

Copyright  
by  
Javier Rodríguez-Fernández  
2020

The Dissertation Committee for Javier Rodríguez-Fernández certifies that this is the approved version of the following dissertation:

**Millimeter Wave Link Configuration with Hybrid  
MIMO Architectures**

Committee:

---

Prof. Brian L. Evans, Supervisor

---

Prof. Jeffrey G. Andrews

---

Prof. Ahmed H. Tewfik

---

Prof. Murat Torlak (UT Dallas)

---

Prof. Haris Vikalo

**Millimeter Wave Link Configuration with Hybrid  
MIMO Architectures**

by

**Javier Rodríguez-Fernández, B.S., M.S.**

**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 2020

Dedicated to my parents, my brother, and my girlfriend.

## Acknowledgments

The path to arrive at this current point in my Ph.D. journey has been long and full of roadblocks. I would not like to forget any professor who accompanied in any of the stages of my education and helped me become the person I am today, especially during the last years of my academic education. First and foremost, I would like to thank my family, my father Odilo Rodríguez-Núñez, my mother Concepción Fernández-Arias, and my brother Martín Rodríguez-Fernández, for having always supported me, given me strength, and feeling their warmth, both during happy moments and the most difficult moments.

I would like to extend a warm thank you to the following professors from the University of Vigo (UVIGO): to Prof. Roberto López-Valcarce, who awakened my research interest in the field of signal processing for communications; to Prof. Artemio Mojón-Ojea, for his dedication and generosity, and the close treatment he has had with me; to Prof. Carlos Mosquera-Nartallo and Prof. Fernando Pérez-González, for having laid the foundations of digital communications and multiple-antenna communications; to Prof. Pedro Comesaña-Alfaro, for his exciting lectures on information and coding theory; and to Prof. Nuria González-Prelcic, for having hired me to work at ATLANTIC research center, for having funded my trips to conferences, and for having widened my expectations when I transferred to The University of Texas at

Austin (UT Austin); and to every colleague and labmate I had during my years at AtlantTIC.

I would like to thank: to Prof. Robert W. Heath Jr., for recommending me to the Ph.D. program at UT Austin; and to Prof. Renato Zanetti, for his patience during the lectures in advanced estimation theory, and for having widened my research field.

I would like to close my acknowledgments to professors from UT Austin by mentioning my Ph.D. committee members Prof. Jeffrey G. Andrews, Prof. Brian L. Evans, Prof. Prof. Murat Torlak, Prof. Ahmed H. Tewfik and Prof. Haris Vikalo. I would like to acknowledge to my Ph.D. committee who, as a whole, have made it possible for me to be currently writing these lines, and especially to my supervisor Prof. Brian L. Evans. I would like to extend a capitalized thank you to him, for his generosity, his professionalism, his humanity, and his selfless help and support. I know that there have been difficult moments for him, but in spite of everything, he never left me alone, and I will forever be indebted to him.

I would also like to acknowledge the emotional help and support I received from my colleagues Pooja Nuti, Elyes Balti, and Kevin Jinho Joe. I know that the most common thing to do is to name professors and study labmates in the acknowledgments, but I would not like to forget about the administrative staff who help us navigate through bureaucracies countless, whose valuable help makes the difference, especially: Ms. Carmen Touriño from UVIGO, and Ms. Melanie Gulick, Dr. Verónica Vera-Vasquez, Ms.

Karen Little, Ms. Apipol Piman, and Ms. Jaymie Udan from UT Austin.

# Millimeter Wave Link Configuration with Hybrid MIMO Architectures

Publication No. \_\_\_\_\_

Javier Rodríguez-Fernández, Ph.D.  
The University of Texas at Austin, 2020

Supervisor: Prof. Brian L. Evans

The use of multiple antennas, widely known as Multiple-Input Multiple-Output (MIMO) technology, is a key feature to deploy Millimeter Wave (mmWave) communication systems enabling high-data-rate applications. With more than two decades of global experience in deploying Wi-Fi and cellular communication using sub-6 GHz frequency bands, simply repurposing these designs for millimeter wave (mmWave) bands would fail to account for additional propagation impairments and circuit design constraints at these higher frequencies. A solution to overcome the propagation challenges is the use of multiple directional communication beams, whereby proper alignment between transceivers provides sufficient link quality to enable reliable decoding of the transmitted data.

In this dissertation, efficient link configuration solutions suitable for mmWave cellular communications are developed. To gain some insight into

the achievable performance of mmWave systems, two broadband channel-estimation-based link configuration solutions are proposed for MIMO-Orthogonal Frequency Division Multiplexing (OFDM) systems, in which both the transmitter and receiver are assumed to be perfectly synchronized. The proposed solution exploits the spatially common sparsity in the mmWave channel and enables efficient acquisition of the Channel State Information (CSI) while allowing the use of multiple Radio-Frequency (RF) chains on both the transmitter and receiver sides. In a simplified scenario, the Cramér-Rao Lower Bound (CRLB) for the channel estimation problem is derived, and the proposed channel estimation algorithms are shown to both outperform prior work in communication performance and exhibit excellent estimation performance. Furthermore, the proposed algorithms are assessed in a more challenging scenario with realistic channel parameters, and it is shown that both near-optimal spectral efficiency and low Bit Error Rate (BER) can be attained with lower overhead and computational complexity than prior solutions.

Next, the impact of imperfect Carrier Frequency Offset (CFO) synchronization on the channel estimation problem is analyzed under a narrowband channel model. The CRLB for the estimation of the different unknown parameters involved in the problem is theoretically analyzed, and closed-form expressions are provided for the estimation of the different parameters. Under a joint estimation-theoretic and Compressed Sensing (CS) framework, a low-complexity multi-stage solution is proposed to estimate both the different unknown synchronization parameters and the large-dimensional mmWave MIMO

channel. Different trade-offs between estimation, spectral efficiency, and overhead performance are exposed, and the proposed estimators are shown to be asymptotically optimal in the low Signal-to-Noise Ratio (SNR) regime. The proposed solution is assessed under a channel model with several clusters and rays per cluster, and is shown to attain near-optimal spectral efficiency values in both the low and high SNR regimes. The computational complexity of the proposed solution is also analyzed, in which it is shown to achieve a marginal increase in computational complexity with respect to the solution proposed in the previous contribution.

Finally, the impact of Timing Offset (TO), CFO, and Phase Noise (PN) impairments on the channel estimation problem is analyzed under a broadband channel model. The problem of time-frequency synchronization under PN impairments is theoretically analyzed, and the proposed solutions to the synchronization problem are exploited to estimate the frequency-selective mmWave MIMO channel. The hybrid CRLB for the estimation of the different synchronization impairments is analyzed, and closed-form expressions leveraging the information coupling between the different impairments are provided. The previously proposed joint estimation-theoretic and CS framework is extended to frequency-selective scenarios, and two low-complexity multi-stage solutions are proposed to estimate both the different synchronization impairments and the large-dimensional mmWave MIMO channel. The first solution relies on a batch-processing Linear Minimum Mean Square Error (LMMSE)-based Expectation-Maximization (EM) algorithm to estimate the different

synchronization impairments, while the second solution uses a sequential-processing Extended Kalman Filter (EKF)-Rauch-Tung-Striebel (RTS)-based EM algorithm, thereby reducing computational complexity. Thereafter, both the hybrid CRLB for the estimation of the equivalent beamformed complex channels and the estimates for these parameters are exploited to estimate the large-dimensional frequency-selective mmWave MIMO channel. Finally, a joint PN and data detection algorithm is proposed for data transmission under the 5G New Radio (NR) frame structure. The proposed solutions are evaluated using a 5G NR-based channel model, and different trade-offs between estimation performance, computational complexity, overhead, achievable spectral efficiency and BER are exposed, and comparisons with prior work are also provided. The results show that mmWave link configuration using hybrid MIMO architectures can be established with low overhead without assuming synchronization, even in the low SNR regime.

# Table of Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Abstract</b>	<b>viii</b>
<b>List of Tables</b>	<b>xvi</b>
<b>List of Figures</b>	<b>xvii</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Transceiver Structure of mmWave Systems . . . . .	2
1.2 Channel Models in mmWave MIMO Systems . . . . .	6
1.2.1 Why Using MIMO at mmWave? . . . . .	6
1.2.2 Space-Time MIMO Characteristics . . . . .	7
1.3 Channel Estimation in mmWave Systems . . . . .	10
1.4 Machine Learning in 5G New Radio . . . . .	13
1.5 Synchronization at mmWave . . . . .	17
1.6 Dissertation Summary . . . . .	22
1.6.1 Thesis Statement . . . . .	24
1.6.2 Contributions . . . . .	24
1.6.3 Thesis Organization . . . . .	27
1.6.4 Notation . . . . .	28
1.6.5 List of Acronyms . . . . .	29
<b>Chapter 2. Millimeter Wave Compressive Channel Estimation in the Frequency Domain</b>	<b>38</b>
2.1 Introduction . . . . .	38
2.1.1 Prior Work and Motivation . . . . .	39
2.1.2 Contributions . . . . .	42
2.2 System model . . . . .	45

2.3	Compressive Channel Estimation in the Frequency Domain . . .	48
2.3.1	Problem formulation . . . . .	49
2.3.2	Simultaneous Weighted - Orthogonal Matching Pursuit .	53
2.3.2.1	Support computation with correlated noise . . .	54
2.3.2.2	Computation of the channel gains . . . . .	55
2.3.3	Computation of the residual and noise variance estimation	57
2.3.4	Subcarrier Selection - Simultaneous Weighted - Orthogonal Matching Pursuit + Thresholding . . . . .	60
2.3.5	Convergence Analysis . . . . .	62
2.4	Numerical Results . . . . .	65
2.4.1	Normalized Mean Square Error (NMSE) Comparison . .	67
2.4.2	Spectral efficiency comparison . . . . .	75
2.4.3	Computational complexity . . . . .	77
2.4.4	Bit Error Rate . . . . .	81
2.5	Conclusions . . . . .	84
<b>Chapter 3. Millimeter Wave Compressive Channel Estimation with Carrier Frequency Offset Uncertainties</b>		<b>86</b>
3.1	Introduction . . . . .	86
3.1.1	Prior Work and Motivation . . . . .	87
3.1.2	Contributions . . . . .	90
3.2	System model . . . . .	91
3.3	Theoretical analysis of the estimation problem . . . . .	98
3.4	Estimation of beamformed channels and high-dimensional MIMO channel . . . . .	104
3.4.1	Dictionary-Constrained Channel Estimation . . . . .	109
3.5	Numerical Results . . . . .	114
3.5.1	Performance analysis of ML estimators . . . . .	116
3.5.2	NMSE of channel estimator . . . . .	122
3.5.3	Spectral efficiency . . . . .	124
3.5.4	Computational complexity . . . . .	130
3.6	Conclusions . . . . .	131

<b>Chapter 4. Millimeter Wave Broadband Synchronization, Compressive Channel Estimation and Data Transmission</b>	<b>132</b>
4.1 Introduction . . . . .	132
4.1.1 Prior work and Motivation . . . . .	133
4.2 Contributions . . . . .	136
4.3 System model with synchronization impairments . . . . .	139
4.4 Theoretical analysis of the estimation problem . . . . .	144
4.4.1 Computation of the Hybrid Information Matrix (HIM) . . . . .	148
4.5 Estimation of equivalent beamformed channels and high-dimensional MIMO channel . . . . .	156
4.5.1 LMMSE-EM Algorithm . . . . .	158
4.5.2 EKF-RTS-EM Algorithm . . . . .	165
4.5.3 Initialization and Convergence . . . . .	168
4.5.4 Dictionary-Constrained Channel Estimation . . . . .	170
4.6 Joint Data Detection and PN Mitigation . . . . .	176
4.7 Numerical Results . . . . .	183
4.7.1 Performance analysis of EM-based algorithms . . . . .	185
4.7.2 NMSE of channel estimator . . . . .	196
4.7.3 Spectral efficiency . . . . .	199
4.7.4 Computational complexity analysis . . . . .	206
4.7.5 Bit Error Rate and Modulation Error Ratio . . . . .	212
4.8 Conclusions . . . . .	213
<b>Chapter 5. Conclusions and Future Work</b>	<b>218</b>
5.1 Summary . . . . .	218
5.2 Future Work . . . . .	222
<b>Appendices</b>	<b>229</b>
<b>Appendix A. Derivation of the Cramér-Rao Lower Bound</b>	<b>230</b>
<b>Appendix B. Fundamentals of Compressive Sensing</b>	<b>233</b>

<b>Appendix C. Fundamentals of Estimation Theory</b>	<b>239</b>
C.1 Fisher Information and Cramér-Rao Lower Bound . . . . .	240
C.2 Finding Optimal Estimators . . . . .	242
<b>Bibliography</b>	<b>246</b>
<b>Vita</b>	<b>274</b>

## List of Tables

2.1	Summary of typical system configuration parameters . . . . .	66
2.2	Online computational complexity of proposed algorithms . . . . .	78
2.3	Online computational complexity of previously proposed algorithms . . . . .	79
3.1	Summary of typical system configuration parameters . . . . .	117
3.2	Online computational complexity of proposed algorithm . . . . .	130
4.1	Summary of typical system configuration parameters . . . . .	184
4.2	Online computational complexity of proposed synchronization and channel estimation algorithms . . . . .	207
4.3	Online computational complexity of previously proposed algorithm . . . . .	208
4.4	Asymptotic complexity from Tables 4.2 and 4.3 and number of measured Floating Point Operations (FLOPs) required by the proposed LMMSE-EM, EKF-RTS-EM, and PN mitigation and data detection algorithms for the set of system parameters in Table 4.1 of $K = 256$ , $L_r = 4$ , and $N_{tr} = 2$ , and algorithm parameters $N_{\Delta f} = 129$ , and $N_{DMRS} = 2$ . . . . .	209
4.5	Asymptotic complexity and number of measured FLOPs required by the proposed LMMSE-EM, EKF-RTS-EM, and PN mitigation and data detection algorithms for the set of System Parameters 2: $K = 1024$ , $L_r = 4$ , $N_{tr} = 2$ , $N_{\Delta f} = 129$ , and $N_{DMRS} = 2$ . . . . .	210

## List of Figures

1.1	Illustration of a fully-connected hybrid Zero Prefix (ZP)-MIMO-OFDM architecture including analog RF precoders and combiners, and digital baseband frequency-selective precoders and combiners. This figure has been taken from [1]. . . . .	3
1.2	Illustration of a fully-connected hybrid MIMO architecture including analog RF precoders and combiners, and digital baseband frequency-flat precoders and combiners. This figure has been taken from [2]. . . . .	4
1.3	Illustration of a partially-connected hybrid MIMO-OFDM architecture including analog RF precoders and combiners, and digital baseband frequency-selective precoders and combiners. This figure has been taken from [3]. . . . .	5
1.4	Illustration of the space-time clustering behavior of multipath components in mmWave MIMO channel models. The different multipath components the mmWave channel comprises arrive closely spaced in both the spatial and temporal domains, thereby being organized in clusters. . . . .	9
2.1	Summary and comparison of prior work on broadband channel estimation algorithms as well as the approaches proposed in this Chapter. . . . .	40
2.2	Illustration of the structure of a hybrid MIMO architecture, which include analog and digital precoders and combiners. This structure was already introduced in Fig. 1.1. . . . .	45
2.3	Detailed steps of the first proposed SW-OMP algorithm. . . . .	59
2.4	Detailed steps of the second proposed SS-SW-OMP+Th algorithm. . . . .	63
2.5	Evolution of the NMSE versus SNR for the different frequency-domain algorithms when the AoD/AoA are assumed to lie on the dictionary grid. The number of training frames is set to $M = 80$ (a) and $M = 120$ (b). . . . .	68
2.6	Comparison of evolution of the NMSE versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 80$ . The number of subcarriers is set to $K = 64$ . . . . .	71

2.7	Evolution of the NMSE versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$ (a) and $M = 100$ (b). The channel realizations are taken from NYUSIM channel simulator. . . . .	72
2.8	Comparison of evolution of the NMSE versus $M$ at different SNR. The SNR is set to $-10$ dB (a) and (d), $-5$ dB (b) and (e) and $0$ dB (c) and (f). Plots (a), (b), (c) consider on-grid angular parameters on the channel realizations, while (d), (e), (f) consider the off-grid case with channel realizations extracted from the NYUSIM channel simulator [4]. . . . .	73
2.9	(a) Evolution of the spectral efficiency versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$ . (b) Evolution of the spectral efficiency versus number of training frames $M$ at different SNR for the different frequency-domain algorithms. . . . .	76
2.10	(a) Evolution of the coded BER versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$ . (b) Evolution of the uncoded BER versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$ . . . . .	83
3.1	Summary and comparison of prior work on narrowband channel estimation algorithms with impairments as well as the approach proposed in this Chapter. . . . .	88
3.2	Illustration of the structure of a hybrid MIMO architecture, which includes analog and digital precoders and combiners (same as Fig. 1.2). . . . .	92
3.3	Block diagram of the different components in the proposed estimation framework. . . . .	105
3.4	Detailed steps of the proposed W-OMP algorithm. . . . .	115
3.5	Normalized sample variance of the ML estimators from Section 3.4 as a function of SNR. The normalized CRLB is also shown to compare the performance of the Maximum Likelihood (ML) estimators to the minimum achievable estimation variance. The FFT size of the CFO estimator is set to $4L_rN$ . . . . .	118
3.6	Normalized sample variance of the ML estimators from Section 3.4 as a function of SNR. The normalized CRLB is also shown to compare the performance of the ML estimators to the minimum achievable estimation variance. The FFT size of the CFO estimator is set to $K = 4N = 1024$ points. . . . .	121

3.7	Normalized sample variance of the ML estimators from Section 3.4 as a function of $N$ , for SNR = -10 dB ((a) and (d)), SNR = -5 dB ((b) and (e)), and SNR = 0 dB ((c) and (f)). The normalized CRLB is also shown to compare the performance of the ML estimators to the minimum achievable estimation variance. The FFT size of the CFO estimator is set to $4L_rN$ . . . . .	123
3.8	Evolution of the NMSE achieved by the joint synchronization and channel estimation algorithm as a function of SNR (a) and $M$ (b). The number of training frames is $M = \{16, 32, 64, 128, 256, 512\}$ , and the length of the time-domain training sequence is $N = 128$ . . . . .	125
3.9	Evolution of the spectral efficiency versus SNR for different precoders and combiners. The number of time-domain training symbols is $N = 128$ , and the number of training frames is $M = 128$ . . . . .	128
3.10	Evolution of the spectral efficiency versus $M$ for different precoders and combiners. The number of time-domain training symbols is $N = 128$ . The SNR is set to SNR = $\{-10, 0\}$ dB (a) and SNR = $\{-5, 5\}$ dB (b). . . . .	129
4.1	Summary and comparison of prior work on broadband channel estimation algorithms with synchronization impairments, as well as the approaches proposed in this Chapter. . . . .	134
4.2	Illustration of the structure of a partially-connected hybrid MIMO architecture, which includes analog and digital precoders and combiners (same as Fig. 1.3). . . . .	141
4.3	Block diagram of the proposed estimation-theoretic framework for joint synchronization and compressive channel estimation under broadband mmWave MIMO channels. . . . .	158
4.4	Detailed steps of the first proposed LMMSE-EM algorithm. . . . .	164
4.5	Detailed steps of the second proposed EKF-RTS-EM algorithm. . . . .	169
4.6	Evolution of the NMSE of the CFO estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound. . . . .	186
4.7	Asymptotic evolution of the NMSE of the CFO estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound. . . . .	190
4.8	Evolution of the NMSE of the beamformed channel estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound. . . . .	191

4.9	Evolution of the NMSE of the PN estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound. . . . .	193
4.10	Asymptotic evolution of the NMSE of the PN estimates obtained using the proposed algorithms versus SNR. The curves in (a) show this asymptotic evolution for SNR values between $-20$ and $40$ dB. Magnified curves of the asymptotic evolution are also shown for $\text{SNR} \in [-20, 0]$ dB in (b), $\text{SNR} \in [0, 20]$ dB in (c), and $\text{SNR} \in [20, 40]$ dB in (d). The hybrid CRLB is also provided as a performance bound. . . . .	194
4.11	Evolution of the NMSE of the mmWave MIMO channel estimates obtained using the proposed algorithms versus $M$ , for $\text{SNR} = -10$ dB (a) and $\text{SNR} = 0$ dB (b). . . . .	197
4.12	Evolution of the ergodic spectral efficiency achieved by the proposed algorithms and the compressive beamforming algorithm in [5] under PN impairments versus SNR, for $N_{\text{tr}} = 1$ OFDM training symbol. The curves in (a) are obtained using the mmMAGIC Line-Of-Sight (LOS) channel model, while those in (b) are obtained using the mmMAGIC Non Line-Of-Sight (NLOS) channel model with a Rician factor of $-10$ dB. . . . .	200
4.13	Evolution of the ergodic spectral efficiency achieved by the proposed algorithms versus SNR, for $N_{\text{tr}} = 2$ OFDM training symbol. The curves are obtained using the Third Generation Partnership Protocol (3GPP) Urban Microcell (UMi) NLOS channel model [6]. . . . .	203
4.14	Evolution of the ergodic spectral efficiency achieved by the proposed algorithms versus the number of training frames $M$ , for $\text{SNR} = -10$ dB (a) and $\text{SNR} = 0$ dB (b), and $N_{\text{tr}} = 2$ OFDM training symbols. The curves are obtained using the 3GPP UMi NLOS channel model [6]. . . . .	205
4.15	Evolution of the BER versus SNR (a) and Modulation Error Ratio (MER) versus number of iterations in the proposed PN mitigation and data detection algorithm (b), for $\text{SNR} = \{-14, -11, -8, -5, -2\}$ dB before beamforming, using a 16-Quadrature Amplitude Modulation (QAM) constellation and $N_s = 2$ data streams. . . . .	214
4.16	Evolution of the BER versus SNR (a) and MER versus number of iterations in the proposed PN mitigation and data detection algorithm (b), for $\text{SNR} = \{-14, -11, -8, -5, -2, 1\}$ dB before beamforming, using a 64-QAM constellation and $N_s = 2$ data streams. . . . .	215

4.17	Evolution of the BER versus SNR (a) and MER versus number of iterations in the proposed PN mitigation and data detection algorithm (b), for SNR = $\{-14, -11, -8, -5, -2, 1\}$ dB before beamforming, using a 256-QAM constellation and $N_s = 2$ data streams. . . . .	216
5.1	Illustration of qualitative performance of proposed broadband channel estimation and synchronization algorithms. . . . .	222

# Chapter 1

## Introduction

The mmWave band holds promise for developing communication systems that enable high data rates in wireless local area networks and fifth generation (5G) cellular networks [7–10]. The small carrier wavelengths characterizing mmWave frequencies enable synthesis of compact antenna arrays having a large number of antenna elements, which are capable of providing the necessary array gain to establish high-quality communication links while reducing mean interference levels [7, 8, 10]. To obtain high link quality, transmit and receive antenna arrays need to be properly configured to compensate for the small antenna aperture of antenna elements at mmWave frequencies. In this dissertation, I focus on two critical issues of implementing mmWave MIMO systems in practice, which are: (i) channel estimation to enable high data rate communications, and (ii) synchronization to enable channel estimation at the low SNR regime.

In this chapter, I explain the motivation of the research problems addressed in this dissertation and provide a summary of my contributions. In Section 1.1, I illustrate the transceiver structure of mmWave systems. In Section 1.3, I highlight the challenges of estimating the CSI in mmWave sys-

tems. In Section 1.5, I point out the importance of achieving synchronization at mmWave before performing channel estimation. Finally, Section 1.6 summarizes the dissertation, including presenting the thesis statement, the contributions, organization, a list of abbreviations and the notation used in the dissertation.

## 1.1 Transceiver Structure of mmWave Systems

Unlike fully-digital architectures employed in traditional sub-6 GHz MIMO systems, hybrid analog and digital precoding has become an attractive candidate to exploit both beamforming and spatial multiplexing gains in hardware-constrained mmWave communication systems [11]. The idea of a hybrid analog-digital solution for the precoders and combiners was first proposed in [12], and then developed in [11] for sparse narrowband MIMO channels at mmWave frequencies. In Figure 1.1, a precoded MIMO-OFDM system with ZP is depicted, which employs  $K$  subcarriers and a hybrid precoding and combining transceiver structure.

In a downlink scenario, a Base Station (BS) or a transmitter is equipped with  $N_t$  transmit antennas and  $L_t$  RF chains. Likewise, a User Equipment (UE) or a receiver uses  $N_r$  receive antennas and  $L_r$  RF chains. A similar definition would apply for the uplink. The example in Figure 1.1 shows a fully-connected array architecture, in which every antenna element is jointly controlled by the different RF chains sharing the same network of phase-shifters [13]. In this example, both the transmitter and receiver use Uniform Linear

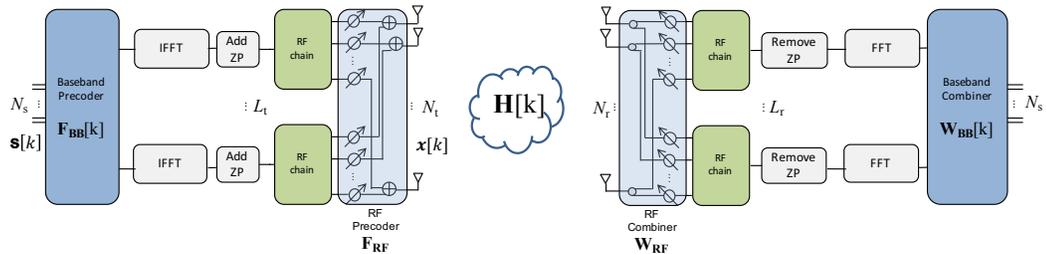


Figure 1.1: Illustration of a fully-connected hybrid ZP-MIMO-OFDM architecture including analog RF precoders and combiners, and digital baseband frequency-selective precoders and combiners. This figure has been taken from [1].

Array (ULA). Other array geometries (e.g., Uniform Planar Array (UPA), non-uniform arrays), and hybrid architectures (e.g., partially-connected architectures) are also possible.

In the hybrid transceiver structure, the number of RF chains is usually much smaller than the number of antennas to reduce power consumption. The number of data streams  $N_s$  is limited by  $N_s \leq \min(L_t, L_r)$ . This splits mmWave precoding/combining into the analog and digital domains. The analog precoder and combiner are subject to the hardware constraints imposed by the particular analog networks. In the example of Figure 1.1, the analog precoding and combining networks are based on phase-shifters, which impose constant-modulus constraints for each phase-shifter, and only the phases are configured, typically using a finite number of possible phase-shifts, which is dictated by the number of quantization bits used to configure such phase-shifts. In context of hybrid mmWave MIMO systems, the problem of designing hybrid analog-digital precoders and combiners for data transmission has been exten-

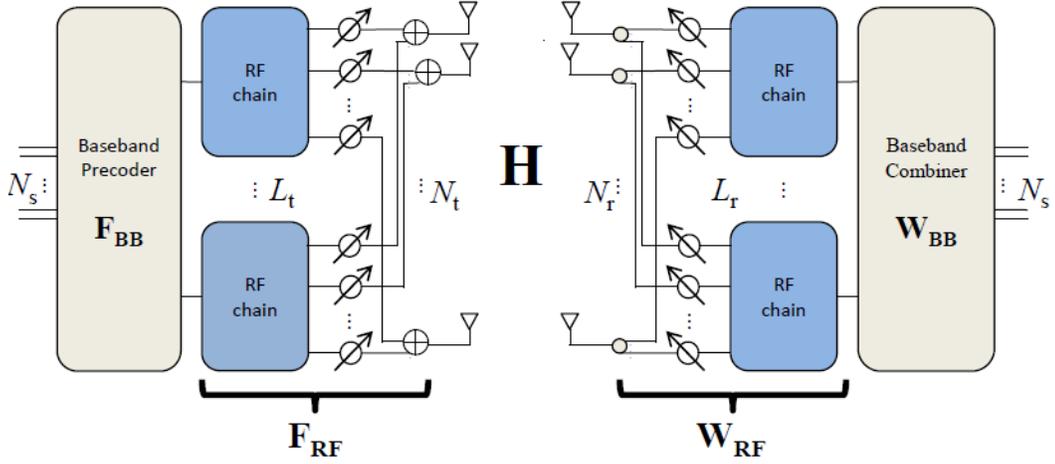


Figure 1.2: Illustration of a fully-connected hybrid MIMO architecture including analog RF precoders and combiners, and digital baseband frequency-flat precoders and combiners. This figure has been taken from [2].

sively studied, both for narrowband [11, 13–19], and broadband channel models [20–27]. In Figure 1.3, an example of a partially-connected hybrid MIMO architecture is shown, in which each RF chain is connected to a subset of antennas, rather than being connected to every antenna as in Figure 1.1, through a serial combination of phase-shifters and switches. In this case, the absence of connection between every RF chain and every antenna makes hybrid precoding/combining design more difficult due to the reduction in flexibility that the partially-connected structure introduces. A benefit of this architecture, however, is the increase in energy efficiency during communication, which comes from the additional reduction in power consumption owing to using a reduced number of phase-shifters [28]. Another important benefit of this architecture is the ease of design and fabrication, since partially-connected architectures

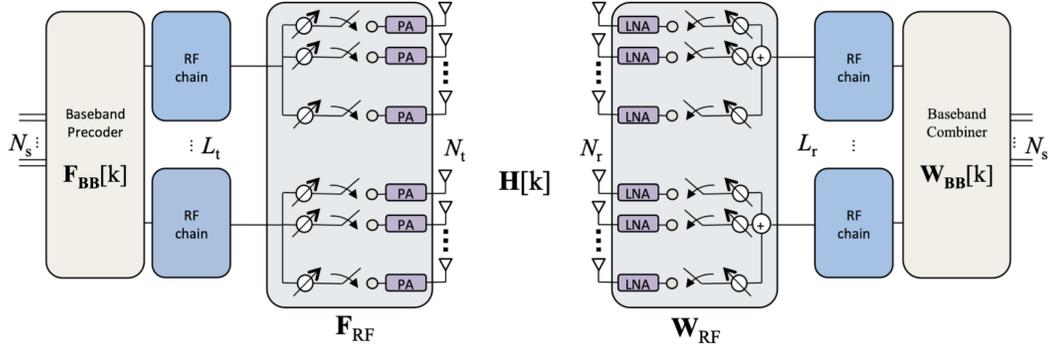


Figure 1.3: Illustration of a partially-connected hybrid MIMO-OFDM architecture including analog RF precoders and combiners, and digital baseband frequency-selective precoders and combiners. This figure has been taken from [3].

result in a smaller number of connections between RF chains and antennas, which reduces possible electromagnetic coupling between phase-shifters. Due to the use of analog components, the spatial processing performed by analog precoders/combiners is modeled using frequency-flat matrices  $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times L_t}$ ,  $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times L_r}$  with entries having constraints depending on each particular analog network. The digital precoders/combiners, however, are not limited by constant-modulus constraints and can be optimized by leveraging the available CSI. They can be modeled as either frequency flat, as shown in Figure 1.2, or as frequency-selective spatial filters, as depicted in Figure 1.1 and Figure 1.3.

In this dissertation, I develop synchronization and channel estimation techniques for mmWave MIMO systems employing hybrid architectures. For broadband channel estimation algorithm design, I use fully-connected hybrid architectures as shown in Figure 1.1. In my proposed CFO estimation and

channel estimation design, I also use fully-connected hybrid architectures, yet the baseband precoders and combiners are frequency flat owing to the mmWave channel being modeled as narrowband. Finally, motivated by the use of array-of-subarrays-based hybrid architectures in 5G NR, I use partially connected hybrid architectures for joint synchronization and channel estimation design, as depicted in Figure 1.3.

## 1.2 Channel Models in mmWave MIMO Systems

A unique aspect of mmWave communications stems from the very particular propagation characteristics present at such high frequencies. Understanding these features plays a fundamental role in designing and developing signal processing algorithms for mmWave transceivers [8]. In this section, I highlight two different aspects of mmWave channel modeling, namely the relation between the path loss during propagation and the use of multiple antennas, and the spatial characteristics of multipath channel models.

### 1.2.1 Why Using MIMO at mmWave?

For free-space communication, the received power  $P_r$  is related to the transmit power  $P_t$  by Friis' law [29] as

$$P_r = G_r G_t \left( \frac{\lambda}{4\pi d} \right)^2 P_t, \quad (1.1)$$

where  $d$  denotes the distance between transmitter and receiver,  $\lambda$  is the wavelength, and  $G_t$ ,  $G_r$  are the transmit and receive antenna gains. Friis' law in

(1.1) indicates that, under unity antenna gains, the received power decreases as the squared wavelength decreases, thereby implying that mmWave propagation will experience higher path loss than sub-6 GHz communication systems in the absence of directional antenna gains [8]. While this is true for omnidirectional antennas, the theoretical maximum directional antenna gains are also dependent on the wavelength, so that the increased path loss can be compensated, as illustrated shortly. The physical antenna aperture,  $A_e$ , is related to the wavelength and antenna gain  $G$  as

$$A_e = \frac{\lambda^2}{4\pi} G. \quad (1.2)$$

The antenna aperture in (1.2) shows that, for a given physical antenna aperture, the maximum directional gain  $G$  scales as  $G \propto \lambda^{-2}$ , which indicates that a larger number of antenna elements can be fit into the same physical area [8]. Therefore, incorporating a large number of antenna elements can compensate for the increased free-space path loss at mmWave frequencies. This solution requires directional transmission with high-dimensional antenna arrays, which makes incorporating MIMO into mmWave communication a natural choice.

### 1.2.2 Space-Time MIMO Characteristics

In sub-6 GHz communications, the propagation environment typically features rich scattering owing to both the significant diffraction present in these frequency bands and the small path loss in comparison with mmWave communications. The richness in scattering justifies the adoption of analytical

channel models such as the Rayleigh, Rice, and Kronecker models [29], which attempt to recreate the statistical behavior of the channel matrix [29].

In mmWave communications, however, the propagation environment does not feature rich scattering, which usually makes these analytical models not applicable at higher frequencies. The main principle that drives the modeling of mmWave MIMO channels, however, is an attempt to recreate the geometry of the propagation environment, which is already used in sub-6 GHz communication systems and yields the geometric channel model [6, 30]. Since diffraction is far from being a predominant propagation effect in mmWave bands, recreating the geometry of the environment is usually justified by the reduced number of multipath components present in the channel.

From a geometric perspective, one of the predominant propagation effects at mmWave is space-time clustering, which makes mmWave channels usually consist of a few clusters with several rays per cluster such that rays belonging to the same cluster arrive closely spaced in both the temporal and spatial domains, as shown in Fig. 1.4. Following this idea, statistical channel models used for system simulation usually describe the multipath components as arriving in clusters having certain distributions on the Delay Spread (DS), power, and Angular Spread (AS) which are used to statistically characterize each multipath [4], [6], [31], [32]. Physically, the different clusters correspond to different macro-level paths, and the DS and AS within each cluster captures the scattering from diffuse reflections along those paths [8]. The effect of a reduced number of multipath components in the mmWave channel is commonly

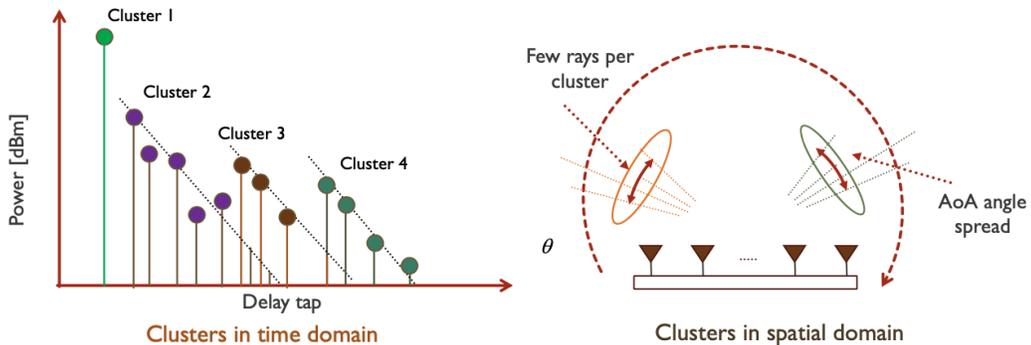


Figure 1.4: Illustration of the space-time clustering behavior of multipath components in mmWave MIMO channel models. The different multipath components the mmWave channel comprises arrive closely spaced in both the spatial and temporal domains, thereby being organized in clusters.

referred to as sparsity [33], [34], [17], [14], [11], [4], [35], which is the effect whereby most of the energy contained in the channel is concentrated around a few components in both the temporal and spatial domains, as shown in Fig. 1.4. The fact that the different multipath components in the channel are organized in spatial and temporal clusters justifies making the distinction between sparsity in the delay domain and sparsity in the spatial (angular) domain, which are crucial features to take into account in order to design and develop channel estimation algorithms, as discussed in the following sections. In narrowband mmWave channel models, sparsity is only significant in the spatial (angular) domain, while for frequency-selective channel models, sparsity arises in both the delay and the spatial domains.

### 1.3 Channel Estimation in mmWave Systems

Knowledge of the wireless propagation channel is crucial to exploit the full benefits of MIMO techniques in mmWave cellular systems. At sub-6 GHz frequency bands, MIMO systems rely on classical channel estimation techniques (e.g. Least Squares (LS), LMMSE), which are not applicable for mmWave frequencies owing to the deployment of large antenna arrays at mmWave transceivers and the use of hybrid precoding and combining [8].

At mmWave, the main motivation of link configuration is to increase the received Signal-to-Interference plus Noise Ratio (SINR) after spatial processing. Two main approaches to accomplish this task are beam training and channel estimation [8]. On the one hand, beam training, adopted in IEEE 802.11ad [36] and 5G NR [37], is a technique that avoids explicit estimation of the channel by iteratively searching for transmit and receive beam pairs that maximize the received SINR, thereby enabling reliable decoding of transmitted data [17, 38–49]. Directional beam training is the most popular link configuration strategy, whose popularity stems from its simplicity and low complexity. Even though beam training methods can increase link quality in both narrowband and broadband systems, these algorithms generally share the disadvantage of converging towards only one communication beam. Extensions to multi-stream and multi-user communication are possible, but they generally require much higher overhead.

Channel estimation, on the other hand, is more flexible since it allows both multi-stream and multi-user communication, thereby overcoming

the limitation of beam training strategies. Channel estimation is significantly more challenging in mmWave systems than in sub-6 GHz systems. In addition to the large training overhead associated with the large antenna arrays [14], the SNR is typically low before hybrid precoders and combiners can be designed for data transmission. Further, the hardware constraints, that result from RF/hybrid signal processing, make the channel at baseband seen only through the RF analog network, which acts as a compression stage for the received signal. Consequently, received samples at antenna level are not directly available to estimate the channel. Exploiting spatial sparsity has been critical in formulating practical channel estimation algorithms for hybrid MIMO architectures [11, 17]. The main reason is that, when sparsity is exploited, the high-dimensional channel can be decomposed in terms of dictionaries under which the channel is sparse, such that an optimization problem can be formulated to estimate the locations of the sparse coefficients and their values. Channel estimation is different from beam training [39, 40, 50], which attempts to find the beams pointing in the most promising angular direction instead of estimating the channel.

Initial solutions to estimate the channel using hybrid architectures were based on exploiting the angular sparsity of the mmWave channel [14, 17, 51], [52–55], [56–60]. In the frequency-flat, narrowband channel model, the sparsity in the angular domain is leveraged by using the extended virtual channel model [8]. Essentially, the MIMO channel is written in terms of dictionary matrices built from transmit and receive array steering vectors evaluated on

uniform grids for the Angle of Departure (AoD) and Angle of Arrival (AoA) of each multipath component. These dictionaries operate as a sparsifying basis for the channel matrix, thereby enabling the formulation of the channel estimation problem as a sparse recovery optimization problem. Using this formulation, several channel estimation algorithms have been developed for hybrid architectures [14, 17, 51–53]. These algorithms differ in how the measurement matrix is designed to search for the dominant AoD and AoA. The measurement matrix can be designed using adaptive CS [17, 54, 55], random CS [14, 51, 53], a mixture of random and adaptive CS [53], and deterministic CS [61–64]. Other non-compressive techniques were also developed for mmWave channel estimation, using either subspace estimation [65], overlapped beams [66, 67], and auxiliary beams [68]. The main limitation of these techniques is that they are tailored to narrowband channel models, hence not directly applicable to frequency-selective scenarios.

Recently, some approaches for channel estimation in frequency-selective mmWave channels have been proposed. These techniques aim at exploiting additional features of the mmWave channel, such as sparsity in the delay domain and spatial congruence between frequency subbands [34, 69, 70]. In [34], a time-domain approach aiming at estimating the wideband mmWave MIMO channel was devised, exploiting space-time sparsity to estimate the channel using dictionaries in both the angular domain and the delay domain. The main limitation of [34] is the high computational complexity of the proposed algorithm. Based on the Structured Sparsity-Adaptive Matching Pur-

suit (SSAMP) algorithm [71], another approach to estimate the wideband channel exploiting spatial congruence in the frequency domain was proposed in [69]. Another approach exploiting common sparsity was proposed in [70]. Prior work in [69], [70] share the limitations of not considering a realistic band-limited channel response [33], and employing Gaussian measurement matrices to estimate the channel, which cannot be implemented using hybrid architectures. Furthermore, the algorithms are only evaluated in the high SNR regime. Hence, it is necessary to devise new wideband channel estimation solutions that are compatible with hybrid architectures, realistic channel models, and exhibiting reduced computational complexity in comparison with current solutions.

## 1.4 Machine Learning in 5G New Radio

Most prior work on CSI acquisition at mmWave, encompassing both beam-training-based and channel-estimation-based link configuration, are model-driven methods with the ultimate goal of effectively configuring antenna arrays in mmWave transceivers to enable high-quality data transmission. During the last few years, data-driven methods have also been studied to solve this problem, which generally fall within the realm of machine learning [72,73]. Machine learning is a suitable framework for problems which are too complex owing to high non-linearity and/or non-convexity which often results in the problem becoming intractable through model-based approaches. Furthermore, machine learning can be an appropriate approach in problems requiring intensive man-

ual hand-tuning of parameters. These difficult problems can be dealt with by learning a function based on available data, which relates the inputs of a system to its outputs. Thereby, conventional model-based solutions can be replaced by machine learning algorithms that automatically learn from previous data [74].

There are four main different subcategories within the realm of machine learning, namely supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [72–74]. These different categories are named according to the level of supervision that the machine learning procedure requires on the training stage.

In supervised learning, a learning model is trained with samples in which inputs and outputs are labeled and paired, thereby enabling finding the optimum solution to the problem at hand for the training samples [74], by using a decision tree, neural network, etc. [72]. Then, the learning model is used to predict solutions from new samples. This machine learning category consistently provides excellent accuracy results across of variety of problems in wireless communications, such as dynamic frequency and bandwidth allocation [75], path-loss prediction modeling [76], channel learning [77], and prediction of beamforming vectors [78, 79], but it exhibits very high computational complexity. To reduce computational complexity, unsupervised learning was born to solve complex problems without supervision.

Under unsupervised learning, the data used to train the machine learning algorithms is an unlabeled collection of features, and the machine learn-

ing algorithm attempts to separate subgroups or clusters exhibiting similar characteristics among the different variables without any guidance or supervision [72, 74]. Unsupervised learning has found several applications within the realm of wireless communications, including cooperative spectrum sensing [80], coverage planning of Heterogeneous Networks (HetNets) with dynamic clusters [81], and resource management [82].

Finally, reinforcement learning is a variant of machine learning that aims at finding optimum policies in stochastic environments under uncertainty. The behavior of wireless networks, for instance, evolves according to stochastic dynamics that can be modeled using a Markov Decision Process (MDP) [74] comprising several states. Reinforcement learning aims at finding the best policy to maximize the rewards by selecting the properest action in a given state [74], which has been applied to a variety of wireless communication problems such as load balancing [83, 84], mobility management [85], and resource allocation [86], and joint beamforming, power control and interference coordination [87].

In the context of 5G NR communications, machine-learning-based techniques have been investigated to solve the problem of beam selection [78, 88–91], channel estimation/tracking [92–94], downlink Coordinated MultiPoint (CoMP) transmission [95], and joint beamforming, power control and interference coordination [87]. These methods generally rely on training neural networks to predict either the optimal beam pairs or the mmWave channel, in which explicit feedback from the UE to the BS may be required [78, 87, 90].

A main limitation that these papers share is that, similar to prior work on channel estimation at mmWave, perfect synchronization is assumed at the receiver side. Furthermore, the proposed approaches are tailored to narrowband channel models except for the approach in [87].

With regards to channel estimation, in [92], a deep CS-aided feedforward network is used to estimate the mmWave Multiple-Input Single-Output (MISO) channel in a massive MISO network. Thereafter, a convolutional neural network is proposed to configure the hybrid precoders given the predicted channels. The channel tracking problem is studied in [93], in which past predicted CSI is used to predict the future channels using a Long Short Term Memory (LSTM) structure. In [94], a convolutional neural network is proposed to predict the mmWave MIMO channel by treating the sparse channel as a natural image. While these approaches comprise initial interesting alternatives to model-based channel estimation solutions, the algorithms in [92–94] are assessed using simplistic channel models comprising very few paths [92,93], or the system is assumed to work under a high SNR assumption [94], which is impractical at mmWave. Furthermore, a common limitation of these papers, similar to machine learning-based beam training solutions, is that perfect synchronization is assumed at the receiver side, which is also impractical since mmWave systems are expected to work in the low SNR regime.

## 1.5 Synchronization at mmWave

Channel estimation is crucial to design hybrid precoders and combiners maximizing performance metrics such as the spectral efficiency or the received SINR at mmWave. Estimating the channel in practice, however, requires TO, CFO, and PN estimation and compensation. The problem of PN estimation has been studied in [96] for Single-Input Single-Output (SISO) systems under an Additive White Gaussian Noise (AWGN) channel model, and the problem of joint CFO, channel and PN estimation have been studied for SISO communication systems under OFDM signaling in [97], [98], [99]. In the context of MIMO-OFDM systems, the effects of PN in MIMO-OFDM systems were studied in [100], [101], [102], but there is a lack of understanding of theoretical limits and performance of signal processing algorithms in the context of mmWave MIMO systems, even under the assumption of a narrowband channel model.

Compared to TO, OFDM is very sensitive to CFO and PN, which arise due to instabilities and thermal noise in the local oscillator [103]. CFO and time-varying PN result in both Common Phase Error (CPE) and Inter Carrier Interference (ICI) at the receiver, which degrade the performance of OFDM systems [104], [105], [106], [97], [98]. In particular, the impact of PN in systems operating at mmWave frequency bands can be even more profound [107]. Thus, as wireless communication systems and standards, e.g., IEEE 802.11ad [36], IEEE 802.11ay [108], and 5G NR [37] migrate to mmWave frequencies to take advantage of the large bandwidth and adopt higher order modulations and

closely spaced subcarriers to achieve higher spectral efficiencies, it is increasingly important to develop efficient and accurate estimation and detection algorithms to compensate for the effect of CFO and PN in MIMO-OFDM systems.

In the context of mmWave MIMO systems, synchronization parameters need to be properly estimated and compensated for before CSI can be acquired. This sets new challenges as synchronization acquisition must be performed at the low SNR regime, before transmit and receive communication beams can be aligned for data transmission. Joint beam training and synchronization is a popular low-complexity solution to the beam alignment problem, and it has been incorporated into the 5G NR beam management procedure for initial access [42, 109]. The main advantage of joint beam training and synchronization stems from the fact that, when the beam pair that maximizes the received SNR is probed, synchronization can be performed at the high SNR regime, thereby enabling the application of well-known CFO retrieval algorithms such as the Moose [110] or the Schmidl-Cox [111] algorithms for TO and CFO estimation and compensation.

The problem of CS based beam training and synchronization is studied in [5, 112]. In [112], the problem of beam training under PN errors and unknown CFO was studied for narrowband MIMO systems using analog architectures. In [112], an EKF is proposed to track the joint phase of the unknown PN and beamformed narrowband channel, the phase of the received signal is compensated, and then Matching Pursuit (MP) is used to estimate the dom-

inant AoD and AoA. In [5], a compressive initial access approach based on omnidirectional pseudorandom analog beamforming is proposed as an alternative to the directional initial access procedure used during beam management in 5G NR, and the effects of imperfect TO and CFO are studied therein.

Although [5] is an interesting initial solution, the proposed algorithm is tailored to LOS channel models and presence of phase measurement errors due to CFO, thereby ignoring space-time clustering behavior and the PN impairment. While the application of joint beam training and synchronization techniques is interesting and can provide reasonable performance while keeping complexity low, beam training techniques generally converge to a single communication path, and are difficult to scale to multi-stream and multi-user settings without dramatically increasing the training overhead and receiver complexity. Besides performing synchronization at the high SNR regime, an advantage of beam training algorithms is that phase information is not explicitly needed for beam configuration, while channel estimation algorithms generally depend on both the magnitude and phase of the received signal. The problem of channel estimation without phase measurements was solved using Received Signal Strength (RSS) matching pursuit [113], Hash table [114], and sparse phase retrieval [115], although these phase-free measurements were associated with a particular testbed, a constraint that does not necessarily apply to mmWave systems in general.

Prior work on joint channel estimation and synchronization for mmWave MIMO is limited, since most of prior work on channel estimation assumes

perfect synchronization at the receiver side [34], [60], [58], [69], [70]. Prior work on channel estimation at mmWave considering synchronization impairments is limited, both for narrowband [116, 117] and broadband channel models [118–121]. In [118], the problem of joint channel and PN estimation for a SISO system is considered, which is unrealistic at mmWave, and the proposed algorithms are only evaluated in very high SNR regime. In [116], a tensor-based joint CFO and channel estimation algorithm is proposed for analog-only mmWave MIMO systems, which exhibits very high computational complexity, and it assumes that analog beamformers and combiners can be reconfigured for each transmitted sample, which is unlikely to happen in practice [36]. In [119], analog-only architectures with a single RF chain are assumed, and an autocorrelation-based iterative algorithm is proposed to jointly estimate the CFO and the mmWave channel. Similarly to [116], [119] assumes that analog beamformers and combiners can be instantaneously reconfigured for two consecutive transmitted time-domain samples. Further, the algorithm proposed in [119] has only been evaluated for mmWave channels having a very reduced number of non-clustered multipath components, which is not realistic at mmWave [35]. In addition, owing to the nature of the autocorrelation function, the proposed algorithm does not perform well both when the CFO is considerably large and the SNR is low.

