad\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})\Big(\frac{\partial \overline{\mathbf{A}}_{\mathrm{T}}}{\partial \epsilon_{\mathrm{t},j}}\star\mathbf{A}_{\mathrm{R}}\Big)\mathbf{g}[c]\Big\},\ \forall i\in\mathcal{I}(N_{\mathrm{r}}), j\in\mathcal{I}(N_{\mathrm{t}}),
\end{aligned}
\tag{A.0.19}
$$

$$
\begin{aligned}
[\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t},i},\epsilon_{\mathrm{t},j}} &= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Big\{\mathbf{g}^*[c]\Big(\frac{\partial \overline{\mathbf{A}}_{\mathrm{T}}}{\partial \epsilon_{\mathrm{t},i}}\star\mathbf{A}_{\mathrm{R}}\Big)^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}} \\
&\quad\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})\Big(\frac{\partial \overline{\mathbf{A}}_{\mathrm{T}}}{\partial \epsilon_{\mathrm{t},j}}\star\mathbf{A}_{\mathrm{R}}\Big)\mathbf{g}[c]\Big\},\ \forall i,j\in\mathcal{I}(N_{\mathrm{t}}),
\end{aligned}
\tag{A.0.20}
$$

$$
\begin{aligned}
[\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r},i},\epsilon_{\mathrm{r},j}} &= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Big\{\mathbf{g}^*[c]\Big(\overline{\mathbf{A}}_{\mathrm{T}}\star\frac{\partial \mathbf{A}_{\mathrm{R}}}{\partial \epsilon_{\mathrm{r},i}}\Big)^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}} \\
&\quad\cdot(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\mathbf{C}_{\mathrm{R}})(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})\Big(\overline{\mathbf{A}}_{\mathrm{T}}\star\frac{\partial \mathbf{A}_{\mathrm{R}}}{\partial \epsilon_{\mathrm{r},j}}\Big)\mathbf{g}[c]\Big\},\ \forall i,j\in\mathcal{I}(N_{\mathrm{r}}).
\end{aligned}
\tag{A.0.21}
$$

Finally, similar computations for the elements of FIM $\mathbf{I}(\boldsymbol{\xi})$ w.r.t. the coupling matrix coefficients yield

$$[\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t},i,j},c_{\mathrm{r},m,n}} = \frac{2}{\sigma^2}\Re\left\{\frac{\partial {\boldsymbol{\mu}_{\mathrm{w}}}^*(\boldsymbol{\xi})}{\partial c_{\mathrm{t},i,j}}\frac{\partial \boldsymbol{\mu}_{\mathrm{w}}(\boldsymbol{\xi})}{\partial c_{\mathrm{r},m,n}}\right\}$$
$$= \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Bigg\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*\Big(\frac{\partial\overline{\mathbf{C}}_{\mathrm{T}}}{\partial c_{\mathrm{t},i,j}}\otimes\mathbf{C}_{\mathrm{R}}\Big)^*{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}}$$
$$\cdot\Big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\frac{\partial\mathbf{C}_{\mathrm{R}}}{\partial c_{\mathrm{r},m,n}}\Big)(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\Bigg\},$$
$$\forall 1<i<j<N_{\mathrm{t}}, 1<m<n<N_{\mathrm{r}}, \quad \text{(A.0.22)}$$

where the derivative of $\mathbf{C}_{\mathrm{R}}$ w.r.t. $c_{\mathrm{r},m,n}$ is an $N_{\mathrm{r}}\times N_{\mathrm{r}}$ matrix with ones at the indices of $c_{\mathrm{r},m,n}$ and zeros otherwise, and the derivative of $\mathbf{C}_{\mathrm{T}}$ w.r.t. $c_{\mathrm{t},i,j}$ can be expressed accordingly. Besides, we have

$$[\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r},i,j},c_{\mathrm{t},m,n}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Bigg\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*\Big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\frac{\partial\mathbf{C}_{\mathrm{R}}}{\partial c_{\mathrm{r},i,j}}\Big)^*$$
$$\cdot{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}}\Big(\frac{\partial\overline{\mathbf{C}}_{\mathrm{T}}}{\partial c_{\mathrm{t},m,n}}\otimes\mathbf{C}_{\mathrm{R}}\Big)(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\Bigg\},$$
$$\forall 1<i<j<N_{\mathrm{r}}, 1<m<n<N_{\mathrm{t}}, \quad \text{(A.0.23)}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t},i,j},c_{\mathrm{t},m,n}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Bigg\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*\Big(\frac{\partial\overline{\mathbf{C}}_{\mathrm{T}}}{\partial c_{\mathrm{t},i,j}}\otimes\mathbf{C}_{\mathrm{R}}\Big)^*$$
$$\cdot{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}}\Big(\frac{\partial\overline{\mathbf{C}}_{\mathrm{T}}}{\partial c_{\mathrm{t},m,n}}\otimes\mathbf{C}_{\mathrm{R}}\Big)(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\Bigg\},$$
$$\forall 1<i<j<N_{\mathrm{t}}, 1<m<n<N_{\mathrm{t}}, \quad \text{(A.0.24)}$$

$$[\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r},i,j},c_{\mathrm{r},m,n}} = \frac{2}{\sigma^2}\sum_{c=0}^{N_{\mathrm{c}}-1}\Re\Bigg\{\mathbf{g}^*[c](\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})^*(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})^*\Big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\frac{\partial\mathbf{C}_{\mathrm{R}}}{\partial c_{\mathrm{r},i,j}}\Big)^*$$
$$\cdot{\boldsymbol{\Phi}_{\mathrm{w}}}^*\boldsymbol{\Phi}_{\mathrm{w}}\Big(\overline{\mathbf{C}}_{\mathrm{T}}\otimes\frac{\partial\mathbf{C}_{\mathrm{R}}}{\partial c_{\mathrm{r},m,n}}\Big)(\overline{\boldsymbol{\Gamma}}_{\mathrm{T}}\otimes\boldsymbol{\Gamma}_{\mathrm{R}})(\overline{\mathbf{A}}_{\mathrm{T}}\star\mathbf{A}_{\mathrm{R}})\mathbf{g}[c]\Bigg\},$$
$$\forall 1<i<j<N_{\mathrm{r}}, 1<m<n<N_{\mathrm{r}}. \quad \text{(A.0.25)}$$

Note that the off-diagonal blocks of the FIM $\mathbf{I}(\boldsymbol{\xi})$ between two different types of parameters can be obtained similarly following (2.4.4) and the derivations of above diagonal blocks, and thus are omitted for space limitation. The complete FIM $\mathbf{I}(\boldsymbol{\xi})$ is obtained as follows

$$\mathbf{I}(\boldsymbol{\xi}) = \begin{bmatrix} \mathbf{I}^{(1,1)}(\boldsymbol{\xi}) & \mathbf{I}^{(1,2)}(\boldsymbol{\xi}) \\ \mathbf{I}^{(2,1)}(\boldsymbol{\xi}) & \mathbf{I}^{(2,2)}(\boldsymbol{\xi}) \end{bmatrix}, \tag{A.0.26}$$

in which the sub-matrix $\mathbf{I}^{(1,1)}(\boldsymbol{\xi})$ is defined as

$$\mathbf{I}^{(1,1)}(\boldsymbol{\xi}) \triangleq \begin{bmatrix} [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\theta},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\theta},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\theta},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\theta},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\phi},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\phi},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\phi},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\boldsymbol{\phi},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[0],\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[0],\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[0],\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[0],\mathbf{g}[N_{\mathrm{c}}-1]} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[N_{\mathrm{c}}-1],\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[N_{\mathrm{c}}-1],\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[N_{\mathrm{c}}-1],\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\mathbf{g}[N_{\mathrm{c}}-1],\mathbf{g}[N_{\mathrm{c}}-1]} \end{bmatrix}, \tag{A.0.27}$$

containing the Fisher information between AOAs/AODs and channel gains. Similarly, $\mathbf{I}^{(1,2)}(\boldsymbol{\xi})$ is given as

$$\mathbf{I}^{(1,2)}(\boldsymbol{\xi}) \triangleq \begin{bmatrix} [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\mathbf{g}[N_{\mathrm{c}}-1]} \\ [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\boldsymbol{\theta}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\boldsymbol{\phi}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\mathbf{g}[0]} & \cdots & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\mathbf{g}[N_{\mathrm{c}}-1]} \end{bmatrix}, \tag{A.0.28}$$

which gathers the Fisher information between AOAs/AODs, channel gains and the remaining disturbance parameters. Moreover, $\mathbf{I}^{(2,1)}(\boldsymbol{\xi}) = \left[\mathbf{I}^{(1,2)}(\boldsymbol{\xi})\right]^T$. Lastly, $\mathbf{I}^{(2,2)}(\boldsymbol{\xi})$ contains the Fisher information between remaining disturbance

parameters and is expressed as

$$\mathbf{I}^{(2,2)}(\boldsymbol{\xi}) \triangleq \begin{bmatrix} [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{t}},c_{\mathrm{r}}} \\ [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{g_{\mathrm{r}},c_{\mathrm{r}}} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{t}},c_{\mathrm{r}}} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\phi_{\mathrm{r}},c_{\mathrm{r}}} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{t}},c_{\mathrm{r}}} \\ [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{\epsilon_{\mathrm{r}},c_{\mathrm{r}}} \\ [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{t}},c_{\mathrm{r}}} \\ [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},g_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},g_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\phi_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\phi_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\epsilon_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},\epsilon_{\mathrm{r}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},c_{\mathrm{t}}} & [\mathbf{I}(\boldsymbol{\xi})]_{c_{\mathrm{r}},c_{\mathrm{r}}} \end{bmatrix}. \tag{A.0.29}$$

# Appendix B

# Proof of equation (3.3.10)

Let the delayed filter be

$$r(t) = p(t - \tau). \tag{B.0.1}$$

If we have a set of measurements $\mathbf{r} = [r(0), r(T_\mathrm{s}), \ldots, r((K-1)T_\mathrm{s})]^T$, the $k$-th entry after DFT can be computed as

$$\hat{r}[k] = \sum_{d=0}^{K-1} [\mathbf{r}]_d \cdot e^{-\mathrm{j}\frac{2\pi kd}{K}}. \tag{B.0.2}$$

Equivalently, we can use the continuous interpretation of the DFT to represent this as

$$\hat{r}[k] = \mathcal{F}\Big( \sum_d \delta_{dT_\mathrm{s}} (p * \delta_\tau) \Big)[f_k], \tag{B.0.3}$$

where $\sum_d \delta_{dT_\mathrm{s}} = \sum_{d=-\infty}^{\infty} \delta(t - dT_\mathrm{s})$ is the sampling function with period $T_\mathrm{s}$ and $\delta_\tau \triangleq \delta(t - \tau)$. Moreover, $*$ denotes convolution operation and $\mathcal{F}(\cdot)[f_k]$ is the Fourier transform evaluated at frequency $f_k$. Next we can make use of Fourier product and convolution identities to get to

$$\hat{r}[k] = \Big( \mathcal{F}\big( \sum_d \delta_{dT_\mathrm{s}} \big) * \big( \mathcal{F}(p)\mathcal{F}(\delta_\tau) \big) \Big)[f_k]. \tag{B.0.4}$$

Using the Dirac comb Fourier identity formula we reach

$$\hat{r}[k] = \Big( \big( \frac{1}{T_\mathrm{s}} \sum_i \delta_{i/T_\mathrm{s}} \big) * \big( \mathcal{F}(p) \mathcal{F}(\delta_\tau) \big) \Big) [f_k]. \tag{B.0.5}$$

Now we can express the convolution in terms of an integral

$$\hat{r}[k] = \frac{1}{T_\mathrm{s}} \int \mathcal{F}(p)[f] e^{-\mathrm{j}2\pi f \tau} \sum_i \delta_{i/T_\mathrm{s}}(f_k - f) df, \tag{B.0.6}$$

and this can be solved by evaluating the Dirac comb as follows

$$\begin{aligned} \hat{r}[k] &= \sum_i \frac{\mathcal{F}(p)[f_k + i/T_\mathrm{s}]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi (f_k + i/T_\mathrm{s})\tau} \\ &= \Big( \sum_i \frac{\mathcal{F}(p)[f_k + i/T_\mathrm{s}]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi i \tau / T_\mathrm{s}} \Big) e^{-\mathrm{j}2\pi f_k \tau}. \end{aligned} \tag{B.0.7}$$