In [120], a CFO-robust beam alignment technique is developed to find the beam pairs maximizing the received SNR. The main limitation of [120] is that the algorithm proposed therein can only be applied to analog MIMO

architectures, and its CFO correction capability is limited by both the number of delay taps in the mmWave MIMO channel, as well as the length of the training sequence, thereby making the algorithm impractical for practical mmWave deployments with more significant CFO. In [121], the joint CFO and broadband channel estimation problem is formulated as a sparse bilinear optimization problem, which is solved using the parametric bilinear generalized approximate message passing (PBiGAMP) algorithm in [122]. The main limitation of [121] is that the proposed estimation strategy is tailored to all-digital MIMO architectures with low-resolution Analog-to-Digital Converter (ADC) converters, thereby not being directly applicable to hybrid MIMO architectures. In [117], a similar strategy to the one in [121] is followed, in which the joint CFO and channel estimation problem is studied for all-digital MIMO architectures. The problem is formulated as a quantized sparse bilinear optimization problem, which is solved using sparse lifting to increase the dimension of the CFO and channel estimation problem [123], and then applying the generalized approximate message passing (GAMP) algorithm in [124] to solve the lifted problem.

In summary, prior work on joint synchronization and channel estimation for mmWave MIMO is limited to analog-only architectures and all-digital MIMO architectures with low-resolution ADCs. Hence, it is necessary to devise practical synchronization schemes that are compatible with hybrid MIMO architectures and are applicable to scenarios with realistic channel models. Furthermore, it is crucial that these synchronization strategies exhibit low

overhead and computational complexity, to ensure that CSI can be properly acquired at the low SNR regime in practice.

## 1.6 Dissertation Summary

Enabling broadband mmWave MIMO systems in practice requires addressing the key challenges discussed in Sections 1.1-1.5. With this motivation, the problems tackled in this dissertation lie on the intersection of these challenges. I propose low-overhead and low-complexity channel estimation and synchronization solutions that are robust at low SNR and yield efficient performance while respecting the hardware constraints imposed by hybrid MIMO architectures. For a brief discussion and background on CS and estimation theory, the reader is referred to Appendix B and Appendix C, respectively.

In the first part of this dissertation, I first propose two CS-based algorithms to estimate frequency-selective mmWave MIMO channels in the frequency domain using hybrid architectures. These algorithms overcome the limitations of prior work discussed in Section 1.3 by combining the property of spatially common sparsity present in the channel with the concept of Successive Interference Cancellation (SIC), and have convergence guarantees that are also analyzed. The first proposed algorithm exploits information coming from every subcarrier and iteratively estimates the AoA, AoD, and frequency-domain channel gains for different multipath components. In contrast, the second proposed algorithm only exploits information coming from a reduced number of subcarriers to estimate the AoA and AoD corresponding to the dif-

ferent channel components, thereby reducing online computational complexity when compared to current channel estimation approaches.

In the second part of this dissertation, I focus on the joint CFO and channel estimation problem in mmWave MIMO systems using hybrid MIMO architectures. I theoretically analyze the problem of narrowband channel estimation under CFO impairments and accounting for the reconfiguration time of phase-shifters imposed by hybrid architectures, and develop a multi-stage solution to estimate the high-dimensional mmWave MIMO channel. I theoretically calculate the CRLB for frame-wise estimation of the CFO, baseband equivalent beamformed channels, and noise variance, and find the optimal ML estimates of the different parameters. Then, both of these estimates and their corresponding CRLBs are used as proxy signals to estimate the mmWave MIMO channel using a variant of the Orthogonal Matching Pursuit (OMP) algorithm, without requiring prior knowledge of the channel's sparsity level or the noise variance.

In the final part of this dissertation, I extend the proposed synchronization framework to frequency-selective hybrid mmWave MIMO systems. I study the problem of time-frequency synchronization under the influence of PN impairments, and extend the proposed multi-stage synchronization solution to estimate the frequency-selective mmWave MIMO channel. I theoretically analyze the hybrid CRLB for frame-wise estimation of the CFO, baseband equivalent frequency-selective beamformed channels, and PN synchronization parameters, and then develop two EM-based algorithms to find

the ML estimates of the CFO, baseband equivalent beamformed channels, and the LMMSE estimator of the PN samples. Then, these estimates and their hybrid CRLB are leveraged to estimate the mmWave MIMO channel using the channel estimation algorithms developed in the first part of this dissertation. Last, I propose to design a joint PN tracking and data detection algorithm leveraging the frame structure of the 5G NR wireless standard, and enabling spatial multiplexing of parallel data streams according to the spatial degrees of freedom in the mmWave MIMO channel.

### 1.6.1 Thesis Statement

In this dissertation, I defend the following statement:

*Advanced hybrid analog-digital signal processing techniques can enable unprecedented communication performance while keeping training overhead low, even in the practical scenario of link configuration in the low SNR regime.*

### 1.6.2 Contributions

Hereafter, I summarize my contributions in this dissertation as follows.

- Chapter 2: Millimeter Wave Compressive Channel Estimation in the Frequency Domain
  1. I propose two novel algorithms for estimation of frequency-selective mmWave MIMO channels in the frequency domain, overcoming the limitations of prior work and providing different trade-

offs between spectral efficiency and computational complexity for a fixed training overhead. Further, I provide theoretical convergence guarantees for both algorithms and show that they converge to a local optimum.

2. I derive the CRLB for the estimation of the mmWave MIMO channel accounting for the spatially common sparsity. To compute this metric, I assume that the sparse channel support is known and show that our proposed strategies are asymptotically efficient when the AoD/AoD are distributed on spatially quantized grids. Further, I show that the CRLB can be attained without using frequency-selective baseband precoders and combiners, thereby significantly reducing computational complexity.

3. I introduce the concept of subcarrier selection as an approach to further reduce computational complexity during estimation of the sparse channel support, and show that a reduced number of subcarriers is sufficient to asymptotically attain the CRLB.

- Chapter 3: Millimeter Wave Compressive Channel Estimation with Carrier Frequency Offset Uncertainties

1. I formulate and propose a multi-stage solution to the problem of joint CFO and channel estimation using a data-aided approach based on forwarding several training frames, and accounting for the reconfiguration time of phase-shifters in analog precoding and

combining networks.

2. For every training frame, I theoretically calculate the CRLB for the estimation of the CFO, equivalent beamformed channel, and noise variance for hybrid MIMO architectures with several RF chains, and find the optimal ML estimators for the different unknown parameters.

3. Using both estimates of the unknown parameters for every training frame and their corresponding CRLB, I formulate the problem of estimating the high-dimensional MIMO channel and propose an OMP-based algorithm to find the sparse channel coefficients.

- Chapter 4: Millimeter Wave Broadband Synchronization, Compressive Channel Estimation, and Data Transmission

1. I formulate and propose a data-aided multi-stage solution to the problem of synchronization and compressive channel estimation for hybrid frequency-selective mmWave MIMO systems under TO, CFO, and PN impairments. I propose to forward several training frames using Zadoff-Chu (ZC)-based beamforming in combination with random subarray switching and antenna selection in order to both acquire synchronization and enable compressive channel estimation at the low SNR regime.

2. For every training frame, comprising of several OFDM symbols, I theoretically analyze the hybrid CRLB for the problem of estimat-

ing the CFO, PN, and equivalent frequency-selective beamformed channels.

3. I propose two novel iterative algorithms based on the EM method, which aim at finding the optimal ML estimates for the CFO and beamformed equivalent channels, as well as the LMMSE estimates for the PN samples that impair the received signal.

4. Using both estimates of the unknown parameters for every training frame and their hybrid CRLB, I formulate the problem of estimating the high-dimensional frequency-selective mmWave MIMO channel, and find a solution to this problem using a variant of the Simultaneous Weighted - Orthogonal Matching Pursuit (SW-OMP) algorithm proposed in my first contribution.

5. Last, I propose to design a low-complexity joint PN tracking and data detection algorithm leveraging the frame structure of the 5G NR wireless standard, and enabling spatial multiplexing of parallel data streams according to the spatial degrees of freedom in the mmWave MIMO channel.

### **1.6.3 Thesis Organization**

I organize the remainder of this dissertation as follows. In Chapter 2, I present the proposed broadband channel estimation algorithms for hybrid mmWave MIMO systems. In Chapter 3, I introduce the proposed joint CFO and channel estimation strategy for mmWave MIMO systems. In Chapter 4,

I extend the joint CFO and channel estimation strategy to frequency-selective scenarios, and include the additional TO and PN impairments. Finally, I conclude this dissertation and summarize potential future research directions in Chapter 5.

#### 1.6.4 Notation

I use the following notation throughout this dissertation. Bold lowercase  $\mathbf{x}$  is used for column vectors, bold uppercase  $\mathbf{X}$  is used for matrices, non-bold letters  $x, X$  are used for scalars.  $[\mathbf{x}]_i$ ,  $[\mathbf{X}]_{i,j}$ ,  $[\mathbf{X}]_{i,:}$ , and  $[\mathbf{X}]_{:,j}$  denote  $i$ th entry of  $\mathbf{x}$ , entry at the  $i$ th row and  $j$ th column of  $\mathbf{X}$ ,  $i$ th row of  $\mathbf{X}$ , and  $j$ th column of  $\mathbf{X}$ , respectively. I use the serif font, e.g.,  $\mathbf{x}$ , for the frequency-domain variables (the vectors (matrices) in the frequency domain are represented using bold sans serif font, i.e.,  $\mathbf{x}, \mathbf{X}$ ).  $\mathbf{A}^T$ ,  $\mathbf{A}^C$ ,  $\mathbf{A}^*$ , and  $\mathbf{A}^\dagger$  represent the transpose, conjugate, conjugate transpose, and Moore-Penrose pseudo-inverse of a matrix  $\mathbf{A}$ .  $\mathbf{0}_N$ ,  $\mathbf{1}_N$ , and  $\mathbf{I}_N$  denote the  $N$ -dimensional vector comprising of zero-valued entries, the  $N$ -th dimensional vector with unit entries, and the  $N$ th order identity matrix. The matrix containing the entries of a vector  $\mathbf{a}$  in its main diagonal is denoted by  $\text{diag}\{\mathbf{a}\}$ , and the block diagonal matrix containing the matrices  $\mathbf{A}_1, \dots, \mathbf{A}_N$  in its block diagonal is denoted by  $\bigoplus_{n=0}^{N-1} \mathbf{A}_n$ .  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes a complex circularly symmetric Gaussian random vector with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . I use  $\mathbb{E}\{\cdot\}$ ,  $\|\cdot\|_p$ , and  $\|\cdot\|_F$  to denote expectation,  $\ell_p$ -norm, and Frobenius norm, respectively.  $\mathbf{X} \otimes \mathbf{Y}$  is the Kronecker product of  $\mathbf{X}$  and  $\mathbf{Y}$ , and  $\mathbf{X} \circ \mathbf{Y}$  denotes the

Khatri-Rao product of  $\mathbf{X}$  and  $\mathbf{Y}$ . Calligraphic letter  $\mathcal{X}$  denotes a set. Finally,  $|\cdot|$  is the absolute value of its argument or the cardinality of a set, and  $\text{vec}\{\cdot\}$  yields a vector for a matrix argument.

### 1.6.5 List of Acronyms

**3GPP** Third Generation Partnership Protocol.

**ADC** Analog-to-Digital Converter.

**AoA** Angle of Arrival.

**AoD** Angle of Departure.

**AS** Angular Spread.

**AWGN** Additive White Gaussian Noise.

**BER** Bit Error Rate.

**BLUE** Best Linear Unbiased Estimator.

**BPDN** Basis Pursuit De-Noising.

**BS** Base Station.

**CFO** Carrier Frequency Offset.

**CoMP** Coordinated MultiPoint.

**CoSaMP** Compressive Sampling Matching Pursuit.

**CP** Cyclic Prefix.

**CPE** Common Phase Error.

**CRLB** Cramér-Rao Lower Bound.

**CS** Compressed Sensing.

**CSI** Channel State Information.

**CSI-RS** Channel State Information - Reference Signal.

**DFT** Discrete Fourier Transform.

**DGMP** Distributed Grid Matching Pursuit.

**DMRS** Demodulation Reference Signal.

**DS** Delay Spread.

**EKF** Extended Kalman Filter.

**EM** Expectation-Maximization.

**FET** Field-Effect Transistor.

**FFT** Fast Fourier Transform.

**FIM** Fisher Information Matrix.

**FLOPs** Floating Point Operations.

**GHP** Greedy Hybrid Precoding.

**GLM** General Linear Model.

**GPS** Global Positioning System.

**GSF** Gaussian-Sum Filtering.

**HARQ** Hybrid Automatic Repeat Request.

**HetNets** Heterogeneous Networks.

**HIM** Hybrid Information Matrix.

**ICI** Inter Carrier Interference.

**ISI** Inter Symbol Interference.

**KF** Kalman Filter.

**KLT** Karhunen-Loeve Transform.

**LASSO** Least Absolute Shrinkage and Selection Operator.

**LDPC** Low Density Parity Check.

**LF** Likelihood Function.

**LLF** Log-Likelihood Function.

**LLR** Log-Likelihood Ratio.

**LMMSE** Linear Minimum Mean Square Error.

**LOS** Line-Of-Sight.

**LS** Least Squares.

**MAP** Maximum A Posteriori.

**MCS** Modulation and Coding Scheme.

**MDP** Markov Decision Process.

**MER** Modulation Error Ratio.

**MIMO** Multiple-Input Multiple-Output.

**MISO** Multiple-Input Single-Output.

**ML** Maximum Likelihood.

**MMSE** Minimum Mean Square Error.

**mmWave** Millimeter Wave.

**MP** Matching Pursuit.

**MPF** Marginalized Particle Filtering.

**MSE** Mean Squared Error.

**MVUE** Minimum Variance Unbiased Estimator.

**NLOS** Non Line-Of-Sight.

**NMSE** Normalized Mean Square Error.

**NR** New Radio.

**OFDM** Orthogonal Frequency Division Multiplexing.

**OMP** Orthogonal Matching Pursuit.

**PC** Per-Antenna Constrained.

**PDCCH** Physical Downlink Control Channel.

**PDF** Probability Density Function.

**PDSCH** Physical Downlink Shared Channel.

**PN** Phase Noise.

**PSD** Power Spectral Density.

**PTRS** Phase-Tracking Reference Signal.

**QAM** Quadrature Amplitude Modulation.

**QMMSE** Quadratic Minimum Mean Square Error.

**QuaDRiGa** Quasi Deterministic Radio channel Generator.

**RF** Radio-Frequency.

**RIP** Restricted Isometry Property.

**RSS** Received Signal Strength.

**RTS** Rauch-Tung-Striebel.

**S-OMP** Simultaneous - Orthogonal Matching Pursuit.

**SIC** Successive Interference Cancellation.

**SINR** Signal-to-Interference plus Noise Ratio.

**SISO** Single-Input Single-Output.

**SNR** Signal-to-Noise Ratio.

**SOCP** Second-Order Cone Program.

**SS** Synchronization Signal.

**SS-SW-OMP+Th** Subcarrier Selection - Simultaneous Weighted - Orthogonal Matching Pursuit + Thresholding.

**SSAMP** Structured Sparsity-Adaptive Matching Pursuit.

**SSP-OMP** Spatially Sparse Precoding - Orthogonal Matching Pursuit.

**SVD** Singular Value Decomposition.

**SW-OMP** Simultaneous Weighted - Orthogonal Matching Pursuit.

**TO** Timing Offset.

**UE** User Equipment.

**ULA** Uniform Linear Array.

**UMi** Urban Microcell.

**UPA** Uniform Planar Array.

**VCO** Voltage-Controlled Oscillator.

**W-OMP** Weighted - Orthogonal Matching Pursuit.

**WLS** Weighted Least Squares.

**ZC** Zadoff-Chu.

**ZP** Zero Prefix.

## Chapter 2

# Millimeter Wave Compressive Channel Estimation in the Frequency Domain

### 2.1 Introduction

In this chapter, I propose and evaluate new channel estimation algorithms for frequency-selective hybrid mmWave MIMO systems. Frequency-flat pseudorandom training precoders and combiners are employed by the transmitter and receiver to estimate the channel, and two novel algorithms exploiting spatially common sparsity in the angular domain are devised. The first proposed algorithm exploits information coming from every subcarrier, thereby yielding the best performance in terms of spectral efficiency, estimation error, and BER. The second proposed algorithm, however, exploits information from only a reduced number of subcarriers to further reduce computational complexity. I evaluate different tradeoffs between communication performance and computational complexity of the proposed algorithms, and show that the proposed algorithms outperform prior work on broadband channel estimation in terms of spectral efficiency, estimation error, BER, and computational complexity. Further, I show that the proposed algorithms asymptotically attain the CRLB for the estimation of the high-dimensional mmWave MIMO channel when the AoD and AoA are distributed on quantized angular grids. This work

was published in [1], [125].

### 2.1.1 Prior Work and Motivation

Two different approaches to increase SNR after spatial processing are beam training and channel estimation [8]. On the one hand, beam training, first adopted in IEEE 802.11ad [36], is a technique that relies on searching for transmit and receive beam pairs that maximize the received SNR, thereby enabling reliable decoding of transmitted data [38–40]. Though they can increase link quality, beam training strategies are typically restricted to single-stream communication, which disables spatial multiplexing capabilities to obtain high data rate communications [?]. Recently, some efforts have been made towards extending beam training protocols to multi-stream communication [126, 127]. The main limitation of these algorithms is that the beam detection capabilities are sensitive to the beamwidth of the transmit and receive candidate beams [126]. Therefore, to obtain higher spectral efficiency, narrower beams need to be deployed, especially for medium transmit-receive distances, which significantly increases training overhead and detection complexity [127].

Channel estimation, on the other hand, allows transmission of several data streams, overcoming the limitation of beam training strategies. Most of these strategies exploit the spatially sparse structure in the mmWave MIMO channel, formulating its estimation as a sparse recovery problem. The support of the estimated sparse vector identifies the pairs of AoD and AoA for each multipath component in the mmWave channel, while the amplitudes of

Approach	Computes/ exploits noise covariance	Pulse- shaping	Application to mmWave	Hybrid architecture	Online Complexity	Training Overhead	Communication Performance
SSAMP [69,71]					High	High	Medium
DGMP [70]					Low	High	Low
OMP [34]		✓	✓	✓	Medium	High	Medium
<b>Proposed SW-OMP</b>	✓	✓	✓	✓	Medium	Low	Very High
<b>Proposed SS-SW-OMP+Th</b>	✓	✓	✓	✓	Low	Medium	High

Figure 2.1: Summary and comparison of prior work on broadband channel estimation algorithms as well as the approaches proposed in this Chapter.

the non-zero coefficients provide the channel gains for such multipath components. Compressive estimation leads to a reduction in the channel training length when compared to conventional approaches such as those based on LS estimation [14]. The main limitation of most of this prior work with hybrid MIMO architectures comes from considering a frequency-flat channel model [14, 17, 53, 56–60, 128], since the mmWave channel is frequency-selective.

Recently, some approaches for channel estimation in frequency-selective mmWave channels have been proposed, which are summarized in Fig. 2.1.

In [34], a time-domain approach was designed to estimate the wideband mmWave channel assuming a hybrid MIMO architecture. This algorithm exploits the sparsity of the wideband mmWave channel in both the angular and delay domains. The sparse formulation of the problem in [34] includes the effect of non-integer sampling of the transmit pulse-shaping filter, with the subsequent leakage effect and increase of sparsity level in the channel matrix. The main limitation of [34] is the high computational complexity of

the algorithm. A frequency-domain strategy to estimate frequency-selective mmWave channels was also proposed in [34]. A sparse reconstruction problem was formulated there to estimate the channel independently for every subcarrier, without exploiting spatial congruence between subbands. Another approach in the frequency domain was designed in [129], but only exploiting the information from a reduced number of subcarriers.

Based on the SSAMP algorithm, first proposed in [71], another approach to estimate the mmWave channel was proposed in [69]. Exploiting the fact that spatial propagation characteristics do not change significantly within the system bandwidth, [69] assumed spatially common sparsity between the channels corresponding to the different subcarriers. The SSAMP algorithm from [71] was then considered to reconstruct the channels in the frequency domain. Thus, [69] is an interesting initial solution to the problem, but has several limitations when applied to a mmWave communications system:

1. The effect of sampling the pulse-shaping filter delayed by a non integer factor was not considered in the channel model for a given delay tap. As shown in [33], not accounting for this effect leads to virtual MIMO matrices with an artificially enhanced sparsity.
2. The algorithm was evaluated only for medium and high SNR regimes, not realistic at mmWave, where the expected SNR is below 0 dB.
3. The reconstruction algorithm provides accurate results when Gaussian measurement matrices are employed; to generate the Gaussian matrices,

unquantized phases were considered in the training precoders, which is unrealistic in a practical implementation of a mmWave system based on a hybrid architecture.

Another algorithm exploiting common sparsity in the frequency domain at mmWave was proposed in [70]. Unlike the SSAMP algorithm proposed in [71], the algorithm in [70] is proposed to estimate mmWave wideband MU-MIMO channels. Besides the limitations 1)-3) described above, which also hold in this case, this algorithm exhibits another problem that makes it less feasible to be applied in a real mmWave communication system. A LOS Rician channel model with  $K_{\text{factor}} = 20$  dB was considered, which is only applicable when there is a strong LOS path. Owing to this artifact of the channel model, the algorithm in [70] estimates only a single path for each user, such that the task of channel estimation in a general mmWave system cannot be successfully accomplished.

### 2.1.2 Contributions

In this chapter, I propose two novel frequency-domain approaches to estimate frequency-selective mmWave MIMO channels. These approaches overcome the limitations of prior work and provide different tradeoffs between complexity and achievable rate for a fixed training length. As in recent work on hybrid architectures for frequency-selective mmWave channels [69, 130], I also consider a MIMO-OFDM communications system. Similar to [34], I use ZP as a cyclic prefix to avoid loss and/or distortion of training data during

reconfiguration of RF circuitry. A geometric channel is considered to model the different scattering clusters as in [34], [70], [33], including the bandlimiting property in the overall channel response. The contributions of this work are listed hereafter:

- I formulate the problem of compressive channel estimation in the frequency domain, and propose two novel algorithms to solve this problem. The two approaches explained in this chapter exploit the spatially common sparsity within the system bandwidth. The first algorithm aims at exploiting the information on the support coming from every subcarrier in the MIMO-OFDM system and provides the best performance. In contrast, the second algorithm uses less information to estimate the different frequency-domain subchannels, thereby managing to significantly reduce computational complexity.
- I theoretically compute the CRLB for the estimation of the high-dimensional mmWave MIMO channel assuming perfect retrieval of the AoD and AoA to assess the robustness of the estimation of the channel support. I show that the two proposed strategies are asymptotically efficient when the channel AoD/AoA lie within quantized angular grids, since they both attain the CRLB. Further, I show that asymptotic efficiency can be achieved without using frequency-selective baseband precoders and combiners during the training stage, thereby further reducing computational complexity.

- I analyze and provide convergence guarantees for the proposed algorithms, and show that the proposed algorithms are guaranteed to converge to a local optimum when the angular grid sizes are of sufficient size.

Simulation results in the low SNR regime show that the two proposed algorithms significantly outperform the approach in the frequency domain developed in [34]. Comparisons with the algorithms proposed in [69] and [70] are also provided to show their performance in terms of estimation error in the SNR regime where mmWave systems are expected to work. To the best of my knowledge, there is no prior work that evaluates channel estimation algorithms with realistic frequency-selective channel samples. I evaluate the proposed algorithms with channel realizations obtained from the NYUSIM channel simulator [4], which was developed based on the statistical model in [131]. I show that the proposed algorithms are suitable for estimation of more general mmWave frequency-selective channels than the model in [33], thereby demonstrating their practical application. The two proposed channel estimation algorithms provide a good tradeoff between communication performance and overhead. Results show that using a reasonably small training length, approximately in the range of 60 – 100 frames, leads to low estimation errors. The computational complexity of the proposed algorithms and previous strategies is also analyzed to compare the tradeoffs between performance-complexity provided by the different algorithms. Finally, I also show that it is not necessary to exploit the information on the support coming from ev-

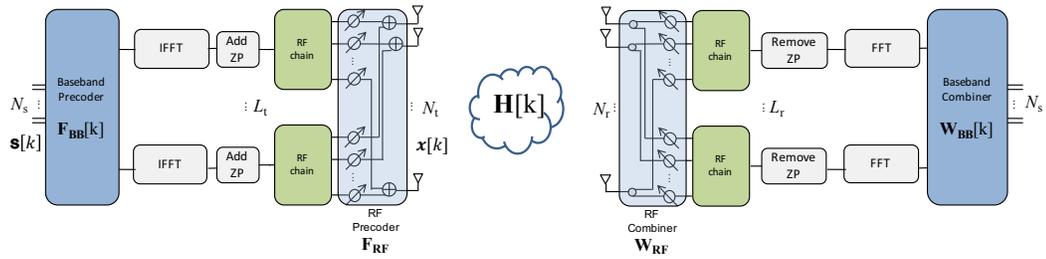


Figure 2.2: Illustration of the structure of a hybrid MIMO architecture, which include analog and digital precoders and combiners. This structure was already introduced in Fig. 1.1.

ery OFDM subcarrier to estimate the different mmWave subchannels. Yet, a reduced number of subcarriers is enough to asymptotically attain the CRLB.

I organize the rest of this chapter as follows. In Section 2.2, I introduce the system and channel models. In Section 2.3, I describe the proposed frequency-domain compressive channel estimation approaches, and include the derivation of the CRLB. Thereafter, Section 2.4 provides the main simulation results for the two proposed algorithms, and comparisons with the OMP-based compressive approach proposed in [34], the SSAMP algorithm in [69] and the Distributed Grid Matching Pursuit (DGMP) algorithm proposed in [70], respectively. Finally, I draw the conclusions extracted from numerical results in Section 2.5.

## 2.2 System model

I consider mmWave MIMO-OFDM link employing  $K$  subcarriers to send  $N_s$  data streams using a transmitter with  $N_t$  antennas and a receiver

with  $N_r$  antennas. The system is based on a hybrid MIMO architecture as shown in Fig. 2.2, with  $L_t$  and  $L_r$  RF chains at the transmitter and receiver sides. For a general exposition, a frequency-selective hybrid precoder is used, with  $\mathbf{F}[k] = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}[k] \in \mathbb{C}^{N_t \times N_s}$ ,  $k = 0, \dots, K - 1$ , where  $\mathbf{F}_{\text{RF}}$  is the analog precoder and  $\mathbf{F}_{\text{BB}}[k]$  the digital one. Note that the analog precoder is frequency-flat, while the digital precoder is different for every subcarrier. The RF precoder and combiner are implemented using a fully-connected network of phase-shifters, as described in [14]. The symbol blocks are transformed into the time domain using  $L_t$  parallel  $K$ -point IFFTs. As in [34, 132], I consider ZP to both suppress Inter Symbol Interference (ISI) and account for the RF circuitry reconfiguration time, which can be upper bounded by 64 discrete samples at a sampling rate of 1760 MHz [36]. The discrete-time complex baseband signal at subcarrier  $k$  can be written as

$$\mathbf{x}[k] = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}[k]\mathbf{s}[k], \quad (2.1)$$

where the transmitted symbol sequence at subcarrier  $k$  of size  $N_s \times 1$  is denoted as  $\mathbf{s}[k]$ .

The MIMO channel between the transmitter and the receiver is assumed to be frequency selective, having a delay tap length  $D$  in the time domain. The  $d$ -th delay tap of the channel is represented by a  $N_r \times N_t$  matrix denoted as  $\mathbf{H}[d]$ ,  $d = 0, 1, \dots, D - 1$ , which, assuming a geometric channel model [33], can be written as

$$\mathbf{H}[d] = \sqrt{\frac{N_t N_r}{L \rho_L}} \sum_{\ell=1}^L \alpha_{\ell} p_{\text{rc}}(dT_s - \tau_{\ell}) \mathbf{a}_{\text{R}}(\phi_{\ell}) \mathbf{a}_{\text{T}}^*(\theta_{\ell}), \quad (2.2)$$

where  $\rho_L$  denotes the path loss between the transmitter and the receiver,  $L$  denotes the number of paths,  $T_s$  denotes the sampling period,  $p_{rc}(\tau)$  is a filter that includes the effects of pulse-shaping and other lowpass filtering evaluated at  $\tau$ ,  $\alpha_\ell \in \mathbb{C}$  is the complex gain of the  $\ell$ th path,  $\tau_\ell \in \mathbb{R}$  is the delay of the  $\ell$ th path,  $\phi_\ell \in [0, 2\pi)$  and  $\theta_\ell \in [0, 2\pi)$  are the AoA and AoD, of the  $\ell$ th path, and  $\mathbf{a}_R(\phi_\ell) \in \mathbb{C}^{N_r \times 1}$  and  $\mathbf{a}_T(\theta_\ell) \in \mathbb{C}^{N_t \times 1}$  are the array steering vectors for the receive and transmit antennas evaluated at the AoA and AoD of the  $\ell$ -th path. Each one of these matrices can be written in a more compact way as

$$\mathbf{H}[d] = \mathbf{A}_R \mathbf{G}[d] \mathbf{A}_T^*, \quad (2.3)$$

where  $\mathbf{G}[d] \in \mathbb{C}^{L \times L}$  is diagonal with non-zero complex entries, and  $\mathbf{A}_R \in \mathbb{C}^{N_r \times L}$  and  $\mathbf{A}_T \in \mathbb{C}^{N_t \times L}$  contain the receive and transmit array steering vectors  $\mathbf{a}_R(\phi_\ell)$  and  $\mathbf{a}_T(\theta_\ell)$ , respectively. The channel  $\mathbf{H}[d]$  can be approximated using the extended virtual channel model defined in [8] as

$$\mathbf{H}[d] \approx \tilde{\mathbf{A}}_R \mathbf{G}^v[d] \tilde{\mathbf{A}}_T^*, \quad (2.4)$$

where  $\mathbf{G}^v[d] \in \mathbb{C}^{G_r \times G_t}$  is a sparse matrix which contains the path gains at the quantized spatial frequencies in the non-zero elements. The dictionary matrices  $\tilde{\mathbf{A}}_T$  and  $\tilde{\mathbf{A}}_R$  contain the transmit and receive array response vectors evaluated on angular grids of sizes  $G_t$  and  $G_r$ . Due to the few scattering clusters in mmWave channels, the sparse assumption for  $\mathbf{G}^v[d]$  is commonly accepted [4, 35]. Finally, the channel at subcarrier  $k$  can be written in terms of the different delay taps as

$$\mathbf{H}[k] = \sum_{d=0}^{D-1} \mathbf{H}[d] e^{-j \frac{2\pi k}{K} d} = \mathbf{A}_R \mathbf{G}[k] \mathbf{A}_T^*. \quad (2.5)$$

It is also useful to write this matrix in terms of the sparse matrices  $\mathbf{G}^v[d]$  and the dictionaries

$$\mathbf{H}[k] \approx \tilde{\mathbf{A}}_R \left( \sum_{d=0}^{D-1} \mathbf{G}^v[d] e^{-j \frac{2\pi k}{K} d} \right) \tilde{\mathbf{A}}_T^* \approx \tilde{\mathbf{A}}_R \mathbf{G}^v[k] \tilde{\mathbf{A}}_T^* \quad (2.6)$$

to help expose the sparse structure in Section 2.3.

Assuming that the receiver applies a hybrid combiner  $\mathbf{W}[k] \in \mathbb{C}^{N_r \times N_s}$ ,  $\mathbf{W}[k] = \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}[k]$ , the received signal at subcarrier  $k$  can be written as

$$\begin{aligned} \mathbf{y}[k] &= \mathbf{W}_{\text{BB}}^*[k] \mathbf{W}_{\text{RF}}^* \mathbf{H}[k] \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}[k] \mathbf{s}[k] \\ &\quad + \mathbf{W}_{\text{BB}}^*[k] \mathbf{W}_{\text{RF}}^* \mathbf{n}[k], \end{aligned} \quad (2.7)$$

where  $\mathbf{n}[k] \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$  is the circularly symmetric complex Gaussian distributed additive noise vector. The receive signal model in (2.7) corresponds to the data transmission phase. As it will be seen in Section 2.3, during the channel acquisition phase, I will consider frequency-flat training precoders and combiners to reduce complexity during channel estimation.

## 2.3 Compressive Channel Estimation in the Frequency Domain

In this section, I formulate a compressed sensing problem to estimate the wideband mmWave MIMO channel in the frequency domain. I also propose two algorithms to solve this problem that leverage the common support between the channel matrices for every subcarrier, providing different trade-offs between performance and computational complexity. The first algorithm leverages the common support between the  $K$  different subchannels providing

very good performance, while the second one only exploits information from a reduced number of subcarriers, thereby keeping computational complexity at a lower level.

### 2.3.1 Problem formulation

I assume that  $L_t$  and  $L_r$  RF chains are used at the transmitter and receiver. During the training phase,  $M$  training frames are forwarded from the transmitter to the receiver. For the  $m$ -th frame, the transmitter and the receiver use a training precoder  $\mathbf{F}_{\text{tr}}^{(m)} \in \mathbb{C}^{N_t \times L_t}$  and a training combiner  $\mathbf{W}_{\text{tr}}^{(m)} \in \mathbb{C}^{N_r \times L_r}$ . This means that during the training phase, frequency-flat precoders and combiners are considered to keep the complexity of the sparse recovery algorithms low. I assume that the transmitted symbols satisfy  $\mathbb{E}\{\mathbf{s}^{(m)}[k]\mathbf{s}^{(m)*}[k]\} = \frac{P}{N_s}\mathbf{I}_{N_s}$ , with  $P$  the total transmitted power and  $N_s = L_t$ . To reduce computational complexity, I decompose the transmitted symbol  $\mathbf{s}^{(m)}[k]$  as  $\mathbf{s}^{(m)}[k] = \mathbf{q}^{(m)}\mathbf{t}^{(m)}[k]$ , with  $\mathbf{q}^{(m)} \in \mathbb{C}^{L_t \times 1}$  a frequency-flat vector and  $\mathbf{t}^{(m)}[k]$  a pilot symbol known at the receiver. This choice is motivated to simultaneously enable exploitation of the  $L_t$  spatial degrees of freedom coming from  $L_t$  RF chains and to allow channel estimation with a single subcarrier-independent measurement matrix, as it will soon become apparent. This enables both online and offline complexity reduction when the noise statistics are used to estimate the MIMO channel at the different subcarriers. Furthermore, each entry in  $\mathbf{F}_{\text{tr}}^{(m)}$ ,  $\mathbf{W}_{\text{tr}}^{(m)}$  is normalized to have squared-modulus  $N_t^{-1}$  and  $N_r^{-1}$ , respectively. Then, the received samples in the frequency domain

for the  $m$ -th training frame can be written as

$$\mathbf{y}^{(m)}[k] = \mathbf{W}_{\text{tr}}^{(m)*} \mathbf{H}[k] \mathbf{F}_{\text{tr}}^{(m)} \mathbf{q}^{(m)} \mathbf{t}^{(m)}[k] + \mathbf{n}_c^{(m)}[k], \quad (2.8)$$

where  $\mathbf{H}[k] \in \mathbb{C}^{N_r \times N_t}$  is the frequency-domain MIMO channel response at the  $k$ -th subcarrier and  $\mathbf{n}_c^{(m)}[k] \in \mathbb{C}^{L_r \times 1}$ ,  $\mathbf{n}_c^{(m)}[k] = \mathbf{W}_{\text{tr}}^{(m)*} \mathbf{n}^{(m)}[k]$ , is the frequency-domain combined noise vector received at the  $k$ -th subcarrier. The average received SNR is given by  $\text{SNR} = \frac{P}{\rho_L \sigma^2}$ . I assume that the channel coherence time is larger than the frame duration and that the same channel can be considered for several consecutive frames. To enable sparse reconstruction with a single, subcarrier-independent measurement matrix, I will invert the effect of the scalar  $\mathbf{t}^{(m)}[k]$  by means of multiplying the received signal by  $(\mathbf{t}^{(m)}[k])^{-1}$ . Using the result  $\text{vec}\{\mathbf{A}\mathbf{X}\mathbf{C}\} = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}\{\mathbf{X}\}$ , the vectorized received signal after compensating for  $\mathbf{s}^{(m)}[k]$  is

$$\text{vec}\{\mathbf{y}^{(m)}[k]\} = (\mathbf{q}^{(m)T} \mathbf{F}_{\text{tr}}^{(m)T} \otimes \mathbf{W}_{\text{tr}}^{(m)*}) \text{vec}\{\mathbf{H}[k]\} + \mathbf{n}_c^{(m)}[k]. \quad (2.9)$$

Taking into account the expression in (2.6), the vectorized channel matrix can be written as  $\text{vec}\{\mathbf{H}[k]\} = (\tilde{\mathbf{A}}_T^C \otimes \tilde{\mathbf{A}}_R) \text{vec}\{\mathbf{G}^V[k]\}$ . Therefore, if I define the measurement matrix  $\Phi^{(m)} \in \mathbb{C}^{L_r \times N_t N_r}$  as

$$\Phi^{(m)} = (\mathbf{q}^{(m)T} \mathbf{F}_{\text{tr}}^{(m)T} \otimes \mathbf{W}_{\text{tr}}^{(m)*}), \quad (2.10)$$

and the dictionary  $\Psi \in \mathbb{C}^{N_t N_r \times G_t G_r}$  as

$$\Psi = \tilde{\mathbf{A}}_T^C \otimes \tilde{\mathbf{A}}_R, \quad (2.11)$$

(2.9) can be rewritten as

$$\text{vec}\{\mathbf{y}^{(m)}[k]\} = \Phi^{(m)} \Psi \mathbf{g}^V[k] + \mathbf{n}_c^{(m)}[k], \quad (2.12)$$

where  $\mathbf{g}^v[k] = \text{vec}\{\mathbf{G}^v[k]\} \in \mathbb{C}^{G_r G_t \times 1}$  is the sparse vector containing the complex channel gains. To have enough measurements and accurately reconstruct the sparse vector  $\mathbf{g}^v[k]$ , it is necessary to use several training frames, especially in the very-low SNR regime. If the transmitter and receiver communicate during  $M$  training steps using different pseudorandomly built precoders and combiners, (2.12) can be extended to

$$\underbrace{\begin{bmatrix} \mathbf{y}^{(1)}[k] \\ \vdots \\ \mathbf{y}^{(M)}[k] \end{bmatrix}}_{\mathbf{y}[k]} = \underbrace{\begin{bmatrix} \Phi^{(1)} \\ \vdots \\ \Phi^{(M)} \end{bmatrix}}_{\Phi} \Psi \mathbf{g}^v[k] + \underbrace{\begin{bmatrix} \mathbf{n}_c^{(1)}[k] \\ \vdots \\ \mathbf{n}_c^{(M)}[k] \end{bmatrix}}_{\mathbf{n}_c[k]}. \quad (2.13)$$

Finally, the vector  $\mathbf{g}^v[k]$  can be found by solving the sparse reconstruction problem

$$\min_{\{\mathbf{g}^v[k]\}_{k=0}^{K-1}} \sum_{k=0}^{K-1} \|\mathbf{g}^v[k]\|_1 \quad \text{subject to} \quad \frac{1}{K} \sum_{k=0}^{K-1} \|\mathbf{y}[k] - \Phi \Psi \mathbf{g}^v[k]\|_2^2 < \epsilon, \quad (2.14)$$

where  $\epsilon$  is a tunable parameter defining the maximum error between the measurement and the received signal assuming the reconstructed channel between the transmitter and the receiver. Since the sparsity (number of channel paths) is usually unknown, the choice of  $\epsilon$  is critical to solve (2.14). The choice of this parameter will be explained in Section 2.3.3.

There is a great variety of algorithms to solve (2.14). For example, OMP was considered in [34]. However, this requires running the algorithm as many times as the number of OFDM subcarriers. In the next subsections

I consider an additional assumption to solve this problem, which avoids the need to run  $K$  OMP algorithms in parallel as proposed in [34].

The matrices  $\mathbf{G}^v[k]$  exhibit an interesting property that can be exploited when solving the compressed channel estimation problems defined in (2.14). Let us define the  $G_t G_r \times 1$  vectorized virtual channel matrix for a given delay tap as

$$\mathbf{g}^v[d] \triangleq \text{vec}\{\mathbf{G}^v[d]\}. \quad (2.15)$$

Let  $\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_{D-1}$  denote the supports of the virtual channel matrices  $\mathbf{G}^v[d]$ ,  $d = 0, \dots, N_c - 1$ . Then, since  $\mathbf{g}^v[k] = \text{vec}\{\mathbf{G}^v[k]\}$ , with

$$\mathbf{G}^v[k] = \sum_{d=0}^{D-1} \mathbf{G}^v[d] e^{-j \frac{2\pi k}{K} d}, \quad k = 0, \dots, K - 1 \quad (2.16)$$

it is clear that, in general

$$\text{supp}\{\mathbf{g}^v[k]\} = \bigcup_{d=0}^{D-1} \text{supp}\{\mathbf{g}^v[d]\} \quad k = 0, \dots, K - 1, \quad (2.17)$$

where the union of the supports of the time-domain virtual channel matrices comes from the additive nature of the Fourier transform. Therefore, the channel model in (2.6) exhibits the same sparse structure for every subcarrier, since the AoA and AoD do not change with frequency in the transmission bandwidth [4, 31, 32]. Notice, however, that the AoA/AoD in (4.7) do not depend on the delay tap  $d$ , such that (2.17) reduces to  $\text{supp}\{\mathbf{g}^v[k]\} = \text{supp}\{\mathbf{g}^v[d]\}$ ,  $k = 0, \dots, K - 1$ . The sparse assumption on the vectorized channel matrix for a given delay tap  $\mathbf{g}^v[d]$  is commonly accepted, since in mmWave channels  $L \ll G_r G_t$ . In general, the vectorized channel matrix  $\mathbf{g}^v[k]$  will have, in

the worst case,  $DL$  non-zero coefficients. Typical values for  $D$  in mmWave channels are usually lower than 64 symbols (for example IEEE 802.11ad has been designed to work robustly for a maximum of 64 delay taps in the channel), while the number of measured paths usually satisfies  $L < 30$  for outdoor and indoor scenarios [133]. From these values, using dictionaries of size  $G_r \geq 64$   $G_t \geq 64$ , allows us to assume a spatially sparse structure for  $\mathbf{g}^v[k]$  as well.

### 2.3.2 Simultaneous Weighted - Orthogonal Matching Pursuit

To develop a channel estimation algorithm that leverages the sparse nature and the common support property for all  $\mathbf{g}^v[k]$ , I propose to modify the Simultaneous - Orthogonal Matching Pursuit (S-OMP) algorithm proposed in [134]. For a given iteration, this algorithm aims at finding a new index of the support exploiting information coming from several signals, thus increasing the reliability of the final support estimate. If the algorithm processes  $K$  received signals simultaneously, the amount of information it obtain is approximately  $K$  times larger. Consequently, the likelihood of the estimated AoD/AoA, or equivalently, the support of the channel, increases accordingly. The S-OMP algorithm in [134] computes the non-zero values of the sparse vector using a LS approach once the support is obtained, assuming that the noise covariance matrix is the identity matrix  $\mathbf{I}_{ML_r}$ . In this section, I generalize the S-OMP algorithm to account for correlated noise after combining, and show that the proposed algorithm attains the CRLB.

### 2.3.2.1 Support computation with correlated noise

Before explicit estimation of the channel gains, it is necessary to compute the atom, i.e., vector in the measurement matrix, which yields the largest sum-correlation with the received signals, since the different sparse vectors share a common support. The S-OMP algorithm is based on the assumption that the perturbation (noise) covariance matrix is diagonal, such that no correlation between the different noise components is present. The correlation vector  $\mathbf{c}[k] \in \mathbb{C}^{ML_r}$  is defined as

$$\mathbf{c}[k] = \mathbf{\Upsilon}^* \mathbf{y}[k], \quad (2.18)$$

in which  $\mathbf{\Upsilon} \in \mathbb{C}^{ML_r \times G_t G_r}$ ,  $\mathbf{\Upsilon} = \mathbf{\Phi} \mathbf{\Psi}$  is the equivalent measurement matrix and  $\mathbf{y}[k] \in \mathbb{C}^{ML_r \times 1}$  is the received signal for a given  $k$ ,  $k = 0, \dots, K - 1$ . If there is correlation between noise components, the atom estimated from the correlation in (2.18) may be different from the actual atom. To introduce the appropriate correction in the correlation, the specific form the noise covariance matrix takes needs to be taken into account. Since the received noise at antenna level is both temporally and spatially white, the noise covariance matrix of  $\mathbf{y}[k]$  can be written as a block diagonal matrix  $\mathbf{C} \in \mathbb{C}^{ML_r \times ML_r}$ ,  $\mathbf{C} = \sigma^2 \mathbf{C}_w$ , where  $\mathbf{C}_w$  is given by  $\mathbf{C}_w = \bigoplus_{m=1}^M \mathbf{W}_{\text{tr}}^{(m)*} \mathbf{W}_{\text{tr}}^{(m)}$ . Now, I use the Cholesky factorization to write  $\mathbf{C}_w = \mathbf{D}_w^* \mathbf{D}_w$ , where  $\mathbf{D}_w \in \mathbb{C}^{ML_r \times ML_r}$  is an upper triangular matrix. The subscript in  $\mathbf{C}_w$  and  $\mathbf{D}_w$  indicates that these matrices only depend on the combiners  $\{\mathbf{W}_{\text{tr}}^{(m)}\}_{m=1}^M$ . Then, the correlation step is performed as

$$\mathbf{c}[k] = \mathbf{\Upsilon}_w^* \mathbf{y}_w[k], \quad (2.19)$$

where  $\mathbf{\Upsilon}_w \in \mathbb{C}^{ML_r \times G_t G_r}$  is the whitened measurement matrix given by  $\mathbf{\Upsilon}_w = \mathbf{D}_w^{-*} \mathbf{\Upsilon}$ . Likewise, the  $ML_r \times 1$  post-whitened received signal  $\mathbf{y}_w[k]$  is given by  $\mathbf{y}_w[k] = \mathbf{D}_w^{-*} \mathbf{y}[k]$ . The matrix  $\mathbf{D}_w^{-1} \in \mathbb{C}^{ML_r \times ML_r}$  can be further expressed as

$$\mathbf{D}_w^{-1} = \bigoplus_{m=1}^M \left( \mathbf{D}_w^{(1)} \right)^{-1} \quad (2.20)$$

where  $\left( \mathbf{D}_w^{(m)} \right)^{-1}$  can be interpreted as the frequency-flat baseband combiner  $\mathbf{W}_{\text{BB,tr}}^{(m)}$  used in the  $m$ -th training step. Thereby, the resulting correlation simultaneously whitens the spatial noise components and estimates the most likely support index in the sparse vectors  $\mathbf{g}^v[k]$ .

### 2.3.2.2 Computation of the channel gains

Once an estimate  $\hat{\mathcal{J}}$  of the support of the sparse channel vectors is found, with  $\hat{L} = |\hat{\mathcal{J}}|$  the estimated sparsity level, I can define the matrix  $[\mathbf{\Upsilon}]_{:, \hat{\mathcal{J}}} \in \mathbb{C}^{ML_r \times \hat{L}}$  as  $[\mathbf{\Upsilon}]_{:, \hat{\mathcal{J}}} = [\mathbf{\Phi} \mathbf{\Psi}]_{:, \hat{\mathcal{J}}}$ . Accordingly, the signal model for the  $k$ -th subcarrier can be written as

$$\mathbf{y}[k] = [\mathbf{\Upsilon}]_{:, \hat{\mathcal{J}}} [\mathbf{g}^v[k]]_{\hat{\mathcal{J}}} + \tilde{\mathbf{n}}_c[k], \quad (2.21)$$

where  $\tilde{\mathbf{n}}_c[k] \in \mathbb{C}^{ML_r \times 1}$  is the residual noise in the linear model after estimating the channel support. If the estimation of the support is accurate,  $\tilde{\mathbf{n}}_c[k]$  will be close to the post-combining noise vector  $\mathbf{n}_c[k]$ . The  $\hat{L} \times 1$  vector  $[\mathbf{g}^v[k]]_{\hat{\mathcal{J}}}$  is the vector of channel gains to be estimated after sparse recovery. It is important to remark that the support estimated by the proposed algorithm may be different from the actual channel support. In general,  $\hat{\mathcal{J}}$  can be different from the actual

support. Therefore, the vector  $[\mathbf{g}^v[k]]_{\hat{\mathcal{T}}} \in \mathbb{C}^{\hat{L} \times 1}$  can be also different from  $\mathbf{g}[k] \in \mathbb{C}^{L \times 1}$ ,  $\mathbf{g}[k] = \text{vec}\{\text{diag}\{\mathbf{G}[k]\}\}$ . Since the model in (2.21) is linear on the parameter vector  $[\mathbf{g}^v[k]]_{\hat{\mathcal{T}}}$ , there is a Minimum Variance Unbiased Estimator (MVUE) that happens to be the Best Linear Unbiased Estimator (BLUE) as well [135].

The equation in (2.21) is usually referred to as the General Linear Model (GLM), for which the estimator of  $[\mathbf{g}^v[k]]_{\hat{\mathcal{T}}}$  for real parameters is provided in [135]. The extension for a complex vector of parameters is straightforward and given by

$$[\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{T}}} = \left( [\mathbf{\Upsilon}]_{:, \hat{\mathcal{T}}}^* \mathbf{C}_w^{-1} [\mathbf{\Upsilon}]_{:, \hat{\mathcal{T}}} \right)^{-1} [\mathbf{\Upsilon}]_{:, \hat{\mathcal{T}}}^* \mathbf{C}_w^{-1} \mathbf{y}[k], \quad (2.22)$$

which can be further reduced to

$$[\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{T}}} = \left( [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{T}}} \right)^\dagger \mathbf{y}_w[k]. \quad (2.23)$$

Therefore,  $[\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{T}}}$  is the MVUE for the parameter vector  $[\mathbf{g}^v[k]]_{\hat{\mathcal{T}}}$ ,  $k = 0, \dots, K-1$ . Hence, it is unbiased and attains the CRLB if the support is estimated correctly. It is interesting to note that this corresponds to a Weighted Least Squares (WLS) estimator, with the corresponding weights given by the inverse noise covariance matrix. An important feature of this estimator is that the difference in performance given by the LS and the WLS estimators is more accentuated as the number of RF chains grows (if and only if the hybrid combiner is not built from orthonormal vectors).

To assess the robustness of the support estimator, it is important to calculate the CRLB for the estimation of the channel matrices at each subcarrier

assuming perfect sparse reconstruction. To that end, taking into account only the non-zero entries in  $\mathbf{g}^v[k]$ , the Fisher Information Matrix (FIM) is derived from the GLM in (2.21) as

$$\mathbf{I}([\mathbf{g}^v[k]]_{\hat{\mathcal{J}}}) = [\boldsymbol{\Upsilon}]_{:, \hat{\mathcal{J}}}^* \mathbf{C}^{-1} [\boldsymbol{\Upsilon}]_{:, \hat{\mathcal{J}}}. \quad (2.24)$$

Note that (2.24) gives the FIM for the vector  $[\mathbf{g}^v[k]]_{\hat{\mathcal{J}}}$ , which contains the actual channel gains. To compute the CRLB for the estimation of the frequency-domain channel matrix  $\mathbf{H}[k]$ , (2.5) can be vectorized as

$$\text{vec}\{\mathbf{H}[k]\} = (\mathbf{A}_T^C \circ \mathbf{A}_R) \text{vec}\{\mathbf{G}[k]\}. \quad (2.25)$$

The decomposition in (2.25) is expressed with equality, since the focus is on finding the CRLB when the estimation of the support is perfect.