Note that since $\mathcal{F}(p)$ is bounded, this sum only includes a few terms. Let us define the delay-frequency distortion as

$$g(k, \tau) = \sum_i \frac{\mathcal{F}(p)[f_k + i/T_\mathrm{s}]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi i \tau / T_\mathrm{s}}, \tag{B.0.8}$$

and then (B.0.7) can be simplified to

$$\hat{r}[k] = g(k, \tau) e^{-\mathrm{j}2\pi f_k \tau}. \tag{B.0.9}$$

If we assume the domain of $\mathcal{F}(p)$ to be in $[f_\mathrm{c} - \frac{1+\beta}{2T_s}, f_\mathrm{c} + \frac{1+\beta}{2T_s}]$ with $\beta \in [0, 1]$, then we have the expression for $g(k, \tau)$ as

$$g(k, \tau) = \begin{cases} \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} + \frac{\mathcal{F}(p)[f_k - 1/T_\mathrm{s}]}{T_\mathrm{s}} e^{\mathrm{j}2\pi\tau/T_\mathrm{s}} & \text{if } \ 2(f_k - f_\mathrm{c}) > (1-\beta)/T_\mathrm{s} \\ \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} & \text{if } \ 2|f_k - f_\mathrm{c}| \leq (1-\beta)/T_\mathrm{s} \\ \frac{\mathcal{F}(p)[f_k]}{T_\mathrm{s}} + \frac{\mathcal{F}(p)[f_k + 1/T_\mathrm{s}]}{T_\mathrm{s}} e^{-\mathrm{j}2\pi\tau/T_\mathrm{s}} & \text{if } \ 2(f_k - f_\mathrm{c}) < -(1-\beta)/T_\mathrm{s} \end{cases} \tag{B.0.10}$$

In the case of a raised cosine filter with parameters $T = T_\mathrm{s}$ and $\beta \in [0, 1]$, it is straightforward to prove that

$$g(k,\tau) = \begin{cases} \frac{1}{2}(1 + e^{\mathrm{j}\frac{2\pi\tau}{T_\mathrm{s}}} + (1 - e^{\mathrm{j}\frac{2\pi\tau}{T_\mathrm{s}}})\cos(\frac{\pi T_\mathrm{s}}{\beta}(|f_k| - \frac{1-\beta}{2T_\mathrm{s}}))), \\ \qquad\qquad \text{if } \ 2(f_k - f_\mathrm{c}) > (1-\beta)/T_\mathrm{s} \\ 1, \qquad\qquad \text{if } \ 2|f_k - f_\mathrm{c}| \leq (1-\beta)/T_\mathrm{s} \\ \frac{1}{2}(1 + e^{-\mathrm{j}\frac{2\pi\tau}{T_\mathrm{s}}} + (1 - e^{-\mathrm{j}\frac{2\pi\tau}{T_\mathrm{s}}})\cos(\frac{\pi T_\mathrm{s}}{\beta}(|f_k| - \frac{1-\beta}{2T_\mathrm{s}}))), \\ \qquad\qquad \text{if } \ 2(f_k - f_\mathrm{c}) < -(1-\beta)/T_\mathrm{s} \end{cases} \tag{B.0.11}$$

# Appendix C

# Derivative of the objective function in (3.4.12) with respect to hardware impairment related dictionary

First, we calculate the derivative of the first sum term in (3.4.12) with respect to $\mathbf{D}_{\mathrm{R},1}$. For $u \in \mathcal{I}(N_{\mathrm{sa}}), k \in \mathcal{K}_{\mathrm{cen}}$, we have

$$\frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{\mathrm{R},1}} = -\overline{(\hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{\mathrm{R},1}\mathbf{X}_{\mathrm{R},1}^{(u)}[k])}(\mathbf{X}_{\mathrm{R},1}^{(u)}[k])^T. \tag{C.0.1}$$

Next, we calculate the derivative of the second sum term in (3.4.12) with respect to $\mathbf{D}_{\mathrm{R},1}$. Recalling the chain rule, we can express the Jacobian matrix of the second sum term with respect to $\mathbf{D}_{\mathrm{R},1}$ as, $\forall u \in \mathcal{I}(N_{\mathrm{sa}}), k \in \mathcal{K}_{\mathrm{side}}$,

$$\begin{aligned}
&\frac{\partial J_k^{(u)}}{\partial \mathrm{vec}(\mathbf{D}_{\mathrm{R},1})^T} = \frac{\partial J_k^{(u)}}{\partial \mathrm{vec}(\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{\mathrm{R},1}\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k])^T} \\
&\cdot \frac{\partial \mathrm{vec}(\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{\mathrm{R},1}\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k])}{\partial \mathrm{vec}(\mathbf{D}_{\mathrm{R},1}\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k])^T} \cdot \frac{\partial \mathrm{vec}\big(\mathbf{D}_{\mathrm{R},1}\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]\big)}{\partial \mathrm{vec}(\mathbf{D}_{\mathrm{R},1})^T} \\
&= -\mathrm{vec}\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \overline{\hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{\mathrm{R},1}\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]}\Big)^T \\
&\cdot \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathrm{diag}\big\{\mathrm{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}))\big\}\big((\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k])^T \otimes \mathbf{I}_{N_{\mathrm{r}}}\big).
\end{aligned} \tag{C.0.2}$$

With the Jocobian matrix, we can expressive the derivative of the second sum term in (3.4.12) with respect to $\mathbf{D}_{\mathrm{R},1}$ as follows

$$\begin{aligned}\frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{\mathrm{R},1}} &= \mathrm{unvec}\Big\{\frac{\partial J_k^{(u)}}{\partial \mathrm{vec}(\mathbf{D}_{\mathrm{R},1})}\Big\} = \mathrm{unvec}\Big\{ -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \big(\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k] \otimes \mathbf{I}_{N_{\mathrm{r}}}\big) \\ &\quad \cdot \mathrm{diag}\big\{\mathrm{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}))\big\} \cdot \mathrm{vec}\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]} \\ &\qquad\qquad - \sum_{\hat{l}=1}^{\hat{L}} \overline{\hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{\mathrm{R},1} \mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]}\Big)\Big\} \\ &= -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \Big[\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \overline{\hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{\mathrm{R},1} \mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]}\Big) \\ &\qquad\qquad \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}})\Big] \big(\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]\big)^T. \end{aligned} \tag{C.0.3}$$

Then combining the derivatives in (C.0.1) for $k \in \mathcal{K}_{\mathrm{cen}}$ and the derivatives in (C.0.3) for $k \in \mathcal{K}_{\mathrm{side}}$, we can obtain the final derivative of the objective function in (3.4.12) with respect to $\mathbf{D}_{\mathrm{R},1}$, i.e.,

$$\begin{aligned}\frac{\partial J}{\partial \mathbf{D}_{\mathrm{R},1}} &= -\sum_{u \in \mathcal{I}(N_{\mathrm{sa}})} \sum_{k \in \mathcal{K}_{\mathrm{cen}}} \overline{(\hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{\mathrm{R},1} \mathbf{X}_{\mathrm{R},1}^{(u)}[k])} (\mathbf{X}_{\mathrm{R},1}^{(u)}[k])^T \\ &\quad - \sum_{u \in \mathcal{I}(N_{\mathrm{sa}})} \sum_{k \in \mathcal{K}_{\mathrm{side}}} \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \Big[\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]} \\ &\quad - \sum_{\hat{l}=1}^{\hat{L}} \overline{\hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{\mathrm{R},1} \mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]}\Big) \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}})\Big] \big(\mathbf{X}_{\mathrm{R},1,\hat{l}}^{(u)}[k]\big)^T. \end{aligned} \tag{C.0.4}$$

# Appendix D

# Derivative of the objective function in (3.4.15) with respect to location error

First, we calculate the derivative of the first sum term in (3.4.15) with respect to the location error $\boldsymbol{\epsilon}_{\mathrm{R}}$. Note that the gradient of any element of $\mathbf{D}_{\mathrm{R},2}$ with respect to the antenna location error $\epsilon_{\mathrm{R},m}$ can be expressed as

$$\frac{\partial[\mathbf{D}_{\mathrm{R},2}]_{m,n}}{\partial \epsilon_{\mathrm{R},m}} = \frac{\partial e^{-\mathrm{j}2\pi f_{\mathrm{c}}\epsilon_{\mathrm{R},m}\cdot\sin(\phi_n^{\mathrm{v}})/c}}{\partial \epsilon_{\mathrm{R},m}} = [\mathbf{D}_{\mathrm{R},2}]_{m,n}\cdot\frac{-\mathrm{j}2\pi f_{\mathrm{c}}\sin(\phi_n^{\mathrm{v}})}{c}. \tag{D.0.1}$$

Therefore, we have the gradient of $J_k^{(u)}$ with respect to $\boldsymbol{\epsilon}_{\mathrm{R}}$ as

$$\frac{\partial J_k^{(u)}}{\partial \boldsymbol{\epsilon}_{\mathrm{R}}} = 2\mathcal{R}\left\{\left\{\frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{\mathrm{R},2}}\odot\mathbf{D}_{\mathrm{R},2}\right\}\cdot\frac{-\mathrm{j}2\pi f_{\mathrm{c}}\sin(\boldsymbol{\phi}^{\mathrm{v}})}{c}\right\}, \tag{D.0.2}$$

We next calculate the derivative of the second sum term in (3.4.15) with respect to $\boldsymbol{\epsilon}_{\mathrm{R}}$. Similar to Appendix C, using the chain rule of Jacobian matrix, we have, $\forall u\in\mathcal{I}(N_{\mathrm{sa}}), k\in\mathcal{K}_{\mathrm{side}}$,

$$\begin{aligned}
\frac{\partial J_k^{(u)}}{\partial \boldsymbol{\epsilon}_{\mathrm{R}}^T} &= 2\mathcal{R}\Bigg\{\frac{\partial J_k^{(u)}}{\partial\mathrm{vec}\Big(\sum_{\hat{l}=1}^{\hat{L}}\hat{\alpha}_{\hat{l}}e^{-\mathrm{j}2\pi f_k\hat{\tau}_{\hat{l}}}\mathbf{G}_k(\hat{\tau}_{\hat{l}},\hat{\phi}_{\hat{l}},\hat{\theta}_{\hat{l}})\odot\big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]\big)\Big)^T}\\
&\cdot\frac{\partial\mathrm{vec}\Big(\sum_{\hat{l}=1}^{\hat{L}}\hat{\alpha}_{\hat{l}}e^{-\mathrm{j}2\pi f_k\hat{\tau}_{\hat{l}}}\mathbf{G}_k(\hat{\tau}_{\hat{l}},\hat{\phi}_{\hat{l}},\hat{\theta}_{\hat{l}})\odot\big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]\big)\Big)}{\partial\mathrm{vec}\big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]\big)^T}\\
&\cdot\frac{\partial\mathrm{vec}\big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]\big)}{\partial\mathrm{vec}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)^T}\cdot\frac{\partial\mathrm{vec}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)}{\partial\mathrm{vec}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\big)^T}\cdot\frac{\partial\mathrm{vec}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\big)}{\partial\boldsymbol{\epsilon}_{\mathrm{R}}^T}\Bigg\}\\
&= 2\mathcal{R}\Bigg\{-\mathrm{vec}\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]}-\sum_{\hat{l}=1}^{\hat{L}}\overline{\hat{\alpha}_{\hat{l}}e^{-\mathrm{j}2\pi f_k\hat{\tau}_{\hat{l}}}\mathbf{G}_k(\hat{\tau}_{\hat{l}},\hat{\phi}_{\hat{l}},\hat{\theta}_{\hat{l}})\odot\big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\odot\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}_{\mathrm{R},2,\hat{l}}^{(u)}[k]\big)}\Big)^T
\end{aligned}$$

$$\cdot \sum_{\hat{l}=1}^{\hat{L}} \alpha_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathrm{diag}\big\{\mathrm{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}))\big\}$$

$$\cdot \big((\mathbf{X}^{(u)}_{\mathrm{R},2,\hat{l}}[k])^T \otimes \mathbf{D}_{\mathrm{R},1}\big) \cdot \mathrm{diag}\big\{\mathrm{vec}(\mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}}))\big\} \cdot \mathrm{diag}\big\{\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}})\big\} \cdot \frac{-\mathrm{j}2\pi f_{\mathrm{c}} \sin(\hat{\phi}_{\hat{l}})}{c} \Bigg\}. \tag{D.0.3}$$

Therefore, we have

$$\begin{aligned}
\frac{\partial J_k^{(u)}}{\partial \boldsymbol{\epsilon}_{\mathrm{R}}} = 2\mathcal{R}\Bigg\{ & -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \cdot \frac{-\mathrm{j}2\pi f_{\mathrm{c}} \sin(\hat{\phi}_{\hat{l}})}{c} \cdot \mathrm{diag}\big\{\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big\} \\
& \cdot \big((\mathbf{X}^{(u)}_{\mathrm{R},2,\hat{l}}[k]) \otimes \mathbf{D}^T_{\mathrm{R},1}\big) \cdot \mathrm{diag}\big\{\mathrm{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}))\big\} \\
& \cdot \mathrm{vec}\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \overline{\hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}^{(u)}_{\mathrm{R},2,\hat{l}}[k]\big)}\Big) \Bigg\} \\
= 2\mathcal{R}\Bigg\{ & -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \cdot \frac{-\mathrm{j}2\pi f_{\mathrm{c}} \sin(\hat{\phi}_{\hat{l}})}{c} \cdot \Bigg[\mathbf{D}^T_{\mathrm{R},1} \\
& \cdot \Bigg(\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \overline{\hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}^{(u)}_{\mathrm{R},2,\hat{l}}[k]\big)}\Big) \\
& \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}})\Bigg)(\mathbf{X}^{(u)}_{\mathrm{R},2,\hat{l}}[k])^T\Bigg] \odot \big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big) \Bigg\}.
\end{aligned} \tag{D.0.4}$$