The overall minimum variance for an unbiased estimator of the  $KN_tN_r$  entries in  $\{\mathbf{H}[k]\}_{k=0}^{K-1}$  is given by the sum of the variances for the estimators of the elements in the MIMO channel matrices. I will denote the overall variance of the estimator for  $\{\mathbf{H}[k]\}_{k=0}^{K-1}$  as  $\gamma(\{\mathbf{H}[k]\}_{k=0}^{K-1})$ . Then,  $\gamma(\{\mathbf{H}[k]\}_{k=0}^{K-1})$  is derived as

$$\gamma(\{\mathbf{H}[k]\}_{k=0}^{K-1}) = \sum_{k=0}^{K-1} \text{trace}\{(\mathbf{A}_T^C \circ \mathbf{A}_R) \mathbf{I}^{-1}([\mathbf{g}^v[k]]_{\hat{\mathcal{J}}}) (\mathbf{A}_T^C \circ \mathbf{A}_R)^*\}. \quad (2.26)$$

### 2.3.3 Computation of the residual and noise variance estimation

After estimating the channel gains, it is necessary to determine whether a sufficient number of paths has already been estimated or not. To solve this

detection problem, some prior information is needed to compare the received signals  $\mathbf{y}[k]$  to the reconstructed signals  $\hat{\mathbf{x}}_{\text{rec}}[k] = [\boldsymbol{\Upsilon}]_{:, \hat{\mathcal{J}}} [\hat{\mathbf{g}}^{\vee}[k]]_{\hat{\mathcal{J}}}$ . For this reason, I assume that the noise variance is known at the receiver. This is a practical assumption since the receiver can accurately estimate this parameter before the training stage takes place. Therefore, the received signal  $\mathbf{y}[k]$  can be approximately modeled as  $\mathbf{y}[k] \approx \hat{\mathbf{x}}_{\text{rec}}[k] + \tilde{\mathbf{n}}_c[k]$ , which is the same model as in (2.7), since  $\hat{\mathbf{x}}_{\text{rec}}[k]$  is an estimate of the mean of  $\mathbf{y}[k]$ . Let us define the complete received signal as  $\mathbf{y} \triangleq \text{vec}\{\mathbf{y}[0], \dots, \mathbf{y}[K-1]\}$ , and the complete reconstructed signal as  $\hat{\mathbf{x}}_{\text{rec}} \triangleq \text{vec}\{\hat{\mathbf{x}}_{\text{rec}}[0], \dots, \hat{\mathbf{x}}_{\text{rec}}[K-1]\}$ . Then, the estimation of the noise variance can be formulated as a ML estimation problem [135],

$$\hat{\sigma}_{\text{ML}}^2 = \arg \max_{\sigma^2} L(\mathbf{y}, \hat{\mathbf{x}}_{\text{rec}}, \sigma^2), \quad (2.27)$$

where  $L(\mathbf{y}, \hat{\mathbf{x}}_{\text{rec}}, \sigma^2)$  denotes the Log-Likelihood Function (LLF) of  $\mathbf{y}$ . This function is given by [135]

$$\begin{aligned} L(\mathbf{y}, \hat{\mathbf{x}}_{\text{rec}}, \sigma^2) = & -ML_r \ln \pi \sigma^2 - \ln \det\{\mathbf{C}_w\} - \\ & - \frac{1}{\sigma^2} \sum_{k=0}^{K-1} (\mathbf{y}[k] - \hat{\mathbf{x}}_{\text{rec}}[k])^* \mathbf{C}_w^{-1} (\mathbf{y}[k] - \hat{\mathbf{x}}_{\text{rec}}[k]). \end{aligned} \quad (2.28)$$

The ML estimator of the noise variance is then obtained by taking the partial derivative of  $L(\mathbf{y}, \hat{\mathbf{x}}_{\text{rec}}, \sigma^2)$  to zero, such that  $\hat{\sigma}_{\text{ML}}^2$  is given by

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{KML_r} \sum_{k=0}^{K-1} \underbrace{(\mathbf{y}[k] - \hat{\mathbf{x}}_{\text{rec}}[k])^* \mathbf{C}_w^{-1} (\mathbf{y}[k] - \hat{\mathbf{x}}_{\text{rec}}[k])}_{\mathbf{r}^*[k]\mathbf{r}[k]}, \quad (2.29)$$

where the  $ML_r \times 1$  vector  $\mathbf{r}[k] \triangleq \mathbf{y}_w[k] - \mathbf{D}_w^{-*} \hat{\mathbf{x}}_{\text{rec}}$  is the residual. Observe that  $\mathbf{r}[k]$  can also be written as  $\mathbf{r}[k] = (\mathbf{I}_{ML_r} - \mathbf{P}) \mathbf{y}_w[k]$ , where  $\mathbf{P} \in \mathbb{C}^{ML_r \times ML_r}$  is

the projection matrix given by  $\mathbf{P} = [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}}^\dagger [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}}$ . Thereby, after a sufficient number of iterations,  $\hat{L}$  different paths are expected to be estimated. Consequently, the estimated noise variance will be comparable to the true noise variance of the received signal, such that the detection process is accomplished by setting  $\epsilon$  in (2.14) to  $\sigma^2$ .

---

```

1: procedure SW-OMP( $\mathbf{y}[k], \Phi, \Psi, \epsilon$ )
2:   Compute the whitened equivalent measurement matrix
3:      $\mathbf{\Upsilon}_w = \mathbf{D}_w^{-*} \Phi \Psi$ 
4:   Initialize the residual vectors to the input signal vectors and
   support estimate
5:      $\mathbf{y}_w[k] = \mathbf{D}_w^{-*} \mathbf{y}[k], \mathbf{r}[k] = \mathbf{y}_w[k], k = 0, \dots, K - 1, \hat{\mathcal{J}} = \{\emptyset\}$ 
6:   while MSE >  $\epsilon$  do
7:     Subcarrier-wise correlation
8:        $\mathbf{c}[k] = \mathbf{\Upsilon}_w^* \mathbf{r}[k], k = 0, \dots, K - 1$ 
9:     Find the maximum sum-correlation
10:       $p^* = \arg \max_p \sum_{k=0}^{K-1} |\{\mathbf{c}[k]\}_p|$ 
11:     Update the current guess of the common support
12:       $\hat{\mathcal{J}} = \hat{\mathcal{J}} \cup p^*$ 
13:     Estimate the channel gains using the support estimate
14:       $[\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{J}}} = \left([\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}}\right)^\dagger \mathbf{y}_w[k], k = 0, \dots, K - 1$ 
15:     Update residual
16:       $\mathbf{r}[k] = \mathbf{y}_w[k] - [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}} [\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{J}}}, k = 0, \dots, K - 1$ 
17:     Compute the current MSE
18:       $\text{MSE} = \frac{1}{KML_r} \sum_{k=0}^{K-1} \mathbf{r}^*[k] \mathbf{r}[k]$ 
19:   end while
20: end procedure

```

---

Figure 2.3: Detailed steps of the first proposed SW-OMP algorithm.

Of particular importance: the larger the number of subcarriers, the smaller the estimation variance the ML estimator can achieve. Thereby, if the

number of averaging subcarriers  $K$  is large enough, the lack of knowledge of the sparsity level is not so critical because of two reasons: 1) the computation of the support is more precise due to noise averaging during the correlation estimation step, and 2) if the support is estimated correctly, the estimate of  $\sigma^2$  will be very close to the true noise variance, such that the halting criterion is optimal from the ML perspective. It should be clear that the higher the correlation between adjacent noise components, the larger the performance gap between the S-OMP and the SW-OMP algorithms, which depends on the ratio between  $N_r$  and  $L_r$ . The modification of the S-OMP algorithm to include the MVUE estimator for the channel gains, as well as the whitening matrix to estimate the support and the residual is provided in Algorithm 2.3.

### 2.3.4 Subcarrier Selection - Simultaneous Weighted - Orthogonal Matching Pursuit + Thresholding

Despite the use of a single, subcarrier-independent measurement matrix  $\Upsilon$  to estimate the frequency-domain MIMO channels, the algorithm presented in the previous section may exhibit high computational complexity in practice. The SW-OMP algorithm computes a correlation metric exploiting every received subcarrier; however, a tradeoff between estimation performance and computational complexity can be achieved if a small number of subcarriers  $K_p \ll K$  is used, instead. The problem amounts as to how to choose those subcarriers, since no quality measure is available beforehand. The ideal situation would require knowledge of the SNR, which is unknown so far. Nonetheless, the different frequency-domain received vectors  $\mathbf{y}[k]$ ,  $k = 0, 1, \dots, K - 1$ ,

can be used to infer an SNR-related metric. Therefore, with the co-authors in [1], I proposed to use the  $\ell_2$ -norm of the different vectors as quality metric. Owing to the triangle inequality,  $\|\mathbf{y}[k]\|_2^2 \leq \|\Phi \mathbf{g}^v[k]\|_2^2 + \|\mathbf{n}_c[k]\|_2^2$ , such that the  $K_p$  selected signals are expected to exhibit the strongest channel response. Thereby, the  $K_p$  subcarrier signals having largest  $\ell_2$ -norm can be used to derive an estimate of the support of the already defined sparse channel vectors  $\mathbf{g}^v[k]$ ,  $k = 0, \dots, K - 1$ .

The main problem concerning MP algorithms comes from the lack of knowledge of the channel sparsity level  $L$ . For this reason, there is usually an iteration in MP algorithms at which  $L$  paths have been detected but the estimate of the average residual energy is a little larger than the noise variance itself. This makes the algorithm find additional paths which are not actually contained in the MIMO channel. These paths usually have low power, and a pruning procedure is needed to filter out these undesired components. An approach is to remove those components whose power falls below a given threshold, which can be related to the average power of the component in the estimated sparse vectors having maximum average power. Let us denote this power by  $P^*$ . Then, the threshold can be defined as  $\eta = \beta P^*$ ,  $\beta \in (0, 1)$ . The value  $P^*$  is taken as  $P^* = \max_{\ell} \frac{1}{K} \sum_{k=0}^{K-1} |[\hat{\mathbf{g}}^v[k]]_{\ell}|^2$ . To keep the common sparsity property it is necessary to ensure that the channel support after thresholding remains invariant across subcarriers. For this purpose, I define a signal  $\hat{\mathbf{p}}_{\text{av}} \in \mathbb{C}^{\hat{L} \times 1}$  whose  $i$ -th component is given by  $\hat{p}_{\text{av},i} = \frac{1}{K} \sum_{k=0}^{K-1} |[\hat{\mathbf{g}}^v[k]]_i|^2$ ,  $i = 1, \dots, \hat{L}$ , such that  $\hat{\mathbf{p}}_{\text{av}}$  measures the average power of each spatial component in the

quantized angle grid across the different subcarriers. The final support after thresholding  $\hat{\mathcal{J}}_{\text{Th}}$  is defined as  $\hat{\mathcal{J}}_{\text{Th}} = \bigcup_{i=1}^{\hat{L}} \{i / \hat{p}_{\text{av},i} \geq \beta P^*\}$ . Therefore, the components in  $\hat{\mathbf{g}}^{\text{v}}[k]$  indexed by  $\hat{\mathcal{J}}_{\text{Th}}$  are the final channel gains estimates for each subcarrier. The modification of the proposed SW-OMP algorithm to reduce computational complexity and implement this pruning procedure is provided in Algorithm 2.4.

### 2.3.5 Convergence Analysis

In this section, the proposed SW-OMP and Subcarrier Selection - Simultaneous Weighted - Orthogonal Matching Pursuit + Thresholding (SS-SW-OMP+Th) algorithms are theoretically analyzed to show its convergence to a local optimum. I assume that the dictionary sizes  $G_t, G_r$  are large enough such that the coarsely quantized AoD/AoD are accurately estimated, which holds for large enough values of  $M$  [14] and  $K$ . Since the computation of the residual is identical for both algorithms, they are analyzed following the same procedure. A sufficient condition for convergence to a local optimum is that the energy of the residual computed at the  $(n+1)$ -th iteration is strictly smaller than that of the  $n$ -th iteration, i.e.,

$$\|\mathbf{r}^{(n+1)}[k]\|_2^2 < \|\mathbf{r}^{(n)}[k]\|_2^2, \quad k = 0, \dots, K-1. \quad (2.30)$$

The residual for a given iteration  $n$  can be written as

$$\mathbf{r}^{(n)}[k] = (\mathbf{I}_{ML_r} - \mathbf{P}^{(n)}) \mathbf{y}_w[k], \quad (2.31)$$

---

```

1: procedure SS-SW-OMP+TH( $\mathbf{y}[k], \Phi, \Psi, K_p, \beta, \epsilon$ )
2:   Initialize counter, set of subcarriers and residual vectors
3:    $i = 0, \mathcal{K} = \{\emptyset\}, \mathbf{y}_w[k] = \mathbf{D}_w^{-*} \mathbf{y}[k], \mathbf{r}[k] = \mathbf{y}_w[k], k = 0, \dots, K - 1$ 
4:   Find the  $K_p$  strongest subcarriers
5:   while  $i \leq K_p$  do
6:      $\mathcal{K} = \mathcal{K} \cup \arg \max_{k \notin \mathcal{K}} \|\mathbf{y}[k]\|_2^2$ 
7:      $i = i + 1$ 
8:   end while
9:   Compute the whitened equivalent measurement matrix
10:   $\Upsilon_w = \mathbf{D}_w^{-*} \Phi \Psi$ 
11:  while  $\text{MSE} > \epsilon$  do
12:    Subcarrier-wise correlation for the  $K_p$  selected subcarriers
13:     $\mathbf{c}[k] = \Upsilon_w^* \mathbf{r}[k], k \in \mathcal{K}$ 
14:    Find the maximum sum-correlation
15:     $p^* = \arg \max_p \sum_{k \in \mathcal{K}} |\{\mathbf{c}[k]\}_p|$ 
16:    Update the current guess of the common support
17:     $\hat{\mathcal{T}} = \hat{\mathcal{T}} \cup p^*$ 
18:    Estimate the channel gains using the support estimate
19:     $[\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{T}}} = \left( [\Upsilon_w]_{:, \hat{\mathcal{T}}} \right)^\dagger \mathbf{y}_w[k], k = 0, \dots, K - 1$ 
20:    Update residual
21:     $\mathbf{r}[k] = \mathbf{y}_w[k] - [\Upsilon_w]_{:, \hat{\mathcal{T}}} [\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{T}}}, k = 0, \dots, K - 1$ 
22:    Compute the current MSE
23:     $\text{MSE} = \frac{1}{KML_r} \sum_{k=0}^{K-1} \mathbf{r}^*[k] \mathbf{r}[k]$ 
24:  end while
25:  Thresholding based on maximum average power
26:   $P^* = \max_\ell \frac{1}{K} \sum_{k=0}^{K-1} |[\hat{\mathbf{g}}^v[k]]_\ell|^2$ 
27:   $\hat{p}_{\text{av}, i} = \frac{1}{K} \sum_{k=0}^{K-1} |[\hat{\mathbf{g}}^v[k]]_i|^2, i = 1, \dots, \hat{L}$ 
28:   $\hat{\mathcal{T}}_{\text{Th}} = \bigcup i / \hat{p}_{\text{av}, i} \geq \beta P^*, i \in \hat{\mathcal{T}}$ 
29:   $\hat{\mathbf{g}}^v[k] = [\hat{\mathbf{g}}^v[k]]_{\hat{\mathcal{T}}_{\text{Th}}}, k = 0, \dots, K - 1$ 
30: end procedure

```

Figure 2.4: Detailed steps of the second proposed SS-SW-OMP+Th algorithm.

with  $\mathbf{P}^{(n)} \in \mathbb{C}^{ML_r \times ML_r}$  a projection matrix given by  $\mathbf{P}^{(n)} \triangleq [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}^{(n)}} [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}^{(n)}}^\dagger$ . Accordingly,  $\mathbf{r}^{(n)}[k]$  is the vector resulting from projecting  $\mathbf{y}_w[k]$  onto the subspace orthogonal to the column space of  $[\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}^{(n)}}$ . Therefore, the condition in (2.30) can be rewritten as a function of the projection onto the column space of  $[\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}^{(n)}}$  as

$$\|\mathbf{P}^{(n+1)} \mathbf{y}_w[k]\|_2^2 > \|\mathbf{P}^{(n)} \mathbf{y}_w[k]\|_2^2. \quad (2.32)$$

Observe that the term inside the  $\ell_2$ -norm on the left side of (2.32) can be expressed as

$$\mathbf{P}^{(n+1)} \mathbf{y}_w[k] = \begin{bmatrix} [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}^{(n)}} & [\mathbf{\Upsilon}_w]_{:, \hat{p}^{(n+1)*}} \end{bmatrix} \begin{bmatrix} [\mathbf{\Upsilon}_w]_{:, \hat{\mathcal{J}}^{(n)}} & [\mathbf{\Upsilon}_w]_{:, \hat{p}^{(n+1)*}} \end{bmatrix}^\dagger \mathbf{y}_w[k], \quad (2.33)$$

with  $\hat{p}^{(n+1)*}$  the estimate for the support index found during the  $(n+1)$ -th iteration, satisfying  $\hat{p}^{(n+1)*} \notin \hat{\mathcal{J}}^{(n)}$ . Thereby, the projection matrix  $\mathbf{P}^{(n+1)}$  can be recursively written as a function of  $\mathbf{P}^{(n)}$  using the formula for the inverse of a  $2 \times 2$  block matrix [135] as

$$\mathbf{P}^{(n+1)} = \mathbf{P}^{(n)} + \underbrace{\frac{(\mathbf{I}_{ML_r} - \mathbf{P}^{(n)}) [\mathbf{\Upsilon}_w]_{:, \hat{p}^{(n+1)*}} [\mathbf{\Upsilon}_w]_{:, \hat{p}^{(n+1)*}}^* (\mathbf{I}_{ML_r} - \mathbf{P}^{(n)})}{[\mathbf{\Upsilon}_w]_{:, \hat{p}^{(n+1)*}}^* (\mathbf{I}_{ML_r} - \mathbf{P}^{(n)}) [\mathbf{\Upsilon}_w]_{:, \hat{p}^{(n+1)*}}}}_{\Delta \mathbf{P}^{(n+1)}}, \quad (2.34)$$

with  $\Delta \mathbf{P}^{(n+1)} \in \mathbb{C}^{ML_r \times ML_r}$  another projection matrix accounting for the relation between the projection matrices at the  $n$ -th and  $(n+1)$ -th iterations. The equation in (2.34) can be noticed to fulfill the orthogonality principle,  $\mathbf{P}^{(n+1)} \Delta \mathbf{P}^{(n+1)} = \mathbf{0}$ . The left-handed term in (2.32) can then be expressed as

$$\begin{aligned} \|\mathbf{P}^{(n+1)} \mathbf{y}_w[k]\|_2^2 &= \|\mathbf{P}^{(n)} \mathbf{y}_w[k] + \Delta \mathbf{P}^{(n+1)} \mathbf{y}_w[k]\|_2^2 \\ &= \|\mathbf{P}^{(n)} \mathbf{y}_w[k]\|_2^2 + \|\Delta \mathbf{P}^{(n+1)} \mathbf{y}_w[k]\|_2^2, \end{aligned} \quad (2.35)$$

thereby satisfying the triangle equality. Finally, from (2.35), it is immediate that  $\|\mathbf{P}^{(n+1)}\mathbf{y}_w[k]\|_2^2 > \|\mathbf{P}^{(n)}\mathbf{y}_w[k]\|_2^2$  since  $\Delta\mathbf{P}^{(n+1)}$  has a non-zero unity eigenvalue. Thereby, the condition in (2.32) is satisfied and convergence of the proposed algorithms to a local optimum is guaranteed.

## 2.4 Numerical Results

This section includes the main numerical results obtained with the two proposed algorithms, SW-OMP and SS-SW-OMP+Th, and comparisons with other frequency-domain channel estimation algorithms including SSAMP [69] and DGMP [70] are also provided. To obtain these results, I perform Monte Carlo simulations averaged over 100 trials to evaluate the NMSE and the ergodic rate as a function of SNR and number of training frames  $M$ . I also provide calculations of the computational complexity for the proposed algorithms in Table 2.2 and prior work in Table 2.3.

The typical parameters for the system configuration are summarized as follows and included in Table 2.1. Both the transmitter and the receiver are assumed to use a ULA with half-wavelength separation. Such a ULA has steering vectors obeying the expressions  $[\mathbf{a}_T(\theta_\ell)]_n = \sqrt{\frac{1}{N_t}}e^{jn\pi\cos(\theta_\ell)}$ ,  $n = 0, \dots, N_t - 1$  and  $[\mathbf{a}_R(\phi_\ell)]_m = \sqrt{\frac{1}{N_r}}e^{jm\pi\cos(\phi_\ell)}$ ,  $m = 0, \dots, N_r - 1$ . I take  $N_t = N_r = 32$  and  $G_t = G_r = 64$  for illustration. The phase-shifters used in both the transmitter and the receiver are assumed to have  $N_Q$  quantization bits, so that the entries of the analog training precoders and combiners  $\mathbf{F}_{\text{tr}}^{(m)}$ ,  $\mathbf{W}_{\text{tr}}^{(m)}$ ,  $m = 1, 2, \dots, M$  are drawn from a set  $\mathcal{A} = \left\{0, \frac{2\pi}{2^{N_Q}}, \dots, \frac{2\pi(2^{N_Q}-1)}{2^{N_Q}}\right\}$ . The

Table 2.1: Summary of typical system configuration parameters

Simulation Parameters		
Description	Parameter	Value
Number of transmit antennas	$N_t$	32
Number of receive antennas	$N_r$	32
Number of transmit RF chains	$L_t$	4
Number of receive RF chains	$L_r$	4
Transmit angular grid size	$G_t$	64
Receive angular grid size	$G_r$	64
Number of phase-shifter quantization bits	$N_Q$	2
Number of OFDM subcarriers	$K$	16
Number of ZP samples	$Z_p$	4
Number of OFDM training symbols	$M$	80
Sampling period	$T_s$	$1/1760 \mu\text{s}$
Number of channel paths	$L$	4
Channel tap length	$D$	4
Pulse-shape roll-off factor	$\beta$	0.8

number of quantization bits is set to  $N_Q = 2$ . The number of RF chains is set to  $L_t = 4$  at the transmitter and  $L_r = 4$  at the receiver. The number of OFDM subcarriers is set to  $K = 16$ .

I initially generate channels according to (2.2) with the following parameters:

- A number of  $L = 4$  channel paths are assumed to be independent and identically distributed, with delay  $\tau_\ell$  chosen uniformly from  $[0, (D-1)T_s]$ , with  $T_s = \frac{1}{1760} \mu\text{s}$ , as in the IEEE 802.11ad wireless standard.
- The AoD/AoA are assumed to be uniformly distributed in  $(0, \pi)$ .
- The gains of each path are zero-mean complex Gaussian distributed such

that  $\mathbb{E}_k\{\|\mathbf{H}[k]\|_F^2\} = \frac{N_r N_t}{\rho_L}$ .

- The band-limiting filter  $p_{rc}(t)$  is assumed to be a raised-cosine pulse-shape with roll-off factor of 0.8.
- The number of delay taps of the channel is set to  $D = 4$  symbols.

In the simulations, I consider channel realizations in which the AoD/AoA are off-grid, i.e. do not correspond to the angles used to build the dictionary, and also the on-grid case, to analyze the loss due to the model mismatch.

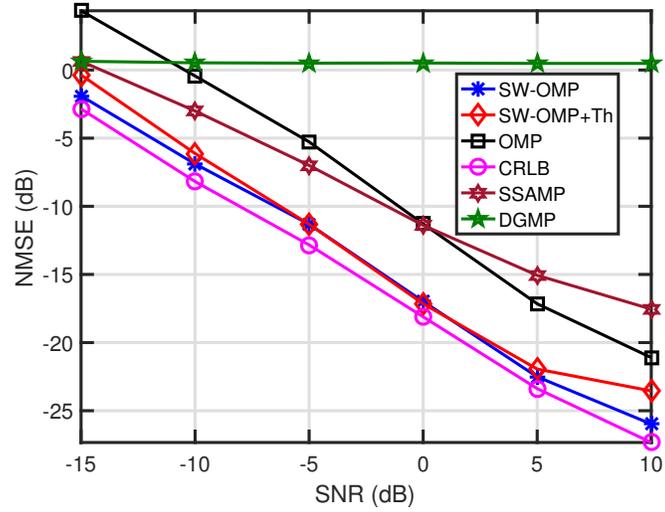
#### 2.4.1 NMSE Comparison

An important estimation performance metric is the NMSE of a channel estimate  $\{\hat{\mathbf{H}}[k]\}_{k=0}^{K-1}$  for a given realization, defined as

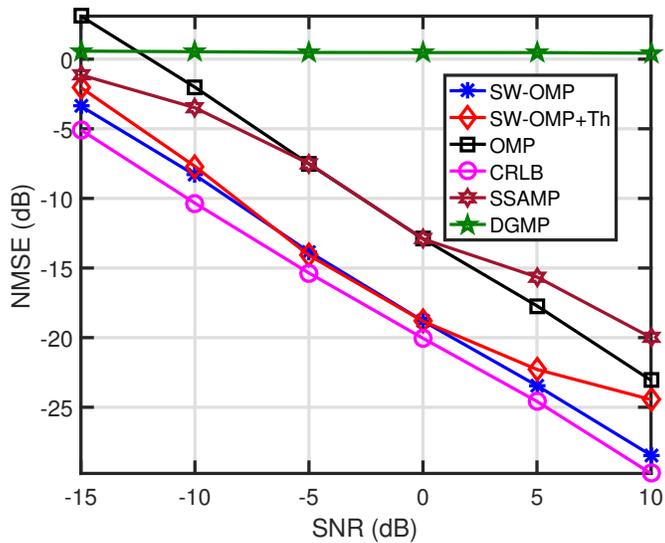
$$\text{NMSE} = \frac{\sum_{k=0}^{K-1} \|\hat{\mathbf{H}}[k] - \mathbf{H}[k]\|_F^2}{\sum_{k=0}^{K-1} \|\mathbf{H}[k]\|_F^2}. \quad (2.36)$$

The NMSE will be the baseline metric to compute the performance of the different algorithms, and will be averaged over 100 channel realizations. The normalized CRLB is also provided to compare the average performance of each algorithm with the lowest achievable NMSE, and will also be averaged over many channel realizations.

I compare the average NMSE versus SNR obtained for the different channel estimation algorithms in Fig. 2.5 for a practical SNR range of  $-15$  dB to  $10$  dB, on-grid AoD/AoA, and two different values for the number of training frames.



(a)



(b)

Figure 2.5: Evolution of the NMSE versus SNR for the different frequency-domain algorithms when the AoD/AoA are assumed to lie on the dictionary grid. The number of training frames is set to  $M = 80$  (a) and  $M = 120$  (b).

The first proposed SW-OMP algorithm exhibits the best estimation error performance, achieving NMSE values very close to the CRLB. The second proposed SS-SW-OMP+Th algorithm performs similarly to SW-OMP, although there is some performance loss due to the fact that SS-SW-OMP+Th does not employ every subcarrier to estimate the common support of the sparse channel vectors. In SS-SW-OMP+Th, the number of selected subcarriers for the estimation of the support is set to  $K_p = 4$  for illustration, and the parameter  $\beta$  is chosen as  $\beta = 0.025$ , which is a reasonably small value to filter out undesired components in the sparse channel estimate. From the curves shown in Fig. 2.5, OMP performs poorly over all the SNR range since it is not designed to process several vectors which are sparse in a common vector basis. Exploiting common spatial sparsity provides an NMSE reduction of approximately 7 dB, although there are slight variations depending on the SNR value. This improvement comes at the cost of a higher offline computational complexity in the proposed algorithms in comparison with OMP, as I show in Section 2.4.3.

It can also be observed that the DGMP algorithm from [70] exhibits the worst estimation error performance, which is expected since it was designed to estimate near-LOS mmWave MIMO channels. Since it only estimates a single path, the estimation error for NLOS channels is large. The algorithm SSAMP is also shown for comparison. At low SNR regime, the information on the common support is enough to outperform the OMP algorithm, but not at the high SNR regime. This comes from the fact that the SSAMP algorithm

estimates more than a single path per iteration. Since the dictionary matrices are not square in this setting, the redundancy between columns in the transmit and receive array matrices makes it difficult to properly estimate more than a single support index per iteration.

Using a larger number of training frames  $M$  enhances estimation performance, but at the cost of both higher overhead and computational complexity, since the complexity of estimating the support, channel gains and noise variance grows linearly with  $L_r M$ . Nonetheless, there is an SNR-increasing performance gap between SS-SW-OMP+Th and SW-OMP. Even further, this gap increases with the number of training frames, which may seem counter-intuitive. This effect comes the variance of the MVUE for the channel gains depends on the SNR. Clearly, in high SNR regime, the estimates of the weakest paths have smaller estimation variance than in low SNR regime. Therefore, since the threshold (dependent on  $\beta$ ) is not adapted to the noise variance, these paths are more likely to be removed at higher SNR. When  $M$  increases, the estimation variance also decreases, which further increases this gap, as observed in Fig. 2.5(a) and 2.5(b).

I show in Fig. 2.6 the performance of the different frequency-domain algorithms when increasing the number of subcarriers. The parameters for the simulation scenario are the same as in Fig. 2.5, however, the number of subcarriers is set to  $K = 64$  in this case.  $K_p$  is set to 32 subcarriers and  $\beta = 0.025\sigma^2$ . Interestingly, both SW-OMP and SS-SW-OMP+Th are asymptotically efficient in SNR since they are both unbiased and attain the

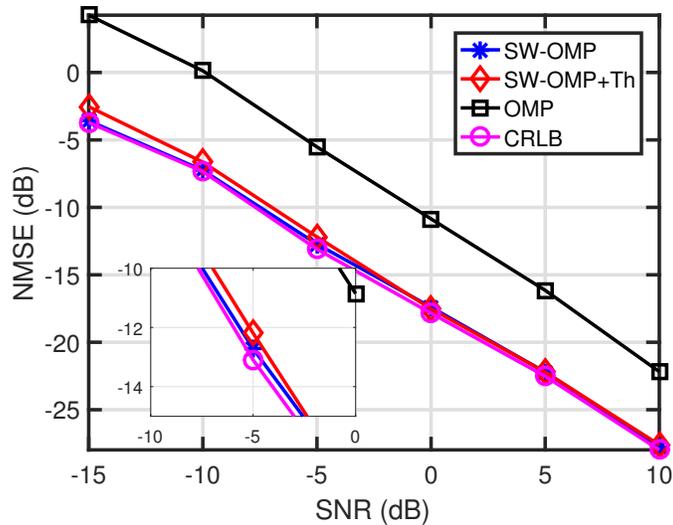
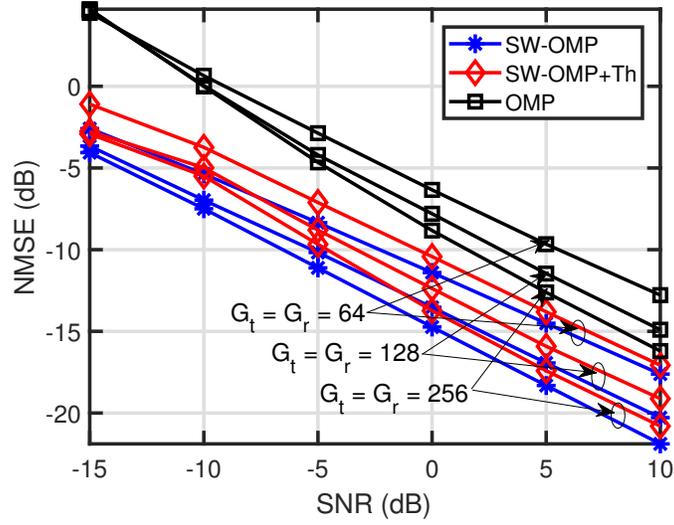


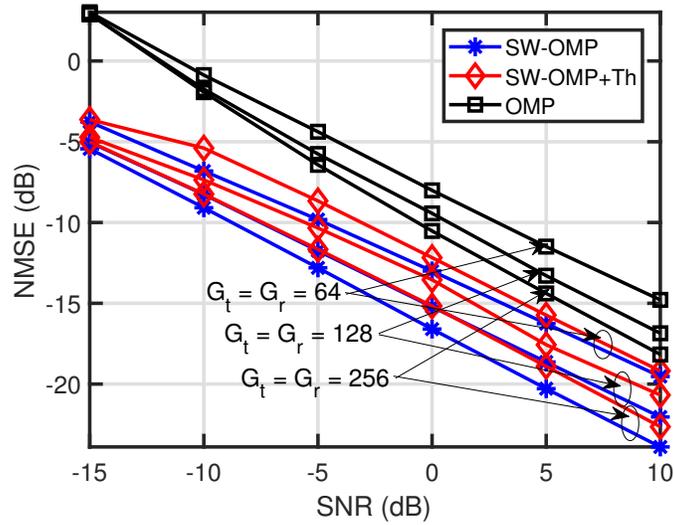
Figure 2.6: Comparison of evolution of the NMSE versus SNR for the different frequency-domain algorithms. The number of training frames is set to  $M = 80$ . The number of subcarriers is set to  $K = 64$ .

NCRLB. A magnified plot for  $\text{SNR} = -5$  dB is also shown to clearly see the performance gap between the different algorithms and also the CRLB.

The previous simulations showed the performance of the algorithms when the channel fits the on-grid model, but it is also important to analyze the performance in a practical scenario, when the AoD/AoA do not fall within the quantized spatial grid. Fig. 2.7 shows the performance of the different algorithms under a more challenging scenario with channel realizations extracted from the NYUSIM channel simulator [4]. The simulation parameters for this scenario are chosen as  $K = 256$ ,  $K_p = K/4 = 64$ ,  $\beta = 0.01\sigma^2$  and  $L_t = L_r = 4$ . The remaining parameters are the same as in Fig. 2.5. Results are shown for  $M = 60$  frames (a), and  $M = 100$  frames (b).

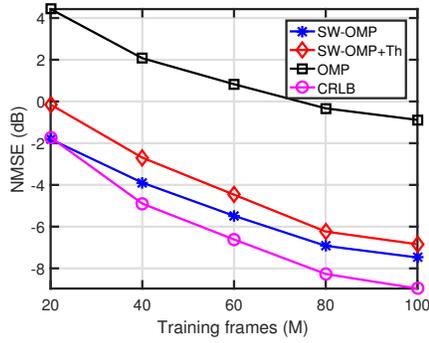


(a)

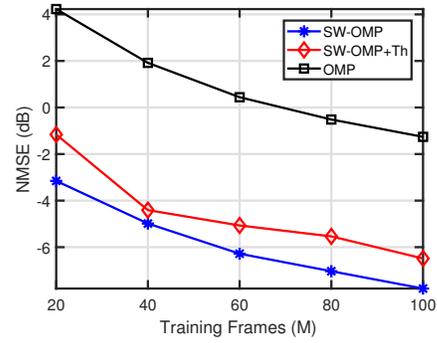


(b)

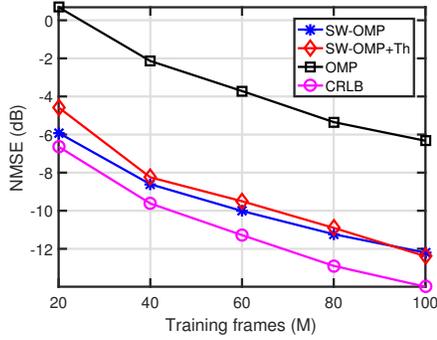
Figure 2.7: Evolution of the NMSE versus SNR for the different frequency-domain algorithms. The number of training frames is set to  $M = 60$  (a) and  $M = 100$  (b). The channel realizations are taken from NYUSIM channel simulator.



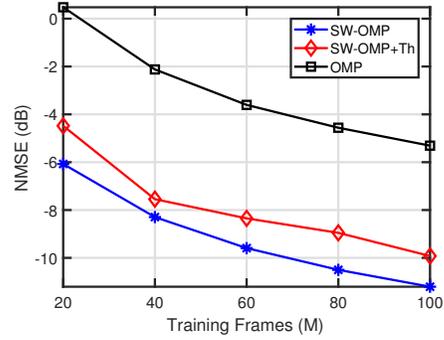
(a)



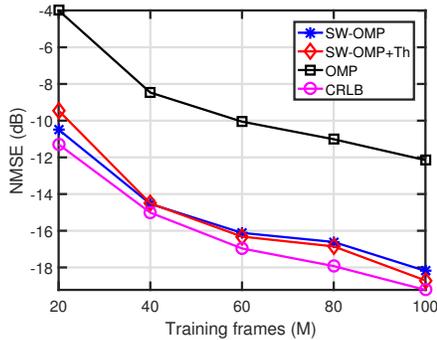
(d)



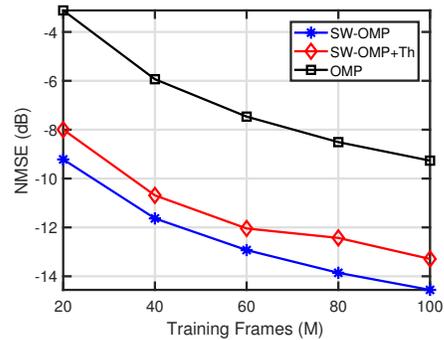
(b)



(e)



(c)



(f)

Figure 2.8: Comparison of evolution of the NMSE versus  $M$  at different SNR. The SNR is set to  $-10$  dB (a) and (d),  $-5$  dB (b) and (e) and  $0$  dB (c) and (f). Plots (a), (b), (c) consider on-grid angular parameters on the channel realizations, while (d), (e), (f) consider the off-grid case with channel realizations extracted from the NYUSIM channel simulator [4].

The estimation error is observed to be below  $-10$  dB for values of SNR in the order of 0 and beyond. On the other hand, since the SNR expected in mmWave communication systems is in the order of  $-20$  dB up to 0 dB, the attained NMSE should be reduced. Increasing the size of the dictionary is one of the possible solutions, as shown by the curves in Fig. 2.7 corresponding to  $G_t = G_r = 128$  and  $G_t = G_r = 256$ .

Fig. 2.8 shows the average NMSE vs number of training frames  $M$ . The number of training frames  $M$  is increased from 20 to 100. The remaining parameters in the simulation scenario are the same as in Fig. 2.7, with  $G_t = G_r = 128$ . Results are shown for channel realizations in which the angular parameters fall within the quantized angle grid and when they do not.

The average performance of the OMP algorithm is poor for all the considered cases, which comes from its inability to exploit the common support property shared by the different subchannels. The first proposed SW-OMP algorithm can be observed to provide the best performance for the different values of  $M$  and SNR. The larger the number of training frames and the higher the SNR, the estimation of the support is more robust and gets closer to the actual one. In the on-grid case, if the number of training frames is large enough and the SNR is not low, the performance gap between SW-OMP and the CRLB is smaller than 1 dB. The difference in performance between SS-SW-OMP+Th and SW-OMP reduces when either  $M$  or SNR is increased. As in Figs. 2.5 and 2.7, there is a big difference in performance between OMP and SW-OMP, depending on both the SNR and the number of frames.

## 2.4.2 Spectral efficiency comparison

Another performance metric is the spectral efficiency, which is computed assuming fully-digital precoding and combining using estimates of the  $N_s$  dominant left and right singular vectors of the channel estimate. This gives  $K$  parallel effective channels  $\mathbf{H}_{\text{eff}}[k] = \left[ \hat{\mathbf{U}}[k] \right]_{:,1:N_s}^* \mathbf{H}[k] \left[ \hat{\mathbf{V}}[k] \right]_{:,1:N_s}$ . Accordingly, the average spectral efficiency can be expressed as [136]

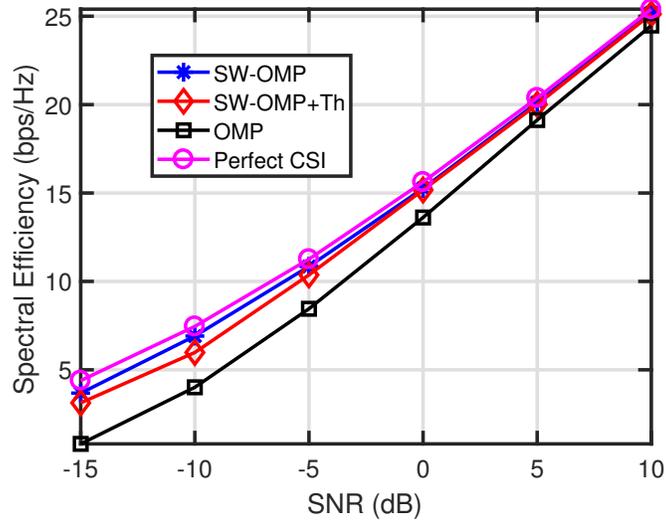
$$\mathcal{R}(\text{SNR}) = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{n=1}^{N_s} \log \left( 1 + \frac{\text{SNR}}{N_s} \lambda_n(\mathbf{H}_{\text{eff}}[k])^2 \right), \quad (2.37)$$

with  $\lambda_n(\mathbf{H}_{\text{eff}}[k])$ ,  $n = 1, \dots, N_s$  denoting the singular values of the effective channel  $\mathbf{H}_{\text{eff}}[k]$ .

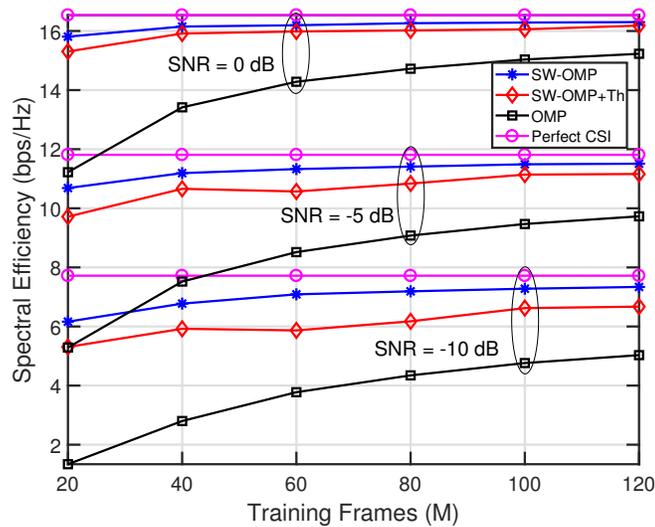
In Fig. 2.9(a), I show the achievable spectral efficiency as a function of the SNR for the different channel estimation algorithms when realistic channel realizations are considered. The simulation parameters are the same as in Fig. 2.7.

The difference in performance between OMP and the two proposed algorithms is noticeable, which comes from the fact that OMP does not exploit spatially common sparsity. The two proposed algorithms perform similarly for all the range of SNR, which is an indicator that  $K_p < K$  subcarriers provide sufficient information to obtain a reliable channel estimate. Therefore, SS-SW-OMP+Th can be claimed to be a good tradeoff between performance and computational complexity.

Finally, Fig. 2.9(b) shows the spectral efficiency as a function of the number of training frames  $M$  for the different channel estimation algorithms.



(a)



(b)

Figure 2.9: (a) Evolution of the spectral efficiency versus SNR for the different frequency-domain algorithms. The number of training frames is set to  $M = 60$ . (b) Evolution of the spectral efficiency versus number of training frames  $M$  at different SNR for the different frequency-domain algorithms.

Comparisons are provided for  $\text{SNR} = \{-10, -5, 0\}$  dB. It is observed that SW-OMP is the algorithm providing the best performance, followed closely by SS-SW-OMP+Th, whilst OMP performs the worst. For low values of SNR, there is a noticeable performance gap between SW-OMP and the perfect CSI case, which becomes smaller as  $M$  increases. It can be noticed that using  $M \geq 40$  frames does not bring about significant improvement in performance, which leverages the robustness of the two proposed approaches. Simulations also show that near-optimal achievable rates can be achieved by using a reasonable number of frames, i.e.,  $40 \leq M \leq 100$ .

### 2.4.3 Computational complexity

The computational complexity for each step in the different proposed algorithms is also provided for the  $j$ -th iteration in Table 2.2. Since some steps can be performed before running the channel estimation algorithms, it is important to distinguish between on-line and off-line operations. Values are provided for a single iteration. In the case of the SSAMP algorithm [69, 71], the DGMP algorithm in [70], and the OMP algorithm in [34], I take the notation used in the corresponding papers for frequency-domain vectors and measurement matrices.

The computational complexity of SS-SW-OMP+Th is lower than its SW-OMP counterpart, since a reduced number of correlations are computed to estimate the channel support. It must be noticed that the complexity of SS-SW-OMP+Th is lower than that of OMP, since the matrix product

Table 2.2: Online computational complexity of proposed algorithms

SW-OMP	
Operation	Complexity
$(K) \times \mathbf{c}[k] = \mathbf{\Upsilon}_w^* \mathbf{r}[k]$	$\mathcal{O}(K(G_r G_t - (j-1))L_r M)$
Maximum of $\sum_{k=0}^{N-1}  \{\mathbf{c}[k]\}_p $	$\mathcal{O}(K(G_r G_t - (j-1)))$
$(K) \times \mathbf{x}_{\hat{j}}[k] = \left([\mathbf{\Upsilon}_w]_{:, \hat{j}}\right)^\dagger \mathbf{y}_w[k]$	$\mathcal{O}(j^2 L_r M + j^3)$
$(K) \times \mathbf{r}[k] = \mathbf{y}_w[k] - [\mathbf{\Upsilon}_w]_{:, \hat{j}} \hat{\boldsymbol{\xi}}[k]$	$\mathcal{O}(K L_r M)$
$\text{MSE} = \frac{1}{K M L_r} \sum_{k=0}^{K-1} \mathbf{r}^*[k] \mathbf{r}[k]$	$\mathcal{O}(K L_r M)$
Overall	$\mathcal{O}(L_r M (K(G_r G_t - (j-1))))$
SS-SW-OMP+Th	
Operation	Complexity
Find the $K_p$ strongest subcarriers	$\mathcal{O}(K L_r M)$
$(K_p) \times \mathbf{c}[k] = \mathbf{\Upsilon}_w^* \mathbf{r}[k]$	$\mathcal{O}(K_p (G_r G_t - (j-1)) L_r M)$
Maximum of $\sum_{k \in \mathcal{X}}  \{\mathbf{c}[k]\}_p $	$\mathcal{O}(K_p (G_r G_t - (j-1)))$
$(K) \times \mathbf{x}_{\hat{j}}[k] = \left([\mathbf{\Upsilon}_w]_{:, \hat{j}}\right)^\dagger \mathbf{y}_w[k]$	$\mathcal{O}(j^2 L_r M + j^3)$
$(K) \times \mathbf{r}[k] = \mathbf{y}_w[k] - [\mathbf{\Upsilon}_w]_{:, \hat{j}} \hat{\boldsymbol{\xi}}[k]$	$\mathcal{O}(K L_r M)$
$\text{MSE} = \frac{1}{K M L_r} \sum_{k=0}^{K-1} \mathbf{r}^*[k] \mathbf{r}[k]$	$\mathcal{O}(K L_r M)$
Thresholding	$\mathcal{O}(K \hat{L})$
Overall	$\mathcal{O}(L_r M (K_p (G_r G_t - (j-1))))$

$\mathbf{\Upsilon}_w = \mathbf{D}_w^{-*} \mathbf{\Upsilon}$  can be computed before explicit channel estimation. The online computational complexity of SW-OMP is lower than that of OMP, since OMP computes  $K$  matrix pseudoinverses while SW-OMP only computes one. Conversely, the offline computational complexity of both the proposed SW-OMP and SS-SW-OMP+Th algorithms is higher than those for the other algorithms, since the matrix  $\mathbf{\Upsilon}_w$  must be computed before explicit channel estimation.

The computational complexity of SW-OMP is lower than its SSAMP counterpart. SSAMP exhibits an increase in complexity of at most  $\mathcal{O}(4j^2)$

Table 2.3: Online computational complexity of previously proposed algorithms

OMP in [34]	
Operation	Complexity
$(K) \times \mathbf{c}[k] = \mathbf{\Upsilon}^*[k]\mathbf{r}[k]$	$\mathcal{O}(K(G_r G_t - (j-1))L_r M)$
$(K) \times$ Maximum of $ \{\mathbf{c}[k]\}_p $	$\mathcal{O}(K(G_r G_t - (j-1)))$
$(K) \times \mathbf{x}_{\hat{\gamma}}[k] = [\mathbf{\Upsilon}]_{:, \hat{\gamma}}^\dagger [k]\mathbf{z}[k]$	$\mathcal{O}(K(2j^2 L_r M + j^3))$
$(K) \times \mathbf{r}[k] = \mathbf{y}[k] - [\mathbf{\Upsilon}]_{:, \hat{\gamma}} [k]\hat{\boldsymbol{\xi}}[k]$	$\mathcal{O}(K L_r M)$
$(K) \times \text{MSE} = \frac{1}{M L_r} \mathbf{r}^*[k]\mathbf{r}[k]$	$\mathcal{O}(K L_r M)$
Overall	$\mathcal{O}(K L_r M(G_r G_t - (j-1) + 2j^2))$
SSAMP in [69, 71]	
Operation	Complexity
$(K) \times \mathbf{a}_p = \mathbf{\Phi}_p^* \mathbf{b}_p^{i-1}$	$\mathcal{O}(K G_r G_t L_r M)$
Maximum of $\sum_{p=0}^{K-1} \ \mathbf{a}_p\ _2^2$	$\mathcal{O}(N G_t G_r)$
$(K) \times \{\mathbf{t}_p\}_{\Omega^{i-1} \cup \Gamma} = ((\mathbf{\Phi}_p)_{\Omega^{i-1} \cup \Gamma}^\dagger \mathbf{r}_p$	$\mathcal{O}(K(4j^2 M L_r + (2j)^3))$
Prune support	$\mathcal{O}(2Kj)$
$(K) \times \{\mathbf{c}_p\}_\Omega = (\{\mathbf{\Phi}_p\}_\Omega)^\dagger \mathbf{r}_p$	$\mathcal{O}(K(2j^2 M L_r + j^3))$
$(K) \times \mathbf{b}_p = \mathbf{r}_p - \mathbf{\Phi}_p \mathbf{c}_p$	$\mathcal{O}(K L_r M G_t G_r)$
Computation of total error $\sum_{p=0}^{K-1} \ \mathbf{b}_p\ _2^2$	$\mathcal{O}(K M L_r)$
Overall	$\mathcal{O}(2K L_r M(G_r G_t + 6j^2))$
DGMP in [70]	
Operation	Complexity
$(K) \times \mathbf{a}_p = \mathbf{\Upsilon}_p^* \mathbf{r}_p$	$\mathcal{O}(K G_r G_t L_r M)$
Maximum of $\rho = \arg \max_{\hat{\rho}} \sum_{p=0}^{N-1} \ \mathbf{a}_p\ _2^2$	$\mathcal{O}(K G_t G_r)$
$(K) \times \{\alpha_p\}_\rho = (\{\mathbf{\Phi}_p\}_\rho)^\dagger \mathbf{r}_p$	$\mathcal{O}(K(2j^2 M L_r + j^3))$
Overall	$\mathcal{O}(K L_r M(G_r G_t + 2j^2))$

owing to the estimation of  $j$  paths at the  $j$ -th iteration of SSAMP. More specifically, this algorithm uses an iteration index  $i$  to estimate the sparsity level  $L$ , and a stage index  $j$  to estimate the  $j$  channel paths found at the current iteration. Afterwards, the support of the channel estimate is pruned

to select the  $j$  most likely channel paths. Therefore, at a given iteration  $i$  and stage  $j$ , at most  $2j = |\tilde{\Omega}^{i-1} \cup \Gamma|$  paths are estimated and then pruned, such that only  $j$  paths are selected among the  $2j$  candidates. The union of the sets  $\Omega$  and  $\Gamma$  comes from the possibility of finding new potential paths at the  $i$ -th iteration and the  $j$ -th stage. This is done by jointly considering the paths found at  $(i - 1)$ -th iteration and the ones found in the  $j$ -th stage within the  $i$ -th iteration. While both SW-OMP, SS-SW-OMP+Th and OMP estimate a single path at a given iteration  $j$ , SSAMP estimates at most  $2j$  different paths by using LS. When computing the pseudoinverse during LS estimation, this results in an additional increase in complexity of  $\mathcal{O}(4j^2)$ , as shown in Table 2.3. By contrast, as shown in Table 2.2, the proposed ML estimator for the channel gains exhibits computational complexity in the order of  $\mathcal{O}(2j^2)$ , thereby slightly reducing the number of operations.