Then combining the derivatives in (D.0.2) for $k \in \mathcal{K}_{\mathrm{cen}}$ and the derivatives in (D.0.4) for $k \in \mathcal{K}_{\mathrm{side}}$, we can obtain the final derivative of the objective function in (3.4.15) with respect to $\boldsymbol{\epsilon}_{\mathrm{R}}$, i.e.,

$$\begin{aligned}
\frac{\partial J}{\partial \boldsymbol{\epsilon}_{\mathrm{R}}} = & \sum_{u \in \mathcal{I}(N_{\mathrm{sa}})} \sum_{k \in \mathcal{K}_{\mathrm{cen}}} 2\mathcal{R}\Bigg\{\Bigg\{\frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{\mathrm{R},2}} \odot \mathbf{D}_{\mathrm{R},2}\Bigg\} \cdot \frac{-\mathrm{j}2\pi f_{\mathrm{c}} \sin(\boldsymbol{\phi}^{\mathrm{v}})}{c}\Bigg\} \\
& - \sum_{u \in \mathcal{I}(N_{\mathrm{sa}})} \sum_{k \in \mathcal{K}_{\mathrm{side}}} 2\mathcal{R}\Bigg\{ \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \cdot \frac{-\mathrm{j}2\pi f_{\mathrm{c}} \sin(\hat{\phi}_{\hat{l}})}{c} \cdot \Bigg[\mathbf{D}^T_{\mathrm{R},1} \\
& \cdot \Bigg(\Big(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \overline{\hat{\alpha}_{\hat{l}} e^{-\mathrm{j}2\pi f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \big(\mathbf{D}_{\mathrm{R},1}\big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\mathbf{X}^{(u)}_{\mathrm{R},2,\hat{l}}[k]\big)}\Big) \\
& \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}})\Bigg)(\mathbf{X}^{(u)}_{\mathrm{R},2,\hat{l}}[k])^T\Bigg] \odot \big(\mathbf{e}_{\mathrm{R}}(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{\mathrm{R},k}(\hat{\phi}_{\hat{l}})\big)\Bigg\}.
\end{aligned} \tag{D.0.5}$$

# Bibliography

[1] 3GPP, "Physical channels and modulation (release 15)," *Tech. Rep. v15.1.0*, 2017.

[2] Yonghee Han and Jungwoo Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. of IEEE ISIT'16*, July 2016, pp. 860–864.

[3] Junho Lee, Gye-Tae Gil, and Yong H Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. of IEEE GLOBECOM'14*, Dec. 2014, pp. 3326–3331.

[4] Robert W Heath, Nuria González-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[5] Javier Rodríguez-Fernández, Nuria González-Prelcic, Kiran Venugopal, and Robert W Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.

[6] Kiran Venugopal, Ahmed Alkhateeb, Nuria González Prelcic, and Robert W Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sept. 2017.

[7] Roi Méndez-Rial, Cristian Rusu, Nuria González-Prelcic, Ahmed Alkhateeb, and Robert W Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, , no. 99, Jan 2016.

[8] Zhenyu Xiao, Pengfei Xia, and Xiang-Gen Xia, "Channel estimation and hybrid precoding for millimeter-wave MIMO systems: A low-complexity overall solution," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2017.

[9] Sicong Liu, Fang Yang, Wenbo Ding, Xianbin Wang, and Jian Song, "Two-dimensional structured-compressed-sensing-based NBI cancelation exploiting spatial and temporal correlations in MIMO systems," *IEEE Trans. Vehi. Tech.*, vol. 65, no. 11, pp. 9020–9028, Nov 2016.

[10] Zhen Gao, Linglong Dai, and Zhaocheng Wang, "Channel estimation for mmwave massive MIMO based access and backhaul in ultra-dense network," in *Proc. of IEEE ICC'16*, May 2016, pp. 1–6.

[11] Javier Rodríguez-Fernández, Nuria González-Prelcic, and Robert W Heath, "A compressive sensing-maximum likelihood approach for off-grid wideband channel estimation at mmwave," in *Proc. of IEEE CAMSAP'17*, Dec 2017, pp. 1–5.

[12] José P González-Coma, Javier Rodriguez-Fernandez, Nuria González-Prelcic, Luis Castedo, and Robert W Heath, "Channel estimation and hybrid precoding for frequency selective multiuser mmwave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.

[13] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (release 14.3.0)," *Technical Report*, Dec 2017.

[14] Joan Palacios, Danilo De Donno, and Joerg Widmer, "Tracking mm-wave channel dynamics: Fast beam training strategies under mobility," in *Proc. of IEEE INFOCOM'17*, May 2017, pp. 1–9.

[15] George C Alexandropoulos and Symeon Chouvardas, "Low complexity channel estimation for millimeter wave systems with hybrid A/D antenna processing," in *Proc. of IEEE GC Wkshps'16*, Dec 2016, pp. 1–6.

[16] Jiguang He, Taejoon Kim, Hadi Ghauch, Kunpeng Liu, and Guangjian Wang, "Millimeter wave mimo channel tracking systems," in *Proc. of IEEE GC Wkshps'14*, Dec 2014, pp. 416–421.

[17] Linglong Dai and Xinyu Gao, "Priori-aided channel tracking for millimeter-wave beamspace massive MIMO systems," in *Proc. of IEEE URSI AP-RASC'16*, Aug 2016, pp. 1493–1496.

[18] Jin Hyeok Yoo, Jisu Bae, Sun Hong Lim, Sunwoo Kim, Jun Won Choi, and Byonghyo Shim, "Sampling-based tracking of time-varying channels

for millimeter wave-band communications," in *Proc. of IEEE ICC'17*, May 2017, pp. 1–6.

[19] Karthik Upadhya, Robert W Heath, and Sergiy A Vorobyov, "Tracking abruptly changing channels in mmwave systems using overlaid data and training," in *Proc. of IEEE CAMSAP'17*, Dec 2017, pp. 1–5.

[20] Chuang Zhang, Dongning Guo, and Pingyi Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *Proc. of IEEE ICC'16*, May 2016, pp. 1–6.

[21] Jianwei Zhao, Feifei Gao, Weimin Jia, Shun Zhang, Shi Jin, and Hai Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, Oct 2017.

[22] Han Yan, Shailesh Chaudhari, and Danijela Cabric, "Wideband channel tracking for mmwave MIMO system with hybrid beamforming architecture: (invited paper)," in *Proc. of IEEE CAMSAP'17*, Dec 2017, pp. 1–5.

[23] Javier Rodríguez-Fernández, Nuria Gonzáiez-Prelcic, and Robert W Heath, "Frequency-domain wideband channel estimation and tracking for hybrid MIMO systems," in *Proc. of IEEE Asilomar'17*, Oct 2017, pp. 1829–1833.

[24] Javier Rodriguez-Fernandez, Nuria Gonzalez-Prelcic, and Takayuki Shimizu, "Position-aided compressive channel tracking for wideband millimeter wave multi-user communication," in *Proc. of IEEE ICC'19*, Dec 2019.

[25] George C Alexandropoulos, Evangelos Vlachos, and John Thompson, "Wideband channel tracking for millimeter wave massive MIMO systems with hybrid beamforming reception," in *Proc. of IEEE ICASSP'20*, 2020, pp. 8698–8702.

[26] Bolei Wang, Feifei Gao, Shi Jin, Hai Lin, and Geoffrey Ye Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, July 2018.

[27] John H Brady and Akbar M Sayeed, "Wideband communication with high-dimensional arrays: New results and transceiver architectures," in *Proc. of IEEE ICCW'15*, June 2015, pp. 1042–1047.

[28] Ta-Shun Chu and Hossein Hashemi, "True-time-delay-based multi-beam arrays," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 8, pp. 3072–3082, Aug. 2013.

[29] Qibo Qin, Lin Gui, Bo Gong, and Sheng Luo, "Sparse channel estimation for massive MIMO-OFDM systems over time-varying channels," *IEEE Access*, vol. 6, pp. 33740–33751, June 2018.

[30] Javier Rodríguez-Fernández and Nuria González-Prelcic, "Position-aided compressive channel estimation and tracking for millimeter wave multi-user MIMO air-to-air communications," in *Proc. of IEEE ICC'18*, 2018.

[31] Javier Rodriguez-Fernandez, Nuria Gonzalez-Prelcic, and Takayuki Shimizu, "Position-aided compressive channel tracking for wideband millimeter wave multi-user communication," in *Proc. of IEEE ICC'19*, 2019, pp. 1–6.

[32] Nuria González-Prelcic, Hongxiang Xie, Joan Palacios, and Takayuki Shimizu, "Wideband channel tracking and hybrid precoding for mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2161–2174, Oct. 2020.

[33] Chuang Zhang, Dongning Guo, and Pingyi Fan, "Mobile millimeter wave channel acquisition, tracking, and abrupt change detection," *arXiv:1610.09626*, 2016.

[34] Karthik Upadhya, Robert W Heath, and Sergiy A Vorobyov, "Tracking abruptly changing channels in mmwave systems using overlaid data and training," in *Proc. of IEEE CAMSAP'17*, 2017, pp. 1–5.

[35] Yue Wu, Yuchen Jiao, Feifei Gao, and Yuantao Gu, "Pilot-free channel change detection for mmwave massive MIMO system," in *Proc. of IEEE GLOBECOM'19*, 2019, pp. 1–6.

[36] Hongxiang Xie and Nuria González-Prelcic, "Dictionary learning for channel estimation in hybrid frequency-selective mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7407–7422, Nov. 2020.

[37] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[38] Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. of IEEE CVPR'14*, June 2014, pp. 3842–3849.

[39] Mehdi Bahri, Yannis Panagakis, and Stefanos Zafeiriou, "Robust kronecker component analysis," *arXiv:1801.06432*, 2018.

[40] Julien Mairal, Jean Ponce, Guillermo Sapiro, Andrew Zisserman, and Francis R. Bach, "Supervised dictionary learning," in *Proc. of NIPS'09*, pp. 1033–1040. 2009.

[41] Feiping Nie, Heng Huang, Xiao Cai, and Chris H. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. of NIPS'10*, pp. 1813–1821. 2010.

[42] Yang Liu, Wei Chen, Qingchao Chen, and Ian Wassell, "Support discrimination dictionary learning for image classification," in *Proc. of ECCV'16*, Cham, 2016, pp. 375–390.

[43] Simon Hawe, Matthias Seibert, and Martin Kleinsteuber, "Separable dictionary learning," in *Proc. of IEEE CVPR'13*, 2013, pp. 438–445.

[44] Yacong Ding and Bhaskar D Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, 2018.

[45] Thomas Wiese, Lorenz Weiland, and Wolfgang Utschick, "Low-rank approximations for spatial channel models," in *Proc. of IEEE WSA'16*, Mar. 2016, pp. 1–5.

[46] Eldad Perahia and Robert Stacey, *Next Generation Wireless LANs 802.11n and 802.11ac*, Cambridge University Press, 2 edition, 2013.

[47] Zhen Gao, Linglong Dai, Wei Dai, Byonghyo Shim, and Zhaocheng Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb. 2016.

[48] Zhen Gao, Linglong Dai, Zhaocheng Wang, and Sheng Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[49] Hongxiang Xie, Nuria González-Prelcic, and Robert W Heath, "Separable dictionary learning for channel estimation in hybrid mmWave MIMO

systems," *Proc. of ITA workshop'19*, pp. 1–6, 2019.

[50] Philip Schniter and Akbar Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. of Asilomar'14*, Nov. 2014, pp. 273–277.

[51] Michael Eberhardt, Philipp Eschlwech, and Erwin Biebl, "Investigations on antenna array calibration algorithms for direction-of-arrival estimation," *Advances in Radio Science: ARS*, vol. 14, pp. 181–190, 2016.

[52] Boon Chong Ng and Chong Meng Samson See, "Sensor-array calibration using a maximum-likelihood approach," *IEEE Trans. Antennas Propag.*, vol. 44, no. 6, pp. 827–835, 1996.

[53] Boon Poh Ng, Joni Polili Lie, Meng Hwa Er, and Aigang Feng, "A practical simple geometry and gain/phase calibration technique for antenna array processing," *IEEE Trans. Antennas Propag.*, vol. 57, no. 7, pp. 1963–1972, 2009.

[54] Philipp Heidenreich, Abdelhak M Zoubir, and Michael Rubsamen, "Joint 2-D DOA estimation and phase calibration for uniform rectangular arrays," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4683–4693, 2012.

[55] Muhammad Z Ikram, Murtaza Ali, and Dan Wang, "Joint antenna-array calibration and direction of arrival estimation for automotive radars," in *Proc. of IEEE Radar Conference (RadarConf)*. IEEE, 2016, pp. 1–5.