While OMP does not require any offline operation, both SW-OMP and SS-SW-OMP+Th need to compute the matrix  $\mathbf{\Upsilon}_w$ . The offline computation of  $\mathbf{D}_w^{-1}$  has complexity of  $\mathcal{O}(\frac{M}{3}L_r^3)$ , since  $\mathbf{C}_w^{-1}$  is a block diagonal matrix containing  $M$  Hermitian matrices. It is important to remark that this cost comes from the use of frequency-flat precoders/combiners. This entails a reduction in computational complexity with respect to the case in which frequency-selective baseband combiners were used during the channel estimation stage. In summary, the proposed algorithms reduce the computational complexity by approximately a factor of  $K$  vs. the OMP algorithm.

#### 2.4.4 Bit Error Rate

The last performance measure I consider in this chapter is the BER. Simulation results on the BER achieved by the proposed algorithms are provided, as well as comparisons with the baseline approach in [34]. I assume that channel estimation has already been performed, and a data transmission phase takes places. I provide both uncoded and coded BER results. For the latter, I use Low Density Parity Check (LDPC) codes, inspired by the IEEE 802.11ad mmWave communications standard [36], to encode the data bits to be transmitted. I consider the OFDM-PHY transmission frame specified in the aforementioned standard, and use the Modulation and Coding Scheme (MCS) 18 for illustration. Such MCS considers a 16-QAM constellation using dual carrier modulation [36], in which two constellation symbols are shared between two subcarriers, and LDPC coding with rate  $R_{\text{LDPC}} = \frac{1}{2}$ . To estimate the equivalent (beamformed) channel, I use known symbols from the pilot subcarriers contained in this frame, and then use spline interpolation to estimate the beamformed channel at the data subcarriers. I denote the set of positions of pilot subcarriers as  $\mathcal{P}$ . Likewise, the set of positions of data subcarriers is denoted as  $\mathcal{D}$ . Let us express the Singular Value Decomposition (SVD) of the MIMO channel estimate  $\hat{\mathbf{H}}[k]$  as  $\hat{\mathbf{H}}[k] = \hat{\mathbf{U}}[k]\hat{\mathbf{\Sigma}}[k]\hat{\mathbf{V}}^*[k]$ . For the data phase, information is transmitted using a precoder  $\mathbf{f}[k] \in \mathbb{C}^{N_t \times 1}$ ,  $\mathbf{f}[k] = \begin{bmatrix} \hat{\mathbf{V}}[k] \\ \vdots \\ 1 \end{bmatrix}$ , and combine the received MIMO signal using a combiner  $\mathbf{w}[k] \in \mathbb{C}^{N_r \times 1}$ ,  $\mathbf{w}[k] = \begin{bmatrix} \hat{\mathbf{U}}[k] \\ \vdots \\ 1 \end{bmatrix}$ . Transmitted data in IEEE 802.11ad is partitioned into  $B_{\text{data}}$  OFDM symbols with  $K = 512$  subcarriers, with  $N_{\text{SD}} = 336$

data subcarriers, and  $N_{\text{SP}} = 16$  pilot subcarriers [36].

Let  $\mathbf{s}^{(\ell)}[p] \in \mathbb{C}$  be the transmitted pilot in the  $\ell$ -th data block at subcarrier  $p \in \mathcal{P}$ ,  $\mathbf{s}^{(\ell)}[d] \in \mathbb{C}$  the data constellation symbol at subcarrier  $d \in \mathcal{D}$ ,  $\mathbf{g}[k] \in \mathbb{C}$  be the equivalent channel  $\mathbf{g}[k] = \mathbf{w}^*[k]\mathbf{H}[k]\mathbf{f}[k]$ ,  $\mathbf{g}[k] = \alpha[k]e^{j\phi[k]}$ , and  $\mathbf{n}^{(\ell)}[k] \in \mathbb{C}$  the combined Gaussian noise sample, at subcarrier  $k$ . Then, the  $l$ -th received block at subcarrier  $k$  is given by

$$\mathbf{r}^{(\ell)}[k] = \mathbf{g}[k]\mathbf{s}^{(\ell)}[k] + \mathbf{n}^{(\ell)}[k], \quad k = 0, \dots, K - 1. \quad (2.38)$$

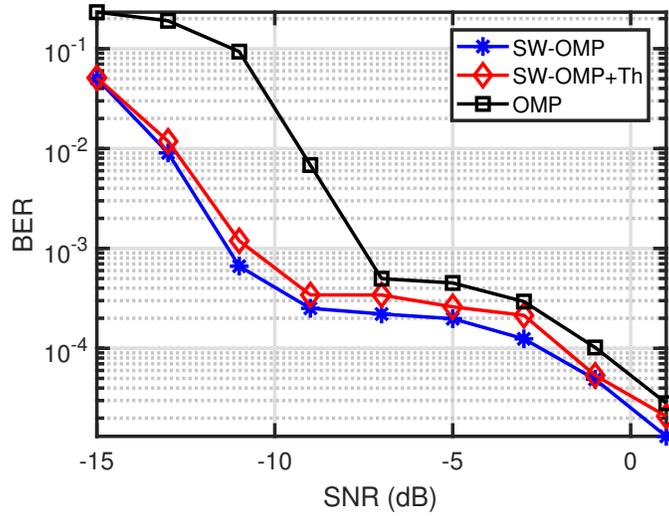
The unknown parameters  $\alpha[k]$ ,  $\phi[k]$  are estimated using the ML criterion, for which the ML estimators are [135]

$$\begin{aligned} \hat{\phi}_{\text{ML}}[p] &= \tan^{-1} \left\{ \frac{\sum_{\ell=1}^{B_{\text{data}}} \text{Im}\{\mathbf{r}^{(\ell)}[p]\bar{\mathbf{s}}^{(\ell)}[k]\}}{\sum_{\ell=1}^{B_{\text{data}}} \text{Re}\{\mathbf{r}^{(\ell)}[p]\bar{\mathbf{s}}^{(\ell)}[k]\}} \right\} \\ \hat{\alpha}_{\text{ML}}[p] &= \frac{1}{B_{\text{data}}} \sum_{\ell=1}^{B_{\text{data}}} \left| \mathbf{r}^{(\ell)}[p]\bar{\mathbf{s}}^{(\ell)}[p]e^{-j\hat{\phi}_{\text{ML}}[p]} \right|. \end{aligned} \quad (2.39)$$

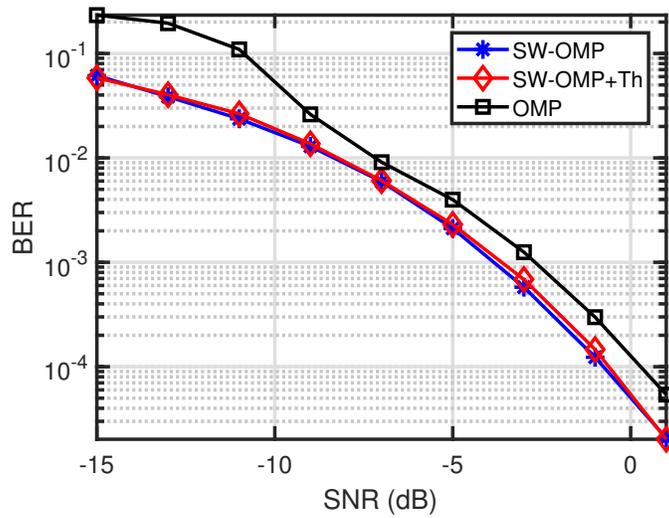
After interpolating the estimates  $\mathbf{g}[k]$  at  $k \in \mathcal{D}$  using  $\mathbf{g}[p]$ ,  $p \in \mathcal{P}$ , the receiver estimates the Log-Likelihood Ratio (LLR) for the coded bits, and the transmitted data bits are estimated using the belief propagation algorithm.

I show in Fig. 2.10 the empirical BER obtained with the proposed algorithms and the baseline approach, as a function of the SNR, for the same simulation parameters as with Fig. 2.7, and with  $K = 512$ ,  $M = 60$ , and  $G_{\text{t}} = G_{\text{r}} = 128$ .

It is observed that the proposed approaches outperform the baseline channel estimation technique for both uncoded and coded BER, and for all



(a)



(b)

Figure 2.10: (a) Evolution of the coded BER versus SNR for the different frequency-domain algorithms. The number of training frames is set to  $M = 60$ . (b) Evolution of the uncoded BER versus SNR for the different frequency-domain algorithms. The number of training frames is set to  $M = 60$ .

values of SNR. At the very low SNR regime, the proposed algorithms show a noticeable performance gain, which reduces as the SNR increases because estimating the channel support becomes easier. There is, however, still a BER performance gain in the medium-high SNR regime, which comes from the use of the noise covariance matrix in the proposed algorithms. This results in more accurate estimates of the channel gains, and also better estimates of the sparsity level.

## 2.5 Conclusions

In this chapter, I developed and evaluated two compressive channel estimation algorithms suitable for MIMO-OFDM-based communication systems operating at the mmWave frequency band. In the proposed estimation strategies, by leveraging both the spatially common sparsity present in frequency-selective mmWave MIMO channels and the statistics of the received signal, high-quality estimates of the large-dimensional MIMO channels can be obtained. I exposed several tradeoffs in the designs of the proposed algorithms, including offline and online computational complexity, achievable estimation error, spectral efficiency and BER performance.

To evaluate the proposed algorithms, I first presented numerical results in a simplified scenario to evaluate the efficiency of the estimates for the common support of the sparse channel vectors. This scenario considered a geometric channel model with a reduced number of multipath components, wherein no space-time clustering is considered. Thereafter, I performed nu-

merical simulations in a more realistic scenario in which channel realizations were taken from the NYUSIM channel simulator, in which space-time clustering is considered and the AoD/AoA do not fall within the quantized angular grids. I compared the achievable NMSE attained by the proposed algorithms to the CRLB for this problem when the channel support is assumed known and showed that the algorithms are asymptotically efficient when there is no grid quantization error and the estimation of the channel support is correct. I also compared the performance of the proposed strategies to prior work, and showed that the proposed approaches significantly outperform previous channel estimation algorithms, in terms of NMSE, spectral efficiency, online computational complexity, and BER performance for both uncoded and coded communication. Finally, I showed that near-optimum spectral efficiency can be attained even when information from only a reduced number of subcarriers is used to compute the channel support, thereby reducing computational complexity with respect to prior estimation algorithms.

## Chapter 3

# Millimeter Wave Compressive Channel Estimation with Carrier Frequency Offset Uncertainties

### 3.1 Introduction

In the previous chapter, I proposed two novel solutions to the problem of channel estimation for frequency-selective hybrid mmWave MIMO systems assuming perfect synchronization at the receiver side. As discussed in Chapter 1, time-frequency synchronization cannot be taken for granted at mmWave, since link configuration occurs in the low SNR regime before beamforming. To gain insight into the impact of synchronization uncertainties on the channel estimation problem, I now turn the focus on the channel estimation problem under CFO synchronization impairments assuming a narrowband channel model.

In this chapter, I propose a multi-stage solution to the problem of mmWave channel estimation when the received signal is impaired by Carrier Frequency Offset (CFO) uncertainties. Different from the proposed work in Chapter 2 and [1], [125], I focus on narrowband channels and turn the attention into the problem of joint channel estimation and frequency synchronization in the low SNR regime. Based on a training protocol similar to that

in Chapter 2, I develop an estimation-theoretic algorithm to infer the CFO, the low-dimensional mmWave channels seen by every receive RF chain for every training frame, and the noise variance. Building upon this estimation approach, I develop an OMP-based estimation algorithm to reconstruct the large-dimensional mmWave MIMO channel, leveraging both the statistics of the unknown parameters, and their corresponding CRLB. Further, I analyze the computational complexity of the proposed technique and show that near-optimal spectral efficiency can be attained at similar computational complexity to that exhibited by a near-optimal compressive channel estimator when there is no CFO impairment. The design target here is to reduce the training overhead required to estimate the mmWave MIMO channel, while ensuring near-optimum channel estimation performance. This work has been published in [2], [137]

### 3.1.1 Prior Work and Motivation

A significant number of papers have proposed solutions to the problem of narrowband channel estimation [14, 17, 53, 56–60, 128], assuming both perfect synchronization at the receiver side and zero-delay response for the phase-shifter network that operates in the analog domain. This is not practical in mmWave WiFi systems, for example, where both synchronization and channel estimation are performed on a burst-by-burst basis.

In the 5G NR standard [37], synchronization and beamforming is also performed on a frame-by-frame basis, such that joint synchronization and

Approach	Computes/ exploits noise covariance	Requires no prior info.	Application to mmWave	Hybrid architecture	Online Complexity	Training Overhead	Communication Performance
Narrowband w/ PN [118]					Low	Medium	Medium
Tensor approach [116]			✓		High	Medium	Low
Channel est. w/ CFO [119]			✓		Low	Medium	Low
Swift-Link [120]		✓	✓		Medium	Medium	Medium
Beam training w/ CFO and PN [112]		✓	✓		Medium	Medium	Medium
<b>Proposed</b>	✓	✓	✓	✓	Low	Medium	High

Figure 3.1: Summary and comparison of prior work on narrowband channel estimation algorithms with impairments as well as the approach proposed in this Chapter.

beamforming needs to be addressed. To the best of my knowledge, only [118], [116], [119], and [120] consider non-ideal synchronization effects. These papers have however several limitations, which are summarized in Fig. 3.1.

In [118], a SISO communications system is considered, in which the PN and the channel are iteratively estimated using a LS-based algorithm. In [116], a joint solution to the problem of MIMO channel and CFO is provided for analog MIMO architectures. The main limitations of [118] are twofold: i) it only considers a SISO communications system, which is not realistic for mmWave communication systems; and ii) the proposed iterative method has been evaluated only in the very high SNR regime, i.e.  $\text{SNR} \geq 10$  dB. The main limitations of [116] are the following: i) it assumes that analog beamformers and combiners can be reconfigured for each transmitted sample. At a sampling rate of 1760 MHz, phase-shifters need around 64 samples to be

updated for a new configuration [36]; and ii) the proposed joint estimation of the tensor containing the channel response and the frequency offset exhibits high complexity, and sets both range and on-grid constraints on the CFO to be estimated.

In [120], a compressive beam alignment algorithm is proposed to estimate the channel under CFO synchronization errors. The main limitations of [120] are: i) similar to [116], it assumes that analog beamformers and combiners can be instantaneously reconfigured; ii) the proposed algorithm is targeted at analog-only MIMO architectures, thereby making it inapplicable to hybrid MIMO systems; iii) the proposed channel estimation algorithm heavily depends on a sparse structure in the MIMO channel, with a small delay tap length; and iv) the CFO correction range of the proposed algorithm is limited by the length of the training sequence and the channel delay tap length, thereby making it difficult to apply the algorithm in practice. Finally, [119] deals with the problem of channel estimation in frequency-unsynchronized mmWave networks.

In [119], analog-only architectures with a single RF chain are assumed, and an autocorrelation-based iterative algorithm is proposed to jointly estimate the CFO and the mmWave channel. The limitations of [119] are: i) similar to [116], it assumes that analog beamformer and combiners can be instantaneously reconfigured for two consecutive transmitted time-domain samples; ii) the algorithm has only been evaluated for mmWave channels having a very reduced number of non-clustered multipath components, which is not

realistic for mmWave communication systems; iii) owing to the nature of the autocorrelation function, the proposed algorithm does not perform well both when the CFO is considerably large and the SNR is low; and iv) the proposed algorithm assumes that the channel's sparsity level is known, and that the AoD/AoA lie on a spatial grid, which does not hold in practice. In summary, the joint CFO and channel estimation techniques in [118], [116], [120], and [119] are developed based on different assumptions that make the proposed algorithms both not robust to low SNR and not applicable to MIMO systems employing hybrid architectures.

### 3.1.2 Contributions

In this chapter, I theoretically analyze and develop a multi-stage solution to the problem of channel estimation with CFO imperfections and hybrid MIMO architectures. The contributions of this chapter are listed hereafter:

- I formulate and find a solution to the problem of CFO and channel estimation accounting for the reconfiguration time of phase-shifters, which sets the notion of *training frame* in the remaining of this chapter.
- For every training frame, I theoretically analyze and find the optimum solution to the problem of estimating the CFO, equivalent channel, and noise variance for an arbitrary hybrid architecture with several RF chains. I theoretically calculate the CRLB for the estimation of each of these parameters and also find the corresponding ML estimators.

- Using estimates of the unknown parameters for every training frame, I formulate the problem of estimating the high-dimensional MIMO channel. I propose an iterative low-complexity algorithm to find the sparse channel coefficients based on OMP. It is noteworthy to mention that, unlike prior work, **the proposed strategy does not require either knowledge of the channel’s sparsity level or the noise variance.**

I assess the performance of the proposed estimators both in terms of estimation error and spectral efficiency. I use both all-digital and hybrid precoders and combiners to show the effectiveness of the proposed approaches. Simulation results obtained from the estimated channel show that the CFO may be estimated with sufficient reliability despite the low SNR before beamforming. Moreover, the proposed CFO synchronization and channel estimation method, along with the designed hybrid precoders and combiners, is shown to offer excellent performance, even though the MIMO channel has a large number of multipath components and the noise variance is considered to be unknown.

### 3.2 System model

I consider a single-user mmWave MIMO communications link in which a transmitter equipped with  $N_t$  antennas sends  $N_s$  data streams to a receiver having  $N_r$  antennas. Both transmitter and receiver are assumed to use fully-connected hybrid MIMO architectures as shown in Fig. 3.2, with  $L_t$  and  $L_r$  RF chains. A hybrid precoder is used, with  $\mathbf{F} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_t \times N_s}$ ,

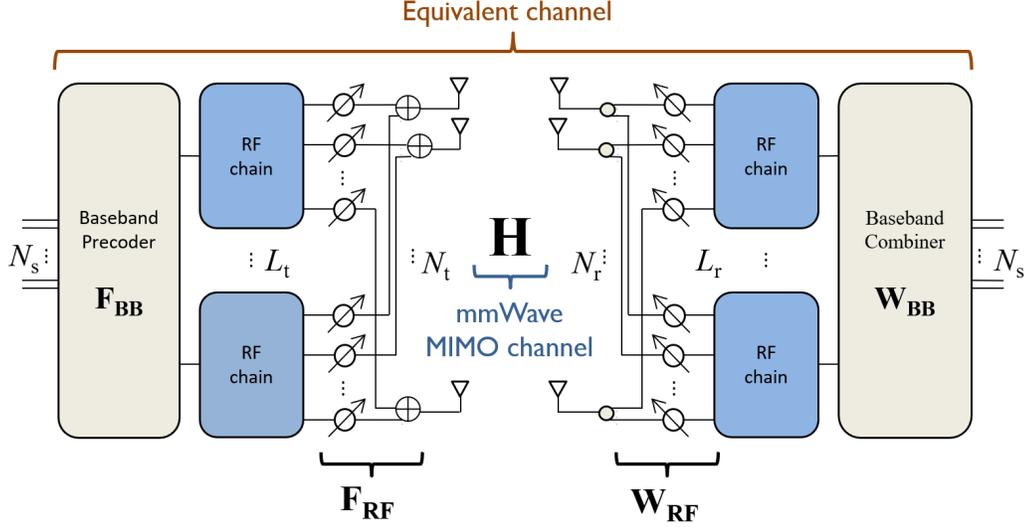


Figure 3.2: Illustration of the structure of a hybrid MIMO architecture, which includes analog and digital precoders and combiners (same as Fig. 1.2).

where  $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times L_t}$  is the analog precoder and  $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{L_t \times N_s}$  is the digital one. The RF precoder and combiner are implemented using a fully-connected network of phase-shifters, as described in [14]. The MIMO channel between the transmitter and the receiver is modeled as an  $N_r \times N_t$  matrix denoted as  $\mathbf{H}$ , which is assumed to be a sum of the contributions of  $C$  spatial clusters, each contributing with  $R_c$  rays,  $c = 1, \dots, C$  [14], [138], [60]. I use  $\rho_L$  to denote the pathloss,  $\alpha_{c,r} \in \mathbb{C}$  is the complex gain of the  $r$ -th ray within the  $c$ -th cluster,  $\phi_{c,r}, \theta_{c,r} \in [0, 2\pi)$  are the angles of arrival and departure (AoA/AoD), and  $\mathbf{a}_{\text{R}}(\phi_{c,r}) \in \mathbb{C}^{N_r \times 1}$  and  $\mathbf{a}_{\text{T}}(\theta_{c,r}) \in \mathbb{C}^{N_t \times 1}$  denote the receive and transmit array steering vectors. Using this notation, the channel matrix is given by

$$\mathbf{H} = \sqrt{\frac{N_r N_t}{\rho_L \sum_{c=1}^C R_c}} \sum_{c=1}^C \sum_{r=1}^{R_c} \alpha_{c,r} \mathbf{a}_{\text{R}}(\phi_{c,r}) \mathbf{a}_{\text{T}}^*(\theta_{c,r}). \quad (3.1)$$

I define  $\mathbf{G} \in \mathbb{C}^{\sum_{c=1}^C R_c \times \sum_{c=1}^C R_c}$  as the matrix containing the complex channel gains,  $\mathbf{G} = \sqrt{\frac{N_r N_t}{\rho \sum_{c=1}^C R_c}} \text{diag}\{\alpha_{1,1}, \dots, \alpha_{1,R_1}, \dots, \alpha_{C,1}, \dots, \alpha_{C,R_C}\}$  and  $\mathbf{A}_R \in \mathbb{C}^{N_r \times \sum_{c=1}^C R_c}$  and  $\mathbf{A}_T \in \mathbb{C}^{N_t \times \sum_{c=1}^C R_c}$  the matrices containing the receive and transmit array steering vectors  $\mathbf{a}_R(\phi_{c,r})$  and  $\mathbf{a}_T(\theta_{c,r})$  in their columns. Using this convention,  $\mathbf{H}$  can be written in a more compact way as

$$\mathbf{H} = \mathbf{A}_R \mathbf{G} \mathbf{A}_T^*. \quad (3.2)$$

Now, the matrix  $\mathbf{H}$  in (3.2) can be approximated using the extended virtual channel representation [8] as

$$\mathbf{H} \approx \tilde{\mathbf{A}}_R \mathbf{G}^v \tilde{\mathbf{A}}_T^*, \quad (3.3)$$

where  $\mathbf{G}^v \in \mathbb{C}^{G_r \times G_t}$  is a sparse matrix which contains the path gains of the quantized spatial frequencies in the non-zero elements, and the dictionary matrices  $\tilde{\mathbf{A}}_T$  and  $\tilde{\mathbf{A}}_R$  contain the transmit and receive array response vectors evaluated on grids of sizes  $G_t$  and  $G_r$ . Assuming that the receiver applies a hybrid combiner  $\mathbf{W} = \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_r \times L_r}$ , with  $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times L_r}$  being the analog combiner, and  $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{L_r \times N_s}$  the baseband combiner, the received signal at discrete time instant  $n$  can be written as

$$\mathbf{y}[n] = \mathbf{W}^* \mathbf{H} \mathbf{f} \mathbf{s}[n] e^{j2\pi \Delta f n} + \mathbf{v}[n], \quad (3.4)$$

for  $n = 0, \dots, N - 1$ . The signal  $\mathbf{s}[n] \in \mathbb{C}^{N_s \times 1}$  is a training sequence known to the receiver during the channel estimation phase,  $0 \leq n \leq N - 1$ ,  $\Delta f$  is the unknown CFO normalized to the sampling frequency  $f_s = 1/T_s$ , and

$\mathbf{v}[n] \sim \mathcal{CN}(0, \sigma^2 \mathbf{W}^* \mathbf{W})$  is the circularly symmetric complex Gaussian distributed additive noise vector.

Similar to prior work in [14], I consider a training protocol in which the transmitter sends  $M$  consecutive training frames to the receiver in order to perform both CFO synchronization and compressive estimation of the MIMO channel. The transmitted signal corresponding to the  $m$ -th frame,  $1 \leq m \leq M$ , is given by

$$\mathbf{x}^{(m)}[n] = \mathbf{F}_{\text{RF}}^{(m)} \mathbf{F}_{\text{BB}}^{(m)} \mathbf{s}^{(m)}[n]. \quad (3.5)$$

I design now the training signal as  $\mathbf{s}^{(m)}[n] = \tilde{\mathbf{q}}^{(m)} s^{(m)}[n]$ , with  $\tilde{\mathbf{q}}^{(m)} \in \mathbb{C}^{N_s \times 1}$  a frequency-flat spatial modulation vector, static for every frame, and  $s^{(m)}[n] \in \mathbb{C}$  a scalar time-domain training sequence. With this choice for  $\tilde{\mathbf{q}}^{(m)}$ , I will show in Section 3.3 that *the information the receiver is given only depends on the energy of  $\mathbf{s}^{(m)}[n]$ , but not on its particular design*. Now, letting  $\mathbf{q} = \mathbf{F}_{\text{BB}} \tilde{\mathbf{q}} \in \mathbb{C}^{L_t \times 1}$ , which operates as an equivalent baseband precoder for this particular design of the training sequence, the signal in (3.5) can be expressed as

$$\mathbf{x}^{(m)}[n] = \mathbf{F}_{\text{RF}}^{(m)} \mathbf{q}^{(m)} s^{(m)}[n], \quad (3.6)$$

In this chapter, I choose to design  $\mathbf{q}^{(m)}$  such that its entries are independent and identically distributed energy-normalized QPSK symbols. The analog training precoder  $\mathbf{F}_{\text{RF}}^{(m)}$ , however, is designed to be pseudorandomly built such that every entry  $[\mathbf{F}_{\text{RF}}^{(m)}]_{i,j} \in \left\{ e^{j \frac{2\pi k}{N_{\text{T},\text{Q}}}} \right\}_{k=0}^{N_{\text{T},\text{Q}}-1}$  with uniform probability, and  $N_{\text{T},\text{Q}} = 2^{B_{\text{T},\text{Q}}}$  is the number of possible states for an arbitrary phase-shifter, which depends on the number of bits  $B_{\text{T},\text{Q}}$ .

As to the hybrid receive combiner  $\mathbf{W}^{(m)} = \mathbf{W}_{\text{RF}}^{(m)} \mathbf{W}_{\text{BB}}^{(m)}$ , I propose to design the analog combiner such that its entries follow the same distribution as the analog precoder, although the number of quantization bits  $B_{\text{R,Q}}$  may be different in general, and so would the number of states  $N_{\text{R,Q}} = 2^{B_{\text{R,Q}}}$  be as well. The design of the baseband combiner is motivated by both the Slepian-Bangs formula [135] and the ML principle. Using an analog-only combiner  $\mathbf{W}_{\text{RF}}^{(m)}$  would result in a receive noise covariance matrix  $\mathbf{C}^{(m)} = \sigma^2 \mathbf{C}_{\text{w}}^{(m)} = \sigma^2 \mathbf{W}_{\text{RF}}^{(m)*} \mathbf{W}_{\text{RF}}^{(m)}$ , which is not diagonal. This entails an additional difficulty when estimating different components in a vector under spatially correlated noise. In fact, the optimum estimator would need to estimate all the components in that vector at the same time, for which it may be difficult to find a closed-form solution. To overcome this difficulty, I propose to set the baseband combiner such that the post-combining noise is spatially white. Letting the Cholesky decomposition of  $\mathbf{C}_{\text{w}}^{(m)}$  be  $\mathbf{C}_{\text{w}}^{(m)} = \mathbf{D}_{\text{w}}^{(m)*} \mathbf{D}_{\text{w}}^{(m)}$ , with  $\mathbf{D}_{\text{w}}^{(m)} \in \mathbb{C}^{L_r \times L_r}$  an upper triangular matrix, then the ML philosophy establishes that the optimum baseband combiner is given by  $\mathbf{W}_{\text{BB}}^{(m)} = \mathbf{D}_{\text{w}}^{(m)-1}$ , such that  $\mathbf{W}_{\text{BB}}^{(m)*} \mathbf{W}_{\text{RF}}^{(m)*} \mathbf{W}_{\text{RF}}^{(m)} \mathbf{W}_{\text{BB}}^{(m)} = \mathbf{I}_{L_r}$ .

The MIMO signal observed at the  $N_r$  received antennas is impaired by both temporally and spatially white circularly symmetric Additive White Gaussian Noise (AWGN) [139]. After processing this received signal using a linear combiner, the post-combining noise vector is Gaussian owing to linear operators preserving Gaussianity. Accordingly, for the  $m$ -th transmitted

frame, the received signal in (4.2) can be expressed as

$$\mathbf{y}^{(m)}[n] = \underbrace{\mathbf{D}_w^{(m)*} \mathbf{W}_{\text{RF}}^{(m)*} \mathbf{H}\mathbf{F}_{\text{RF}}^{(m)} \mathbf{q}^{(m)}}_{\tilde{\boldsymbol{\alpha}}^{(m)}} s^{(m)}[n] e^{j2\pi\Delta f^{(m)}n} + \mathbf{v}^{(m)}[n], \quad (3.7)$$

with  $\mathbf{v}^{(m)}[n] \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{L_r})$ , and  $\tilde{\boldsymbol{\alpha}}^{(m)} = [\alpha_1^{(m)} e^{j\beta_1^{(m)}}, \dots, \alpha_{L_r}^{(m)} e^{j\beta_{L_r}^{(m)}}]^T$  is the complex equivalent beamformed channel for the  $m$ -th training step. Therefore, the interest in this chapter lies on estimating the vector of parameters  $\boldsymbol{\xi}^{(m)} \triangleq [\alpha_1^{(m)}, \dots, \alpha_{L_r}^{(m)}, \beta_1^{(m)}, \dots, \beta_{L_r}^{(m)}, \Delta f^{(m)}, \sigma^2]^T$ , prior to performing compressive MIMO channel estimation. In (3.7), the CFO is considered to be time-varying and may change for every training frame. This is justified by the instability of high-frequency oscillators at mmWave frequencies. In [140], it was shown that, for a Field-Effect Transistor (FET)-based Voltage-Controlled Oscillator (VCO) at 60 GHz, small variations in temperature may result in changes of the gate-source controlling voltage, which bring about significant changes in the VCO's local frequency. For this reason, the CFO is considered to be time-varying. Furthermore, it is well-known that phase-shifters have a reconfiguration time when the different weights are adjusted. The signaling needed to do this reconfiguration may lead to changes in temperature, which would also lead to changes in the local frequency of the VCO. Owing to such changes, it is reasonable to consider a constant CFO within a given training frame, and changes in the CFO are considered on a frame-by-frame basis.

Besides the practical assumptions adopted in the proposed framework, there are some assumptions that need be taken into consideration for the

proposed solution to be applicable. These assumptions are listed and properly justified hereinafter:

- The system bandwidth is small in comparison with the carrier frequency. More specifically, if the channel's coherence bandwidth is larger than the bandwidth of the transmitted signal, then the narrowband channel model assumption holds. Practical channels are not exactly frequency-flat, but this is a reasonable assumption if the channel's delay spread is sufficiently small in comparison with the time duration of the transmitted signal. This is not necessarily a limitation of the proposed framework, as the transmitted bandwidth is not large in comparison with the carrier frequency. Further, the time duration of the transmitted signal is likely to be longer than the channel's delay spread under Line-of-Sight (LoS) propagation, which holds with high probability under the 5G New Radio (NR) channel model when the transmit-receive distance is sufficiently small [6].
- Both integer and fractional timing offsets (with respect to the sampling period  $T_s$ ) are perfectly known and have already been compensated for. In other words, both symbol and frame synchronization have already been performed. This is a practical assumption if the number of receive RF chains  $L_r \geq 2$  and the SNR is not too low [141]. In [141], probability of perfect detection (time synchronization) is close to 1 even at SNR < 0 dB, which is the expected scenario in upcoming mmWave deployments.

- At the receiver side, there is no interference coming from other users or base stations. This is a reasonable assumption for indoor and certain outdoor channels given the high oxygen absorption present at 60 GHz, besides the already existent small antenna aperture, which requires the use of directional beamforming for communications. This fact also contributes to further reducing interference, thereby making the proposed signal model applicable.

Regarding computational complexity, the proposed framework is based on a frame-wise estimation of the different unknown parameters. Therefore, computational complexity is reduced with respect to the case in which the signals received from the different training frames were processed all at once. It is the asymptotic invariance property of ML estimators that allows keeping computational complexity at a reasonable level without compromising optimality guarantees, as it will be shown in the numerical results section.

### **3.3 Theoretical analysis of the estimation problem**

In this section, I present the theoretical analysis for the problem of estimating the low-dimensional beamformed channels, the CFO and the noise variance. More specifically, I theoretically calculate the FIM and the CRLB associated to the unknown parameters. Further, I also provide an analysis on the asymptotic behavior of the CRLB for the different parameters to gain additional insight into how the system parameters affect estimation performance.

Since the signal model in (3.7) defines a Gaussian distribution on  $\mathbf{y}^{(m)}[n]$ , the regularity condition for the CRLB to exist is satisfied. Let  $\mathbf{\Omega}_n \in \mathbb{C}^{L_r \times L_r}$  be a diagonal matrix given by  $\mathbf{\Omega}_n \triangleq e^{j2\pi\Delta f^{(m)}n} \mathbf{I}_{L_r}$ . Then, stacking the different  $N$  samples of the received signal in (3.7) yields

$$\underbrace{\begin{bmatrix} \mathbf{y}^{(m)}[0] \\ \vdots \\ \mathbf{y}^{(m)}[N-1] \end{bmatrix}}_{\mathbf{y}^{(m)}} = \underbrace{\begin{bmatrix} \mathbf{\Omega}_0^{(m)} \tilde{\boldsymbol{\alpha}}^{(m)} & & \\ & \ddots & \\ & & \mathbf{\Omega}_{N-1}^{(m)} \tilde{\boldsymbol{\alpha}}^{(m)} \end{bmatrix}}_{\mathbf{T}(\boldsymbol{\xi}^{(m)})} \times \underbrace{\begin{bmatrix} s^{(m)}[0] \\ \vdots \\ s^{(m)}[N-1] \end{bmatrix}}_{\mathbf{s}^{(m)}} + \underbrace{\begin{bmatrix} \mathbf{v}^{(m)}[0] \\ \vdots \\ \mathbf{v}^{(m)}[N-1] \end{bmatrix}}_{\mathbf{v}^{(m)}}, \quad (3.8)$$

such that  $\mathbf{y}^{(m)}$  is distributed according to  $\mathbf{y}^{(m)} \sim \mathcal{CN}(\mathbf{T}(\boldsymbol{\xi}^{(m)}) \mathbf{s}^{(m)}, \sigma^2 \mathbf{I}_{L_r})$ , where  $\mathbf{T}(\boldsymbol{\xi}^{(m)}) \in \mathbb{C}^{NL_r \times N}$  is the transfer matrix that relates the transmitted training signal  $\mathbf{s}^{(m)}$  with  $\mathbf{y}^{(m)}$ .

Let  $\mathbf{I}(\boldsymbol{\xi}^{(m)}) \in \mathbb{C}^{2(L_r+1) \times 2(L_r+1)}$  denote the FIM for the estimation of the unknown vector of parameters  $\boldsymbol{\xi}^{(m)}$ . Let the mean of  $\mathbf{y}^{(m)}$  be denoted by  $\boldsymbol{\mu}(\boldsymbol{\xi}^{(m)}) = \mathbf{T}(\boldsymbol{\xi}^{(m)}) \mathbf{s}^{(m)}$ . The  $\{i, j\}$ -th element in  $\mathbf{I}(\boldsymbol{\xi}^{(m)})$  is given by [135]

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\xi}^{(m)})]_{i,j} &= \frac{2}{\sigma^2} \operatorname{Re} \left\{ \frac{\partial \boldsymbol{\mu}(\boldsymbol{\xi}^{(m)})}{\partial \xi_i^{(m)}} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\xi}^{(m)})}{\partial \xi_j^{(m)}} \right\} \\ &+ \frac{1}{\sigma^4} \operatorname{trace} \left\{ \frac{\partial \mathbf{C}(\boldsymbol{\xi}^{(m)})}{\partial \xi_i^{(m)}} \frac{\partial \mathbf{C}(\boldsymbol{\xi}^{(m)})}{\partial \xi_j^{(m)}} \right\}. \end{aligned} \quad (3.9)$$

The diagonal element in the FIM corresponding to  $\alpha_i^{(m)}$  can be found

as

$$\mathbf{I}_{\alpha_i^{(m)}, \alpha_i^{(m)}}(\boldsymbol{\xi}^{(m)}) = \frac{2}{\sigma^2} \operatorname{Re} \left\{ \mathbf{s}^{(m)*} \frac{\partial \mathbf{T}^*(\boldsymbol{\xi}^{(m)})}{\partial \alpha_i^{(m)}} \frac{\partial \mathbf{T}(\boldsymbol{\xi}^{(m)})}{\partial \alpha_i^{(m)}} \mathbf{s}^{(m)} \right\}. \quad (3.10)$$

The term  $\frac{\partial \mathbf{T}(\boldsymbol{\xi}^{(m)})}{\partial \alpha_i^{(m)}}$  is given by

$$\frac{\partial \mathbf{T}(\boldsymbol{\xi}^{(m)})}{\partial \alpha_i^{(m)}} = \bigoplus_{n=0}^{N-1} \Omega_n \mathbf{E}_{ii} e^{j\angle\{\tilde{\boldsymbol{\alpha}}^{(m)}\}}, \quad (3.11)$$

where the  $L_r \times L_r$  matrix  $\mathbf{E}_{ii}$  is given by  $[\mathbf{E}_{ii}]_{k,l} = \delta[i-k]\delta[i-l]$ . Then, the term  $\mathbf{I}_{\alpha_i^{(m)}, \alpha_i^{(m)}}(\boldsymbol{\xi}^{(m)})$  in (3.10) yields

$$\begin{aligned} \mathbf{I}_{\alpha_i^{(m)}, \alpha_i^{(m)}}(\boldsymbol{\xi}^{(m)}) &= \frac{2}{\sigma^2} \sum_{n=0}^{N-1} \operatorname{Re} \left\{ |s^{(m)}[n]|^2 e^{-j\angle\{\tilde{\boldsymbol{\alpha}}^{(m)}\}} \mathbf{E}_{ii}^2 e^{j\angle\{\tilde{\boldsymbol{\alpha}}^{(m)}\}} \right\} \\ &= \frac{2}{\sigma^2} \sum_{n=0}^{N-1} |s^{(m)}[n]|^2. \end{aligned} \quad (3.12)$$

For the phase offset parameter  $\beta_i^{(m)}$ , the diagonal FIM element is given by

$$\mathbf{I}_{\beta_i^{(m)}, \beta_i^{(m)}}(\boldsymbol{\xi}^{(m)}) = \frac{2}{\sigma^2} \operatorname{Re} \left\{ \mathbf{s}^{(m)*} \frac{\partial \mathbf{T}^*(\boldsymbol{\xi}^{(m)})}{\partial \beta_i^{(m)}} \frac{\partial \mathbf{T}(\boldsymbol{\xi}^{(m)})}{\partial \beta_i^{(m)}} \mathbf{s}^{(m)} \right\}, \quad (3.13)$$

where the matrix partial derivative with respect to  $\beta_i^{(m)}$  is given by

$$\frac{\partial \mathbf{T}(\boldsymbol{\xi}^{(m)})}{\partial \beta_i^{(m)}} = \bigoplus_{n=0}^{N-1} \Omega_n j \mathbf{E}_{ii} \tilde{\boldsymbol{\alpha}}^{(m)}. \quad (3.14)$$

Then, the term  $\mathbf{I}_{\beta_i^{(m)}, \beta_i^{(m)}}(\boldsymbol{\xi}^{(m)})$  reads as

$$\begin{aligned} \mathbf{I}_{\beta_i^{(m)}, \beta_i^{(m)}}(\boldsymbol{\xi}^{(m)}) &= \frac{2}{\sigma^2} \sum_{n=0}^{N-1} \operatorname{Re} \left\{ |s^{(m)}[n]|^2 \tilde{\boldsymbol{\alpha}}^{(m)*} \mathbf{E}_{ii}^2 \tilde{\boldsymbol{\alpha}}^{(m)} \right\} \\ &= \frac{2N\alpha_i^2}{\sigma^2}. \end{aligned} \quad (3.15)$$

For the carrier frequency offset  $\Delta f^{(m)}$ , the diagonal element in the FIM is given by

$$\mathbf{I}_{\Delta f^{(m)}, \Delta f^{(m)}} = \frac{2}{\sigma^2} \operatorname{Re} \left\{ \mathbf{s}^{(m)*} \frac{\partial \mathbf{T}^* (\boldsymbol{\xi}^{(m)})}{\partial \Delta f^{(m)}} \frac{\partial \mathbf{T} (\boldsymbol{\xi}^{(m)})}{\partial \Delta f^{(m)}} \mathbf{s}^{(m)} \right\}. \quad (3.16)$$

The derivative  $\frac{\partial \mathbf{T}(\boldsymbol{\xi}^{(m)})}{\partial \Delta f^{(m)}}$  is given by

$$\frac{\partial \mathbf{T} (\boldsymbol{\xi}^{(m)})}{\partial \Delta f^{(m)}} = \bigoplus_{n=0}^{N-1} (2\pi n)^2 \boldsymbol{\Omega}_n \tilde{\boldsymbol{\alpha}}^{(m)} \quad (3.17)$$

such that (3.16) yields

$$\begin{aligned} \mathbf{I}_{\Delta f^{(m)}, \Delta f^{(m)}} &= \frac{2}{\sigma^2} \sum_{n=0}^{N-1} \operatorname{Re} \left\{ |s^{(m)C}[n]|^2 (2\pi n)^2 \|\tilde{\boldsymbol{\alpha}}^{(m)*}\|_2^2 \right\} \\ &= \frac{2 \operatorname{trace}\{\mathbf{P}^{(m)}\}}{\sigma^2} \sum_{n=0}^{N-1} |s^{(m)C}[n]|^2 (2\pi n)^2, \end{aligned} \quad (3.18)$$

with  $\mathbf{P}^{(m)} \in \mathbb{R}^{L_r \times L_r}$  a diagonal matrix containing the energy received at the different RF chains,  $\mathbf{P}^{(m)} \triangleq \operatorname{diag}\{\alpha_1^2, \dots, \alpha_{L_r}^2\}$ . Finally, the diagonal element corresponding to the noise variance  $\sigma^2$  is given by

$$\mathbf{I}_{\sigma^2, \sigma^2} (\boldsymbol{\xi}^{(m)}) = \frac{NL_r}{\sigma^4}. \quad (3.19)$$

Regarding the off-diagonal terms in the FIM, they can be checked to be zero-valued except for  $\mathbf{I}_{\Delta f^{(m)}, \beta_i^{(m)}} (\boldsymbol{\xi}^{(m)})$ , which is given by

$$\mathbf{I}_{\Delta f^{(m)}, \beta_i^{(m)}} (\boldsymbol{\xi}^{(m)}) = \frac{2}{\sigma^2} \operatorname{Re} \left\{ \mathbf{s}^{(m)*} \frac{\partial \mathbf{T}^* (\boldsymbol{\xi}^{(m)})}{\partial \Delta f^{(m)}} \frac{\partial \mathbf{T} (\boldsymbol{\xi}^{(m)})}{\partial \beta_i^{(m)}} \mathbf{s}^{(m)} \right\}. \quad (3.20)$$

The inner product  $\frac{\partial \mathbf{T}^* (\boldsymbol{\xi}^{(m)})}{\partial \Delta f^{(m)}} \frac{\partial \mathbf{T} (\boldsymbol{\xi}^{(m)})}{\partial \beta_i^{(m)}}$  can be written as

$$\frac{\partial \mathbf{T}^* (\boldsymbol{\xi}^{(m)})}{\partial \Delta f^{(m)}} \frac{\partial \mathbf{T} (\boldsymbol{\xi}^{(m)})}{\partial \beta_i^{(m)}} = \alpha_i^{(m)*} e^{j\beta_i^{(m)}} \bigoplus_{n=0}^{N-1} (2\pi n) \tilde{\boldsymbol{\alpha}}^{(m)*} \mathbf{e}_i \quad (3.21)$$

where  $\mathbf{e}_i \in \mathbb{R}^{L_r \times 1}$  is the  $i$ -th element of the canonical basis,  $[\mathbf{e}_i]_j = \delta[i - j]$ .

Thereby, the term in (3.20) can be expressed as

$$\mathbf{I}_{\Delta f^{(m)}, \beta_i^{(m)}}(\boldsymbol{\xi}^{(m)}) = \frac{2}{\sigma^2} \sum_{n=0}^{N-1} |s^{(m)}[n]|^2 (2\pi n) \alpha_i^2. \quad (3.22)$$

Finally, the total FIM is found to be

$$\mathbf{I}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_1(\boldsymbol{\xi}^{(m)}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (3.23)$$

where  $\mathbf{I}_1(\boldsymbol{\xi}^{(m)}) \in \mathbb{C}^{(L_r+1) \times (L_r+1)}$ ,  $\mathbf{I}_2(\boldsymbol{\xi}^{(m)}) \in \mathbb{C}^{(L_r+1) \times (L_r+1)}$  are given by

$$\mathbf{I}_1(\boldsymbol{\xi}^{(m)}) = \text{diag} \left\{ \frac{2N}{\sigma^2} \mathbf{1}_{L_r}^T, \frac{L_r N}{\sigma^4} \right\} \quad (3.24)$$

and

$$\mathbf{I}_2(\boldsymbol{\xi}^{(m)}) = \frac{2}{\sigma^2} \begin{bmatrix} \text{trace}\{\mathbf{P}^{(m)}\} \sum_{n=0}^{N-1} (2\pi n)^2 & \mathbf{1}_{L_r}^T \mathbf{P}^{(m)} \sum_{n=0}^{N-1} 2\pi n \\ \mathbf{P}^{(m)} \mathbf{1}_{L_r} \sum_{n=0}^{N-1} 2\pi n & N \mathbf{P}^{(m)} \end{bmatrix}. \quad (3.25)$$

Upon inverting the block diagonal matrix  $\mathbf{I}(\boldsymbol{\xi}^{(m)})$ , the CRLB for the parameters is given by

$$\text{var} \left\{ \hat{\alpha}_i^{(m)} \right\} \geq \frac{\sigma^2}{N} \quad (3.26)$$

$$\text{var} \left\{ \hat{\sigma}^2 \right\} \geq \frac{\sigma^4}{N L_r} \quad (3.27)$$

$$\text{var} \left\{ \widehat{\Delta f}^{(m)} \right\} \geq \frac{6\sigma^2}{\text{trace}\{\mathbf{P}^{(m)}\} (2\pi)^2 N (N^2 - 1)} \quad (3.28)$$

$$\text{var} \left\{ \hat{\beta}_i^{(m)} \right\} \geq \frac{\sigma^2 \left( N^2 (N^2 - 1) / 12 + \frac{N^2 (N-1)^2 \alpha_i^{(m)2}}{4 \text{trace}\{\mathbf{P}^{(m)}\}} \right)}{2\alpha_i^{(m)2} N^3 (N^2 - 1) / 12}. \quad (3.29)$$

The proof concerning the derivation of these bounds can be found in Appendix A. Although the expressions in (3.28) and (3.29) may seem difficult to analyze,

we can notice that: i) the denominator in (3.28) is a third order polynomial in  $N$ . Therefore, for  $N$  large, (3.28) behaves according to

$$\text{var} \left\{ \widehat{\Delta f}^{(m)} \right\} \propto \frac{3\sigma^2}{2\pi^2 \text{trace} \{ \mathbf{P}^{(m)} \} N^3}. \quad (3.30)$$

By direct inspection of (3.30), it is observed that the CRLB decreases with  $N^3$  and with the summation of SNR measured at the output of the different RF chains,  $\text{SNR}_{\text{total}} = \text{trace} \{ \mathbf{P}^{(m)} \} / \sigma^2$ . Accordingly, better estimates for  $\Delta f^{(m)}$  can be found as the number of RF chains increases. Consequently, mmWave transceivers with larger antenna arrays are capable of achieving better estimates for this parameter. The terms in (3.29), however, are somewhat harder to analyze. The numerator is a fourth-order polynomial in  $N$ , whereas the denominator is a fifth-order polynomial. Then, for large values of  $N$ , (3.29) may be expressed as

$$\text{var} \left\{ \hat{\beta}_i^{(m)} \right\} \propto \frac{3\sigma^2}{2N \text{trace} \{ \mathbf{P}^{(m)} \}}. \quad (3.31)$$

As to (3.31), however, it is noticed that as  $\text{SNR}_{\text{total}}$  increases, the estimation of this parameter becomes easier, since the CFO can be more accurately estimated and the phase offset can be thereby more successfully dealt with. Furthermore, the SNR at the output of every RF chain,  $\gamma_i^{(m)} = \frac{\alpha_i^{(m)2}}{\sigma^2}$ , can also be estimated, for which the CRLB can be found using the *Transformed Parameters Property* [135] as

$$\text{var} \left\{ \hat{\gamma}_i^{(m)} \right\} \geq \frac{4\gamma_i^{(m)}}{N} + \frac{\gamma_i^{(m)2}}{N}. \quad (3.32)$$

The expression in (3.32) provides a clear insight into how the estimation of  $\gamma_i^{(m)}$  is influenced by the length of the training sequence  $t[n]$  and the true SNR

$\gamma_i^{(m)}$ . For small values of  $\gamma_i^{(m)}$ , the CRLB is dominated by the term growing linearly with this parameter, whereas for larger values the term that evolves quadratically with  $\gamma_i^{(m)}$  dominates the achievable variance for the estimation of  $\gamma_i^{(m)}$ , thereby hardening the estimation of the SNR.

### 3.4 Estimation of beamformed channels and high-dimensional MIMO channel

In this section, I formulate and present a solution to the problem of estimating both the CFO, the equivalent channel given by the joint effect of the hybrid precoder, channel matrix, and hybrid combiner, and the noise variance in (4.2). Then, I formulate the problem of estimating the high-dimensional mmWave MIMO channel  $\mathbf{H}$  from the estimates of the equivalent channel accounting for the correlated received signal, the statistical distribution of the ML estimators, and the lack of prior knowledge on both the channel sparsity or noise variance. The estimation approach adopted in this chapter is shown in Fig. 3.3.

First, the ML estimators for the different parameters in  $\boldsymbol{\xi}^{(m)}$  need to be found. The problem of finding these statistics can be formalized as

$$\left\{ \left\{ \hat{\alpha}_{i,\text{ML}}^{(m)} \right\}_{i=1}^{L_r}, \left\{ \hat{\beta}_{i,\text{ML}}^{(m)} \right\}_{i=1}^{L_r}, \widehat{\Delta f}_{\text{ML}}^{(m)}, \widehat{\sigma}_{\text{ML}}^2 \right\} = \arg \max_{\left\{ \alpha_i^{(m)} \right\}_{i=1}^{L_r}, \left\{ \beta_i^{(m)} \right\}_{i=1}^{L_r}, \Delta f^{(m)}, \sigma^2} \log p(\mathbf{y}^{(m)}; \boldsymbol{\xi}^{(m)}), \quad (3.33)$$

with  $p(\mathbf{y}^{(m)}; \boldsymbol{\xi}^{(m)})$  the Probability Density Function (PDF) of  $\mathbf{y}^{(m)}$  in (3.8)

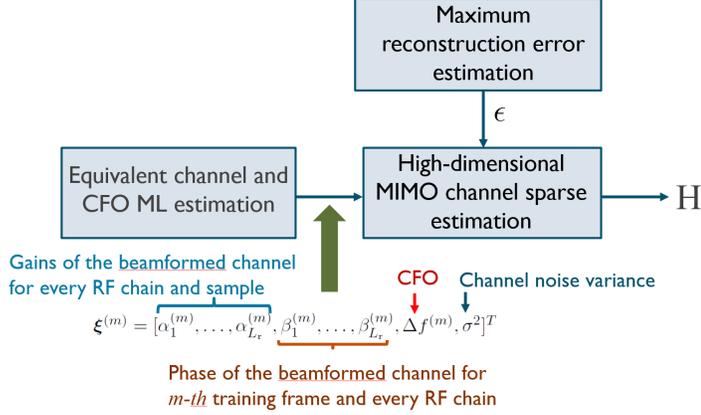


Figure 3.3: Block diagram of the different components in the proposed estimation framework.

parameterized by  $\boldsymbol{\xi}^{(m)}$ . The LLF is given by

$$\begin{aligned} \log p(\mathbf{y}^{(m)}; \boldsymbol{\xi}^{(m)}) = & -NL_r \log \pi \sigma^2 - \frac{1}{\sigma^2} \left( \mathbf{y}^{(m)*} \mathbf{y}^{(m)} \right. \\ & \left. - 2 \operatorname{Re} \{ \mathbf{s}^{(m)*} \mathbf{T}^* (\boldsymbol{\xi}^{(m)}) \mathbf{y}^{(m)} \} + N \operatorname{trace} \{ \mathbf{P}^{(m)} \} \right). \end{aligned} \quad (3.34)$$

It is clear that, in general, each variable in  $\boldsymbol{\xi}^{(m)}$  may be coupled with the others. For this reason, this problem is solved in four different steps by splitting the original optimization problem in (3.33) into four connected separate problems. This can be formalized as

$$\begin{aligned} & \left\{ \{ \hat{\alpha}_{i,\text{ML}}^{(m)} \}_{i=1}^{L_r}, \{ \hat{\beta}_{i,\text{ML}}^{(m)} \}_{i=1}^{L_r}, \hat{\Delta f}_{\text{ML}}^{(m)}, \hat{\sigma}_{\text{ML}}^2 \right\} = \\ & \arg \max_{\sigma^2} \max_{\Delta f^{(m)}} \max_{\{ \alpha_i^{(m)} \}_{i=1}^{L_r}} \max_{\{ \beta_i^{(m)} \}_{i=1}^{L_r}} \log p(\mathbf{y}^{(m)}; \boldsymbol{\xi}^{(m)}). \end{aligned} \quad (3.35)$$

The optimization problem in (3.35) consists of the maximization of a multivariate function with respect to  $2(L_r+1)$  parameters. Since a multivariate complex Gaussian probability density function is increasing for support values below

the mean, and decreasing for support values above the mean, it follows that  $\mathcal{L}(\boldsymbol{\xi}^{(m)}) = \log\{p(\mathbf{y}^{(m)}; \boldsymbol{\xi}^{(m)})\}$  is quasi-concave (log-concave). Thereby, searching for the maximum of (3.35) along each component of  $\boldsymbol{\xi}^{(m)}$  as a function of the other components leads to finding the global maximizer of (3.33). Consequently, (3.33) and (3.35) will produce the same optimizer, and optimality is preserved. Accordingly, first the statistics  $\{\hat{\beta}_{i,\text{ML}}^{(m)}\}_{i=1}^{L_r}$  will be found. Then, they can be merged into (3.35) to find the statistics  $\{\hat{\alpha}_{i,\text{ML}}^{(m)}\}_{i=1}^{L_r}$ . Thereafter, the statistic  $\widehat{\Delta f}_{\text{ML}}^{(m)}$  can be found and, finally,  $\widehat{\sigma}_{\text{ML}}^2$ .

To find  $\{\hat{\beta}_{i,\text{ML}}^{(m)}\}_{i=1}^{L_r}$ , notice that the only term in (3.34) that depends on  $\{\beta_i^{(m)}\}_{i=1}^{L_r}$  is the second scalar inside brackets. Therefore, the optimum  $\{\hat{\beta}_{i,\text{ML}}^{(m)}\}_{i=1}^{L_r}$  are given by

$$\{\hat{\beta}_{i,\text{ML}}^{(m)}\}_{i=1}^{L_r} = \arg \max_{\{\beta_i^{(m)}\}_{i=1}^{L_r}} \text{Re} \left\{ \mathbf{s}^{(m)*} \mathbf{T}^* (\boldsymbol{\xi}^{(m)}) \mathbf{y}^{(m)} \right\}. \quad (3.36)$$

The first derivative of the cost function in (3.36) is given by

$$\begin{aligned} \frac{\partial \log p(\mathbf{y}^{(m)}, \boldsymbol{\xi}^{(m)})}{\partial \beta_i^{(m)}} &= \sum_{n=0}^{N-1} \text{Re} \left\{ s^{(m)\text{C}}[n] \alpha_i^{(m)} (-j) e^{-j\beta_i^{(m)}} \mathbf{e}_i^T \boldsymbol{\Omega}_n \mathbf{y}^{(m)}[n] \right\} \\ &= \sum_{n=0}^{N-1} \text{Im} \left\{ s^{(m)\text{C}}[n] \alpha_i^{(m)} e^{-j\beta_i^{(m)}} e^{-j2\pi\Delta f^{(m)}n} y_i^{(m)}[n] \right\}. \end{aligned} \quad (3.37)$$

Setting the derivative in (3.37) to zero allows obtaining that the optimum estimator for the phase-offset is given by

$$\hat{\beta}_{i,\text{ML}}^{(m)} = \tan^{-1} \left\{ \frac{\sum_{n=0}^{N-1} \text{Im} \left\{ s^{(m)\text{C}}[n] y_i^{(m)}[n] e^{-j2\pi\Delta f^{(m)}n} \right\}}{\sum_{n=0}^{N-1} \text{Re} \left\{ s^{(m)\text{C}}[n] y_i^{(m)}[n] e^{-j2\pi\Delta f^{(m)}n} \right\}} \right\}. \quad (3.38)$$

The statistic in (3.38) can be interpreted as a matched-filter with the sequence  $s^{(m)}[n]$  after compensating for the carrier frequency offset  $\Delta f^{(m)}$ . It is important to notice that the optimum estimator for  $\beta_i^{(m)}$  requires knowledge of the true frequency offset  $\Delta f^{(m)}$ . Since it is impossible to know its value exactly, the optimum estimator for  $\Delta f^{(m)}$  is to be plugged in (3.38) for the estimator to be applicable in practice.

As to the amplitude parameters,  $\{\alpha_i^{(m)}\}_{i=1}^{L_r}$ , ML estimation problem can be formalized as

$$\begin{aligned} \{\hat{\alpha}_{i,\text{ML}}^{(m)}\}_{i=1}^{L_r} = \arg \max_{\alpha_i^{(m)}} & 2 \operatorname{Re} \{ \mathbf{s}^{(m)*} \mathbf{T}^* (\boldsymbol{\xi}^{(m)}) \mathbf{y}^{(m)} \} \\ & - N \operatorname{trace}\{\mathbf{P}^{(m)}\}. \end{aligned} \quad (3.39)$$

Similar to the previous parameter, the first partial derivative can be computed as

$$\begin{aligned} \frac{\partial \log p(\mathbf{y}^{(m)}, \boldsymbol{\xi}^{(m)})}{\partial \alpha_i^{(m)}} = & \\ & 2 \sum_{n=0}^{N-1} \operatorname{Re} \left\{ y_i^{(m)}[n] s^{(m)\text{C}}[n] e^{-j\beta_i^{(m)}} e^{-j2\pi\Delta f^{(m)}n} \right\} - 2\alpha_i^{(m)} N \end{aligned} \quad (3.40)$$

and set to zero to obtain

$$\hat{\alpha}_i^{(m)} = \frac{1}{N} \sum_{n=0}^{N-1} \operatorname{Re} \left\{ y_i^{(m)}[n] s^{(m)\text{C}}[n] e^{-j\beta_i^{(m)}} e^{-j2\pi\Delta f^{(m)}n} \right\}. \quad (3.41)$$

Similar to the statistic  $\hat{\beta}_{i,\text{ML}}^{(m)}$  in (3.38), the statistic in (3.41) depends on the true value of  $\beta_i^{(m)}$  and  $\Delta f^{(m)}$ . Therefore, the estimators for these parameters must be substituted in (3.41) to build the final estimator.

Regarding the carrier frequency offset  $\Delta f^{(m)}$ , the optimum estimator can be found as the solution to the problem

$$\widehat{\Delta f}_{\text{ML}}^{(m)} = \arg \max_{\Delta f^{(m)}} \text{Re} \left\{ \mathbf{s}^{(m)*} \mathbf{T}^* (\boldsymbol{\xi}^{(m)}) \mathbf{y}^{(m)} \right\}. \quad (3.42)$$

The objective function  $f(\Delta f^{(m)}) = \text{Re} \left\{ \mathbf{s}^{(m)*} \mathbf{T}^* (\boldsymbol{\xi}^{(m)}) \mathbf{y}^{(m)} \right\}$  can be developed as

$$f(\Delta f^{(m)}) = \text{Re} \left\{ \sum_{n=0}^{N-1} \sum_{i=1}^{L_r} \alpha_i^{(m)} y_i^{(m)} s^{(m)\text{C}}[n] e^{-j\beta_i^{(m)}} e^{-j2\pi\Delta f^{(m)}n} \right\}. \quad (3.43)$$

Now, the already calculated statistics in (3.38) and (3.41) can be plugged in (3.43). If the statistic  $\hat{\beta}_i^{(m)}$  in (3.38) is substituted in (3.41), the estimators can then be rewritten as

$$\hat{\beta}_i^{(m)} = \angle \left\{ \sum_{n=0}^{N-1} y_i^{(m)}[n] s^{(m)\text{C}}[n] e^{-j2\pi\Delta f^{(m)}n} \right\} \quad (3.44)$$

and

$$\hat{\alpha}_i^{(m)} = \frac{1}{N} \left| \sum_{n=0}^{N-1} y_i^{(m)}[n] s^{(m)\text{C}}[n] e^{-j2\pi\Delta f^{(m)}n} \right|. \quad (3.45)$$

Then, the ML estimator for  $\Delta f^{(m)}$  can be expressed after substituting (3.44) and (3.45) into (3.43), thereby obtaining

$$\widehat{\Delta f}_{\text{ML}}^{(m)} = \arg \max_{\Delta f} \left\{ \frac{1}{N} \sum_{i=1}^{L_r} \left| \sum_{n=0}^{N-1} y_i^{(m)}[n] s^{(m)\text{C}}[n] e^{-j2\pi\Delta f^{(m)}n} \right|^2 \right\}. \quad (3.46)$$

The estimator in (3.46) can be interpreted as follows. After applying a matched-filter with the training sequence  $s^{(m)}[n]$ , a periodogram is computed for each of the matched-filtered signals, and the parameter  $\Delta f^{(m)}$  is estimated as the maximum of the averaged periodograms. The periodograms can be computed

using a Fast Fourier Transform, although there is no guarantee that the true  $\Delta f^{(m)}$  will fall within one of the discrete bins in the transform. For this reason, I propose to find the maximum of the function in (3.46) using quadratic interpolation with the three frequency bins yielding the largest absolute value.

Regarding the noise variance  $\sigma^2$ , the optimum estimator can be found by again setting the derivative of the LLF to zero

$$\frac{\partial \log p(\mathbf{y}^{(m)}; \boldsymbol{\xi}^{(m)})}{\partial \sigma^2} = -\frac{NL_r}{\sigma^2} + \frac{1}{\sigma^4} \left\| \mathbf{y}^{(m)} - \mathbf{T} \left( \hat{\boldsymbol{\xi}}^{(m)} \right) \mathbf{s}^{(m)} \right\|_2^2, \quad (3.47)$$

which allows obtaining

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{NL_r} \left\| \mathbf{y}^{(m)} - \mathbf{T} \left( \hat{\boldsymbol{\xi}}^{(m)} \right) \mathbf{s}^{(m)} \right\|_2^2. \quad (3.48)$$

Finally, the optimum estimator for the complex amplitudes  $\tilde{\alpha}_i^{(m)} = [\tilde{\boldsymbol{\alpha}}^{(m)}]_i$ ,  $1 \leq i \leq L_r$  can be found using the property of *Asymptotic Invariance* from ML estimation [135] as

$$\hat{\alpha}_{i,\text{ML}}^{(m)} = \hat{\alpha}_i^{(m)} e^{j\hat{\beta}_i^{(m)}}, \quad (3.49)$$

which can be more explicitly expressed as

$$\hat{\alpha}_{i,\text{ML}}^{(m)} = \frac{1}{N} \sum_{n=0}^{N-1} y_i^{(m)}[n] s^{(m)\text{C}}[n] e^{-j2\pi \hat{\Delta} f_{\text{ML}}^{(m)} n} \quad (3.50)$$

### 3.4.1 Dictionary-Constrained Channel Estimation

In this section, I formulate the problem of estimating the high-dimensional mmWave MIMO channel based on the statistics already estimated using the ML criterion, as well as their corresponding CRLB. Once  $M$  training symbols

$s^{(m)}[n]$ ,  $1 \leq m \leq M$ ,  $0 \leq n \leq N - 1$  are processed and the  $M$  statistics  $\hat{\alpha}_{i,\text{ML}}^{(m)}$  are estimated, these statistics can be stacked into the signal model

$$\begin{aligned} \begin{bmatrix} \hat{\alpha}_{\text{ML}}^{(1)} \\ \vdots \\ \hat{\alpha}_{\text{ML}}^{(M)} \end{bmatrix} &= \begin{bmatrix} \mathbf{D}_{\mathbf{w}}^{(1)*} (\mathbf{q}^{(1)T} \mathbf{F}^{(1)T} \otimes \mathbf{W}^{(1)*}) \\ \vdots \\ \mathbf{D}_{\mathbf{w}}^{(M)*} (\mathbf{q}^{(M)T} \mathbf{F}^{(M)T} \otimes \mathbf{W}^{(M)*}) \end{bmatrix} \text{vec}\{\mathbf{H}\} \\ &+ \begin{bmatrix} \tilde{\mathbf{v}}^{(1)} \\ \vdots \\ \tilde{\mathbf{v}}^{(M)} \end{bmatrix}, \end{aligned} \quad (3.51)$$

with  $\tilde{\mathbf{v}}^{(m)}$  the estimation error of  $\hat{\alpha}_{i,\text{ML}}^{(m)}$ ,  $1 \leq m \leq M$ . Now, the channel matrix in (3.3) can be vectorized and plugged into (3.52) to yield

$$\begin{aligned} \underbrace{\begin{bmatrix} \hat{\alpha}_{\text{ML}}^{(1)} \\ \vdots \\ \hat{\alpha}_{\text{ML}}^{(M)} \end{bmatrix}}_{\hat{\alpha}_{\text{ML}}} &\approx \underbrace{\begin{bmatrix} \mathbf{D}_{\mathbf{w}}^{(1)*} (\mathbf{q}^{(1)T} \mathbf{F}^{(1)T} \otimes \mathbf{W}^{(1)*}) \\ \vdots \\ \mathbf{D}_{\mathbf{w}}^{(M)*} (\mathbf{q}^{(M)T} \mathbf{F}^{(M)T} \otimes \mathbf{W}^{(M)*}) \end{bmatrix}}_{\Phi_{\mathbf{w}}} \\ &\times \underbrace{\left( \tilde{\mathbf{A}}_{\text{T}}^{\text{C}} \otimes \tilde{\mathbf{A}}_{\text{R}} \right)}_{\Psi} \underbrace{\text{vec}\{\mathbf{G}^{\text{v}}\}}_{\mathbf{g}^{\text{v}}} + \underbrace{\begin{bmatrix} \tilde{\mathbf{v}}^{(1)} \\ \vdots \\ \tilde{\mathbf{v}}^{(M)} \end{bmatrix}}_{\tilde{\mathbf{v}}}, \end{aligned} \quad (3.52)$$

where  $\Phi_{\mathbf{w}} \in \mathbb{C}^{ML_{\text{r}} \times N_{\text{r}}N_{\text{t}}}$  is the post-whitened measurement matrix, and  $\Psi \in \mathbb{C}^{N_{\text{r}}N_{\text{t}} \times G_{\text{r}}G_{\text{t}}}$  is the dictionary matrix. Since there is no prior information about the channel, the design of the precoders and combiners needs to be such that the equivalent measurement matrix  $\Phi_{\mathbf{w}}$  in (3.51)-(3.52) has as small correlation (inner product in Hilbert space) between any two different columns as possible. As proven in the compressed sensing literature [142], [143], this is necessary to ensure that the estimation of the support will be robust. With the proposed

design, and a reasonable number of training frames, the correlation between columns in  $\Phi_w$  is small enough so that the channel can be both accurately and efficiently reconstructed, as will be shown in the numerical results. To estimate the  $G_r G_t \times 1$  sparse vector  $\mathbf{g}^v$ , it is necessary to have some prior knowledge on either the sparsity level of the channel or the noise variance in the linear model in (3.52). Knowing the sparsity level is unrealistic in practice, and it may not be helpful as there is no guarantee that the best sparse approximation of  $\mathbf{g}^v$  has  $C \sum_{c=1}^C R_c$  non-zero entries. For this reason, I will focus on finding the variance of  $\tilde{\mathbf{v}}$ . From the property of *asymptotic efficiency* of ML estimators [135] it can be established that, if the SNR is not too low, and the number of time-domain samples  $N$  is large enough, the estimation errors  $\hat{\mathbf{v}}^{(m)}$ ,  $1 \leq m \leq M$  are Gaussian-distributed, with zero mean and covariance given by the CRLB matrix for the estimation of the complex amplitudes  $\hat{\boldsymbol{\alpha}}_{\text{ML}}^{(m)}$ . Since the received noise vectors  $\mathbf{v}^{(m)}[n]$  in (3.8) are independent and identically distributed, the estimation errors for the complex amplitude vectors  $\tilde{\boldsymbol{\alpha}}^{(m)}$  are independent as well, although not identically distributed. Let us define the CRLB matrix corresponding to  $\mathbf{I}_1(\boldsymbol{\xi}^{(m)})$  and  $\mathbf{I}_2(\boldsymbol{\xi}^{(m)})$  in (3.24), (3.25) as

$$\mathbf{I}_1^{-1}(\boldsymbol{\xi}^{(m)}) = \text{diag} \left\{ \frac{\sigma^2}{2N} \mathbf{1}_{L_r}^T, \frac{\sigma^4}{L_r N} \right\} \quad (3.53)$$

and

$$\mathbf{I}_2^{-1}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \text{var} \left\{ \widehat{\Delta f}^{(m)} \right\} & \dots & \dots & \text{covar} \left\{ \widehat{\Delta f}^{(m)}, \hat{\beta}_{L_r}^{(m)} \right\} \\ \text{covar} \left\{ \widehat{\Delta f}^{(m)}, \hat{\beta}_1^{(m)} \right\} & \text{var} \left\{ \hat{\beta}_1^{(m)} \right\} & \dots & \text{covar} \left\{ \hat{\beta}_1^{(m)}, \hat{\beta}_{L_r}^{(m)} \right\} \\ \vdots & \vdots & \ddots & \vdots \\ \text{covar} \left\{ \widehat{\Delta f}^{(m)}, \hat{\beta}_{L_r}^{(m)} \right\} & \dots & \dots & \text{var} \left\{ \hat{\beta}_{L_r}^{(m)} \right\} \end{bmatrix}. \quad (3.54)$$

Let us also define the covariance matrix for any estimator of the vector  $\mathbf{x}^{(m)} = [\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)}]$  and denote it as  $\mathbf{C}_{\mathbf{x}^{(m)}\mathbf{x}^{(m)}} \in \mathbb{C}^{2L_r \times 2L_r}$ . This matrix satisfies

$$\mathbf{C}_{\mathbf{x}^{(m)}\mathbf{x}^{(m)}} \geq \begin{bmatrix} [\mathbf{I}_1^{-1}(\boldsymbol{\xi}^{(m)})]_{1:L_r, 1:L_r} & \mathbf{0} \\ \mathbf{0} & [\mathbf{I}_2^{-1}(\boldsymbol{\xi}^{(m)})]_{2:L_r+1, 2:L_r+1} \end{bmatrix}, \quad (3.55)$$

where  $\mathbf{A} \geq \mathbf{B}$  indicates that the matrix  $\mathbf{A} - \mathbf{B}$  is positive semi-definite.

Therefore, the covariance matrix for any estimator of  $\tilde{\boldsymbol{\alpha}}^{(m)} = \mathbf{f}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)}) = \boldsymbol{\alpha}^{(m)} e^{j\boldsymbol{\beta}^{(m)}}$  can be denoted as  $\mathbf{C}_{\tilde{\boldsymbol{\alpha}}^{(m)}, \tilde{\boldsymbol{\alpha}}^{(m)}}$  and satisfies [135]

$$\mathbf{C}_{\tilde{\boldsymbol{\alpha}}^{(m)}, \tilde{\boldsymbol{\alpha}}^{(m)}} \geq \underbrace{\begin{bmatrix} \frac{\partial \mathbf{f}^*(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\alpha}^{(m)}} & \frac{\partial \mathbf{f}^*(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}^{(m)}} \end{bmatrix} \mathbf{C}_{\mathbf{x}^{(m)}\mathbf{x}^{(m)}} \begin{bmatrix} \frac{\partial \mathbf{f}^T(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\alpha}^{(m)}} \\ \frac{\partial \mathbf{f}^T(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}^{(m)}} \end{bmatrix}}_{\mathbf{C}_{\tilde{\boldsymbol{\alpha}}_{\text{ML}}, \tilde{\boldsymbol{\alpha}}_{\text{ML}}}}. \quad (3.56)$$

Let the Jacobian matrix of  $\mathbf{f}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)}) : \mathbb{R}^{2L_r} \rightarrow \mathbb{C}^{L_r}$  be denoted as  $\mathbf{J}_{\mathbf{f}} \in \mathbb{C}^{L_r \times 2L_r}$ . This matrix can be calculated as

$$\mathbf{J}_{\mathbf{f}} = \begin{bmatrix} e^{j \text{diag}\{\boldsymbol{\beta}^{(m)}\}} & j \text{diag}\{\tilde{\boldsymbol{\alpha}}\} \end{bmatrix}. \quad (3.57)$$

Accordingly, the estimation error vector  $\tilde{\mathbf{v}}^{(m)}$  in (3.52) is distributed as  $\tilde{\mathbf{v}}^{(m)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\tilde{\boldsymbol{\alpha}}_{\text{ML}}, \tilde{\boldsymbol{\alpha}}_{\text{ML}}})$ . Consequently, the covariance matrix of  $\tilde{\mathbf{v}}$  in (3.8) is  $\mathbf{C}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}} =$

$\bigoplus_{m=1}^M \mathbf{C}_{\hat{\boldsymbol{\alpha}}_{\text{ML}}^{(m)} \hat{\boldsymbol{\alpha}}_{\text{ML}}^{(m)}}$ . Thereby, the problem of estimating  $\mathbf{g}^v$  can be formulated as

$$\begin{aligned} \hat{\mathbf{g}}^v = \arg \min_{\mathbf{g}^v} \|\mathbf{g}^v\|_1, \quad \text{subject to} \\ \left( \hat{\boldsymbol{\alpha}}_{\text{ML}} - \Phi_w \Psi \mathbf{g}^v \right)^* \mathbf{C}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}^{-1} \left( \hat{\boldsymbol{\alpha}}_{\text{ML}} - \Phi_w \Psi \mathbf{g}^v \right) \leq \epsilon, \end{aligned} \quad (3.58)$$

where  $\epsilon \in \mathbb{R}$  is a design parameter defining the maximum reconstruction error for the vector  $\mathbf{g}^v$ . Since the training precoders and combiners  $\mathbf{F}^{(m)}$ ,  $\mathbf{W}^{(m)}$  will, in general, result in different covariance matrices for the estimation error  $\tilde{\mathbf{v}}^{(m)}$ , it is crucial to find a good enough representative for the variance of the entire vector  $\tilde{\mathbf{v}}$ . To overcome this issue, I propose to design  $\epsilon$  as a convex combination of the CRLB for the different  $\tilde{\alpha}_i^{(m)}$  using estimates of the SNR per RF chain. These estimates can be found by using again the property of *asymptotic invariance* from ML estimators as  $\hat{\gamma}_{i,\text{ML}}^{(m)} = \hat{\alpha}_{i,\text{ML}}^{(m)2} / \hat{\sigma}_{\text{ML}}^2$ . Then, the parameter  $\epsilon$  is given by

$$\epsilon = \sum_{m=1}^M \frac{\hat{\gamma}_{i,\text{ML}}^{(m)} \text{trace} \{ \mathbf{C}_{\hat{\boldsymbol{\alpha}}^{(m)} \hat{\boldsymbol{\alpha}}^{(m)}} \}}{\sum_{i=1}^{L_r} \sum_{m=1}^M \hat{\gamma}_{i,\text{ML}}^{(m)}}. \quad (3.59)$$

Similar to the problem of estimating  $\boldsymbol{\xi}^{(m)}$ , I choose to decorrelate the estimates  $\hat{\boldsymbol{\alpha}}^{(m)}$  using the Cholesky decomposition of estimates for the CRLB matrix  $\mathbf{C}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}$ . Let us express the Cholesky factorization of  $\mathbf{C}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}$  as  $\mathbf{C}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}} = \mathbf{D}_{\tilde{\mathbf{v}}}^* \mathbf{D}_{\tilde{\mathbf{v}}}$ . Then, we can use a similar approach to the SW-OMP algorithm in [1] can be used to solve the problem in (3.58), which is henceforth referred to as Weighted - Orthogonal Matching Pursuit (W-OMP). The detailed steps the algorithm follows are outlined in Algorithm 3.4. The W-OMP algorithm is different from the standard OMP in that standard OMP does assume spatially white noise components in the received signal vector. The W-OMP algorithm is especially

designed to account for correlated noise at the output of the analog combiner. The weights the proposed W-OMP algorithm uses are given by the Cholesky factor of  $\mathbf{C}_{\hat{\boldsymbol{\alpha}}^{(m)}\hat{\boldsymbol{\alpha}}^{(m)}}$  in (3.56).

It is noteworthy to mention that the CRLB of the channel matrix  $\mathbf{H}$  is not computed because, for realistic mmWave channels having  $C$  clusters with  $R_c$  rays per cluster, the grid quantization error makes it impossible to assume that any two different rays will fall within two different spatial bins in the dictionary matrix  $\boldsymbol{\Psi}$ . For this reason, the estimate of the sparse vector  $\mathbf{g}^v$  will, in general, satisfy  $\|\hat{\mathbf{g}}^v\|_1 \neq \|\mathbf{g}^v\|_1$ . Consequently, the theory of CRLB cannot be applied to this problem since unbiasedness does not hold.

### 3.5 Numerical Results

In this section, I show the main numerical results concerning the proposed estimation strategies. I analyze the performance of the proposed estimators derived in Section 3.4, showing that they are asymptotically unbiased and asymptotically efficient, and compare the achievable estimation variance with the CRLB derived in Section 3.3. I also show the performance of the proposed channel estimation algorithm, both in terms of normalized mean squared error (NMSE) and spectral efficiency. For the latter, I consider both all-digital and hybrid design of precoders and combiners, for which I adopt the solutions developed in [16] and [11]. I define the normalized sample variance

---

**Algorithm 1** W-OMP

---

```
1: procedure W-OMP( $\mathbf{y}, \Phi, \Psi, \mathbf{D}_{\hat{\mathbf{v}}}, \epsilon$ )
2:   Compute the whitened equivalent observation matrix
3:    $\Upsilon_{\mathbf{w}} = \mathbf{D}_{\hat{\mathbf{v}}}^{-*} \Phi \Psi$ 
4:   Initialize the residual vector to the input signal vector and
   support estimate
5:    $\mathbf{y}_{\mathbf{w}} = \mathbf{D}_{\hat{\mathbf{v}}}^{-*} \mathbf{y}$ ,  $\mathbf{z} = \mathbf{y}_{\mathbf{w}}$ ,  $\hat{\mathcal{J}} = \{\emptyset\}$ 
6:   while MSE  $>$   $\epsilon$  do
7:     Distributed Correlation
8:      $\mathbf{c} = \Upsilon_{\mathbf{w}}^* \mathbf{z}$ ,  $k = 0, \dots, K - 1$ 
9:     Find the most likely atom in  $\Upsilon_{\mathbf{w}}$ 
10:     $p^* = \arg \max_p |[\mathbf{c}]_p|$ 
11:    Update the current guess of the support
12:     $\hat{\mathcal{J}} = \hat{\mathcal{J}} \cup p^*$ 
13:    Project the input signal onto the subspace given by the
14:    support using WLS
15:     $\hat{\mathbf{g}}^{\mathbf{v}} = \left( [\Upsilon_{\mathbf{w}}]_{:, \hat{\mathcal{J}}} \right)^{\dagger} \mathbf{y}_{\mathbf{w}}$ 
16:    Update residual
17:     $\mathbf{z} = \mathbf{y}_{\mathbf{w}} - [\Upsilon_{\mathbf{w}}]_{:, \hat{\mathcal{J}}} \hat{\mathbf{g}}^{\mathbf{v}}$ 
18:    Compute the current MSE
19:     $\text{MSE} = \frac{1}{ML_r} \|\mathbf{z}\|_2^2$ 
20:  end while
21: end procedure
```

---

Figure 3.4: Detailed steps of the proposed W-OMP algorithm.

---

of an estimate  $\hat{\boldsymbol{\xi}}$  of a vector of parameters  $\boldsymbol{\xi}$  as

$$\text{nvar} \left( \hat{\boldsymbol{\xi}} \right) = \frac{\text{trace}\{\mathbb{E}\{(\hat{\boldsymbol{\xi}} - \mathbb{E}\{\hat{\boldsymbol{\xi}}\})(\hat{\boldsymbol{\xi}} - \mathbb{E}\{\hat{\boldsymbol{\xi}}\})^*\}\}}{\|\boldsymbol{\xi}\|_2^2}. \quad (3.60)$$

I also define the NMSE of an estimator  $\hat{\mathbf{H}}$  of the channel matrix  $\mathbf{H}$  as  $\text{NMSE}(\hat{\mathbf{H}}) = \frac{\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2}{\|\mathbf{H}\|_F^2}$ . Finally, I define the spectral efficiency as a function of a hybrid precoder  $\mathbf{F} \in \mathbb{C}^{N_t \times N_s}$  and combiner  $\mathbf{W} \in \mathbb{C}^{N_r \times N_s}$  as

$$\mathcal{R}(\mathbf{F}, \mathbf{W}) = \log_2 \det \left\{ \mathbf{I}_{N_s} + \frac{\text{SNR}}{N_s} (\mathbf{W}^* \mathbf{W})^{-1} \mathbf{W}^* \mathbf{H} \mathbf{F} \mathbf{F}^* \mathbf{H}^* \mathbf{W} \right\}. \quad (3.61)$$

As to the training precoders and combiners for channel estimation, each entry in  $\mathbf{F}^{(m)}$ ,  $\mathbf{W}^{(m)}$  has modulus  $(\sqrt{N_t})^{-1}$  and  $(\sqrt{N_r})^{-1}$ , and the corresponding phases follow a discrete uniform distribution with values taken from the sets  $\mathcal{A}_T = \left\{ 0, \frac{2\pi}{2^{N_{T,Q}}}, \dots, \frac{2\pi(2^{N_{T,Q}}-1)}{2^{N_{T,Q}}} \right\}$  and  $\mathcal{A}_R = \left\{ 0, \frac{2\pi}{2^{N_{R,Q}}}, \dots, \frac{2\pi(2^{N_{R,Q}}-1)}{2^{N_{R,Q}}} \right\}$ . The number of quantization bits is set to  $N_{T,Q} = N_{R,Q} = 4$ . The simulation parameters are included in Table 3.1 and, unless otherwise indicated, are  $N_t = 32$ ,  $N_r = 32$ ,  $L_t = 4$ ,  $L_r = 4$ ,  $G_t = 128$ , and  $G_r = 128$ . Simulation results are obtained after averaging over 100 channel realizations. The channel is assumed to consist of  $C = 4$  clusters with  $R = 15$  rays per cluster and each cluster has an angular spread equal to 7.5 degrees. The sampling rate is set to  $T_s = \frac{1}{1760} \mu\text{s}$  [36], and the CFO is uniformly distributed between  $-\frac{1}{2T_s}$  and  $\frac{1}{2T_s}$ .

### 3.5.1 Performance analysis of ML estimators

In this subsection, I analyze the performance of the proposed ML estimators from Section 3.4, and compare their achievable performance to the derived CRLB from Section 3.3.

Table 3.1: Summary of typical system configuration parameters

Simulation Parameters		
Description	Parameter	Value
Number of transmit antennas	$N_t$	32
Number of receive antennas	$N_r$	32
Number of transmit RF chains	$L_t$	4
Number of receive RF chains	$L_r$	4
Transmit angular grid size	$G_t$	64
Receive angular grid size	$G_r$	64
Number of phase-shifter quantization bits	$N_Q$	4
Number of time-domain samples/frame	$N$	128
Number of training frames	$M$	128
Sampling period	$T_s$	1/1760 $\mu$ s
Number of multipath clusters	$C$	4
Number of multipath rays/cluster	$R$	15
Maximum CFO	$\Delta f^{(m)}$	880 MHz

In Fig. 3.5, I show the sample estimation variance and the CRLB of the proposed estimators as a function of SNR for  $N = 128$  time-domain training symbols chosen from an energy-normalized QPSK constellation.

The first behavior observed from Fig. 3.5 is that the different ML estimators attain the CRLB as  $\text{SNR} \rightarrow \infty$ . Therefore, as the theory of CRLB predicts, they are asymptotically efficient. Regarding the amplitude parameters  $\alpha_i^{(m)}$ ,  $1 \leq i \leq L_r$ , both the normalized sample variance and the normalized CRLB are different for every parameter, which is due to the different values the parameters take when the statistics are normalized. As to the phases  $\beta_i^{(m)}$ , the same behavior is observed, which comes from the fact that: i) the phase parameters are different for every RF chain, and ii) since the amplitude param-

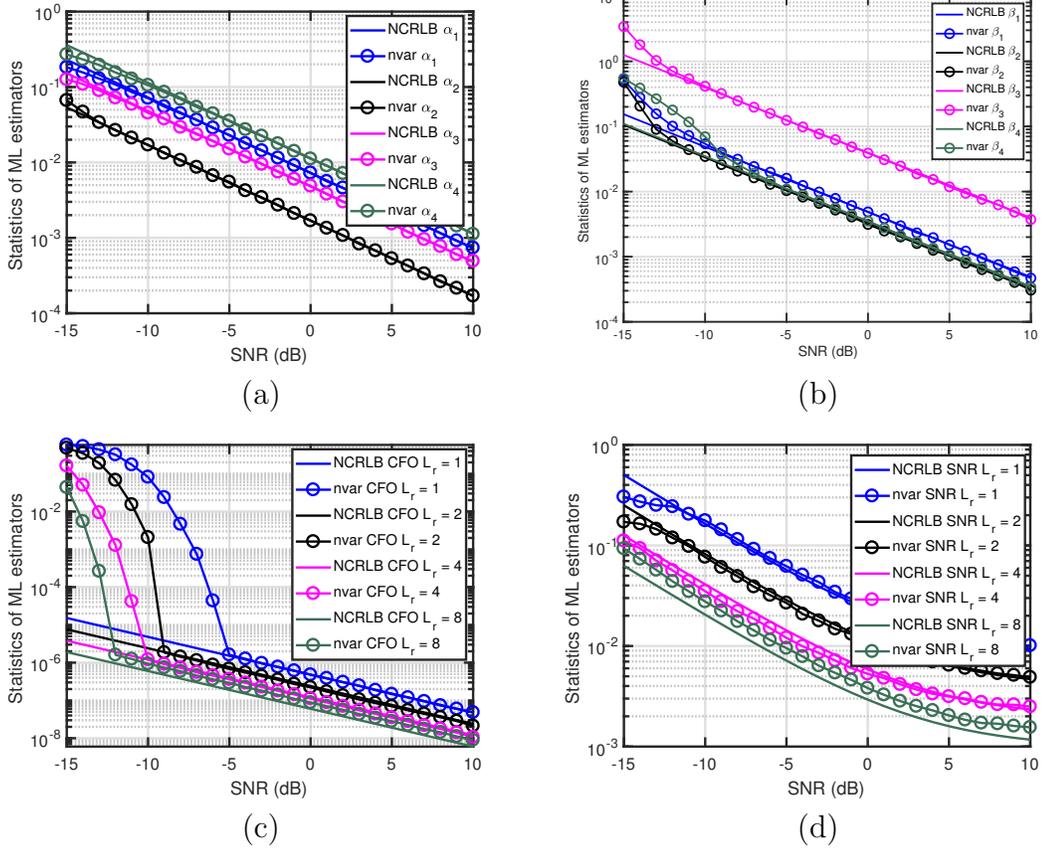


Figure 3.5: Normalized sample variance of the ML estimators from Section 3.4 as a function of SNR. The normalized CRLB is also shown to compare the performance of the ML estimators to the minimum achievable estimation variance. The FFT size of the CFO estimator is set to  $4L_rN$ .

eters  $\alpha_i^{(m)}$  change for every RF chain, the SNR per RF chain is also different. This behavior is also explained by the analytical CRLB previously derived.

Depending on the number of receive RF chains  $L_r$ , the estimator in (3.46) exhibits different performance, and so does the CRLB in (3.28) as well. As  $L_r$  increases, the CRLB for the CFO decreases, as analyzed in (3.30). Of

particular importance: as Fig. 3.5(c) shows, the threshold SNR that sets the boundary of the asymptotic efficiency region for  $\widehat{\Delta f}_{\text{ML}}^{(m)}$  decreases with the number of RF chains  $L_r$ , thereby suggesting that transceivers with a larger number of antennas (i.e. BSs) can be equipped with a larger number of RF chains and are thus capable of obtaining extremely accurate estimates of the CFO even when  $\text{SNR} < -10$  dB. The sample variance of the ML estimator would naturally be expected to decrease by the same amount as the CRLB. Notice, however, that despite the quadratic interpolation of the three largest spectral peaks in the weighted periodogram, this only happens if a larger number of FFT points  $N_{\text{FFT}}$  is chosen to estimate the CFO, such that the ML estimator is asymptotically efficient. To obtain such desirable effect, the FFT size of the CFO estimator should scale with the number of RF chains to better estimate this parameter. Owing to this behavior, the FFT size is set to  $N_{\text{FFT}} = 512L_r$ . Notice that, for  $L_r = 8$ , there is a very small gap between the normalized sample variance and the NCRLB for  $-12 \leq \text{SNR} \leq -5$  dB, which is due to this reason.

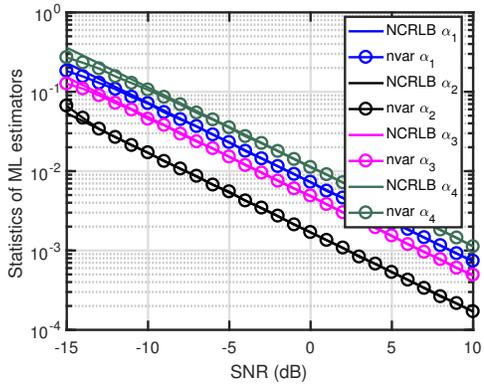
Regarding the SNR parameter, it is observed in Fig. 3.5(d) that, as  $L_r$  increases, both the normalized sample variance and the NCRLB decrease. As  $L_r$  increases, the number of samples to estimate the noise variance also increases, such that better estimates for this parameter can be found. Notice, however, that for the SNR parameter, there is a very small gap between the normalized sample variance and the normalized CRLB, which slightly increases with the SNR itself as  $L_r$  increases. This comes from the fact that, as SNR

increases, the CRLB from Section 3.3 increases quadratically with the SNR. Therefore, it is expected that, as  $\text{SNR} \rightarrow \infty$ , the gap between the normalized sample variance and the CRLB increases, which is due to the difficulty in estimating this parameter.

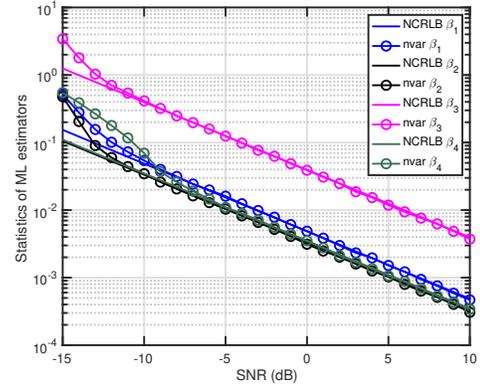
It is important to mention that, if the FFT size  $N_{\text{FFT}}$  is large enough, the only estimators affected by  $N_{\text{FFT}}$  are the CFO and the SNR estimators. This behavior is shown in Fig. 3.6.

Regarding the amplitude parameters, it is clear from the CRLB in (3.26) that the amplitude parameters are uncoupled with the other parameters, such that the behavior in Fig. 3.6 (a) is not surprising and the estimators can therefore be said to be asymptotically efficient. Notice, however, that in spite of the coupling between the phase parameters and the CFO, the normalized sample variance of the phase parameters also attains the corresponding normalized CRLB, such that the phase estimators are also asymptotically efficient. The FFT size, however, does affect the performance of the CFO estimator and the SNR estimator as  $L_r$  increases, although the performance loss can be seen to be negligible. Of particular importance: as  $L_r$  grows, the performance gap between the normalized sample variance of the CFO estimator and the normalized CRLB increases, but the threshold SNR that sets the boundary of the asymptotic efficiency region is invariant. This can be noticed by comparing Fig. 3.5 (c) and Fig. 3.6 (c).

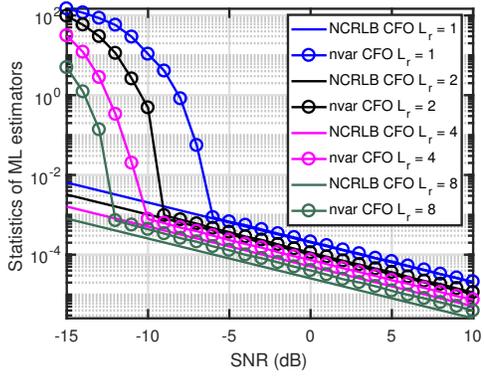
I also show in Fig. 3.7 the sample variance of the proposed SNR and CFO estimators as a function of the number of time-domain samples  $N$  for



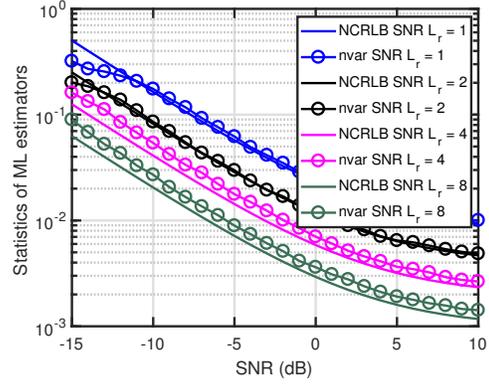
(a)



(b)



(c)



(d)

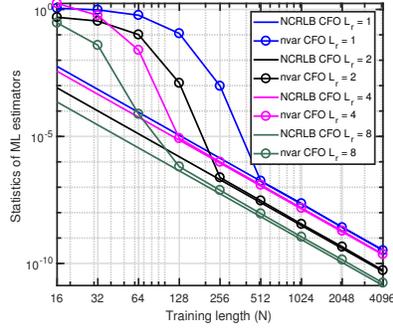
Figure 3.6: Normalized sample variance of the ML estimators from Section 3.4 as a function of SNR. The normalized CRLB is also shown to compare the performance of the ML estimators to the minimum achievable estimation variance. The FFT size of the CFO estimator is set to  $K = 4N = 1024$  points.

SNR =  $\{-10, -5, 0\}$  dB and  $L_r = \{1, 2, 4, 8\}$  RF chains.

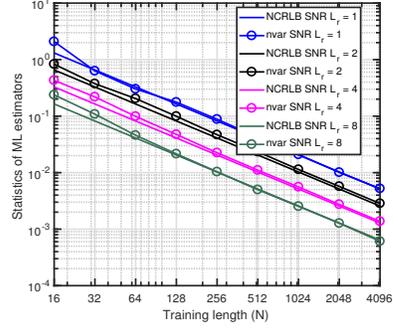
From Fig. 3.7, it can be noticed that larger numbers of RF chains  $L_r$  enhance estimation performance for both the estimation of SNR and  $\Delta f^{(m)}$ , as the CRLB predicted. Furthermore, when  $N = 2^x$ , with  $x$  a positive integer, both the normalized CRLB and the normalized sample variance evolve linearly with  $N$ , which is useful to design the training sequence such that a predefined estimation performance is obtained. For the SNR estimator, it is observed from Fig. 3.7(d-f) that  $N \geq 128$  guarantees asymptotic efficiency, such that it is a reasonable value for the training sequence length. Of course, increasing  $N$  up to  $N = 2^{12} = 4096$  has incredible performance, but at the cost of higher overhead to estimate the complex beamformed channel and the channel matrix itself. As to the CFO parameter, a similar behavior to that in Fig. 3.5 and 3.6 is observed. The threshold  $N$  for the CFO estimator to work on its asymptotic efficiency region is a joint function of  $L_r$  and SNR. For a fixed SNR, the minimum required  $N$  to attain the NCRLB reduces when  $L_r$  increases. Conversely, for a fixed  $N$ , the minimum SNR to attain the CRLB decreases as  $L_r$  increases. Decreasing  $L_r$  results in the opposite effect, such that a joint trade-off for  $N$  and  $L_r$  is needed to guarantee the estimators perform properly for the entire SNR regime the communications system is expected to work on.

### 3.5.2 NMSE of channel estimator

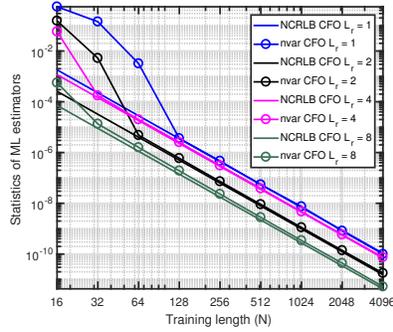
In this subsection, I analyze the performance of the proposed channel estimation strategy from Section 3.4.1. I show in Fig. 3.8 (a) the sample



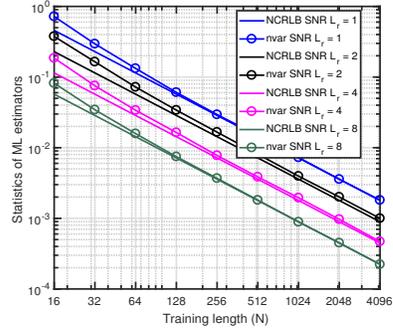
(a)



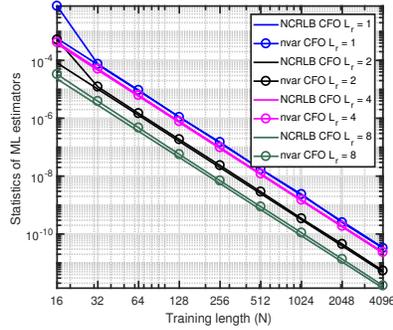
(d)



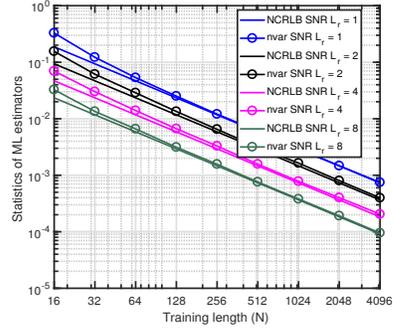
(b)



(e)



(c)



(f)

Figure 3.7: Normalized sample variance of the ML estimators from Section 3.4 as a function of  $N$ , for SNR =  $-10$  dB ((a) and (d)), SNR =  $-5$  dB ((b) and (e)), and SNR =  $0$  dB ((c) and (f)). The normalized CRLB is also shown to compare the performance of the ML estimators to the minimum achievable estimation variance. The FFT size of the CFO estimator is set to  $4L_T N$ .

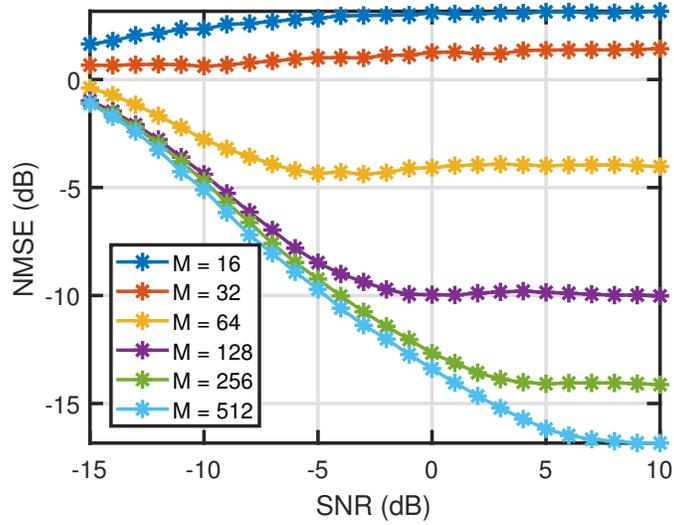
average NMSE obtained with the proposed W-OMP as a function of SNR, for  $M = \{16, 32, 64, 128, 256, 512\}$  training frames and  $N = 128$  time-domain training symbols. It is observed that increasing  $M$  leads to improvement in estimation performance. Notice, however, that the performance for  $M = \{128, 256, 512\}$  is very similar for  $\text{SNR} \leq -5$  dB, and it significantly reduces the estimation error in comparison with  $M \leq 64$ . Thus, at the low and medium SNR regime, the dramatic increase in overhead incurred when  $M > 128$  does not compensate for the marginal decrease in estimation error.