[56] Christian M Schmid, Clemens Pfeffer, Reinhard Feger, and Andreas Stelzer, "An FMCW MIMO radar calibration and mutual coupling compensation approach," in *Proc. of 2013 European Radar Conference*. IEEE, 2013, pp. 13–16.

[57] Michael Stephan, Kuangda Wang, Torsten Reissland, Robert Weigel, Ke Wu, and Fabian Lurz, "Evaluation of antenna calibration and DOA estimation algorithms for FMCW radars," in *Proc. of 2019 49th European Microwave Conference (EuMC)*. IEEE, 2019, pp. 944–947.

[58] Akbar M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct 2002.

[59] Steven M Kay, *Fundamentals of statistical signal processing*, Prentice Hall PTR, 1993.

[60] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[61] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[62] Junfeng Yang and Yin Zhang, "Alternating direction algorithms for $\ell_1$-problems in compressive sensing," *SIAM Journal on Scientific Computing*, vol. 33, no. 1, pp. 250–278, 2011.

[63] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[64] Ron Rubinstein, Michael Zibulevsky, and Michael Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," Tech. Rep., Computer Science Department, Technion, 2008.

[65] Kjersti Engan, Sven Ole Aase, and J Hakon Husoy, "Method of optimal directions for frame design," in *Proc. of IEEE ICASSP'99*, 1999, vol. 5, pp. 2443–2446.

[66] Roland Glowinski and JT Oden, "Numerical methods for nonlinear variational problems," *Journal of Applied Mechanics*, vol. 52, no. 3, pp. 739, 1985.

[67] Cristian Rusu and Bogdan Dumitrescu, "An initialization strategy for the dictionary learning problem," in *Proc. of IEEE ICASSP'14*, May 2014, pp. 6731–6735.

[68] Alekh Agarwal, Animashree Anandkumar, and Praneeth Netrapalli, "A clustering approach to learning sparsely used overcomplete dictionaries,"

*IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 575–592, 2017.

[69] Niladri Chatterji and Peter L Bartlett, "Alternating minimization for dictionary learning with random initialization," in *Proc. of NIPS'17*, 2017, pp. 1997–2006.

[70] Tamara G. Kolda and Brett W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[71] Florian Roemer, Giovanni Del Galdo, and Martin Haardt, "Tensor-based algorithms for learning multidimensional separable dictionaries," in *Proc. of IEEE ICASSP'14*, May 2014, pp. 3963–3967.

[72] HSC Wang, "Performance of phased array antennas with mechanical errors," in *IEEE Conference on Aerospace Applications*, Feb 1990, pp. 317–318.

[73] K-L Du, "Pattern analysis of uniform circular array," *IEEE Trans. Antennas Propag.*, vol. 52, no. 4, pp. 1125–1129, April 2004.

[74] Hongxiang Xie, Joan Palacios, and Nuria González Prelcic, "Dictionary learning and channel estimation for hybrid mmwave mimo systems under hardware impairments and beam squint," *In preparation for submission to IEEE Trans. Wireless Commun.*, 2021.

[75] Javier Rodriguez-Fernandez and Nuria González-Prelcic, "Channel estimation for frequency-selective mmWave MIMO systems with beam-squint," in *Proc. of IEEE GLOBECOM'18*, 2018, pp. 1–6.

[76] José P. González-Coma, Wolfgang Utschick, and Luis Castedo, "Hybrid LISA for wideband multiuser millimeter wave communication systems under beam squint," *CoRR*, vol. abs/1804.09223, 2018.

[77] Mingjin Wang, Feifei Gao, Nir Shlezinger, Mark F Flanagan, and Yonina C Eldar, "A block sparsity based estimator for mmwave massive mimo channels with beam squint," *IEEE Trans. Signal Process.*, vol. 68, pp. 49–64, 2019.

[78] Mengnan Jian, Feifei Gao, Zhi Tian, Shi Jin, and Shaodan Ma, "Angle-domain aided ul/dl channel estimation for wideband mmwave massive mimo systems with beam squint," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3515–3527, 2019.

[79] Bolei Wang, Mengnan Jian, Feifei Gao, Geoffrey Ye Li, and Hai Lin, "Beam squint and channel estimation for wideband mmwave massive MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5893–5908, 2019.

[80] Joan Palacios, Nuria González-Prelcic, and Joerg Widmer, "Managing hardware impairments in hybrid millimeter wave MIMO systems: A dictionary learning-based approach," in *Proc. of IEEE Asilomar'19*, 2019, pp. 168–172.

[81] Peng Tan and Norman C Beaulieu, "Reduced ICI in OFDM systems using the" better than" raised-cosine pulse," *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 135–137, 2004.

[82] Benjamin Friedlander and Anthony J Weiss, "Direction finding in the presence of mutual coupling," *IEEE Trans. Antennas Propag.*, vol. 39, no. 3, pp. 273–284, 1991.

[83] Konstantinos V Stavropoulos and Athanassios Manikas, "Array calibration in the presence of unknown sensor characteristics and mutual coupling," in *Proc. of IEEE EUSIPCO'00*, 2000, pp. 1–4.

[84] Joel A Tropp, Anna C Gilbert, and Martin J Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. of IEEE ICASSP'05*, March 2005, vol. 5, pp. 721–724.

[85] Junil Choi, Vutha Va, Nuria Gonzalez-Prelcic, Robert Daniels, Chandra R Bhat, and Robert W Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016.

[86] Vutha Va, Haris Vikalo, and Robert W Heath, "Beam tracking for mobile millimeter wave communication systems," in *Proc. of IEEE GlobalSIP'16*, Dec 2016, pp. 743–747.

[87] Mathew K Samimi, George R MacCartney, Shu Sun, and Theodore S Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *Proc. of IEEE VTC Spring'16*, May 2016, pp. 1–6.

[88] Junghoon Ko, Yeon-Jea Cho, Sooyoung Hur, Taehwan Kim, Jeongho Park, Andreas F Molisch, Katsuyuki Haneda, Michael Peter, Dong-Jo Park, and Dong-Ho Cho, "Millimeter-wave channel measurements and analysis for statistical spatial channel model in in-building and urban environments at 28 GHz," *IEEE Trans. wireless commun.*, vol. 16, no. 9, pp. 5853–5868, 2017.

[89] Mathew K Samimi and Theodore S Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory*, vol. 64, no. 7, pp. 2207–2225, 2016.

[90] Nitin Jonathan Myers, Yuyang Wang, Nuria González-Prelcic, and Robert W Heath, "Deep Learning-Based Beam Alignment in mmWave Vehicular Networks," in *Proc. of IEEE ICASSP'20*, 2020, pp. 8569–8573.