I also show in Fig. 3.8 (b) the sample average NMSE obtained with the proposed strategy as a function of  $M$ , for  $\text{SNR} = \{-10, -5, 0, 5\}$  dB. The number of time-domain training symbols is set to  $N = 128$ .

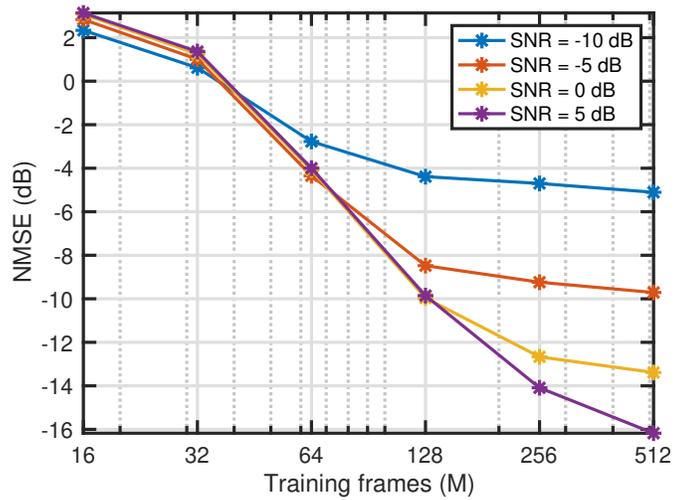
It is observed that increasing  $M$  leads to better estimation performance, but at the cost of higher overhead. As  $M$  grows, the estimation performance presents a noise plateau. This is due to the combination of the grid quantization error (driven by  $G_r = G_t = 128$ ) and the fact that every subpath in the mmWave channel has the same average power. This is the worst case scenario, since the effective size of the sparse channel vector  $\mathbf{g}^v$  to estimate is made the largest, thereby significantly hardening the task of estimating the channel.

### 3.5.3 Spectral efficiency

In this subsection of numerical results, I analyze the spectral efficiency achieved by the proposed estimation strategy. I compare the all-digital SVD-based precoder with both perfect CSI and channel estimates, and I also show



(a)



(b)

Figure 3.8: Evolution of the NMSE achieved by the joint synchronization and channel estimation algorithm as a function of SNR (a) and  $M$  (b). The number of training frames is  $M = \{16, 32, 64, 128, 256, 512\}$ , and the length of the time-domain training sequence is  $N = 128$ .

the performance of different hybrid precoding and combining algorithms to show how the channel estimates impact the final achievable spectral efficiency. As to hybrid precoding and combining algorithms, I consider the Spatially Sparse Precoding - Orthogonal Matching Pursuit (SSP-OMP) algorithm from [11], the Greedy Hybrid Precoding (GHP) algorithm from [16], and the Per-Antenna Constrained (PC) algorithm in [144]. The SSP-OMP algorithm is a dictionary-based algorithm that aims at iteratively finding the RF and baseband precoder (combiner) that minimizes the Euclidean distance between the unconstrained precoder (combiner) and its hybrid counterpart. The GHP algorithm, conversely, aims at minimizing the same metric as the SSP-OMP algorithm but using an SVD-based greedy approach, thereby being a dictionary-free algorithm. The PC algorithm in [144], finally, aims at minimizing the chordal distance between the unconstrained precoder (combiner) and the hybrid factorization. A major difference between the PC algorithm in [144] and both the SSP-OMP and GHP algorithms is that the former is not iterative and has closed-form, while the latter are iterative, thereby exhibiting higher computational complexity.

I show in Fig. 3.9 the achievable spectral efficiency as a function of SNR for  $M = N = 128$ . The number of transmitted data streams is  $N_s = 4$ . The performance of the all-digital precoder and combiner obtained with channel estimates is very close to the optimum perfect CSI solution, even for SNR as low as  $\text{SNR} = -10$  dB. For  $\text{SNR} < -10$  dB, however, the gap between the perfect CSI and all-digital W-OMP based precoders and combiners is more no-

ticeable. This is not surprising, since as Fig. 3.6 shows, the threshold SNR of the asymptotic efficiency region for the CFO estimator is around  $\text{SNR} = -10$  dB for  $L_r = 4$  RF chains. For lower values of SNR, the CFO estimator exhibits poor performance, although it can be certainly improved by increasing either  $L_r$  or  $N$ . Thereby, for  $\text{SNR} < -10$  dB, the poor performance of the CFO estimator affects the remaining ML estimators from Section 3.4, thereby further complicating channel estimation. As to the hybrid precoding and combining strategies, the SSP algorithm performs the worst, having a significant performance gap with respect to both the GHP and PC solutions, which exhibit almost identical performance in terms of achievable spectral efficiency. Furthermore, despite the large number of multipath components of the mmWave channel, the performance of the estimated hybrid precoders and combiners is close to the performance of both the perfect CSI and all-digital channel estimates.

Finally, I show in Fig. 3.10 the spectral efficiency as a function of  $M$  for  $\text{SNR} = \{-10, 0\}$  dB (a) and  $\text{SNR} = \{-5, 5\}$  dB (b). The number of transmitted data streams is  $N_s = 4$ .

As expected, the spectral efficiency increases with  $M$ , up to  $M = 128$  approximately. Beyond this value, the increase in overhead does not compensate for the marginal increase in spectral efficiency. Interestingly, the GHP and the PC algorithms exhibit almost identical performance, and significantly outperform the dictionary-based strategy in [11]. This fact suggests that minimizing the chordal distance between the unconstrained and the hybrid pre-

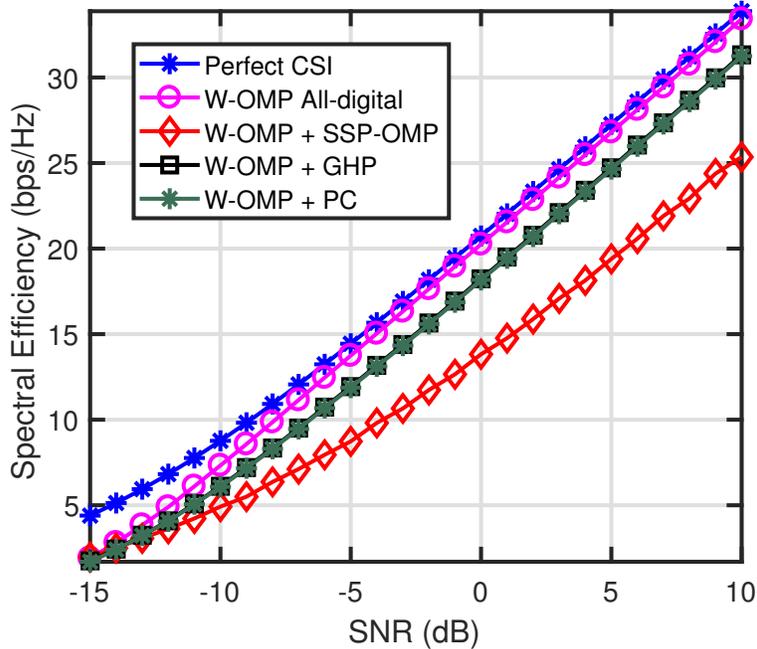
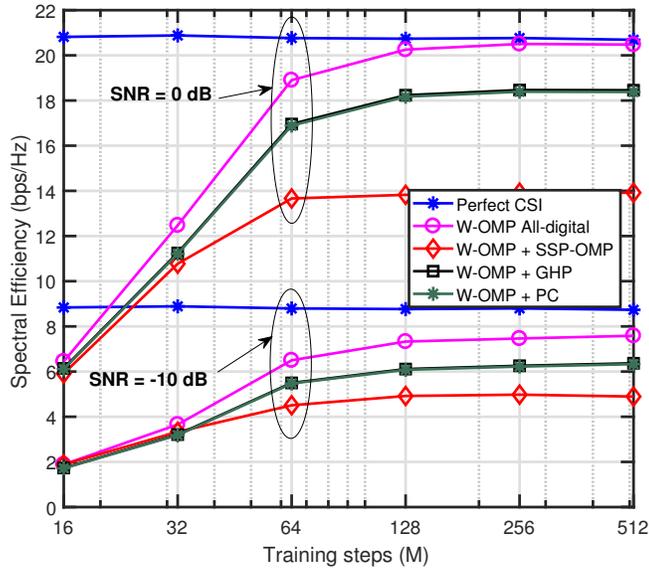
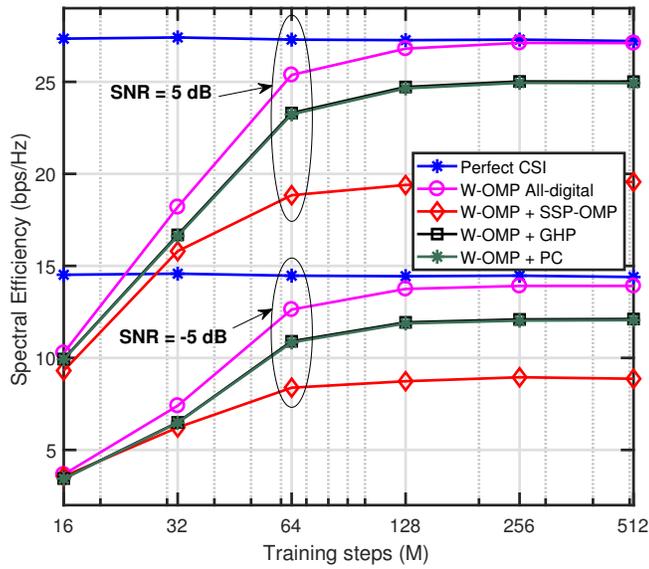


Figure 3.9: Evolution of the spectral efficiency versus SNR for different precoders and combiners. The number of time-domain training symbols is  $N = 128$ , and the number of training frames is  $M = 128$ .

coders (combiners) is approximately equivalent to minimizing their Euclidean distance. This behavior was analyzed in [11] for perfect CSI and assuming near-optimum sphere decoding at the receiver side, thus only focusing on the precoding side. The results in Fig. 3.10 clearly indicate that the same behavior can be expected for the combining side and with channel estimates, not only for perfect CSI. Further, near-optimum values of spectral efficiency are obtained with the proposed joint CFO and channel estimation algorithms, **without assuming any knowledge on either the sparsity level or the noise variance.**



(a)



(b)

Figure 3.10: Evolution of the spectral efficiency versus  $M$  for different precoders and combiners. The number of time-domain training symbols is  $N = 128$ . The SNR is set to  $\text{SNR} = \{-10, 0\}$  dB (a) and  $\text{SNR} = \{-5, 5\}$  dB (b).

Table 3.2: Online computational complexity of proposed algorithm

Joint CFO synchronization and channel estimation	
Operation	Complexity
$M \times$ Compute $\widehat{\Delta f}_{\text{ML}}^{(m)}$	$\mathcal{O}(ML_r(N + N_{\text{FFT}} \log_2(N_{\text{FFT}})))$
$M \times$ Compute $\hat{\tilde{\alpha}}_{\text{ML}}^{(m)}$	$\mathcal{O}(ML_r N)$
Compute $\mathbf{C}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}$	$\mathcal{O}(4ML_r^3)$
Compute $\mathbf{D}_{\tilde{\mathbf{v}}}$	$\mathcal{O}(ML_r^3)$
W-OMP algorithm	$\mathcal{O}(ML_r(G_t G_r - (j - 1)))$
Overall	$\mathcal{O}(ML_r(G_t G_r - (j - 1) + N + N_{\text{FFT}} \log_2(N_{\text{FFT}})))$

### 3.5.4 Computational complexity

Last, I analyze the computational complexity of the proposed joint synchronization and channel estimation strategy, and compare it to that of the state-of-the-art SW-OMP algorithm in [1] for the case of a single subcarrier, so that a fair comparison can be made. The computational complexity of each step in both algorithms is detailed in Table 3.2.

The computational complexity of the SW-OMP algorithm in [1] is of the order of  $\mathcal{O}(ML_r(G_t G_r - (j - 1)))$ . In the proposed joint CFO and channel estimation strategy, the overall complexity is  $\mathcal{O}(ML_r(N + N_{\text{FFT}} \log_2(N_{\text{FFT}})))$  higher, which comes from the additional complexity to estimate  $\Delta f^{(m)}$ . Although the number of operations involved in the joint estimation of the different parameters is larger, the main computationally complex operation is the estimation of the CFO. This signal processing task is, however, performed through computation of the weighted periodograms using the FFT algorithm, which can be computed very efficiently. Further, the estimation of  $\{\tilde{\alpha}^{(m)}\}_{m=1}^M$

is much less complex than estimating  $\Delta f^{(m)}$ , and so the computation of  $\mathbf{C}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}$  is, as well. Thereby, the proposed synchronization plus channel estimation approach can be claimed to be computationally efficient with marginal increase in complexity with respect to the state-of-the-art solution that assumes perfect synchronization in the received signal [1].

### 3.6 Conclusions

In this chapter, I proposed a framework for CFO and channel estimation in mmWave MIMO systems with hybrid architectures. I first formulated the problem of estimating the equivalent channel, CFO and noise variance. Then, I theoretically analyzed the fundamental limits of the estimation problem using the theory of CRLB. Then, I obtained the ML estimators for the unknown parameters and, using both these estimators and their CRLB, I proposed a sparse reconstruction algorithm to recover the high-dimensional mmWave MIMO channel. Simulation results showed that, despite the lack of knowledge on the channel's sparsity level and the noise variance, near-optimum values of spectral efficiency are achieved using both all-digital and hybrid precoders and combiners.

## Chapter 4

# Millimeter Wave Broadband Synchronization, Compressive Channel Estimation and Data Transmission

### 4.1 Introduction

In Chapter 2, I proposed channel estimation solutions for frequency-selective mmWave MIMO systems in which transceivers are equipped with hybrid MIMO architectures and operate under a perfect synchronization assumption. In Chapter 3, I proposed a multi-stage solution to the joint problem of CFO synchronization and channel estimation under a narrowband channel model to shed light on the impact of this synchronization impairment on the channel estimation problem and the resulting spectral efficiency. In this chapter, I extend the framework developed in Chapter 3 to frequency-selective channel models and introduce the additional TO and PN impairments. Under this setting, I show that, even under complete lack of synchronization, robust channel estimation in the low SNR regime can be attained, and multi-stream data transmission can be effectively carried out.

Time-frequency synchronization is one of the most important design aspects in cellular systems. In mmWave systems, however, acquiring synchronization information is significantly more challenging than for traditional

sub-6 GHz MIMO systems. Due to the necessity of using large antenna arrays to obtain the beamforming gain required to compensate for small antenna aperture, time and frequency synchronization must be performed either jointly with beam training as in 5G NR, or at the low SNR regime if the high-dimensional mmWave MIMO channel is to be estimated. Unfortunately, at such high operating frequency bands, synchronization, channel estimation, and data transmission are impacted by PN impairments, which consist of random fluctuations in the phase of the carrier generated by local oscillators. In this last part of my Ph.D. dissertation, I will focus on designing effective and robust time-frequency synchronization and PN compensation methods for compressive MIMO channel estimation at mmWave and subsequent data transmission.

#### 4.1.1 Prior work and Motivation

In the context of mmWave MIMO systems, synchronization parameters need to be properly estimated and compensated for before CSI can be acquired. This sets new challenges as synchronization acquisition must be performed at the low SNR regime, before transmit and receive communication beams can be aligned for data transmission. Most related prior work on broadband channel estimation with impairments is summarized in Fig. 4.1, as well as the proposed algorithms introduced later in this Chapter.

The problem of CS-based joint beam training and synchronization is studied in [5, 112]. In [112], the problem of beam training under PN errors and

Approach	Exploits noise covariance	Considers Timing Offset	Considers phase noise	Considers CFO	Hybrid architecture	Online Complexity	Training Overhead	Communication Performance
Compressive initial access [5]		✓	?	✓		Low	Low-Medium	Low
Swift-Link [120]				✓		Medium	Low-Medium	Low
<b>Proposed LMMSE-EM</b>	✓	✓	✓	✓	✓	Medium-High	Low-Medium	High
<b>Proposed EKF-RTS-EM</b>	✓	✓	✓	✓	✓	Low-Medium	Low-Medium	Medium-High

Figure 4.1: Summary and comparison of prior work on broadband channel estimation algorithms with synchronization impairments, as well as the approaches proposed in this Chapter.

unknown CFO was studied for narrowband mmWave MIMO systems using analog architectures. Therein, an EKF-based solution is proposed to track the joint phase of the unknown PN and beamformed narrowband channel, the phase of the received signal is compensated, and then MP is used to estimate the dominant AoD and AoA, thereby discovering a single communication path. In [5], a compressive initial access approach based on omnidirectional pseudo-random analog beamforming is proposed as an alternative to the directional initial access procedure used during beam management in 5G NR, and the effects of imperfect TO and CFO are studied therein. The main limitations of the algorithm proposed in [5] are: i) the algorithm is targeted at LOS channel models, thereby implicitly ignoring the presence of spatio-temporal clusters in the propagation environment; and ii) the proposed signal model assumes the presence of phase measurement errors only due to CFO, thereby ignoring the PN impairment. Prior work on joint broadband channel estimation and

synchronization for mmWave MIMO is limited, since much (if not most) of the prior work on channel estimation assumes perfect synchronization at the receiver side [34], [60], [58], [69], [70].

In the context of broadband channel models, prior work is limited to [118–121]. In [118], the problem of joint channel and PN estimation for a SISO system is considered, which is unrealistic at mmWave, and the proposed algorithms are only evaluated in very high SNR regime. In [119], analog-only architectures with a single RF chain are assumed, and an autocorrelation-based iterative algorithm is proposed to jointly estimate the CFO and the mmWave channel. Prior work in [119] assumes that analog beamformers and combiners can be instantaneously reconfigured for two consecutive transmitted time-domain samples, which is unrealistic since phase-shifters need an adjustment time for phase reconfiguration [36]. Further, the algorithm proposed in [119] has only been evaluated for mmWave channels having a very small number of non-clustered multipath components, which is not realistic at mmWave [35]. In addition, owing to the nature of the autocorrelation function, the proposed algorithm does not perform both well when the CFO is considerably large and the SNR is low. In [120], a CFO-robust beam alignment technique is developed to find the beam pairs maximizing the received SNR. The main limitation of [120] is that the algorithm proposed therein can only be applied to analog MIMO architectures, and its CFO correction capability is limited by both the number of delay taps in the mmWave MIMO channel, as well as the length of the training sequence, thereby making the algorithm impractical for

practical mmWave deployments with more significant CFO. In [121], the joint CFO and broadband channel estimation problem is formulated as a sparse bilinear optimization problem, which is solved using the parametric bilinear generalized approximate message passing (PBiGAMP) algorithm in [122]. The main limitation of [121] is that the proposed estimation strategy is tailored to all-digital MIMO architectures with low-resolution ADC converters, thereby not being directly applicable to hybrid MIMO architectures. In [117], a similar strategy to the one in [121] is followed, in which the joint CFO and channel estimation problem is studied for all-digital MIMO architectures. The problem is formulated as a quantized sparse bilinear optimization problem, which is solved using sparse lifting to increase the dimension of the CFO and channel estimation problem [123], and then applying the generalized approximate message passing (GAMP) algorithm in [124] to solve the lifted problem.

## 4.2 Contributions

I develop efficient and robust solution to the problem of estimating the TO, CFO, PN, and frequency-selective channel for hybrid mmWave MIMO systems. The proposed solutions can leverage the spatial design degrees of freedom brought by having several RF chains at both the transmitter and receiver to perform synchronization and compressive channel estimation, without relying on any prior channel knowledge. Further, I plan to investigate potential PN synchronization and data detection strategies for data transmission in mmWave MIMO systems enabling spatial multiplexing under the 5G NR

frame structure.

I summarize my contributions as follows:

- Based on a protocol of forwarding several training frames from the transmitter to the receiver [17], [5], [34], [2], I formulate and find a solution to the problem of TO, CFO, PN and frequency-selective mmWave MIMO channel estimation for systems employing hybrid architectures. It is noteworthy to mention that, unlike prior work in [5], I do not consider prior available information on the TO, and I do consider the problem of estimating the PN impairment. Further, the focus is on analyzing the synchronization problem at the low SNR regime.
- I propose to forward several training frames using ZC-based beamforming in combination with random subarray switching and antenna selection in order to both acquire synchronization and enable compressive channel estimation at the low SNR regime.
- For every training frame, which comprises several OFDM symbols, as in the 5G NR wireless standard [37], I theoretically analyze the hybrid CRLB for the problem of estimating the CFO, PN, and equivalent frequency-selective beamformed channel collecting the joint effect of the transmit hybrid precoders, frequency-selective mmWave MIMO channel, receive hybrid combiners, and equivalent transmit-receive pulse-shaping that bandlimits the complex baseband equivalent channel.

- I propose two novel iterative algorithms based on the EM method, which aim at finding the ML estimates for the CFO and beamformed equivalent channels, as well as the LMMSE estimates for the PN samples that impair the receive signals. The first proposed algorithm exhibits very good performance, yet it exhibits high computational complexity. The second proposed algorithm, conversely, offers a trade-off between estimation performance and computational complexity, and exhibits a very small performance gap with respect to the first algorithm.
- Using estimates of the unknown parameters for every training frame, I formulate the problem of estimating the high-dimensional frequency-selective mmWave MIMO channel, and find a solution to this problem using a variation of the SW-OMP algorithm in [1]. Last, using the estimated high-dimensional mmWave MIMO channel, I propose a joint PN tracking and data detection strategy enabling the multiple spatial degrees of freedom in the mmWave MIMO channel.

I evaluate the performance of the proposed algorithms in terms of NMSE, spectral efficiency, BER, and computational complexity. I use all-digital precoders and combiners to show the effectiveness of the proposed algorithms. Simulation results obtained from the estimated channel show that both the TO, CFO, and equivalent channels can be accurately estimated even in the presence of strong PN, and when the MIMO channel has several clusters with non-negligible AS. Furthermore, I show that near-optimum spectral

efficiency can be attained, without incurring in significant overhead and/or computational complexity. To the best of my knowledge, this is the first work that theoretically analyzes and provides solutions to the problem of joint synchronization and compressive channel estimation at mmWave considering hybrid MIMO architectures, and that is robust to both CFO, PN, and low SNR regime.

### 4.3 System model with synchronization impairments

I consider a single-user mmWave MIMO-OFDM communications link in which a transmitter equipped with  $N_t$  antennas sends  $N_s$  data streams to a receiver having  $N_r$  antennas. Both transmitter and receiver are assumed to use partially-connected hybrid MIMO architectures [14], as shown in Fig. 4.2, with  $L_t$  and  $L_r$  RF chains. A frequency-selective hybrid precoder is used at the transmitter, with  $\mathbf{F}[k] = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}[k] \in \mathbb{C}^{N_t \times N_s}$ , where  $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times L_t}$  is the analog precoder and  $\mathbf{F}_{\text{BB}}[k] \in \mathbb{C}^{L_t \times N_s}$  is the digital one at subcarrier  $k$ ,  $0 \leq k \leq K - 1$ . The RF precoder and combiner are implemented using a partially-connected network of phase-shifters and switches, as described in [14]. Likewise, the receiver applies a hybrid linear combiner  $\mathbf{W}[k] = \mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}[k] \in \mathbb{C}^{N_r \times L_r}$ , where  $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times L_r}$  is the analog combiner, and  $\mathbf{W}_{\text{BB}}[k] \in \mathbb{C}^{L_r \times N_s}$  is the baseband combiner at the  $k$ -th subcarrier.

Without loss of generality, I assume that the transmitted signal comprises  $N_{\text{tr}}$  OFDM symbols with a Cyclic Prefix (CP) of length  $L_c$ , similarly to the 5G NR wireless standard [37]. Let us define the rectangular pulse-

shape  $w_N[n] = 1$  for  $n \in [0, N - 1]$ , and  $w_N[n] = 0$  otherwise. Then, the hybrid-precoded transmitted signal can be expressed as

$$\begin{aligned} \mathbf{x}[n] &= \frac{1}{K} \mathbf{F}_{\text{RF}} \sum_{k=0}^{K-1} \sum_{t=0}^{N_{\text{tr}}-1} \mathbf{F}_{\text{BB}}[k] \mathbf{s}_t[k] e^{j \frac{2\pi k(n-L_c-t(K+L_c))}{K}} w_{K+L_c}[n-t(K+L_c)], \\ n &= 0, \dots, (N_{\text{tr}}-1)(K+L_c)-1, \end{aligned} \quad (4.1)$$

Then, let  $n_0 \in \mathbb{K}_+$ ,  $\Delta f \in \mathbb{R}$ ,  $\theta[n] \in \mathbb{R}$  denote the unknown TO, CFO normalized to the sampling rate  $f_s = 1/T_s$ , and  $n$ -th receive PN sample. Also, let  $\{\mathbf{W}[\ell]\}_{\ell=0}^{K-1}$  denote the time-domain hybrid combiner, given by the IFFT of the frequency-selective hybrid combiner  $\{\mathbf{W}[k]\}_{k=0}^{K-1}$ . Then, the received signal at discrete time instant  $n$  can be written as

$$\mathbf{y}[n] = \{\mathbf{W}^*[\ell]\}_{\ell=0}^{K-1} * \left( \sum_{d=0}^{D-1} \mathbf{H}[d] \mathbf{x}[n-d-n_0] e^{j(2\pi\Delta f n + \theta[n])} \right) + \mathbf{v}[n], \quad (4.2)$$

for  $n = 0, \dots, N + D + n_0 - 1$ , with  $N$  being the length of the time-domain transmitted signal  $\mathbf{x}[n]$ , and  $\mathbf{v}[n] \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \sum_{\ell=0}^{K-1} \mathbf{W}^*[\ell] \mathbf{W}[\ell])$  is the post-combining received noise, where  $\sigma^2$  denotes the variance of the noise at any receive antenna.

In this chapter, I focus on the problem of estimating the unknown CFO  $\Delta f$ , PN samples  $\theta[n]$ , and frequency-selective mmWave MIMO channel  $\{\mathbf{H}[d]\}_{d=0}^{D-1}$ . Given the high dimensionality of the channel matrices, I consider a training protocol in which the transmitter forwards  $M$  training frames to the receiver [2], [34], [17], [1], which must estimate the different unknown synchronization parameters. In view of this, for the  $m$ -th training frame,  $1 \leq m \leq M$ , I set  $\mathbf{F}^{(m)}[k] = \mathbf{F}_{\text{tr}}^{(m)}$ , and  $\mathbf{W}^{(m)}[k] = \mathbf{W}_{\text{tr}}^{(m)}$ , for every  $0 \leq k \leq K - 1$ .

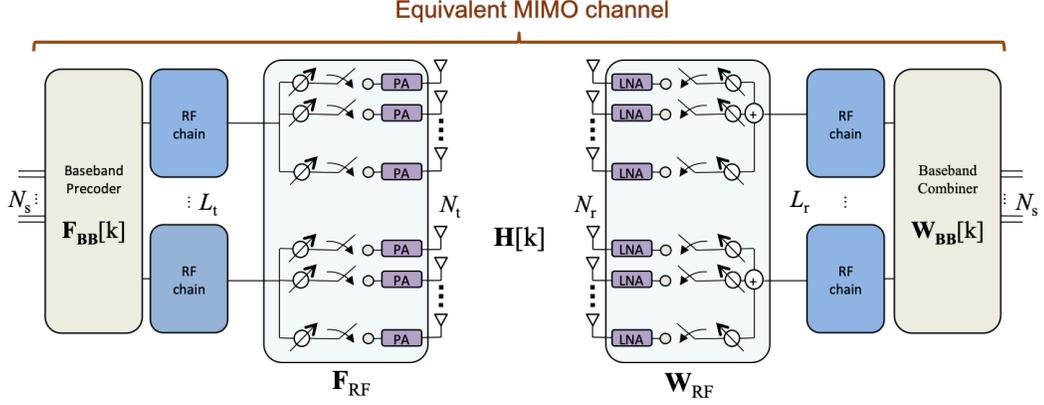


Figure 4.2: Illustration of the structure of a partially-connected hybrid MIMO architecture, which includes analog and digital precoders and combiners (same as Fig. 1.3).

Furthermore, I design the training symbols in (4.1) as  $\mathbf{s}_t^{(m)}[k] = \mathbf{q}^{(m)} \mathbf{s}_t^{(m)}[k]$ , where  $\mathbf{q}^{(m)}$  operates as an equivalent baseband precoder for this particular design of the training sequence [2]. Therefore, for the transmission of the  $m$ -th training frame, comprising of  $N_{\text{tr}}$  OFDM symbols, the received signal reads as

$$\mathbf{y}^{(m)}[n] = \mathbf{W}_{\text{tr}}^{(m)*} \sum_{d=0}^{D-1} \mathbf{H}[d] \mathbf{F}_{\text{tr}}^{(m)} \mathbf{q}^{(m)} s[n-d-n_0] e^{j(2\pi\Delta f^{(m)}n + \theta^{(m)}[n])} + \mathbf{v}^{(m)}[n], \quad (4.3)$$

in which  $\mathbf{v}^{(m)}[n] \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{W}_{\text{tr}}^{(m)*} \mathbf{W}_{\text{tr}}^{(m)})$  is the received post-combining circularly-symmetric complex additive Gaussian noise. As shown in [145], [2], the ML criterion establishes that the baseband combiner must whiten the received signal to estimate the different unknown parameters. For this purpose, let us consider the Cholesky decomposition of  $\mathbf{C}_{\text{w}}^{(m)} = \mathbf{W}_{\text{tr}}^{(m)*} \mathbf{W}_{\text{tr}}^{(m)}$  as  $\mathbf{C}_{\text{w}}^{(m)} = \mathbf{D}_{\text{w}}^{(m)*} \mathbf{D}_{\text{w}}^{(m)}$ , with  $\mathbf{D}_{\text{w}}^{(m)} \in \mathbb{C}^{L_r \times L_r}$  an upper triangular matrix. Now, let us define a vector

$\mathbf{g}^{(m)}[d] \in \mathbb{C}^{L_r \times 1}$ ,  $\mathbf{g}^{(m)}[d] = \mathbf{D}_w^{(m)*} \mathbf{W}_{tr}^{(m)*} \mathbf{H}[d] \mathbf{F}_{tr}^{(m)} \mathbf{q}^{(m)}$ , containing the complex equivalent channel samples for a given training step  $1 \leq m \leq M$ . Accordingly, for the  $m$ -th transmitted frame, the received signal in (4.3) can be expressed as

$$\begin{aligned} \mathbf{y}^{(m)}[n] = & e^{j(2\pi\Delta f^{(m)}n + \theta^{(m)}[n])} \underbrace{\sum_{d=0}^{D-1} \mathbf{g}^{(m)}[d] s^{(m)}[n-d-n_0]}_{\mathbf{x}^{(m)}[n,d,n_0]} \\ & + \mathbf{v}^{(m)}[n], \quad n = 0, \dots, n_0 + (L_c + K)N_{tr} + D - 1 \end{aligned} \quad (4.4)$$

with  $\mathbf{v}^{(m)}[n] \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{L_r})$  being the post-whitened spatially white received noise vector, and  $\mathbf{g}^{(m)}[d] = [\alpha_1[d]e^{j\beta_1[d]}, \dots, \alpha_{L_r}[d]e^{j\beta_{L_r}[d]}]^T$  is the complex equivalent beamformed channel for the  $m$ -th training step and  $d$ -th delay tap. Therefore, the interest here lies on estimating the vector of parameters  $\boldsymbol{\xi}^{(m)} \triangleq \left[ \{\mathbf{g}^{(m)T}[d]\}_{d=0}^{D-1}, \Delta f^{(m)}, \{\theta^{(m)}[n]\}_{n=0}^{N+n_0+D-1}, n_0 \right]^T$  for every training frame. In the next section, I theoretically analyze this estimation problem and find the hybrid CRLB for the different parameters in  $\boldsymbol{\xi}^{(m)}$ .

To model the PN model is taken from the IEEE 802.11ad wireless standard, which is given in [146]. The Power Spectral Density (PSD) of the PN is given in [147] as

$$P(f) = G_\theta \left[ \frac{1 + (f/f_z)^2}{1 + (f/f_p)^2} \right], \quad (4.5)$$

in which  $G_\theta$  is the PSD at  $f = 0$  Hz and it is measured in dBc/Hz,  $f_z = 100$  MHz, and  $f_p = 1$  MHz [147]. Using the inverse Fourier transform of the PSD in (4.5), the autocorrelation of the PN can be obtained as

$$R_{\theta^{(m)}\theta^{(m)}}(\tau) = G_\theta \left[ \frac{f_p^2}{f_z^2} \delta(\tau^{(m)}) + \pi f_p \left( 1 - \frac{f_p^2}{f_z^2} \right) e^{-2\pi f_p |\tau^{(m)}|} \right]. \quad (4.6)$$

From (4.6), the covariance matrix of the PN vector  $\boldsymbol{\theta}^{(m)} \in \mathbb{R}^N$ ,  $\boldsymbol{\theta}^{(m)} = [\theta^{(m)}[0] \dots \theta^{(m)}[N-1]]$  can be obtained by sampling the autocorrelation function as  $[\mathbf{C}_{\boldsymbol{\theta}^{(m)}\boldsymbol{\theta}^{(m)}}]_{i,j} = R_{\boldsymbol{\theta}^{(m)}\boldsymbol{\theta}^{(m)}}(|i-j|T_s)$ . From this, it is clear that the PN variance does not depend on the particular time instant at which the PN sample is observed, but only depends on the absolute time difference  $|i-j|T_s$ .

The discrete-time MIMO channel between the transmitter and the receiver is modeled as a set of  $N_r \times N_t$  matrices denoted as  $\mathbf{H}[d]$ , for a given delay tap  $d = 0, \dots, D-1$ , with  $D$  the delay tap length of the channel. Each of the matrices  $\mathbf{H}[d]$  is assumed to be a sum of the contributions of  $C$  spatial clusters, each contributing with  $R_c$  rays,  $c = 1, \dots, C$ . I use  $\rho_L$  to denote the pathloss,  $\alpha_{c,r} \in \mathbb{C}$  is the complex gain of the  $r$ -th ray within the  $c$ -th cluster,  $\tau_{c,r} \in \mathbb{R}_+$  is the time delay of the  $r$ -th ray within the  $c$ -th cluster,  $\phi_{c,r}, \theta_{c,r} \in [0, 2\pi)$  are the AoA and AoD,  $\mathbf{a}_R(\cdot) \in \mathbb{C}^{N_r \times 1}$  and  $\mathbf{a}_T(\cdot) \in \mathbb{C}^{N_t \times 1}$  denote the receive and transmit array steering vectors,  $p_{\text{RC}}(\tau)$  is the equivalent transmit-receive baseband pulse shape including analog filtering effects evaluated at  $\tau$  [33], and  $T_s$  is the sampling interval. Using this notation, the frequency-domain channel matrix at the  $k$ -th subcarrier is given by

$$\begin{aligned} \mathbf{H}[k] = & \sqrt{\frac{N_r N_t}{\rho_L \sum_{c=1}^C R_c}} \sum_{k=0}^{K-1} \sum_{c=1}^C \sum_{r=1}^{R_c} \alpha_{c,r} p_{\text{RC}}(dT_s - \tau_{c,r}) \times \\ & \times \mathbf{a}_R(\phi_{c,r}) \mathbf{a}_T^*(\theta_{c,r}) e^{-j\frac{2\pi k}{K}}. \end{aligned} \quad (4.7)$$

Taking the inverse Fourier transform of (4.7) allows obtaining the discrete-time MIMO channel used in (4.3). The channel matrix can be represented

more compactly as

$$\mathbf{H}[k] = \mathbf{A}_R \mathbf{G}[k] \mathbf{A}_T^*, \quad (4.8)$$

where  $\mathbf{G}[k] \in \mathbb{C}^{\sum_{c=1}^C R_c \times \sum_{c=1}^C R_c}$  is a diagonal matrix containing the path gains and the equivalent pulse-shaping effect, and  $\mathbf{A}_T \in \mathbb{C}^{N_t \times \sum_{c=1}^C R_c}$ ,  $\mathbf{A}_R \in \mathbb{C}^{N_r \times \sum_{c=1}^C R_c}$  are the frequency-selective array response matrices evaluated on the AoD and AoA, respectively. Finally, the matrix  $\mathbf{H}[k]$  in (4.8) can be approximated using the extended virtual channel model [8] as

$$\mathbf{H}[k] \approx \tilde{\mathbf{A}}_R \mathbf{G}^v[k] \tilde{\mathbf{A}}_T^*, \quad (4.9)$$

where  $\mathbf{G}^v[k] \in \mathbb{C}^{G_r \times G_t}$  is a sparse matrix containing the path gains of the quantized spatial frequencies in the non-zero elements, and the dictionary matrices  $\tilde{\mathbf{A}}_T \in \mathbb{C}^{N_t \times G_t}$ ,  $\tilde{\mathbf{A}}_R \in \mathbb{C}^{N_r \times G_r}$  contain the transmit and receive array response vectors evaluated on spatial grids of sizes  $G_t$  and  $G_r$ , respectively.

#### 4.4 Theoretical analysis of the estimation problem

In this section, I theoretically analyze the problem of estimating the unknown parameters in  $\boldsymbol{\xi}^{(m)}$ . Let us assume, without loss of generality, that  $N_{\text{tr}}$  OFDM symbols are transmitted for the  $m$ -th training frame, and that the number of available received time-domain samples of  $\mathbf{y}^{(m)}[n]$  is given by  $N = n_0 + (L_c + K)N_{\text{tr}} + D$ . Assuming that the received time-domain noise samples in (4.4) are independent and identically distributed, the received signal

in (4.4) has LLF given by

$$\begin{aligned} \log p(\{\mathbf{y}^{(m)}[n]\}_{n=0}^{N-1}; \boldsymbol{\xi}^{(m)}) &= -N \log(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \|\mathbf{y}^{(m)}[n]\|_2^2 \\ &+ 2 \sum_{n=0}^{N-1} \operatorname{Re} \left\{ \mathbf{y}^{(m)*} e^{j(2\pi\Delta f^{(m)}n + \theta^{(m)}[n])} \sum_{d=0}^{D-1} \mathbf{x}^{(m)}[n, d, n_0] \right\} - \sum_{n=0}^{N-1} \left\| \sum_{d=0}^{D-1} \mathbf{x}^{(m)}[n, d, n_0] \right\|_2^2. \end{aligned} \quad (4.10)$$

To ensure robustness of the synchronization algorithm, I will focus on finding the ML estimators for the different unknown parameters in  $\boldsymbol{\xi}^{(m)}$ . From (4.10), it is observed that maximizing  $\mathcal{L}(\boldsymbol{\xi}^{(m)}) = \log p(\{\mathbf{y}^{(m)}[n]\}_{n=0}^{N-1}; \boldsymbol{\xi}^{(m)})$  as a function of  $n_0$  requires knowledge of the other parameters contained in  $\boldsymbol{\xi}^{(m)}$ , which suggests that the ML estimator exhibits high computational complexity. To reduce computational complexity, I propose to exploit the good correlation properties of Golay sequences [36], [108], and append a 64-point  $G_{a,64}$  sequence at the beginning of the training frame, which has been shown to offer excellent performance in the absence of PN [141]. Thereby, a practical TO estimator can be devised by maximizing the third term in (4.10), which is given by [5, 141]

$$\hat{n}_0 = \arg \max_{n_0} \sum_{i=1}^{L_r} \sum_{n=0}^{63} \left| y_i^{(m)*}[n] s^{(m)}[n - d - n_0] \right|, \quad (4.11)$$

which explicitly exploits the information coming from having  $L_r \geq 1$  at the receiver side. Assuming that the TO has been estimated perfectly using (4.11), this parameter can be compensated by advancing the receive signal by  $\hat{n}_0$  as  $\mathbf{y}^{(m)}[n] = \mathbf{y}^{(m)}[n + n_0]$ ,  $n = 0, \dots, (L_c + K)N_{\text{tr}} + D - 1$ . Now, let the initial sample of the  $t$ -th OFDM symbol after CP removal be defined

as  $k_0[t] \triangleq L_c + n_0 + t(K + L_c)$ , let  $\phi^{(m)}[n_0, t] \triangleq e^{j2\pi\Delta f^{(m)}k_0[t]}$  be the common phase change at the  $t$ -th OFDM symbol due to TO, let  $\mathbf{\Omega}_t(\Delta f^{(m)}) \triangleq \phi^{(m)}[n_0, t] \bigoplus_{n=0}^{K-1} e^{j2\pi\Delta f^{(m)}n}$  be the CFO matrix impairing the  $t$ -th OFDM symbol, and let  $\mathbf{s}_t^{(m)} \triangleq \bigoplus_{k=0}^{K-1} \mathbf{s}_t^{(m)}[k]$  and  $\mathbf{S}^{(m)} \triangleq \begin{bmatrix} \mathbf{s}_0^{(m)T} & \dots & \mathbf{s}_{N_{\text{tr}}}^{(m)T} \end{bmatrix}^T$  be matrices containing the  $t$ -th OFDM training symbol and the  $N_{\text{tr}}$  OFDM training symbols, respectively. Also, let  $\mathbf{g}_i^{(m)} \triangleq \begin{bmatrix} \mathbf{g}_i^{(m)}[0] & \dots & \mathbf{g}_i^{(m)}[K-1] \end{bmatrix}$  be the frequency-response of the equivalent beamformed channel seen by the  $i$ -th receive RF chain, let  $\mathbf{v}_{i,t}^{(m)} \triangleq \begin{bmatrix} v_i^{(m)}[k_0[t]] & \dots & v_i^{(m)}[k_0[t] + K - 1] \end{bmatrix}$  contain the time-domain noise samples impairing the  $t$ -th OFDM symbol, let  $\boldsymbol{\theta}_t^{(m)} = \begin{bmatrix} \theta^{(m)}[k_0[t]] & \dots & \theta^{(m)}[k_0[t] + K - 1] \end{bmatrix}^T$  be the PN vector corresponding to the  $t$ -th OFDM symbol, and let  $\mathbf{P}(\boldsymbol{\theta}_t^{(m)}) \triangleq \bigoplus_{n=0}^{K-1} e^{j\theta^{(m)}[k_0[t]+n]}$  be the diagonal PN matrix impairing the received  $t$ -th OFDM symbol. Letting  $\mathbf{F} \in \mathbb{C}^{K \times K}$  denote the  $K$ -point DFT matrix, the received time synchronized signal  $\mathbf{y}^{(m)}[n]$  can be vectorized as

$$\begin{aligned}
\underbrace{\begin{bmatrix} y_i^{(m)}[k_0[t]] \\ \vdots \\ y_i^{(m)}[k_0[t] + K - 1] \end{bmatrix}}_{\mathbf{y}_{i,t}^{(m)}} &= \mathbf{\Omega}_t(\Delta f^{(m)}) \mathbf{P}_t(\boldsymbol{\theta}_t^{(m)}) \mathbf{F}^* \underbrace{\left( \bigoplus_{k=0}^{K-1} \mathbf{s}_t^{(m)}[k] \right)}_{\mathbf{s}_t^{(m)}} \\
&\times \underbrace{\begin{bmatrix} \mathbf{g}_i^{(m)}[0] \\ \vdots \\ \mathbf{g}_i^{(m)}[K-1] \end{bmatrix}}_{\mathbf{g}_i^{(m)}} + \underbrace{\begin{bmatrix} v_i^{(m)}[k_0[t]] \\ \vdots \\ v_i^{(m)}[k_0[t] + K - 1] \end{bmatrix}}_{\mathbf{v}_{i,t}^{(m)}}, \tag{4.12}
\end{aligned}$$

such that the vectorized received signal can be simplified as

$$\mathbf{y}_{i,t}^{(m)} = \boldsymbol{\Omega}_t (\Delta f^{(m)}) \mathbf{P}_t (\boldsymbol{\theta}_t^{(m)}) \mathbf{F}^* \mathbf{S}_t^{(m)} \mathbf{g}_i^{(m)} + \mathbf{v}_{i,t}^{(m)}. \quad (4.13)$$

Now, the  $K \times 1$  random vector  $\mathbf{y}_{i,t}^{(m)}$  can be stacked for the different received OFDM symbols  $1 \leq t \leq N_{\text{tr}}$  and RF chains  $1 \leq i \leq L_r$  as

$$\begin{aligned} \underbrace{\begin{bmatrix} \mathbf{y}_{i,1}^{(m)} \\ \vdots \\ \mathbf{y}_{i,N_{\text{tr}}}^{(m)} \end{bmatrix}}_{\mathbf{y}_i^{(m)}} &= \underbrace{\left( \bigoplus_{t=1}^{N_{\text{tr}}} \boldsymbol{\Omega}_t (\Delta f^{(m)}) \right)}_{\boldsymbol{\Omega}(\Delta f^{(m)})} \underbrace{\left( \bigoplus_{t=1}^{N_{\text{tr}}} \mathbf{P}_t (\boldsymbol{\theta}_t^{(m)}) \right)}_{\mathbf{P}(\boldsymbol{\theta}^{(m)})} \underbrace{(\mathbf{I}_{N_{\text{tr}}} \otimes \mathbf{F}^*)}_{\mathbf{F}_{\otimes}^*} \\ &\times \underbrace{\begin{bmatrix} \mathbf{S}_1^{(m)} \\ \vdots \\ \mathbf{S}_{N_{\text{tr}}}^{(m)} \end{bmatrix}}_{\mathbf{s}^{(m)}} \mathbf{g}_i^{(m)} + \underbrace{\begin{bmatrix} \mathbf{v}_{i,1}^{(m)} \\ \vdots \\ \mathbf{v}_{i,N_{\text{tr}}}^{(m)} \end{bmatrix}}_{\mathbf{v}_i^{(m)}}. \end{aligned} \quad (4.14)$$

Therefore, the received signal  $\mathbf{y}_i^{(m)}$  is distributed according to  $\mathcal{CN}(\boldsymbol{\mu}_i^{(m)}(\boldsymbol{\xi}^{(m)}), \sigma^2 \mathbf{I}_{KN_{\text{tr}}})$ , where  $\boldsymbol{\mu}_i^{(m)}(\boldsymbol{\xi}^{(m)}) = \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P}(\boldsymbol{\theta}^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{g}_i^{(m)}$ . Finally, stacking the received signals  $\mathbf{y}_i^{(m)}$  for the different RF chains yields

$$\begin{bmatrix} \mathbf{y}_1^{(m)} \\ \vdots \\ \mathbf{y}_{L_r}^{(m)} \end{bmatrix} = \left( \mathbf{I}_{L_r} \otimes \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P}(\boldsymbol{\theta}^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \right) \begin{bmatrix} \mathbf{g}_1^{(m)} \\ \vdots \\ \mathbf{g}_{L_r}^{(m)} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1^{(m)} \\ \vdots \\ \mathbf{v}_{L_r}^{(m)} \end{bmatrix}. \quad (4.15)$$

For the purpose of theoretically analyzing the estimation problem of finding the unknown parameters, let  $\mathbf{g}_i^{(m)}[k]$  be further expressed as  $\mathbf{g}_i^{(m)}[k] = \alpha_i^{(m)}[k] e^{j\beta_i^{(m)}[k]}$ , and let  $\tilde{\mathbf{g}}_i^{(m)}[k] \in \mathbb{C}^2$  be defined as  $\tilde{\mathbf{g}}_i^{(m)}[k] \triangleq [\alpha_i^{(m)}[k], \beta_i^{(m)}[k]]^T$ . Finally, let  $\tilde{\mathbf{g}}_i^{(m)} \in \mathbb{C}^{2K \times 1}$  be given by  $\tilde{\mathbf{g}}_i^{(m)} \triangleq [\tilde{\mathbf{g}}_i^{(m)T}[0], \dots, \tilde{\mathbf{g}}_i^{(m)T}[K-1]]^T$ , and

$\tilde{\mathbf{g}}^{(m)} \triangleq [\tilde{\mathbf{g}}_1^{(m)T}, \dots, \tilde{\mathbf{g}}_{L_r}^{(m)T}]^T$ . Likewise, let  $\boldsymbol{\theta}^{(m)} \triangleq [\boldsymbol{\theta}_0^{(m)T}, \dots, \boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)T}]^T$ . Now, the vector of parameters to be estimated is defined as  $\boldsymbol{\xi}^{(m)} \in \mathbb{C}^{(K(L_r+N_{\text{tr}})+1) \times 1}$ , given by  $\boldsymbol{\xi}^{(m)} \triangleq [\Delta f^{(m)}, \tilde{\mathbf{g}}^{(m)T}, \boldsymbol{\theta}^{(m)T}]^T$ .

#### 4.4.1 Computation of the HIM

In this section, I derive the HIM of the vector of parameters  $\boldsymbol{\xi}^{(m)}$  and derive the hybrid CRLB for any unbiased estimator of  $\boldsymbol{\xi}^{(m)}$ . Since there is prior knowledge on the PN parameters in  $\boldsymbol{\theta}_t^{(m)}$ ,  $0 \leq t \leq N_{\text{tr}}$ , the HIM  $\mathbf{H}(\boldsymbol{\xi}^{(m)})$  can be defined as [148]

$$\mathbf{H}(\boldsymbol{\xi}^{(m)}) = \mathbf{I}_D(\boldsymbol{\xi}^{(m)}) + \mathbf{I}_P(\boldsymbol{\xi}^{(m)}), \quad (4.16)$$

where

$$\mathbf{I}_D(\boldsymbol{\xi}^{(m)}) \triangleq \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \{ \mathbf{I}(\boldsymbol{\xi}^{(m)}) \}, \quad (4.17)$$

with  $\mathbf{I}(\boldsymbol{\xi}^{(m)})$  denoting the FIM and

$$\mathbf{I}_P(\boldsymbol{\xi}^{(m)}) \triangleq -\mathbb{E}_{\boldsymbol{\theta}^{(m)} | \tilde{\mathbf{g}}^{(m)}, \Delta f^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)} | \tilde{\mathbf{g}}^{(m)}, \Delta f^{(m)})}{\partial \boldsymbol{\xi}^{(m)} \boldsymbol{\xi}^{(m)T}} \right\} \quad (4.18)$$

is the prior information matrix with  $p(\boldsymbol{\theta}^{(m)} | \tilde{\mathbf{g}}^{(m)}, \Delta f^{(m)})$  denoting the prior distribution of the PN vector given the equivalent beamformed channels  $\tilde{\mathbf{g}}^{(m)}$  and the CFO  $\Delta f^{(m)}$ .

The FIM associated to  $\boldsymbol{\xi}^{(m)}$ ,  $\mathbf{I}(\boldsymbol{\xi}^{(m)}) \in \mathbb{R}^{(K(L_r+N_{\text{tr}})+1) \times (K(L_r+N_{\text{tr}})+1)}$ , can be expressed as [135]

$$[\mathbf{I}(\boldsymbol{\xi}^{(m)})]_{r,c} = \frac{2}{\sigma^2} \sum_{i=1}^{L_r} \text{Re} \left\{ \frac{\partial \boldsymbol{\mu}_i^{(m)*}(\boldsymbol{\xi}^{(m)})}{\partial \xi_r^{(m)}} \frac{\partial \boldsymbol{\mu}_i^{(m)}(\boldsymbol{\xi}^{(m)})}{\partial \xi_c^{(m)}} \right\}, \quad 1 \leq r, c \leq K(L_r+N_{\text{tr}})+1. \quad (4.19)$$

Let  $\mathbf{e}_m^n \in \mathbb{R}^n$  denote the  $m$ -th canonical vector in  $\mathbb{R}^n$ , and let  $p[t, \ell] = n_0 + L_c + (t-1)(K + L_c) + \ell$ . Then, the terms  $\frac{\partial \mu_i^{(m)}(\boldsymbol{\xi}^{(m)})}{\partial \xi_r^{(m)}}$  are given by

$$\frac{\partial \mu_i^{(m)}(\boldsymbol{\xi}^{(m)})}{\partial \xi_r^{(m)}} = \begin{cases} \mathbf{j} \mathbf{M} \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P}(\boldsymbol{\theta}^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{g}_i^{(m)} & \xi_r^{(m)} = \Delta f^{(m)} \\ \mathbf{j} \text{diag} \left\{ \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{g}_i^{(m)} \right\} e^{\mathbf{j} \theta_t^{(m)}[\ell]} \mathbf{e}_{p[t, \ell]}^{K N_{\text{tr}}} & \xi_r^{(m)} = \theta_t^{(m)}[\ell] \\ e^{\mathbf{j} \beta_i^{(m)}[k]} \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P}(\boldsymbol{\theta}^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{e}_k^K & \xi_r^{(m)} = \alpha_i^{(m)}[k] \\ \mathbf{j} \mathbf{g}_i^{(m)}[k] \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P}(\boldsymbol{\theta}^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{e}_k^K & \xi_r^{(m)} = \beta_i^{(m)}[k]. \end{cases} \quad (4.20)$$

where  $\mathbf{M} \in \mathbb{C}^{K \times K}$  is given by  $\mathbf{M} \triangleq \bigoplus_{t=0}^{N_{\text{tr}}-1} \mathbf{M}[t]$ , with  $\mathbf{M}[t]$  given by  $\mathbf{M}[t] = 2\pi \bigoplus_{n=0}^{K-1} (k_0[t] + n)$ . The FIM can be structured as

$$\mathbf{I}(\boldsymbol{\xi}^{(m)}) = \frac{2}{\sigma^2} \text{Re} \left\{ \begin{bmatrix} i_{\Delta f^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{i}_{\Delta f^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \mathbf{i}_{\tilde{\mathbf{g}}^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \mathbf{i}_{\boldsymbol{\theta}^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\boldsymbol{\theta}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) \end{bmatrix} \right\}. \quad (4.21)$$

The element  $i_{\Delta f^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)})$  is given by

$$\begin{aligned} i_{\Delta f^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) &= \sum_{i=1}^{L_r} \mathbf{g}_i^{(m)*} \mathbf{S}^{(m)*} \mathbf{F}_{\otimes} \mathbf{M}^* \mathbf{M} \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{g}_i^{(m)} \\ &= \sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}-1} \mathbf{g}_i^{(m)*} \mathbf{S}_t^{(m)*} \mathbf{F} \mathbf{M}^*[t] \mathbf{M}[t] \mathbf{F}^* \mathbf{S}_t^{(m)} \mathbf{g}_i^{(m)}. \end{aligned} \quad (4.22)$$

The vector  $\mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)})$  can be expressed as

$$\mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_1^{(m)}}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_{L_r}^{(m)}}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (4.23)$$

with  $\mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_i^{(m)}}(\boldsymbol{\xi}^{(m)})$ ,  $1 \leq i \leq L_r$ , being given by

$$\mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_i^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_i^{(m)}[0]}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_i^{(m)}[K-1]}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (4.24)$$

and the vector  $\mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)})$ ,  $0 \leq k \leq K - 1$ , is given by

$$\begin{aligned} \mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) &= \begin{bmatrix} i_{\Delta f^{(m)}, \alpha_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) & i_{\Delta f^{(m)}, \beta_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) \end{bmatrix} \\ &= \sum_{t=0}^{N_{\text{tr}}-1} \begin{bmatrix} -j e^{j\beta_i^{(m)}[k]} & \mathbf{g}_i^{(m)}[k] \end{bmatrix} \mathbf{g}_i^{(m)*} \mathbf{S}_t^{(m)*} \mathbf{F} \mathbf{M}[t] \mathbf{F}^* \mathbf{S}_t^{(m)} \mathbf{e}_k^K. \end{aligned} \quad (4.25)$$

Furthermore, the vector  $\mathbf{i}_{\Delta f^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)})$  can be written as

$$\mathbf{I}_{\Delta f^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_{\Delta f^{(m)}, \boldsymbol{\theta}_0^{(m)}}(\boldsymbol{\xi}^{(m)}) & \dots & \mathbf{I}_{\Delta f^{(m)}, \boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)}}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (4.26)$$

where  $\mathbf{i}_{\Delta f^{(m)}, \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)})$ ,  $0 \leq t \leq N_{\text{tr}} - 1$ , is given by

$$\mathbf{i}_{\Delta f^{(m)}, \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} i_{\Delta f^{(m)}, \boldsymbol{\theta}_t^{(m)}[0]}(\boldsymbol{\xi}^{(m)}) & \dots & i_{\Delta f^{(m)}, \boldsymbol{\theta}_t^{(m)}[K-1]}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}. \quad (4.27)$$

Each of the terms in (4.27) can be expressed as

$$i_{\Delta f^{(m)}, \boldsymbol{\theta}_t^{(m)}[\ell]}(\boldsymbol{\xi}^{(m)}) = \sum_{i=1}^{L_r} \mathbf{g}_i^{(m)*} \mathbf{S}_t^{(m)*} \mathbf{F} \mathbf{M}[t] \text{diag} \left\{ \mathbf{F}^* \mathbf{S}_t^{(m)} \mathbf{g}_i^{(m)} \right\} \mathbf{e}_\ell^K. \quad (4.28)$$

The submatrix  $\mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)})$  can be expressed as

$$\mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_{\tilde{\mathbf{g}}_1^{(m)}, \tilde{\mathbf{g}}_{L_r}^{(m)}}(\boldsymbol{\xi}^{(m)}) & \dots & \mathbf{I}_{\tilde{\mathbf{g}}_1^{(m)}, \tilde{\mathbf{g}}_1^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{\tilde{\mathbf{g}}_{L_r}^{(m)}, \tilde{\mathbf{g}}_1^{(m)}}(\boldsymbol{\xi}^{(m)}) & \dots & \mathbf{I}_{\tilde{\mathbf{g}}_{L_r}^{(m)}, \tilde{\mathbf{g}}_{L_r}^{(m)}}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (4.29)$$

which has a particularly interesting structure from an estimation theoretic perspective. By observing the Kronecker structure in (4.15), as well as the derivatives in (4.20), it is clear that  $\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}, \tilde{\mathbf{g}}_j^{(m)}}(\boldsymbol{\xi}^{(m)}) = \mathbf{0}$  for  $i \neq j$ . Therefore,

only the terms  $\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}, \tilde{\mathbf{g}}_i^{(m)}}(\boldsymbol{\xi}^{(m)})$  are non-zero valued, which can be computed as

$$\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}, \tilde{\mathbf{g}}_i^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[0], \tilde{\mathbf{g}}_i^{(m)}[0]}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[0], \tilde{\mathbf{g}}_i^{(m)}[K-1]}(\boldsymbol{\xi}^{(m)}) \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[K-1], \tilde{\mathbf{g}}_i^{(m)}[0]}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[K-1], \tilde{\mathbf{g}}_i^{(m)}[K-1]}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}. \quad (4.30)$$

The matrix in (4.30) has a similar structure to that of (4.29). From (4.20) and (4.14), it is observed that  $\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[k_1], \tilde{\mathbf{g}}_i^{(m)}[k_2]}(\boldsymbol{\xi}^{(m)}) = \mathbf{0}$  if  $k_1 \neq k_2$ . The non-zero matrices  $\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[k], \tilde{\mathbf{g}}_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)})$  can be written as

$$\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[k], \tilde{\mathbf{g}}_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_{\alpha_i^{(m)}[k], \alpha_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\alpha_i^{(m)}[k], \beta_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) \\ \mathbf{I}_{\beta_i^{(m)}[k], \alpha_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\beta_i^{(m)}[k], \beta_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}. \quad (4.31)$$

By plugging the corresponding partial derivatives in (4.20) into (4.19), the matrix in (4.31) is given by

$$\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[k], \tilde{\mathbf{g}}_i^{(m)}[k]}(\boldsymbol{\xi}^{(m)}) = \sum_{t=0}^{N_{\text{tr}}-1} \left| [\mathbf{s}_t^{(m)}]_{k,k} \right|^2 \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \alpha_i^{(m)2}[k] \end{bmatrix}}_{\boldsymbol{\Lambda}_i^{(m)}[k]}, \quad (4.32)$$

which shows that estimation of both the amplitude and phase of a given subchannel  $\tilde{\mathbf{g}}_i^{(m)}[k]$  do not interfere with each other, a result that has been shown in [2]. Let  $\mathbf{s}_k^{(m)} \in \mathbb{C}^{N_{\text{tr}} \times 1}$  be the vector containing the training pilots for a given subcarrier and all the transmitted OFDM symbols,  $\mathbf{s}_k^{(m)} = [\mathbf{s}_0^{(m)}[k], \dots, \mathbf{s}_{N_{\text{tr}}-1}^{(m)}[k]]^T$ . Then, the matrix in (4.30) can be expressed as

$$\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}, \tilde{\mathbf{g}}_i^{(m)}}(\boldsymbol{\xi}^{(m)}) = \bigoplus_{k=0}^{K-1} \left\| \mathbf{s}_k^{(m)} \right\|_2^2 \boldsymbol{\Lambda}_i^{(m)}[k], \quad (4.33)$$

and (4.29) can be written as

$$\mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \bigoplus_{i=1}^{L_r} \bigoplus_{k=0}^{K-1} \left\| \mathbf{s}_k^{(m)} \right\|_2^2 \boldsymbol{\Lambda}_i^{(m)}[k]. \quad (4.34)$$

Now, the block  $\mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)})$  can be expressed as

$$\mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_{\tilde{\mathbf{g}}_1^{(m)}, \boldsymbol{\theta}_0^{(m)}}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{I}_{\tilde{\mathbf{g}}_1^{(m)}, \boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{\tilde{\mathbf{g}}_{L_r}^{(m)}, \boldsymbol{\theta}_0^{(m)}}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{I}_{\tilde{\mathbf{g}}_{L_r}^{(m)}, \boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)}}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (4.35)$$

wherein the blocks  $\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}, \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)})$  are of the form

$$\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}, \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[0], \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \vdots \\ \mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[K-1], \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (4.36)$$

with  $\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[k], \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)})$  given by

$$\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[k], \boldsymbol{\theta}_t^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{i}_{\tilde{\mathbf{g}}_i^{(m)}[k], \boldsymbol{\theta}_t^{(m)}[0]}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{i}_{\tilde{\mathbf{g}}_i^{(m)}[k], \boldsymbol{\theta}_t^{(m)}[K-1]}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}. \quad (4.37)$$

Let  $\mathbf{s}_t^{(m)} \in \mathbb{C}^{K \times K}$  be the column vector containing the training pilots for the  $t$ -th transmitted OFDM symbol. This vector is given by  $\mathbf{s}_t^{(m)} = \text{vec}\{\text{diag}\{\mathbf{S}_t^{(m)}\}\}$ . Furthermore, let  $\mathbf{f}_\ell \in \mathbb{C}^{K \times 1}$  be the  $\ell$ -th column in the Discrete Fourier Transform (DFT) matrix  $\mathbf{F}$ . Then, plugging the corresponding derivatives from (4.20) into (4.19) allows expressing each column in (4.37) as

$$\mathbf{I}_{\tilde{\mathbf{g}}_i^{(m)}[k], \boldsymbol{\theta}_t^{(m)}[l]}(\boldsymbol{\xi}^{(m)}) = \text{Re} \left\{ \begin{bmatrix} j e^{-j\beta_i^{(m)}[k]} \\ \mathbf{g}_i^{(m)\text{C}}[k] \end{bmatrix} (\mathbf{e}_k^K)^T (\mathbf{s}_t^{(m)\text{C}} \circ \mathbf{f}_\ell) (\mathbf{s}_t^{(m)T} \circ \mathbf{f}_\ell^*) \mathbf{g}_i^{(m)} \right\}. \quad (4.38)$$

Finally, the block matrix  $\mathbf{I}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)})$  can be expressed in the form

$$\mathbf{I}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} \mathbf{I}_{\boldsymbol{\theta}_0^{(m)}, \boldsymbol{\theta}_0^{(m)}}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{I}_{\boldsymbol{\theta}_0^{(m)}, \boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{\boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)}, \boldsymbol{\theta}_0^{(m)}}(\boldsymbol{\xi}^{(m)}) & \cdots & \mathbf{I}_{\boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)}, \boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)}}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}. \quad (4.39)$$

Owing to the structure of the partial derivative of  $\boldsymbol{\mu}_i^{(m)}(\boldsymbol{\xi}^{(m)})$  with respect to  $\theta_t^{(m)}[\ell]$  in (4.20), the different matrices  $\mathbf{I}_{\theta_t^{(m)}, \theta_u^{(m)}}(\boldsymbol{\xi}^{(m)})$  can be checked to be zero-valued for  $t \neq u$ . Further, the matrices in the main block diagonal of  $\mathbf{I}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)})$  can be expressed as

$$\mathbf{I}_{\theta_t^{(m)}, \theta_t^{(m)}}(\boldsymbol{\xi}^{(m)}) = \begin{bmatrix} i_{\theta_t^{(m)}[0], \theta_t^{(m)}[0]}(\boldsymbol{\xi}^{(m)}) & \cdots & i_{\theta_t^{(m)}[0], \theta_t^{(m)}[K-1]}(\boldsymbol{\xi}^{(m)}) \\ \vdots & \ddots & \vdots \\ i_{\theta_t^{(m)}[K-1], \theta_t^{(m)}[0]}(\boldsymbol{\xi}^{(m)}) & \cdots & i_{\theta_t^{(m)}[K-1], \theta_t^{(m)}[K-1]}(\boldsymbol{\xi}^{(m)}) \end{bmatrix}, \quad (4.40)$$

which again, due to the structure of the partial derivative of  $\boldsymbol{\mu}_i^{(m)}(\boldsymbol{\xi}^{(m)})$  with respect to  $\theta_t^{(m)}[\ell]$  in (4.20), is a diagonal matrix given by

$$\mathbf{I}_{\theta_t^{(m)}, \theta_t^{(m)}}(\boldsymbol{\xi}^{(m)}) = \bigoplus_{k=0}^{K-1} \left( \sum_{i=1}^{L_r} \left\| \text{diag} \left\{ \mathbf{F}^* \mathbf{S}_t^{(m)} \mathbf{g}_i^{(m)} \right\} \mathbf{e}_{p[t,k]}^{KN_{\text{tr}}} \right\|_2^2 \right). \quad (4.41)$$

Using (4.41), the matrix in (4.39) can be expressed as

$$\mathbf{I}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \bigoplus_{t=0}^{N_{\text{tr}}-1} \bigoplus_{k=0}^{K-1} \left( \sum_{i=1}^{L_r} \left\| \text{diag} \left\{ \mathbf{F}^* \mathbf{S}_t^{(m)} \mathbf{g}_i^{(m)} \right\} \mathbf{e}_{p[t,k]}^{KN_{\text{tr}}} \right\|_2^2 \right). \quad (4.42)$$

Due to the structure of the FIM, the matrices below the main block diagonal in (4.21) are given by  $\mathbf{I}_{\bar{\mathbf{g}}^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) = \mathbf{I}_{\Delta f^{(m)}, \bar{\mathbf{g}}^{(m)}}^T(\boldsymbol{\xi}^{(m)})$ ,  $\mathbf{I}_{\boldsymbol{\theta}^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) = \mathbf{I}_{\Delta f^{(m)}, \boldsymbol{\theta}^{(m)}}^T(\boldsymbol{\xi}^{(m)})$ , and  $\mathbf{I}_{\boldsymbol{\theta}^{(m)}, \bar{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \mathbf{I}_{\bar{\mathbf{g}}^{(m)}, \boldsymbol{\theta}^{(m)}}^T(\boldsymbol{\xi}^{(m)})$ .

Finally, to obtain  $\mathbf{I}_D(\boldsymbol{\xi}^{(m)})$ , notice that the terms in (4.22), (4.25), (4.28), (4.34), (4.38) and (4.42) do not depend on  $\boldsymbol{\theta}^{(m)}$  since the PN exponentials get canceled by their conjugates. Hence, there is no need to calculate the explicit expectation of  $\mathbf{I}(\boldsymbol{\xi}^{(m)})$  over  $\boldsymbol{\theta}^{(m)}$ , and  $\mathbf{I}_D(\boldsymbol{\xi}^{(m)}) = \mathbf{I}(\boldsymbol{\xi}^{(m)})$ .

Now, the only matrix left to compute in order to find the HIM  $\mathbf{H}(\boldsymbol{\xi}^{(m)})$  is  $\mathbf{I}_P(\boldsymbol{\xi}^{(m)})$  in (C.7). From the expression in (C.9), since no prior knowledge

on either  $\tilde{\mathbf{g}}^{(m)}$  or  $\Delta f^{(m)}$  is assumed,  $\mathbf{I}_P(\boldsymbol{\xi}^{(m)})$  is structured as

$$\begin{aligned} \mathbf{I}_P(\boldsymbol{\xi}^{(m)}) &\triangleq - \begin{bmatrix} \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \Delta f^{(m)2}} \right\} & \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \Delta f^{(m)} \partial \tilde{\mathbf{g}}^{(m)T}} \right\} & \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \Delta f^{(m)} \boldsymbol{\theta}^{(m)T}} \right\} \\ \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \tilde{\mathbf{g}}^{(m)} \partial \Delta f^{(m)}} \right\} & \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \tilde{\mathbf{g}}^{(m)} \partial \tilde{\mathbf{g}}^{(m)T}} \right\} & \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \tilde{\mathbf{g}}^{(m)} \boldsymbol{\theta}^{(m)T}} \right\} \\ \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}^{(m)} \partial \Delta f^{(m)}} \right\} & \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}^{(m)} \partial \tilde{\mathbf{g}}^{(m)T}} \right\} & \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}^{(m)} \boldsymbol{\theta}^{(m)T}} \right\} \end{bmatrix} \\ &= - \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left\{ \frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}^{(m)} \boldsymbol{\theta}^{(m)T}} \right\} \end{bmatrix}, \end{aligned} \quad (4.43)$$

where the last equality comes from the PDF of the PN being independent of the CFO and equivalent channel gains. The LLF of the PN is given by

$$\log p(\boldsymbol{\theta}^{(m)}) = -\frac{KN_{\text{tr}}}{2} \log(2\pi) - \frac{N_{\text{tr}}}{2} \log \det \{ \mathbf{C}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}} \} - \frac{1}{2} \boldsymbol{\theta}^{(m)T} \mathbf{C}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}^{-1} \boldsymbol{\theta}^{(m)}, \quad (4.44)$$

and its Hessian reads

$$\frac{\partial^2 \log p(\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}^{(m)} \boldsymbol{\theta}^{(m)T}} = -\mathbf{C}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}^{-1}. \quad (4.45)$$

Therefore, by combining  $\mathbf{I}_D(\boldsymbol{\xi}^{(m)})$  and  $\mathbf{I}_P(\boldsymbol{\xi}^{(m)})$ , the HIM  $\mathbf{H}(\boldsymbol{\xi}^{(m)})$  is obtained

as

$$\begin{aligned} \mathbf{H}(\boldsymbol{\xi}^{(m)}) &= \begin{bmatrix} i_{\mathbf{H},1,1}(\boldsymbol{\xi}^{(m)}) & \mathbf{i}_{\mathbf{H},1,2}^T(\boldsymbol{\xi}^{(m)}) & \mathbf{i}_{\mathbf{H},1,3}^T(\boldsymbol{\xi}^{(m)}) \\ \mathbf{i}_{\mathbf{H},2,1}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\mathbf{H},2,2}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\mathbf{H},2,3}(\boldsymbol{\xi}^{(m)}) \\ \mathbf{i}_{\mathbf{H},3,1}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\mathbf{H},3,2}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\mathbf{H},3,3}(\boldsymbol{\xi}^{(m)}) \end{bmatrix} \\ &= \begin{bmatrix} i_{\Delta f^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{i}_{\Delta f^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{i}_{\Delta f^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \mathbf{i}_{\tilde{\mathbf{g}}^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) \\ \mathbf{i}_{\boldsymbol{\theta}^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\boldsymbol{\theta}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) & \mathbf{I}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}(\boldsymbol{\xi}^{(m)}) + \mathbf{C}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}^{-1} \end{bmatrix}. \end{aligned} \quad (4.46)$$

Finally, the hybrid CRLB is given by the inverse of the HIM,  $\mathbf{H}^{-1}(\boldsymbol{\xi}^{(m)})$ . In particular, using the formula for the inverse of block matrices [135], the hybrid CRLB for the CFO can be found as follows. Let  $\tilde{i}_{\Delta f^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) \in \mathbb{R}$ ,  $\tilde{\mathbf{I}}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) \in \mathbb{R}^{2KL_r \times 2KL_r}$  and  $\mathbf{x}(\boldsymbol{\xi}^{(m)}) \in \mathbb{R}^{2KL_r \times 1}$  denote the HIM for the CFO parameter when the channel  $\tilde{\mathbf{g}}$  is known, the HIM for the channels when the CFO is known, and a vector accounting for the coupling between the PN, channel, and CFO parameters. These parameters are given by

$$\tilde{i}_{\Delta f^{(m)}, \Delta f^{(m)}}(\boldsymbol{\xi}^{(m)}) = i_{\text{H},1,1}(\boldsymbol{\xi}^{(m)}) - \mathbf{i}_{\text{H},1,3}^T(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\text{H},3,3}^{-1}(\boldsymbol{\xi}^{(m)}) \mathbf{i}_{\text{H},3,1}(\boldsymbol{\xi}^{(m)}) \quad (4.47)$$

$$\tilde{\mathbf{I}}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}(\boldsymbol{\xi}^{(m)}) = \mathbf{I}_{\text{H},2,2}(\boldsymbol{\xi}^{(m)}) - \mathbf{I}_{\text{H},2,3}(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\text{H},3,3}^{-1}(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\text{H},3,2}(\boldsymbol{\xi}^{(m)}) . \quad (4.48)$$

$$\mathbf{x}(\boldsymbol{\xi}^{(m)}) = \mathbf{i}_{\text{H},1,2}(\boldsymbol{\xi}^{(m)}) - \mathbf{i}_{\text{H},1,3}^T(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\text{H},3,3}^{-1}(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\text{H},2,3}(\boldsymbol{\xi}^{(m)}) . \quad (4.49)$$

Then, the hybrid CRLB for any unbiased estimator of  $\Delta f^{(m)}$ ,  $\tilde{\mathbf{g}}^{(m)}$  are given by

$$\text{var} \left\{ \widehat{\Delta f}^{(m)} \right\} \geq \frac{1}{\tilde{i}_{\Delta f}^{(m)}(\boldsymbol{\xi}^{(m)}) - \mathbf{x}^T(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}^{-1}(\boldsymbol{\xi}^{(m)}) \mathbf{x}(\boldsymbol{\xi}^{(m)})} , \quad (4.50)$$

$$\begin{aligned} \text{covar} \left\{ \hat{\tilde{\mathbf{g}}}^{(m)}, \hat{\mathbf{g}}^{(m)} \right\} &\geq \tilde{\mathbf{I}}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}^{-1}(\boldsymbol{\xi}^{(m)}) \\ &+ \frac{\mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}^{-1}(\boldsymbol{\xi}^{(m)}) \mathbf{x}(\boldsymbol{\xi}^{(m)}) \mathbf{x}^T(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}^{-1}(\boldsymbol{\xi}^{(m)})}{\tilde{i}_{\Delta f}^{(m)}(\boldsymbol{\xi}^{(m)}) - \mathbf{x}^T(\boldsymbol{\xi}^{(m)}) \mathbf{I}_{\tilde{\mathbf{g}}^{(m)}, \tilde{\mathbf{g}}^{(m)}}^{-1}(\boldsymbol{\xi}^{(m)}) \mathbf{x}(\boldsymbol{\xi}^{(m)})} . \end{aligned} \quad (4.51)$$

The main insights from equations (4.50) and (4.51) are two fold: i) The hybrid CRLB for the estimation of the CFO parameter can be interpreted as the hybrid CRLB for CFO estimation plus an additional term that gathers the

information coupling between the CFO, beamformed channels and PN impairments, and ii) the hybrid CRLB for the estimation of the beamformed channels is the hybrid CRLB for beamformed channel estimation in the absence of CFO plus a correction term that comprises the information coupling between the CFO, beamformed channels and PN impairments, scaled by the hybrid CRLB for the estimation of the CFO. Since the hybrid CRLB for CFO estimation is expected to be small (similar to the result in Chapter 3), the hybrid CRLB for the estimation of the beamformed channels will be mainly determined by the first term in (4.51), thereby reducing the impact of the information coupling between the CFO and the PN on the resulting hybrid CRLB.

## 4.5 Estimation of equivalent beamformed channels and high-dimensional MIMO channel

In this section, I formulate and present novel solutions to the problem of estimating both the CFO, the equivalent frequency-selective beamformed channels, and the PN vector for the signal model in Section 4.3. Then, I formulate the problem of estimating the high-dimensional frequency-selective mmWave MIMO channel  $\{\mathbf{H}[k]\}_{k=0}^{K-1}$  from the estimates of the equivalent channel accounting for both the estimates for these parameters and their hybrid CRLB. Since prior statistical information on the PN vector is available, it is well-known that the optimum estimator for the PN is the Minimum Mean Square Error (MMSE) estimator, which is well-known to be unbiased and attain the hybrid CRLB. The main problem concerning applying the MMSE

estimator is that it requires knowledge of the CFO and the equivalent channels, which is not available a priori. Another strategy to find the PN vector relies on using the Maximum A Posteriori (MAP) estimator, which is attractive due to its simplicity, but it presents the drawback of being, in general, biased. Due to this, the application of the MAP estimator may well lead to the different estimates  $\hat{\mathbf{g}}_i^{(m)}$ ,  $1 \leq i \leq L_r$ ,  $1 \leq m \leq M$  having random phase errors that could destroy incoherence in the measurements, thereby invalidating the application of CS-based algorithms to retrieve the frequency-selective channel  $\{\mathbf{H}[k]\}_{k=0}^{K-1}$ . For this reason, it is crucial to consider an unbiased estimator for the different parameters. Owing to the difficulty in finding a closed-form solution for the estimation of  $\Delta f^{(m)}$ ,  $\{\mathbf{g}_i^{(m)}\}_{i=1}^{L_r}$ , and  $\{\boldsymbol{\theta}_t^{(m)}\}_{t=0}^{N_{\text{tr}}-1}$ , I propose to use the EM approach [135] to find these estimators. I will show that this leads to finding the MMSE estimator for the PN impairment, parameterized by the current estimates of the unknown CFO and equivalent channels, which can be computed as it will soon become apparent. The EM method is a well-known iterative approach to find the ML estimators for unknown parameters when the LLF is unknown, and hence impossible to optimize directly. The first proposed algorithm aims at finding the LMMSE estimator for the PN by batch processing the  $L_r K N_{\text{tr}}$  received measurements at once, thereby providing very good performance. The second proposed algorithm also aims at finding the LMMSE estimator for the PN but, unlike the first proposed algorithm, it processes the received measurements in sets of  $L_r$  samples to reduce computational complexity. A block diagram of the proposed estimation approach is shown in Fig.

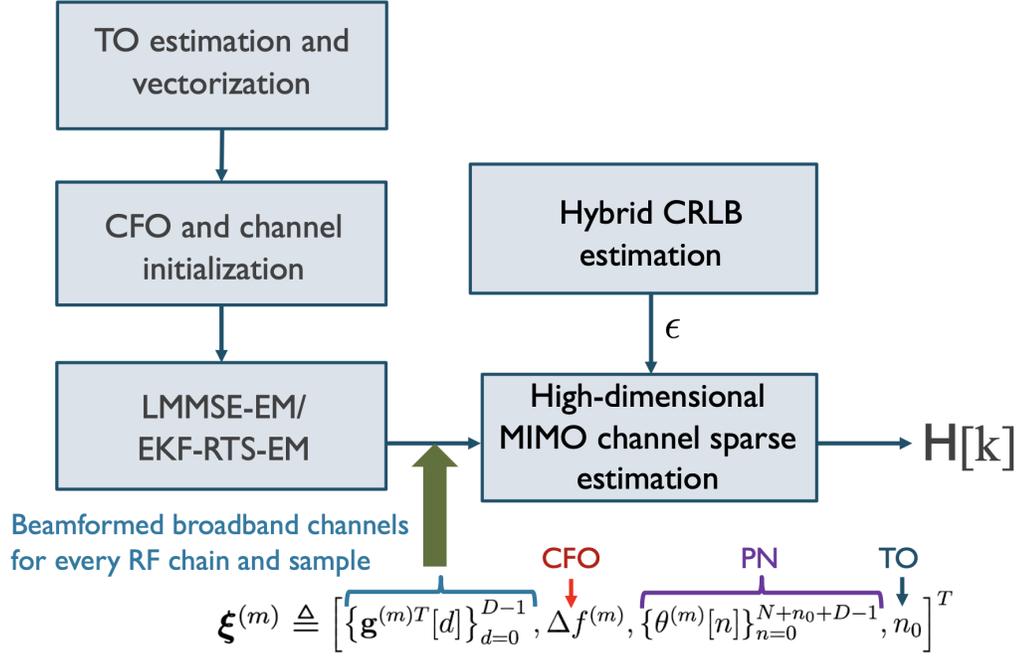


Figure 4.3: Block diagram of the proposed estimation-theoretic framework for joint synchronization and compressive channel estimation under broadband mmWave MIMO channels.

4.3.

#### 4.5.1 LMMSE-EM Algorithm

In this subsection, I present the first proposed algorithm to find the ML estimates for the CFO and the equivalent beamformed channels using the EM iterative estimation approach. At each iteration, this algorithm processes all the  $L_r K N_{tr}$  received measurements using single-shot estimation to find a closed-form solution to the problem of estimating the PN vector. The EM algorithm consists of two steps:

- **E-step:** in the first step of the EM algorithm, the posterior expected value of the joint LLF of  $\mathbf{y}^{(m)}$  and  $\boldsymbol{\xi}_R^{(m)}$  is computed. Let us consider a partition of the vector of parameters to be estimated,  $\boldsymbol{\xi}^{(m)}$ , into a vector of deterministic parameters  $\boldsymbol{\xi}_D^{(m)} = [\Delta f^{(m)}, \mathbf{g}_1^{(m)T}, \dots, \mathbf{g}_{L_r}^{(m)T}]^T$ , and a vector of random parameters  $\boldsymbol{\xi}_R^{(m)} = [\boldsymbol{\theta}_0^{(m)T}, \dots, \boldsymbol{\theta}_{N_{tr}-1}^{(m)T}]^T$ . Then, the expectation step for the  $n$ -th step can be formalized as

$$Q\left(\boldsymbol{\xi}_D^{(m)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)}\right) \triangleq \mathbb{E}_{\boldsymbol{\xi}_R^{(m)} | \mathbf{y}^{(m)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)}} \left\{ \log p\left(\mathbf{y}^{(m)}, \boldsymbol{\xi}_R^{(m)}; \boldsymbol{\xi}_D^{(m)}\right) \right\}, \quad (4.52)$$

where  $\hat{\boldsymbol{\xi}}_D^{(m,n)}$  is the estimate of  $\boldsymbol{\xi}_D^{(m)}$  found at the  $n$ -th iteration of the algorithm.

- **M-step:** this step consists of finding  $\hat{\boldsymbol{\xi}}_D^{(m,n)}$ , which is defined as the maximizer of the function found during the E-step. The maximization step is formalized as

$$\hat{\boldsymbol{\xi}}_D^{(m,n)} = \arg \max_{\boldsymbol{\xi}_D^{(m)}} Q\left(\boldsymbol{\xi}_D^{(m)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)}\right). \quad (4.53)$$

Due to the independence of the PN sequence on the deterministic parameters in  $\boldsymbol{\xi}_D^{(m)}$ , the  $Q$  function in (4.52) can be expressed as

$$Q\left(\boldsymbol{\xi}_D^{(m)}, \hat{\boldsymbol{\xi}}_D^{(m,n)}\right) = -\frac{1}{\sigma^2} \sum_{i=1}^{L_r} \left\| \mathbf{y}_i^{(m)} - \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P}\left(\hat{\boldsymbol{\theta}}_{\text{MMSE}}^{(m,n)}\right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{g}_i^{(m)} \right\|_2^2, \quad (4.54)$$

where  $\hat{\boldsymbol{\theta}}_{\text{MMSE}}^{(m,n)} \triangleq \mathbb{E}_{\boldsymbol{\xi}_R^{(m)} | \mathbf{y}^{(m)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)}} \{\boldsymbol{\theta}^{(m)}\}$  is the MMSE estimator of the PN sequence found during the  $n$ -th E-step.

Finding the MMSE estimator of the PN sequence requires finding the posterior PDF of the PN sequence, given the received measurements  $\mathbf{y}_i^{(m)}$ ,  $1 \leq i \leq L_r$ . Finding this PDF, however, requires multi-dimensional integration over the joint PDF of the received measurements and the PN sequence, which is difficult to find, in general. For this reason, I propose another approach to estimate the PN as follows. Exploiting the fact that the PN sequence typically has small amplitude [5], I use a first-order Taylor series approximation to linearize the received measurement with respect to the PN sequence around the expected value of  $\boldsymbol{\theta}^{(m)}$ , given by  $\boldsymbol{\mu}_{\boldsymbol{\theta}^{(m)}}$ , as

$$\mathbf{y}^{(m)} \approx \underbrace{\left( \mathbf{I}_{L_r} \otimes \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P}(\boldsymbol{\theta}^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \right) \mathbf{g}^{(m)}}_{\mathbf{h}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})} + \nabla_{\mathbf{f}}(\boldsymbol{\theta}^{(m)}) \Big|_{\boldsymbol{\theta}^{(m)} = \boldsymbol{\mu}_{\boldsymbol{\theta}^{(m)}}} \boldsymbol{\theta}^{(m)} + \mathbf{v}^{(m)}, \quad (4.55)$$

where  $\nabla_{\mathbf{f}}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})$  is the Jacobian matrix of  $\mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)}) : \mathbb{R}^{KL_r N_{\text{tr}}} \rightarrow \mathbb{C}^{KL_r N_{\text{tr}}}$ ,  $\mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)}) = (\mathbf{I}_{L_r} \otimes \boldsymbol{\Omega}(\Delta f^{(m)})) \mathbf{P}(\boldsymbol{\theta}^{(m)}) (\mathbf{I}_{L_r} \otimes \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)}) \mathbf{g}^{(m)}$ , which is given by

$$\nabla_{\mathbf{f}}(\boldsymbol{\theta}^{(m)}) = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})}{\partial \boldsymbol{\theta}_0^{(m)T}} \quad \cdots \quad \frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})}{\partial \boldsymbol{\theta}_{N_{\text{tr}}-1}^{(m)T}} \right]. \quad (4.56)$$

Each of the submatrices  $\frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})}{\partial \boldsymbol{\theta}_t^{(m)T}} \in \mathbb{C}^{KL_r N_{\text{tr}} \times K}$  is given by

$$\frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})}{\partial \boldsymbol{\theta}_t^{(m)T}} = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})}{\partial \theta_{t,0}^{(m)}} \quad \cdots \quad \frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})}{\partial \theta_{t,K-1}^{(m)}} \right], \quad (4.57)$$

wherein  $\frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)})}{\partial \theta_{t,k}^{(m)}}$  is given by

$$\begin{aligned} \frac{\partial \mathbf{f}(\boldsymbol{\theta}^{(m)}, \boldsymbol{\xi}_D^{(m)})}{\partial \theta_{t,k}^{(m)}} &= \mathbf{j} e^{\mathbf{j} \theta_{t,k}^{(m)}} \left( \mathbf{I}_{L_r} \otimes \mathbf{e}_{p[t,\ell]}^{KN_{\text{tr}}} \left( \mathbf{e}_{p[t,\ell]}^{KN_{\text{tr}}} \right)^T \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \right) \mathbf{g}^{(m)} \\ &= \mathbf{j} e^{\mathbf{j} \theta_{t,k}^{(m)}} \begin{bmatrix} \text{diag} \left\{ \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{g}_1^{(m)} \right\} \mathbf{e}_{p[t,\ell]}^{KN_{\text{tr}}} \\ \vdots \\ \text{diag} \left\{ \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{g}_{L_r}^{(m)} \right\} \mathbf{e}_{p[t,\ell]}^{KN_{\text{tr}}} \end{bmatrix} \end{aligned} \quad (4.58)$$

Using (4.56)-(4.58), the time and measurement update equations for the estimation of  $\boldsymbol{\xi}_R^{(m)} = \boldsymbol{\theta}^{(m)}$  at the  $n$ -th E-step are given by

- **Time update**

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m,n)} &= \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(m)} \\ \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}, \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}}^{(n)} &= \mathbf{C}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}^{(m)}. \end{aligned} \quad (4.59)$$

- **Measurement update**

$$\begin{aligned} \hat{\mathbf{C}}_{\boldsymbol{\theta}^{(m)}, \mathbf{y}^{(m)}}^{(n)} &= \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}, \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}}^{(n)} \nabla_{\mathbf{f}}^* \left( \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m,n)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)} \right) \\ \hat{\mathbf{C}}_{\mathbf{y}^{(m)}, \mathbf{y}^{(m)}}^{(n)} &= \nabla_{\mathbf{f}} \left( \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m,n)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)} \right) \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}, \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}}^{(n)} \nabla_{\mathbf{f}}^* \left( \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m,n)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)} \right) + \sigma^2 \mathbf{I}_{KN_{\text{tr}} L_r} \\ \hat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n)} &= \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m,n)} + \hat{\mathbf{C}}_{\boldsymbol{\theta}^{(m)}, \mathbf{y}^{(m)}}^{(n)} \left( \hat{\mathbf{C}}_{\mathbf{y}^{(m)}, \mathbf{y}^{(m)}}^{(n)} \right)^{-1} \left( \mathbf{y}^{(m)} - \mathbf{h} \left( \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}, \hat{\boldsymbol{\xi}}_D^{(m,n-1)} \right) \right) \\ \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\text{MU}}^{(m)}, \hat{\boldsymbol{\theta}}_{\text{MU}}^{(m)}}^{(n)} &= \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}, \hat{\boldsymbol{\theta}}_{\text{TU}}^{(m)}}^{(n)} - \hat{\mathbf{C}}_{\boldsymbol{\theta}^{(m)}, \mathbf{y}^{(m)}}^{(n)} \left( \hat{\mathbf{C}}_{\mathbf{y}^{(m)}, \mathbf{y}^{(m)}}^{(n)} \right)^{-1} \hat{\mathbf{C}}_{\mathbf{y}^{(m)}, \boldsymbol{\theta}^{(m)}}^{(n)}. \end{aligned} \quad (4.60)$$

Finally, motivated by the linearization in (4.55) and the assumption that the PN sequence is Gaussian [97, 98], the MMSE estimator for the PN sequence at

the  $n$ -th E-step is substituted by the approximate LMMSE estimate obtained by the EKF recursions in (4.59)-(4.60).

Then, the optimum ML estimator found during the  $n$ -th M-step is found by maximizing (4.52). Optimizing (4.52) directly is, however, computationally complex because of the lack of closed-form solutions for the estimation of  $\Delta f^{(m)}$  [141]. Therefore, to circumvent this issue, I propose to reduce the complexity associated with the M-step by carrying out the optimization in (4.53) with respect to one of the parameters while keeping the remaining parameters at their most recently updated values. First, by using the equivalent channel estimates at the  $(n-1)$ -th E-step,  $\hat{\mathbf{g}}^{(m,n-1)}$ , and the PN vector estimate from the E-step,  $\hat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n)}$ , the function in (4.52) is maximized with respect to  $\Delta f^{(m)}$  to obtain the estimate for the  $n$ -th iteration,  $\widehat{\Delta f}^{(m,n)}$  as

$$\widehat{\Delta f}_{\text{ML}}^{(m,n)} = \arg \min_{\Delta f^{(m)}} \sum_{i=1}^{L_r} \left\| \mathbf{y}_i^{(m)} - \boldsymbol{\Omega}(\Delta f^{(m)}) \mathbf{P} \left( \hat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n)} \right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \hat{\mathbf{g}}_i^{(m,n-1)} \right\|_2^2. \quad (4.61)$$

After simplifying (4.61), it is obtained that

$$\widehat{\Delta f}_{\text{ML}}^{(m,n)} = \arg \max_{\Delta f^{(m)}} \sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}-1} \text{Re} \left\{ \mathbf{y}_{i,t}^{(m)*} \boldsymbol{\Omega}_t(\Delta f^{(m)}) \mathbf{P}_t \left( \hat{\boldsymbol{\theta}}_{t,\text{MU}}^{(m,n)} \right) \mathbf{F}_t^* \mathbf{S}_t^{(m)} \hat{\mathbf{g}}_{i,\text{ML}}^{(m,n-1)} \right\}. \quad (4.62)$$

To resolve the non-linearity in (4.62), I resort to a second-order Taylor series expansion of the function in (4.62) around the previous CFO estimate,  $\widehat{\Delta f}^{(m,n-1)}$ .

For this purpose, let  $\mathbf{h}_{i,t}^{(m,n)} \triangleq \boldsymbol{\Omega}_t \left( \widehat{\Delta f}_{\text{ML}}^{(m,n-1)} \right) \mathbf{P}_t \left( \hat{\boldsymbol{\theta}}_{t,\text{MU}}^{(m,n)} \right) \mathbf{F}_t^* \mathbf{S}_t^{(m)} \hat{\mathbf{g}}_{i,\text{ML}}^{(m,n-1)}$ . Then,

(4.62) can be approximated as

$$\begin{aligned}
\widehat{\Delta f}_{\text{ML}}^{(m,n)} &= \arg \max_{\Delta f^{(m)}} \sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}-1} \text{Re} \left\{ \mathbf{y}_{i,t}^{(m)*} \mathbf{h}_{i,t}^{(m,n)} \right\} \\
&+ \left( \Delta f^{(m)} - \widehat{\Delta f}^{(m,n-1)} \right) \sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}-1} \text{Re} \left\{ \mathbf{y}_{i,t}^{(m)*} \mathbf{J} \mathbf{M} \mathbf{h}_{i,t}^{(m,n)} \right\} \\
&+ \frac{1}{2} \left( \Delta f^{(m)} - \widehat{\Delta f}^{(m,n-1)} \right)^2 \sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}-1} \text{Re} \left\{ \mathbf{y}_{i,t}^{(m)*} \mathbf{J}^2 \mathbf{M}^2 \mathbf{h}_{i,t}^{(m,n)} \right\}.
\end{aligned} \tag{4.63}$$

Setting the partial derivative of (4.63) to zero allows finding the estimate of  $\Delta f^{(m)}$  at the  $n$ -th iteration as

$$\widehat{\Delta f}_{\text{ML}}^{(m,n)} = \widehat{\Delta f}_{\text{ML}}^{(m,n-1)} - \frac{\sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}} \text{Im} \left\{ \mathbf{y}_{i,t}^{(m)*} \mathbf{M} \mathbf{h}_{i,t}^{(m,n)} \right\}}{\sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}} \text{Re} \left\{ \mathbf{y}_{i,t}^{(m)*} \mathbf{M}^2 \mathbf{h}_{i,t}^{(m,n)} \right\}}. \tag{4.64}$$

Finally, using (4.61), we can find the estimator of  $\mathbf{g}_i^{(m)}$  at the  $n$ -th M-step as

$$\widehat{\mathbf{g}}_{i,\text{ML}}^{(m,n)} = \left( \mathbf{S}^{(m)*} \mathbf{S}^{(m)} \right)^{-1} \mathbf{S}^{(m)*} \mathbf{F}_{\otimes} \mathbf{P}^* \left( \widehat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n)} \right) \boldsymbol{\Omega}^* \left( \widehat{\Delta f}_{\text{ML}}^{(m,n)} \right) \mathbf{y}_i^{(m)}. \tag{4.65}$$

Therefore, using (4.59), (4.60), (4.64), and (4.65), the proposed algorithm iteratively updates the PN, CFO, and equivalent channel gains respectively. The algorithm is terminated when the difference between the Likelihood Function (LF) at two iterations is smaller than a threshold  $\eta$ , i.e.,

$$\begin{aligned}
&\left| \sum_{i=1}^{L_r} \left\| \mathbf{y}_i^{(m)} - \boldsymbol{\Omega} \left( \widehat{\Delta f}^{(m,n)} \right) \mathbf{P} \left( \widehat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n)} \right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \widehat{\mathbf{g}}_i^{(m,n)} \right\|_2^2 \right. \\
&\quad \left. - \left\| \mathbf{y}_i^{(m)} - \boldsymbol{\Omega} \left( \widehat{\Delta f}^{(m,n-1)} \right) \mathbf{P} \left( \widehat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n-1)} \right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \widehat{\mathbf{g}}_i^{(m,n-1)} \right\|_2^2 \right| \leq \eta.
\end{aligned} \tag{4.66}$$

The overall LMMSE-EM estimation algorithm is summarized in Algorithm 2.

---

**Algorithm 2** LMMSE-EM algorithm
 

---

- 1: **Initialize CFO, beamformed channel estimates, and initial difference in LLF**
- 2: 
$$\widehat{\Delta f}^{(0)} = \arg \max_{\Delta f} \sum_{i=1}^{L_r} \|\Omega(\Delta f) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{S}^{(m)*} \mathbf{F}_{\otimes} \Omega^*(\Delta f) \mathbf{y}_i^{(m)}\|_2^2$$
- 3: 
$$\hat{\mathbf{g}}_i^{(m,0)} = \left(\mathbf{S}^{(m)*} \mathbf{S}^{(m)}\right)^{-1} \mathbf{S}^{(m)*} \mathbf{F}_{\otimes} \Omega \left(\widehat{\Delta f}^{(0)}\right) \mathbf{y}_i^{(m)}, i = 1, \dots, L_r$$
- 4: 
$$\hat{\eta}^{(m,n)} = \infty, n = 1$$
- 5: **while**  $\hat{\eta}^{(m,n)} > \eta$  **do**
- 6:     **Update PN estimate**
- 7:     (4.56)-(4.60)
- 8:     **Update CFO estimate**
- 9:     
$$\mathbf{h}_{i,t}^{(m,n)} \triangleq \Omega_t \left(\widehat{\Delta f}_{\text{ML}}^{(m,n-1)}\right) \mathbf{P}_t \left(\hat{\boldsymbol{\theta}}_{t,\text{MU}}^{(m,n)}\right) \mathbf{F}_t^* \mathbf{S}_t^{(m)} \hat{\mathbf{g}}_{i,\text{ML}}^{(m,n-1)}$$
- 10:     
$$\widehat{\Delta f}_{\text{ML}}^{(m,n)} = \widehat{\Delta f}_{\text{ML}}^{(m,n-1)} - \frac{\sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}} \text{Im}\{\mathbf{y}_{i,t}^{(m)*} \mathbf{M} \mathbf{h}_{i,t}^{(m,n)}\}}{\sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}} \text{Re}\{\mathbf{y}_{i,t}^{(m)*} \mathbf{M}^2 \mathbf{h}_{i,t}^{(m,n)}\}}$$
- 11:     **Update the beamformed channel estimates**
- 12:     
$$\hat{\mathbf{g}}_{i,\text{ML}}^{(m,n)} = \left(\mathbf{S}^{(m)*} \mathbf{S}^{(m)}\right)^{-1} \mathbf{S}^{(m)*} \mathbf{F} \mathbf{P}^* \left(\hat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n)}\right) \Omega^* \left(\widehat{\Delta f}_{\text{ML}}^{(m,n)}\right) \mathbf{y}_i^{(m)},$$
  

$$i = 1, \dots, L_r$$
- 13:     **Iteration update**
- 14:     
$$n = n + 1$$
- 15:     **Update difference in likelihood function**
- 16:     
$$\hat{\eta}^{(m,n)} = \left\| \sum_{i=1}^{L_r} \left\| \mathbf{y}_i^{(m)} - \Omega \left(\widehat{\Delta f}^{(m,n)}\right) \mathbf{P} \left(\hat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n)}\right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \hat{\mathbf{g}}_i^{(m,n)} \right\|_2^2 - \left\| \mathbf{y}_i^{(m)} - \Omega \left(\widehat{\Delta f}^{(m,n-1)}\right) \mathbf{P} \left(\hat{\boldsymbol{\theta}}_{\text{MU}}^{(m,n-1)}\right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \hat{\mathbf{g}}_i^{(m,n-1)} \right\|_2^2 \right\|_2$$
- 17: **end while**

---

 Figure 4.4: Detailed steps of the first proposed LMMSE-EM algorithm.
 

---

### 4.5.2 EKF-RTS-EM Algorithm

In this subsection, I present an alternative strategy to using our first proposed LMMSE-EM algorithm. Despite the simplicity of (4.61) and (4.65), the algorithm introduced in the previous subsection exhibits high computational complexity. The main computational bottleneck of the LMMSE-EM algorithm is the inversion of  $\hat{\mathbf{C}}_{\mathbf{y}^{(m)}, \mathbf{y}^{(m)}}^{(n)}$  in (4.60), which has complexity  $\mathcal{O}((KN_{\text{tr}}L_r)^3)$  in the worst case. This high complexity comes at the cost of batch processing the  $L_rKN_{\text{tr}}$  measurements at once to find the LMMSE estimator for the PN, which exhibits very good performance but it may not be computationally feasible if the number of subcarriers is in the order of a few thousands. However, a trade-off between estimation performance and computational complexity can be achieved if the size of the matrix inversion in (4.60) is reduced. To reduce computational complexity, I propose to sequentially process every set of  $L_r$  received measurements to reduce complexity to be  $\mathcal{O}(L_r^3)$  at most, which is computationally affordable since  $L_r$  is usually a small number [11]. The PN estimate can be found using a combination of the EKF and the RTS smoother [149], which exploits a first-order linearization of the received measurement vector and uses the RTS smoother on the linearized vector as follows.