[91] Philip Schniter and Akbar Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. of IEEE Asilomar'14*. IEEE, 2014, pp. 273–277.

[92] Javier Rodriguez-Fernandez and Nuria Gonzalez-Prelcic, "Channel estimation for frequency-selective mmwave MIMO systems with beam-squint," in *Proc. of IEEE GLOBECOM'18*, Dec 2018, pp. 1–6.

[93] Kiran Venugopal, Nuria González-Prelcic, and Robert W Heath, "Optimal frequency-flat precoding for frequency-selective millimeter wave channels," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2019.

[94] Imre Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statist. Decisions*, vol. 1, pp. 205–237, 1984.

[95] Steven W Peters and Robert W Heath, "Cooperative algorithms for mimo interference channels," *IEEE Trans. Vehi. Tech.*, vol. 60, no. 1, pp. 206–218, 2011.

[96] Roi Mendez-Rial, Cristian Rusu, Nuria González-Prelcic, and Robert W Heath, "Dictionary-free hybrid precoders and combiners for mmwave MIMO systems," in *Proc. of IEEE SPAWC'15*, June 2015, pp. 151–155.

[97] Jonathan Starr, Omar El Ayach, and Robert W Heath, "Interference alignment with per-antenna power constraints," in *Proc. of 2011 IEEE International Symposium on Information Theory*, 2011, pp. 2746–2750.

[98] Roberto López-Valcarce, Nuria González-Prelcic, Cristian Rusu, and Robert W Heath, "Hybrid precoders and combiners for mmwave mimo systems with per-antenna power constraints," in *Proc. of IEEE GLOBECOM'16*, 2016, pp. 1–6.

[99] Stephan Jaeckel, Leszek Raschkowski, Kai Börner, and Lars Thiele, "QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials," *IEEE Trans. Ant. Propag.*, vol. 62, no. 6, pp. 3242–3256, June 2014.

[100] Rudolph Emil Kalman, "A new approach to linear filtering and prediction problems," *Trans. AMSE J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[101] Thomas Schon, Fredrik Gustafsson, and P-J Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2279–2289, July 2005.

[102] Bishwarup Mondal, Satyaki Dutta, and Robert W Heath, "Quantization on the Grassmann manifold," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4208 – 4216, August 2007.

[103] Hongxiang Xie, Nuria González Prelcic, and Takayuki Shimizu, "Blockage detection and channel tracking in wideband mmwave mimo systems," in *Proc. of IEEE ICC'21*, 2021, pp. 1–1.

[104] Vasanthan Raghavan, Lida Akhoondzadeh-Asl, Vladimir Podshivalov, Joakim Hulten, Mohammad Ali Tassoudji, Ozge Hizir Koymen, Ashwin Sampath, and Junyi Li, "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 7, pp. 3010–3024, 2019.

[105] Vasanthan Raghavan, Tianyang Bai, Ashwin Sampath, Ozge H Koymen, and Junyi Li, "Modeling and combating blockage in millimeter wave systems," in *Proc. of IEEE SPAWC'18*, 2018, pp. 1–5.

[106] Alexander Maltsev, Roman Maslennikov, A Sevastyanov, A Lomayev, and Alexey Khoryaev, "Statistical channel model for 60 GHz WLAN systems in conference room environment," in *Proc. of IEEE Fourth European Conference on Antennas and Propagation*, 2010, pp. 1–5.

[107] George R MacCartney, Sijia Deng, Shu Sun, and Theodore S Rappaport, "Millimeter-wave human blockage at 73 GHz with a simple double knife-edge diffraction model and extension for directional antennas," in *Proc. of IEEE VTC Fall'16*, 2016, pp. 1–6.

[108] Liangtian Wan, Kaihui Liu, Ying-Chang Liang, and Tong Zhu, "DOA and polarization estimation for non-circular signals in 3-D millimeter wave polarized massive MIMO systems," *arXiv:1712.05587v1*, 2017.

[109] Cheng Qian, Xiao Fu, Nicholas D Sidiropoulos, and Ye Yang, "Tensor-based channel estimation for dual-polarized massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6390–6403, 2018.

[110] Dalin Zhu, Junil Choi, and Robert W Heath, "Two-dimensional AoD and AoA acquisition for wideband millimeter-wave systems with dual-polarized MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7890–7905, 2017.

[111] Diego Dupleich, Stephan Häfner, Christian Schneider, Robert Müller, Reiner Thomä, Jian Luo, Naveed Iqbal, Egon Schulz, Xiaofeng Lu, and Guangjian Wang, "Double-directional and dual-polarimetric indoor

measurements at 70 GHz," in *Proc. of IEEE PIMRC'15*, Aug. 2015, pp. 2234–2238.

[112] Xin Su, Dongmin Choi, Xiaofeng Liu, and Bao Peng, "Channel model for polarized MIMO systems with power radiation pattern concern," *IEEE Access*, vol. 4, pp. 1061–1072, 2016.

[113] Ramya Bhagavatula, Claude Oestges, and Robert W Heath, "A new double-directional channel model including antenna patterns, array orientation, and depolarization," *IEEE Trans. Veh. Techno.*, vol. 59, no. 5, pp. 2219–2231, 2010.

[114] Bo Gong, Lin Gui, Qibo Qin, Xiang Ren, and We Chen, "Block distributed compressive sensing-based doubly selective channel estimation and pilot design for large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9149–9161, Oct 2017.

[115] Xu Ma, Fang Yang, Sicong Liu, Jian Song, and Zhu Han, "Doubly selective channel estimation for MIMO systems based on structured compressive sensing," in *Proc. of IEEE IWCMC'17*, June 2017, pp. 610–615.

[116] Xu Ma, Fang Yang, Sicong Liu, Wenbo Ding, and Jian Song, "Structured compressive sensing-based channel estimation for time frequency training OFDM systems over doubly selective channel," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 266–269, Apr. 2017.

[117] Dimitris G Tzikas, Aristidis C Likas, and Nikolaos P Galatsanos, "The variational approximation for bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, 2008.

# Vita

Hongxiang Xie is a Ph.D. candidate in the Department of Electrical and Computer Engineering at The University of Texas at Austin, advised by Professor Nuria González-Prelcic and Professor Robert W. Heath, Jr. in Wireless Networking & Communications Group (WNCG). He received his B.S. degree in 2014 from Southeast University, Nanjing, China, and his M.S. degree in 2017 from Tsinghua University, Beijing, China. His research interests include signal processing for communications in general, and massive MIMO, mmWave communications, and V2X communications in particular.

Permanent address: xiehx@utexas.edu

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.