Using (4.13), the  $L_r$ -dimensional time-domain received measurement

can be expressed as

$$\begin{aligned}
\underbrace{\begin{bmatrix} y_{1,t}^{(m)}[k_0[t] + \ell] \\ \vdots \\ y_{L_r,t}^{(m)}[k_0[t] + \ell] \end{bmatrix}}_{\mathbf{y}_t^{(m)}[k_0[t] + \ell]} &= \underbrace{e^{j2\pi\Delta f^{(m)}(k_0[t] + \ell)} e^{j\theta_t^{(m)}[k_0[t] + \ell]} \begin{bmatrix} \mathbf{f}_\ell^* \mathbf{S}_t^{(m)} \mathbf{g}_1^{(m)} \\ \vdots \\ \mathbf{f}_\ell^* \mathbf{S}_t^{(m)} \mathbf{g}_{L_r}^{(m)} \end{bmatrix}}_{\mathbf{h}_{\ell,t}(\boldsymbol{\xi}_D^{(m)}, \theta_t^{(m)}[k_0[t] + \ell])} \\
&+ \begin{bmatrix} v_{1,t}^{(m)}[k_0[t] + \ell] \\ \vdots \\ v_{L_r,t}^{(m)}[k_0[t] + \ell] \end{bmatrix}. \tag{4.67}
\end{aligned}$$

The RTS smoother consists of a backward filter that follows the EKF recursion given by the following:

- **Forward recursion:** Time Update Equations:

$$\begin{aligned}
\hat{\theta}_{t,\text{TU}}^{(m)}[k_0[t] + \ell] &= \begin{cases} 0 & t = 0, \ell = 0 \\ \hat{\theta}_{t,\text{MU}}^{(m)}[k_0[t] + \ell - 1] & \ell > 0, \\ \hat{\theta}_{t-1,\text{MU}}^{(m)}[k_0[t-1] + K - 1] & \ell = 0, t > 0, \end{cases} \\
\left(\hat{\sigma}_{\hat{\theta}_{t,\text{TU}}^{(m)}}^{(\ell)}\right)^2 &= \begin{cases} \left\| [\mathbf{C}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}]_{1,:} \right\|_2^2 & \ell = 0, t = 0, \\ \left(\hat{\sigma}_{\hat{\theta}_{t,\text{MU}}^{(m)}}^{(\ell-1)}\right)^2 + \left\| \boldsymbol{\Delta} \mathbf{d}_{t,k-1}^{t,k} \right\|_2^2 & \ell > 0 \\ \left(\hat{\sigma}_{t-1,\text{MU}}^{(K-1)}\right)^2 & \ell = 0, t > 0, \end{cases} \tag{4.68}
\end{aligned}$$

Measurement Update Equations:

$$\begin{aligned}
\hat{\theta}_{t,\text{MU}}^{(m)}[k_0[t] + \ell] &= \hat{\theta}_{t,\text{TU}}^{(m)}[k_0[t] + \ell] + \text{Re} \left\{ \hat{\mathbf{c}}_{\boldsymbol{\theta}_t^{(m)}, \mathbf{y}_t^{(m)}}^{(\ell)} \left( \hat{\mathbf{C}}_{\mathbf{y}_t^{(m)}, \mathbf{y}_t^{(m)}}^{(\ell)} \right)^{-1} \right. \\
&\quad \left. \times \left( \mathbf{y}_t^{(m)}[k_0[t] + \ell] - \mathbf{h}_{\ell,t} \left( \hat{\boldsymbol{\xi}}_D^{(m,n-1)}, \hat{\theta}_{t,\text{TU}}^{(m)}[k_0[t] + \ell] \right) \right) \right\} \\
\left(\hat{\sigma}_{\hat{\theta}_{t,\text{MU}}^{(m)}}^{(\ell,t)}\right)^2 &= \text{Re} \left\{ \left(\hat{\sigma}_{\hat{\theta}_{t,\text{TU}}^{(m)}}^{(\ell)}\right)^2 - \left(\hat{\mathbf{c}}_{\boldsymbol{\theta}_t^{(m)}, \theta_t^{(m)}}^{(\ell)}\right)^* \left(\hat{\mathbf{C}}_{\mathbf{y}_t^{(m)}, \mathbf{y}_t^{(m)}}^{(\ell)}\right)^{-1} \hat{\mathbf{c}}_{\mathbf{y}_t^{(m)}, \theta_t^{(m)}}^{(\ell)} \right\}, \tag{4.69}
\end{aligned}$$

where  $\hat{\mathbf{C}}_{\mathbf{y}_t^{(m)}, \theta_t^{(m)}}^{(\ell)} \in \mathbb{C}^{L_r \times 1}$ ,  $\hat{\mathbf{C}}_{\mathbf{y}_t^{(m)}, \mathbf{y}_t^{(m)}}^{(\ell)} \in \mathbb{C}^{L_r \times L_r}$  are the covariance matrix of  $\mathbf{y}_t^{(m)}[k_0[t] + \ell]$  and  $\theta_t^{(m)}[k_0[t] + \ell]$ , and the autocovariance matrix of  $\mathbf{y}_t^{(m)}[k_0[t] + \ell]$ , which are given by

$$\begin{aligned} \hat{\mathbf{C}}_{\mathbf{y}_t^{(m)}, \theta_t^{(m)}}^{(\ell)} &= \mathbf{j} \mathbf{h}_{\ell, t} \left( \hat{\boldsymbol{\xi}}_D^{(m, n-1)}, \hat{\theta}_{t, \text{TU}}^{(m)}[k_0[t] + \ell] \right) \left( \hat{\sigma}_{\hat{\theta}_{\text{MU}}^{(\ell, t)}} \right)^2 \\ \hat{\mathbf{C}}_{\mathbf{y}_t^{(m)}, \mathbf{y}_t^{(m)}}^{(\ell)} &= \mathbf{h}_{\ell, t} \left( \hat{\boldsymbol{\xi}}_D^{(m, n-1)}, \hat{\theta}_{t, \text{TU}}^{(m)}[k_0[t] + \ell] \right) \left( \hat{\sigma}_{\hat{\theta}_{\text{MU}}^{(\ell, t)}} \right)^2 \\ &\quad \times \mathbf{h}_{\ell, t}^* \left( \hat{\boldsymbol{\xi}}_D^{(m, n-1)}, \hat{\theta}_{t, \text{TU}}^{(m)}[k_0[t] + \ell] \right) + \sigma^2 \mathbf{I}_{L_r}. \end{aligned} \quad (4.70)$$

- **Backward recursion:**

$$\begin{aligned} G_t^{(\ell)} &= \frac{\left( \hat{\sigma}_{\hat{\theta}_{\text{MU}}^{(\ell, t)}} \right)^2}{\left( \hat{\sigma}_{\hat{\theta}_{\text{MU}}^{(\ell, t)}} \right)^2 + \left\| \Delta \mathbf{d}_{t, k}^{t, k+1} \right\|_2^2} \\ \hat{\theta}_{t, \text{RTS}}^{(m)}[k_0[t] + \ell] &= \hat{\theta}_{t, \text{MU}}^{(m)}[k_0[t] + \ell] + G_t^{(\ell)} \\ &\quad \times \left( \hat{\theta}_{t, \text{RTS}}^{(m)}[k_0[t] + \ell + 1] - \hat{\theta}_{t, \text{MU}}^{(m)}[k_0[t] + \ell + 1] \right) \\ \left( \hat{\sigma}_{\hat{\theta}_{\text{RTS}}^{(\ell, t)}} \right)^2 &= \left( \hat{\sigma}_{\hat{\theta}_{\text{MU}}^{(\ell, t)}} \right)^2 + \left( G_t^{(\ell)} \right)^2 \\ &\quad \times \left( \left( \hat{\sigma}_{\hat{\theta}_{\text{RTS}}^{(\ell+1, t)}} \right)^2 - \left( \hat{\sigma}_{\hat{\theta}_{\text{MU}}^{(\ell+1, t)}} \right)^2 - \left\| \Delta \mathbf{d}_{t, k}^{t, k+1} \right\|_2^2 \right). \end{aligned} \quad (4.71)$$

In (4.68), the proposed algorithm is initialized as  $\hat{\theta}_{t, \text{TU}}[k_0[0]] = 0$  since the PN vector is assumed to have zero mean, and the predicted variance is initialized as  $\left( \hat{\sigma}_{\hat{\theta}_{0, \text{TU}}^{(0)}} \right)^2 = \left\| [\mathbf{C}_{\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m)}}]_{1, \cdot} \right\|_2^2$ . Also, notice that the CP is removed after timing offset synchronization, which requires properly updating the PN predicted statistics from the last sample of the  $t$ -th OFDM symbol to the first sample of the  $(t+1)$ -th OFDM symbol, as reflected in (4.68).

Thereby, using (4.68)-(4.71), and then (4.64) and (4.65), the second proposed algorithm can iteratively update the PN sample estimates, CFO, and equivalent channel gains, respectively. The termination criterion for the proposed algorithm is analogous to the termination criterion for the first proposed LMMSE-EM algorithm, given in (4.66). The detailed steps the proposed EKF-RTS algorithm follows are summarized in Algorithm 3.

### 4.5.3 Initialization and Convergence

Appropriate initialization of the CFO,  $\Delta f^{(m)}$ , and equivalent beam-formed channels,  $\{\mathbf{g}_i^{(m)}\}_{i=1}^{L_r}$ , are essential to ensure global convergence of the proposed algorithms. The initialization process can be summarized as follows:

- Similar to [98], an initial CFO estimate  $\widehat{\Delta f}^{(m,0)}$  is obtained by applying an exhaustive search for the value of  $\Delta f^{(m)}$  that minimizes the cost function in the absence of PN. This cost function is given in ([141], equation (17)). Simulation results in Section 4.7 show that an exhaustive search with a coarse step size of 0.02 is sufficient to initialize the proposed algorithms.
- Using  $\widehat{\Delta f}^{(m,0)}$ , the initial channel estimates  $\{\widehat{\mathbf{g}}_i^{(m,0)}\}_{i=1}^{L_r}$  are obtained by applying  $\widehat{\mathbf{g}}_i^{(m,0)} = \left(\mathbf{S}^{(m)*}\mathbf{S}^{(m)}\right)^{-1}\mathbf{S}^{(m)*}\mathbf{F}\mathbf{\Omega}^* \left(\widehat{\Delta f}^{(m,0)}\right)\mathbf{y}_i^{(m)}$ .

Based on the equivalent system model in (4.13) and the simulation results in Section 4.7, it can be concluded that the proposed LMMSE-EM and EKF-RTS-EM algorithms converge globally when the PN vector is initialized as

---

**Algorithm 3** EKF-RTS-EM algorithm

---

- 1: **Initialize CFO, beamformed channel estimates, and initial difference in LF**
- 2: 
$$\widehat{\Delta f}^{(0)} = \arg \max_{\Delta f} \sum_{i=1}^{L_r} \|\Omega(\Delta f) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \mathbf{S}^{(m)*} \mathbf{F}_{\otimes} \Omega^*(\Delta f) \mathbf{y}_i^{(m)}\|_2^2$$
- 3: 
$$\hat{\mathbf{g}}_i^{(m,0)} = \left(\mathbf{S}^{(m)*} \mathbf{S}^{(m)}\right)^{-1} \mathbf{S}^{(m)*} \mathbf{F}_{\otimes} \Omega \left(\widehat{\Delta f}^{(0)}\right) \mathbf{y}_i^{(m)}, i = 1, \dots, L_r$$
- 4: 
$$\hat{\eta}^{(m,n)} = \infty, n = 1$$
- 5: **while**  $\hat{\eta}^{(m,n)} > \eta$  **do**
- 6:     **Update PN estimate**
- 7:     (4.68)-(4.71)
- 8:     **Update CFO estimate**
- 9:     
$$\mathbf{h}_{i,t}^{(m,n)} \triangleq \Omega_t \left(\widehat{\Delta f}_{\text{ML}}^{(m,n-1)}\right) \mathbf{P}_t \left(\hat{\boldsymbol{\theta}}_{t,\text{RTS}}^{(m,n)}\right) \mathbf{F}_t^* \mathbf{S}_t^{(m)} \hat{\mathbf{g}}_{i,\text{ML}}^{(m,n-1)}$$
- 10:     
$$\widehat{\Delta f}_{\text{ML}}^{(m,n)} = \widehat{\Delta f}_{\text{ML}}^{(m,n-1)} - \frac{\sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}} \text{Im}\{\mathbf{y}_{i,t}^{(m)*} \mathbf{M} \mathbf{h}_{i,t}^{(m,n)}\}}{\sum_{i=1}^{L_r} \sum_{t=0}^{N_{\text{tr}}} \text{Re}\{\mathbf{y}_{i,t}^{(m)*} \mathbf{M}^2 \mathbf{h}_{i,t}^{(m,n)}\}}$$
- 11:     **Update the beamformed channel estimates**
- 12:     
$$\hat{\mathbf{g}}_{i,\text{ML}}^{(m,n)} = \left(\mathbf{S}^{(m)*} \mathbf{S}^{(m)}\right)^{-1} \mathbf{S}^{(m)*} \mathbf{F} \mathbf{P}^* \left(\hat{\boldsymbol{\theta}}_{\text{RTS}}^{(m,n)}\right) \Omega^* \left(\widehat{\Delta f}_{\text{ML}}^{(m,n)}\right) \mathbf{y}_i^{(m)},$$
  

$$i = 1, \dots, L_r$$
- 13:     **Iteration update**
- 14:     
$$n = n + 1$$
- 15:     **Update difference in likelihood function**
- 16:     
$$\hat{\eta}^{(m,n)} = \left\| \sum_{i=1}^{L_r} \left\| \mathbf{y}_i^{(m)} - \Omega \left(\widehat{\Delta f}^{(m,n)}\right) \mathbf{P} \left(\hat{\boldsymbol{\theta}}_{\text{RTS}}^{(m,n)}\right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \hat{\mathbf{g}}_i^{(m,n)} \right\|_2^2 - \left\| \mathbf{y}_i^{(m)} - \Omega \left(\widehat{\Delta f}^{(m,n-1)}\right) \mathbf{P} \left(\hat{\boldsymbol{\theta}}_{\text{RTS}}^{(m,n-1)}\right) \mathbf{F}_{\otimes}^* \mathbf{S}^{(m)} \hat{\mathbf{g}}_i^{(m,n-1)} \right\|_2^2 \right\|_2$$
- 17: **end while**

---

Figure 4.5: Detailed steps of the second proposed EKF-RTS-EM algorithm.

---

$$\hat{\boldsymbol{\theta}}^{(m,0)} = \mathbf{0}_{KN_{\text{tr}} \times 1}.$$

#### 4.5.4 Dictionary-Constrained Channel Estimation

In this section, I formulate the problem of estimating the frequency-selective mmWave MIMO channel using the ML statistics already estimated using the proposed LMMSE-EM and EKF-RTS-EM algorithms. Once  $M$  training frames are processed, each comprising of  $N_{\text{tr}}$  OFDM symbols, the estimated equivalent beamformed channels can be stacked to form the signal model

$$\underbrace{\begin{bmatrix} \hat{\mathbf{g}}_{\text{ML}}^{(1,N)}[k] \\ \vdots \\ \hat{\mathbf{g}}_{\text{ML}}^{(M,N)}[k] \end{bmatrix}}_{\hat{\mathbf{g}}_{\text{ML}}^{(N)}} = \underbrace{\begin{bmatrix} \mathbf{q}^{(1)T} \mathbf{F}_{\text{tr}}^{(1)T} \otimes \mathbf{D}_{\text{w}}^{(1)-*} \mathbf{W}_{\text{tr}}^{(1)*} \\ \vdots \\ \mathbf{q}^{(M)T} \mathbf{F}_{\text{tr}}^{(M)T} \otimes \mathbf{D}_{\text{w}}^{(M)-*} \mathbf{W}_{\text{tr}}^{(M)*} \end{bmatrix}}_{\boldsymbol{\Phi}_{\text{w}}} \text{vec}\{\mathbf{H}[k]\} + \underbrace{\begin{bmatrix} \tilde{\mathbf{v}}^{(1,N)}[k] \\ \vdots \\ \tilde{\mathbf{v}}^{(M,N)}[k] \end{bmatrix}}_{\tilde{\mathbf{v}}^{(N)}[k]}, \quad (4.72)$$

where  $\tilde{\mathbf{v}}^{(m,N)} \in \mathbb{C}^{L_r \times 1}$  is the estimation error of  $\hat{\mathbf{g}}_{\text{ML}}^{(m,N)}[k]$ ,  $1 \leq m \leq M$ , and  $\boldsymbol{\Phi}_{\text{w}} \in \mathbb{C}^{ML_r \times N_{\text{tr}}N_r}$  is the post-whitened measurement matrix. Now, the channel matrix in (4.9) can be vectorized and plugged into (4.72) to obtain

$$\hat{\mathbf{g}}_{\text{ML}}^{(N)}[k] \approx \boldsymbol{\Phi}_{\text{w}} \underbrace{\left( \tilde{\mathbf{A}}_{\text{T}}^{\text{C}} \otimes \tilde{\mathbf{A}}_{\text{R}} \right)}_{\boldsymbol{\Psi}} \underbrace{\text{vec}\{\mathbf{G}^{\text{v}}[k]\}}_{\mathbf{g}^{\text{v}}[k]} + \tilde{\mathbf{v}}^{(N)}[k], \quad (4.73)$$

where  $\boldsymbol{\Psi} \in \mathbb{C}^{N_{\text{tr}}N_r \times G_{\text{t}}G_{\text{r}}}$  is the angular dictionary matrix, and  $\mathbf{g}^{\text{v}}[k] \in \mathbb{C}^{G_{\text{r}}G_{\text{t}} \times 1}$  is the sparse vector containing the complex channel path gains in its non-zero coefficients [1]. To estimate the frequency-selective sparse vectors  $\{\mathbf{g}^{\text{v}}[k]\}_{k=0}^{K-1}$ , the design of the measurement matrix  $\boldsymbol{\Phi}_{\text{w}}$  in (4.73) needs to be such that this matrix has as small correlation between columns as possible, which is a re-

sult proven in the CS literature to ensure that the estimation of the channel's support will be robust, and this depends on the design of the precoding and combining matrices  $\mathbf{F}_{\text{tr}}^{(m)}$ ,  $\mathbf{q}^{(m)}$ , and  $\mathbf{W}_{\text{tr}}^{(m)}$ . As discussed in [141], the precoders and combiners should be designed accounting for the lack of timing synchronization, such that the equivalent measurement matrix design is suitable for compressive estimation and the estimated timing offset  $\hat{n}_0$  matches the actual timing offset. For this reason, I adopt the design method in [141] to generate hybrid precoders and combiners, which has been shown to offer excellent performance at the low SNR regime.

Another issue to overcome when estimating the sparse channel vectors is how to obtain prior information on either the sparsity level of the channel or the variance of the noise in (4.73). As discussed in [2], knowing the sparsity level is unrealistic in practice, and even if it were known, there is no guarantee that the best sparse approximation of  $\{\mathbf{g}^v[k]\}_{k=0}^{K-1}$  has as many non-zero components as the actual number of multipath components in the frequency-selective channel. This mismatch is even more severe in the frequency-selective scenario, in which transmit and receive pulse-shaping bandlimit the channel, thereby limiting the resolution to detect multipath components at baseband level [150]. For this reason, I will focus on finding the variance of the estimation error in (4.73).

From the property of asymptotic efficiency of ML estimators it is known that, if the SNR is not too low, and the number of samples used to estimate the different parameters is large enough, the estimation errors  $\tilde{\mathbf{v}}^{(m,N)}[k]$  are Gaussian, with zero mean and covariance given by the hybrid CRLB matrix

for the estimation of the complex path gains  $\hat{\mathbf{g}}_{\text{ML}}^{(m,N)}[k]$ . Since the received noise vectors  $\mathbf{v}^{(m)}[n]$  in (4.4) are independent and identically distributed, it is clear that estimation errors for  $\hat{\mathbf{g}}_{\text{ML}}^{(m,N)}[k]$  are independent as well, although not identically distributed. For this reason, it is necessary to compute the covariance matrix for each of the estimation error vectors corresponding to the  $M$  different training frames. Let  $\mathbf{C}_{\hat{\mathbf{g}}_i^{(m)}, \hat{\mathbf{g}}_i^{(m)}} \in \mathbb{C}^{L_c \times L_c}$  denote the hybrid CRLB matrix for the estimation of  $\mathbf{g}_i^{(m)}$ . Using that  $g_i^{(m)}[d] = [\mathbf{g}_i^{(m)}]_d = \alpha_i^{(m)}[d]e^{j\beta_i^{(m)}[d]}$ , it follows that

$$\underbrace{\begin{bmatrix} g_1^{(m)}[0] \\ \vdots \\ g_1^{(m)}[L_c - 1] \\ \vdots \\ g_{L_r}^{(m)}[0] \\ \vdots \\ g_{L_r}^{(m)}[L_c - 1] \end{bmatrix}}_{\mathbf{g}^{(m)}} = \underbrace{\begin{bmatrix} \alpha_1^{(m)}[0]e^{j\beta_1^{(m)}[0]} \\ \vdots \\ \alpha_1^{(m)}[L_c - 1]e^{j\beta_1^{(m)}[L_c - 1]} \\ \vdots \\ \alpha_{L_r}^{(m)}[0]e^{j\beta_{L_r}^{(m)}[0]} \\ \vdots \\ \alpha_{L_r}^{(m)}[L_c - 1]e^{j\beta_{L_r}^{(m)}[L_c - 1]} \end{bmatrix}}_{\mathbf{f}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)})}. \quad (4.74)$$

Using (4.74), the covariance matrix of any unbiased estimator  $\hat{\mathbf{g}}^{(m)}$  of  $\mathbf{g}^{(m)}$  is lower bounded by the hybrid CRLB as

$$\mathbf{C}_{\hat{\mathbf{g}}^{(m)}, \hat{\mathbf{g}}^{(m)}} \geq \mathbf{J}_{\mathbf{f}}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)}) \mathbf{C}_{\hat{\boldsymbol{\alpha}}^{(m)}, \hat{\boldsymbol{\beta}}^{(m)}} \mathbf{J}_{\mathbf{f}}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)})^*, \quad (4.75)$$

where  $\mathbf{J}_{\mathbf{f}}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)}) \in \mathbb{C}^{L_r L_c \times 2L_r L_c}$  is the Jacobian matrix of  $\mathbf{f}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)})$

$$\mathbf{J}_{\mathbf{f}}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}^{(m)}) = \bigoplus_{i=1}^{L_r} \bigoplus_{d=0}^{D-1} \begin{bmatrix} e^{j\beta_i^{(m)}[d]} & jg_i^{(m)}[d] \end{bmatrix}. \quad (4.76)$$

Next, the hybrid CRLB for the estimation of  $\mathbf{g}^{(m)}$  is computed as follows. Notice that the different frequency-domain channel vectors  $\mathbf{g}_i^{(m)}$  are related to

their time-domain counterparts through a Fourier transform, mathematically represented using  $\mathbf{F}_1 \in \mathbb{C}^{K \times L_c}$ , which comprises of the first  $L_c$  in  $\mathbf{F}$ . Thereby, using (4.75) the covariance for any unbiased estimator  $\hat{\mathbf{g}}^{(m)}$  is simply given by

$$\mathbf{C}_{\hat{\mathbf{g}}^{(m)}, \hat{\mathbf{g}}^{(m)}} \geq (\mathbf{I}_{L_r} \otimes \mathbf{F}_1) \mathbf{C}_{\hat{\mathbf{g}}^{(m)}, \hat{\mathbf{g}}^{(m)}} (\mathbf{I}_{L_r} \otimes \mathbf{F}_1)^* . \quad (4.77)$$

Now, the hybrid CRLB for the estimation of  $\mathbf{g}^{(m)}[k] = \left[ \mathbf{g}_1^{(m)}[k] \ \dots \ \mathbf{g}_{L_r}^{(m)}[k] \right]^T$  is related to the hybrid CRLB in (4.77) through a selection matrix as

$$\begin{aligned} \mathbf{g}^{(m)}[k] &= \begin{bmatrix} \mathbf{e}_k^T \mathbf{g}_1^{(m)} \\ \vdots \\ \mathbf{e}_k^T \mathbf{g}_{L_r}^{(m)} \end{bmatrix} \\ &= (\mathbf{I}_{L_r} \otimes \mathbf{e}_k^T) \mathbf{g}^{(m)}, \end{aligned} \quad (4.78)$$

whereby the hybrid CRLB for any unbiased estimator  $\hat{\mathbf{g}}^{(m)}[k]$  of  $\mathbf{g}^{(m)}[k]$  is given by

$$\mathbf{C}_{\hat{\mathbf{g}}^{(m)}[k], \hat{\mathbf{g}}^{(m)}[k]} \geq (\mathbf{I}_{L_r} \otimes \mathbf{e}_k^T) \mathbf{C}_{\hat{\mathbf{g}}^{(m)}, \hat{\mathbf{g}}^{(m)}} (\mathbf{I}_{L_r} \otimes \mathbf{e}_k) . \quad (4.79)$$

Finally, the overall covariance matrix for the estimation error vector  $\tilde{\mathbf{v}}^{(N)}[k]$  in (4.73) needs to be found. Using the fact that the received noise at the antenna level is temporally white, the covariance matrix of  $\tilde{\mathbf{v}}^{(N)}[k]$  is given by the hybrid CRLB for any unbiased estimator of  $\mathbf{g}[k] = \left[ \mathbf{g}^{(1)T}[k] \ \dots \ \mathbf{g}^{(M)T}[k] \right]^T$ . The final hybrid CRLB is given by

$$\mathbf{C}_{\hat{\mathbf{g}}[k], \hat{\mathbf{g}}[k]} \geq \bigoplus_{m=1}^M \mathbf{C}_{\hat{\mathbf{g}}^{(m)}[k], \hat{\mathbf{g}}^{(m)}[k]} . \quad (4.80)$$

Then, the estimation error is distributed as  $\tilde{\mathbf{v}}^{(N)}[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\tilde{\mathbf{v}}^{(N)}[k], \tilde{\mathbf{v}}^{(N)}[k]})$ .

Let  $\mathbf{D}_{\tilde{\mathbf{v}}^{(N)}[k]} \in \mathbb{C}^{ML_r \times ML_r}$  be the Cholesky factor of  $\mathbf{C}_{\tilde{\mathbf{v}}^{(N)}[k], \tilde{\mathbf{v}}^{(N)}[k]}$ , i.e.,  $\mathbf{C}_{\tilde{\mathbf{v}}^{(N)}[k], \tilde{\mathbf{v}}^{(N)}[k]} =$

$\mathbf{D}_{\tilde{\mathbf{v}}^{(N)}[k]}^* \mathbf{D}_{\tilde{\mathbf{v}}^{(N)}[k]}$ . Thereby, the problem of estimating  $\{\mathbf{g}^v[k]\}_{k=0}^{K-1}$  can be formulated as

$$\begin{aligned} \hat{\mathbf{g}}^v[k] &= \arg \min_{\{\mathbf{g}^v[\ell]\}_{\ell=0}^{K-1}} \sum_{k=0}^{K-1} \|\mathbf{g}^v[k]\|^1, \\ \text{subject to } &\frac{1}{K} \sum_{k=0}^{K-1} \left\| \mathbf{D}_{\tilde{\mathbf{v}}^{(N)}[k]}^{-*} \left( \hat{\mathbf{g}}_{\text{ML}}^{(N)}[k] - \Phi_w \Psi \mathbf{g}^v[k] \right) \right\|_2^2 \leq \epsilon, \end{aligned} \quad (4.81)$$

where  $\epsilon \in \mathbb{R}$  is a design parameter defining the maximum allowable reconstruction error for the sparse vectors  $\{\mathbf{g}^v[k]\}_{k=0}^{K-1}$ . From a computational complexity standpoint, the main difficulty in (4.81) comes from the fact that post-whitening the proxy estimates  $\hat{\mathbf{g}}_{\text{ML}}^{(n)}[k]$  results in frequency-dependent measurement matrices  $\Upsilon[k] = \mathbf{D}_{\tilde{\mathbf{v}}^{(N)}[k]}^{-*} \Phi_w \Psi$ , which increases the complexity of sparse recovery algorithms by a factor of  $K$ . Since  $K$  can be in the order of hundreds or thousands of subcarriers, using frequency-dependent measurement matrices results in high-complexity channel estimation algorithms. To circumvent this issue, I propose to find a covariance matrix  $\mathbf{C}_{\mathbf{v}^{(N)}, \tilde{\mathbf{v}}^{(N)}}$  that accurately represents the covariance matrix of the estimation error for every subcarrier in the MMSE sense. Let  $\hat{\mathbf{g}}_{\text{ML}}^{(N)}[k] \in \mathbb{C}^{ML_r \times 1}$  denote an approximate estimate of  $\hat{\mathbf{g}}_{\text{ML}}^{(N)}[k]$  given by

$$\hat{\mathbf{g}}_{\text{ML}}^{(N)}[k] \approx \mathbf{g}[k] + \tilde{\mathbf{v}}^{(N)}[k], \quad (4.82)$$

in which  $\tilde{\mathbf{v}}^{(N)}[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{v}^{(N)}, \tilde{\mathbf{v}}^{(N)}})$ . Then, the problem of finding the covariance matrix  $\mathbf{C}_{\mathbf{v}^{(N)}, \tilde{\mathbf{v}}^{(N)}}$  can be stated as

$$\mathbf{C}_{\mathbf{v}^{(N)}, \tilde{\mathbf{v}}^{(N)}} = \arg \min_{\mathbf{C}} \sum_{k=0}^{K-1} \mathbb{E} \left\{ \left\| \hat{\mathbf{g}}_{\text{ML}}^{(N)}[k] - \hat{\mathbf{g}}_{\text{ML}}^{(N)}[k] \right\|_2^2 \right\}. \quad (4.83)$$

Upon developing the cost function in (4.83), and letting  $\mathbf{D} \in \mathbb{C}^{ML_r \times ML_r}$  be the Cholesky factor of  $\mathbf{C}$ , i.e.,  $\mathbf{C} = \mathbf{D}^* \mathbf{D}$ , the optimal covariance matrix can be found as the solution to the problem

$$\mathbf{C}_{\mathbf{v}^{(N)}, \mathbf{v}^{(N)}} = \arg \min_{\mathbf{C}} \sum_{k=0}^{K-1} \mathbb{E} \left\{ \left\| \mathbf{D} - \mathbf{D}_{\mathbf{v}^{(N)}[k], \mathbf{v}^{(N)}[k]} \right\|_F^2 \right\}, \quad (4.84)$$

which is a LS problem with solution given by

$$\mathbf{C}_{\mathbf{v}^{(N)}, \mathbf{v}^{(N)}} = \left( \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{D}_{\mathbf{v}^{(N)}[k], \mathbf{v}^{(N)}[k]} \right)^* \left( \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{D}_{\mathbf{v}^{(N)}[k], \mathbf{v}^{(N)}[k]} \right). \quad (4.85)$$

The result in (4.85) indicates that the covariance matrix that best represents the covariance of every estimation error vector in the MMSE sense has a Cholesky decomposition with Cholesky factor given by the average of the Cholesky factors for the covariance matrices of the estimation error at the different subcarriers.

The last step to close the estimation problem in (4.81) is the definition of  $\epsilon$ . Since the training precoders and combiners might lead, in general, to different covariance matrices  $\mathbf{C}_{\tilde{\mathbf{v}}^{(N)}[k], \tilde{\mathbf{v}}^{(N)}[k]}$ , an overall representative for the noise variance of the entire vector  $\tilde{\mathbf{v}}^{(N)} = [\tilde{\mathbf{v}}^{(N)T}[0], \dots, \tilde{\mathbf{v}}^{(N)T}[K-1]]^T$  is needed. To overcome this issue, similarly to [2], I propose to design  $\epsilon$  as a convex combination of the hybrid CRLB for the different  $\mathbf{g}_i^{(m)}[k]$ ,  $1 \leq L_r$ ,  $0 \leq k \leq K-1$ ,  $1 \leq m \leq M$ , using estimates of the SNR per RF chain. Using the property of asymptotic invariance of ML estimators, the ML estimate of the SNR per RF chain can be written as  $\hat{\gamma}_{i, \text{ML}}^{(m)}[k] = \hat{\alpha}_{i, \text{ML}}^{(m)2}[k]/\sigma^2$ . Letting  $\hat{\mathbf{\Gamma}}^{(m)}[k] =$

$\bigoplus_{i=1}^{L_r} \hat{\gamma}_{i,\text{ML}}^{(m)}[k]$ , the parameter  $\epsilon$  can be set as

$$\epsilon = \frac{\sum_{m=1}^M \sum_{k=0}^{K-1} \text{trace} \left\{ \hat{\mathbf{\Gamma}}^{(m)}[k] \mathbf{C}_{\hat{\mathbf{g}}^{(m)}[k], \hat{\mathbf{g}}^{(m)}[k]} \right\}}{\sum_{m=1}^M \sum_{k=0}^{K-1} \text{trace} \left\{ \hat{\mathbf{\Gamma}}^{(m)}[k] \right\}}. \quad (4.86)$$

Then, the SW-OMP or SS-SW-OMP+Th algorithms in [1] can be used to solve the problem in (4.81). These algorithms have been shown to offer very good performance even when the mmWave MIMO channel has several clusters with non-negligible AS. It is important to highlight that the hybrid CRLB for the estimation of the channel matrices  $\{\mathbf{H}[k]\}_{k=0}^{K-1}$  is not computed. The reason is that, for realistic mmWave MIMO channel models such as NYUSIM [4], Quasi Deterministic Radio channel Generator (QuaDRiGa) [31], [32], and the 5G NR channel model [6], the finite antenna resolution, bandlimitedness of the baseband equivalent channel, and lack of knowledge of the number of multipath components, make it impossible to assume that an unbiased estimator for the channel can be found. For this reason, the estimates  $\{\hat{\mathbf{g}}^v[k]\}_{k=0}^{K-1}$  will, in general, have a different number of entries than the number of multipath components the channel actually comprises of. Consequently, the theory of CRLB cannot be directly applied to this problem.

## 4.6 Joint Data Detection and PN Mitigation

In this section, I devise and propose a joint data detection and PN mitigation algorithm for data transmission. Although the derivation of the algorithm is general, I will especially focus on the Physical Downlink Shared Channel (PDSCH) of the 5G NR wireless communication standard [37]. The

data frame structure of the 5G NR PDSCH channel depends on the specific numerology used for data transmission, which itself depends on the subcarrier spacing. I assume that the subcarrier spacing is known, as it can be transmitted through the Physical Downlink Control Channel (PDCCH), as well as the MCS.

A 5G NR radio frame comprises 10 subframes, with a variable number of slots within each subframe, and each slot carries 14 OFDM symbols [37]. The number of slots can be easily calculated as  $N_{\text{slots}} = 2^{\frac{\Delta f_s}{15}}$ , where  $\Delta f_s$  is the OFDM subcarrier spacing in kHz units. Within each OFDM symbol, a number of data subcarriers are allocated for data transmission, and others contain PDSCH pilot signals, which include Demodulation Reference Signal (DMRS), Phase-Tracking Reference Signal (PTRS), and Channel State Information - Reference Signal (CSI-RS). A small number of subcarriers is zeroed to facilitate spectral shaping, typically at the edges of the OFDM symbol. Let the positions of the data subcarriers be denoted by the set  $\mathcal{P}_{\text{data}}$ , the number of DMRS training symbols be denoted by  $N_{\text{DMRS}}$ , and the number of OFDM data symbols be denoted by  $N_{\text{data}}$ .

After the channel has been estimated, and assuming that the receiver feeds back relevant CSI information for hybrid precoder and combiner design, both the transmitter and the receiver configure the frequency-flat analog and frequency-selective digital precoders and combiners for data transmission using the algorithm in [151]. Let  $\mathbf{G}[k] \in \mathbb{C}^{L_r \times L_t}$  denote the frequency-selective analog precoded channel at the  $k$ -th subcarrier, which is given by

$\mathbf{G}[k] = \mathbf{W}_{\text{RF}}^* \mathbf{H}[k] \mathbf{F}_{\text{RF}}$ . Using the CFO estimate obtained during training, the effect of the CFO on the received data symbol can be compensated. Then, after CFO compensation, the vectorized received signal at the  $t$ -th received OFDM symbol and the  $i$ -th RF chain reads as (see (4.13))

$$\mathbf{y}_{i,t} = \mathbf{P}(\boldsymbol{\theta}_t) \mathbf{F}^* \mathbf{D}_t \mathbf{g}_i + \mathbf{v}_{i,t} \quad (4.87)$$

where  $\mathbf{P}(\boldsymbol{\theta}_t) \in \mathbb{C}^{K \times K}$  is diagonal and contains the PN impairment for the time-domain received samples, and  $\mathbf{D}_t \in \mathbb{C}^{K \times KN_s}$  is defined as

$$\mathbf{D}_t \triangleq \bigoplus_{k=0}^{K-1} \mathbf{s}_t^T[k], \quad (4.88)$$

and  $\mathbf{g}_i \in \mathbb{C}^{KN_s \times 1}$  is defined as

$$\begin{aligned} \mathbf{g}_i &= \begin{bmatrix} \mathbf{g}_i[0] \\ \vdots \\ \mathbf{g}_i[K-1] \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{F}_{\text{BB}}^T[0] [\mathbf{G}[0]]_{i,:}^T \\ \vdots \\ \mathbf{F}_{\text{BB}}^T[0] [\mathbf{G}[K-1]]_{i,:}^T \end{bmatrix}. \end{aligned} \quad (4.89)$$

Notice that (4.87) can be also expressed as

$$\mathbf{y}_{i,t} = \mathbf{P}(\boldsymbol{\theta}_t) \mathbf{F}^* \mathbf{G}_i \mathbf{s}_t + \mathbf{v}_{i,t}, \quad (4.90)$$

where  $\mathbf{G}_i \in \mathbb{C}^{K \times KN_s}$  is defined as

$$\mathbf{G}_i \triangleq \bigoplus_{k=0}^{K-1} \mathbf{g}_i^T[k], \quad (4.91)$$

and  $\mathbf{s}_t \in \mathbb{C}^{KN_s \times 1}$  is defined as

$$\mathbf{s}_t \triangleq [\mathbf{s}_t^T[0] \quad \dots \quad \mathbf{s}_t^T[K-1]]^T. \quad (4.92)$$

As it will soon become apparent, the expression in (4.87) will be useful for channel estimation, while the alternative expression in (4.90) will be used for data detection. To reduce computational complexity during PN estimation, similarly to the proposed EKF-RTS-EM algorithm, PN tracking is performed on a sample-by-sample basis. The  $n$ -th received sample is  $L_r$ -dimensional and it is given by

$$\mathbf{y}_t[n] = e^{j\theta_t[n]} \begin{bmatrix} \mathbf{f}_n^* \mathbf{D}_t \mathbf{g}_1 \\ \vdots \\ \mathbf{f}_n^* \mathbf{D}_t \mathbf{g}_{L_r} \end{bmatrix} + \begin{bmatrix} v_{1,t}[n] \\ \vdots \\ v_{L_r,t}[n] \end{bmatrix}. \quad (4.93)$$

First, using the received DMRS pilots, the proposed EKF-RTS-EM algorithm is employed to iteratively estimate the frequency-selective low-dimensional channels  $\{bsfg_i\}_{i=1}^{L_r}$  while compensating for the PN impairment. Then, the received data signal in (4.93) undergoes an iterative data detection, and PN estimation and compensation process. Even though channel estimates are readily available after initial CSI acquisition, those have been obtained at the low SNR regime. Motivated by the deployment of especially dense constellations such as 64/256-QAM, it is crucial to obtain high-quality channel estimates to be able to track the PN impairment and detect the transmitted data. To exploit the information coming from every RF chain, it is of interest to express the received signal after stacking the vectors in (4.87), (4.90) for

the different RF chains, which respectively yield

$$\begin{aligned}
\underbrace{\begin{bmatrix} \mathbf{y}_{1,t} \\ \vdots \\ \mathbf{y}_{L_r,t} \end{bmatrix}}_{\mathbf{y}_t} &= (\mathbf{I}_{L_r} \otimes \mathbf{P}(\boldsymbol{\theta}_t)) \mathbf{F}_\otimes^* (\mathbf{I}_{L_r} \otimes \mathbf{D}_t) \underbrace{\begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{L_r} \end{bmatrix}}_{\mathbf{g}} + \begin{bmatrix} \mathbf{v}_{1,t} \\ \vdots \\ \mathbf{v}_{L_r,t} \end{bmatrix} \\
&= (\mathbf{I}_{L_r} \otimes \mathbf{P}(\boldsymbol{\theta}_t)) \mathbf{F}_\otimes^* \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_{L_r} \end{bmatrix} \mathbf{s}_t + \begin{bmatrix} \mathbf{v}_{1,t} \\ \vdots \\ \mathbf{v}_{L_r,t} \end{bmatrix}.
\end{aligned} \tag{4.94}$$

Based on the received signal in (4.94), the LLF for the CFO compensated  $t$ -th received OFDM symbol  $\mathbf{y}_t$  can be written as

$$\begin{aligned}
\log p(\mathbf{y}_t, \mathbf{s}_t, \boldsymbol{\theta}_t) &= -C - \underbrace{\frac{1}{\sigma^2} \sum_{i=1}^{L_r} \|\mathbf{y}_{i,t} - \mathbf{P}(\boldsymbol{\theta}_t) \mathbf{F}_i^* \mathbf{G}_i \mathbf{s}_t\|_2^2}_{\text{Likelihood of data}} - \underbrace{\frac{1}{\sigma_d^2} \|\mathbf{s}_t\|_2^2}_{\text{Prior PDF of } \mathbf{s}_t} \\
&\quad - \underbrace{\frac{1}{2} \boldsymbol{\theta}_t^T \mathbf{C}_{\boldsymbol{\theta}_t, \boldsymbol{\theta}_t}^{-1} \boldsymbol{\theta}_t}_{\text{Prior PDF of PN } \boldsymbol{\theta}_t},
\end{aligned} \tag{4.95}$$

where a Gaussian prior is placed on the transmitted constellation symbols. Even though the transmitted symbols are not Gaussian-distributed, this approximation is useful since it leads to low-complexity linear detectors for the received data [29, 152]. It is important to note that, in the event that a uniform prior is placed on the transmitted constellation symbols, the LMMSE estimator of  $\mathbf{s}_t$  is the solution to maximizing (4.95) obtained after placing a Gaussian prior to the transmitted symbols. While the mathematical solution to both problems is identical, the assumptions are different in both cases, and so the properties of the corresponding estimators of  $\mathbf{s}_t$  are, as well.

To perform PN estimation, and data detection, I propose to follow an EM-based approach, for similar reasons as with the proposed synchronization

algorithms. The ML estimator of  $\mathbf{g}_i$ ,  $\hat{\mathbf{g}}_{i,\text{ML}}$  found using training DMRS pilots, is used to both estimate the  $N_{\text{data}}$  data symbols and compensate for the PN impairment at the data subcarriers. In the event that a dense constellation, i.e. 64/256-QAM is employed, PN tracking depends on the quality of the equalized constellation symbols, which decreases with the order of the constellation. Therefore, for densely packed constellations, there is a potentially high increase in estimation quality while keeping computational complexity low.

Motivated by the linear model with respect to  $\mathbf{s}_t$  in (4.90), the expectation step (E-step) is performed with respect to the PN vector, which yields (see (4.52), (4.90))

$$Q\left(\{\mathbf{g}_i\}_{i=1}^{L_r}, \mathbf{s}_t, \{\hat{\mathbf{g}}_i^{(n)}\}_{i=1}^{L_r}, \hat{\mathbf{s}}_t^{(n)}\right) \propto -\frac{1}{\sigma^2} \sum_{i=1}^{L_r} \left\| \mathbf{y}_{i,t} - \mathbf{P}\left(\hat{\boldsymbol{\theta}}_{t,\text{MMSE}}^{(n)}\right) \mathbf{F}^* \hat{\mathbf{G}}_i \mathbf{s}_t \right\|_2^2 - \frac{1}{\sigma_d^2} \|\mathbf{s}_t\|_2^2. \quad (4.96)$$

Similarly to (4.54), the expected LLF of the received signal depends on the MMSE estimator for the PN impairment, which can be approximated by using the EKF-RTS set of equations (4.68)-(4.71). Let the estimated PN vector at the end of the  $n$ -th E-step be denoted by  $\hat{\boldsymbol{\theta}}_{t,\text{MMSE}}^{(n)}$ .

The final part of the proposed data detection algorithm is the estimation of the data symbols. Since prior information on the data symbols is approximately modeled as Gaussian, it turns out that, after taking the partial derivative of (4.96) and using the estimates  $\hat{\mathbf{g}}_{i,\text{ML}}$ ,  $\hat{\boldsymbol{\theta}}_{t,\text{MMSE}}^{(n)}$ , the optimal maximizer  $\mathbf{s}_t$  of (4.96) at the  $n$ -th iteration is the MAP estimator,  $\hat{\mathbf{s}}_{t,\text{MAP}}^{(n)}$ , given

by

$$\begin{aligned} \hat{\mathbf{s}}_{t,\text{MAP}}^{(n)} &= \left( \left( \bigoplus_{i=1}^{L_r} \hat{\mathbf{G}}_{i,\text{ML}} \hat{\mathbf{G}}_{i,\text{ML}}^* \right) + \frac{\sigma^2}{\sigma_d^2} \mathbf{I}_{KL_r} \right)^{-1} \\ &\quad \times \left( \bigoplus_{i=1}^{L_r} \hat{\mathbf{G}}_{i,\text{ML}}^* \right) \mathbf{F}_{\otimes} \left( \mathbf{I}_{L_r} \otimes \mathbf{P}^* \left( \hat{\boldsymbol{\theta}}_{t,\text{MMSE}}^{(n)} \right) \right) \mathbf{y}_t, \end{aligned} \quad (4.97)$$

which can also be seen as the MMSE estimator for the data symbols conditioned on the current channel and PN estimates. Furthermore, as discussed earlier in this section, if a uniform prior is placed on the transmitted symbols, the LMMSE solution to the data symbols estimation problem coincides with the solution in (4.97), even though the properties of both the MAP and LMMSE estimators are different. Using the data subcarrier positions, the final symbol estimates are given by  $\left[ \hat{\mathbf{s}}_{t,\text{MAP}}^{(n)} \right]_{\mathcal{P}_{\text{data}}}$ . The proposed algorithm iteratively updates the PN and data estimates until the difference between the LF at two iterations is smaller than a threshold  $\eta$ , similarly to (4.66). To obtain the LF, the estimated symbols in (4.97) are replaced by their hard decisions, since the transmitted symbols belong to a particular constellation. Let  $\hat{\mathbf{s}}_{t,\text{MAP}}^{(0)}$  denote the initial estimate of the transmitted data vector. Appropriate initialization of  $\hat{\mathbf{s}}_{t,\text{MAP}}^{(0)}$  results in the proposed iterative detector converging quickly. In the proposed algorithm, the initial data estimate is obtained using the approximate LMMSE estimate for  $\boldsymbol{\theta}_{t-1}$  from the previous OFDM symbol.

This initial estimate can be obtained as

$$\begin{aligned} \hat{\mathbf{s}}_{t,\text{MAP}}^{(0)} &= \left( \left( \bigoplus_{i=1}^{L_r} \hat{\mathbf{G}}_{i,\text{ML}}^{(0)*} \hat{\mathbf{G}}_{i,\text{ML}}^{(0)} \right) + \frac{\sigma^2}{\sigma_d^2} \mathbf{I}_{KL_r} \right)^{-1} \\ &\quad \times \left( \bigoplus_{i=1}^{L_r} \hat{\mathbf{G}}_{i,\text{ML}}^{(0)*} \right) \mathbf{F}_{\otimes} \left( \mathbf{I}_{L_r} \otimes \mathbf{P}^* \left( \hat{\boldsymbol{\theta}}_{t-1,\text{MMSE}}^{(N)} \right) \right) \mathbf{y}_t, \end{aligned} \quad (4.98)$$

which can be seen to comprise a PN cancellation block, Fast Fourier Transform (FFT)-based OFDM demodulation, and an LMMSE detector for the transmitted data.

## 4.7 Numerical Results

This section includes numerical results obtained with the proposed synchronization algorithms. These results are obtained after performing Monte Carlo simulations averaged over 100 trials to evaluate the NMSE, ergodic spectral efficiency, and BER. I also provide calculations of the computational complexity for the proposed algorithms in Table 4.2 and prior work in Table 4.3.

Unless otherwise stated, the typical parameters for the system configuration are as follows and included in Table 4.1. Both the transmitter and the receiver are assumed to use a ULA with half-wavelength separation. Such a ULA has array response vectors given by  $[\mathbf{a}_T(\theta_\ell)]_n = \sqrt{\frac{1}{N_t}} e^{jn\pi \cos(\theta_\ell)}$ ,  $n = 0, \dots, N_t - 1$  and  $[\mathbf{a}_R(\phi_\ell)]_m = \sqrt{\frac{1}{N_r}} e^{jm\pi \cos(\phi_\ell)}$ ,  $m = 0, \dots, N_r - 1$ , for both transmitter and receiver, respectively. The I take  $N_t = 64$  and  $N_r = 32$  for illustration, and  $G_t = G_r = 128$ . The phase-shifters used in both the

Table 4.1: Summary of typical system configuration parameters

Simulation Parameters		
Description	Parameter	Value
Number of transmit antennas	$N_t$	64
Number of receive antennas	$N_r$	32
Number of transmit RF chains	$L_t$	8
Number of receive RF chains	$L_r$	4
Transmit angular grid size	$G_t$	128
Receive angular grid size	$G_r$	128
Number of phase-shifter quantization bits	$N_Q$	6
Number of training frames	$M$	32
Number of OFDM training symbols/frame	$N_{tr}$	2
Number of OFDM subcarriers	$K$	256
Number of CP samples	$L_c$	32
Sampling period	$T_s$	0.509 ns
Number of multipath channel clusters	$C$	40
Number of multipath channel rays/cluster	$R_c$	20
Channel Rician factor		-10 dB
Pulse-shape roll-off factor	$\beta$	0.3
Maximum CFO	$\Delta f^{(m)}$	5 ppm [5]
PN PSD	$G_\theta$	-85 dBc/Hz

transmitter and the receiver are assumed to have  $N_Q$  quantization bits, so that the entries of the analog training precoders and combiners  $\mathbf{F}_{tr}^{(m)}$ ,  $\mathbf{W}_{tr}^{(m)}$ ,  $m = 1, 2, \dots, M$  are drawn from a set  $\mathcal{A} = \left\{0, \frac{2\pi}{2^{N_Q}}, \dots, \frac{2\pi(2^{N_Q}-1)}{2^{N_Q}}\right\}$ . The number of quantization bits is set to  $N_Q = 6$ . The number of RF chains is set to  $L_t = 8$  at the transmitter and  $L_r = 4$  at the receiver. The number of OFDM subcarriers is set to  $K = 256$ , the carrier frequency is set to 60 GHz, and the sampling period is set to  $T_s = 0.509$  ns [37].

The frequency-selective mmWave MIMO channel is generated using

(4.7) with small-scale parameters taken from the QuaDRiGa channel simulator [31], [32], which implements the 3GPP 38.901 UMi channel model in [6]. The channel samples are generated with an average Rician factor of  $-10$  dB, mobility with speed 20 m/s, and the distance between the transmitter and the receiver is set to  $d = 30$  meters for illustration.

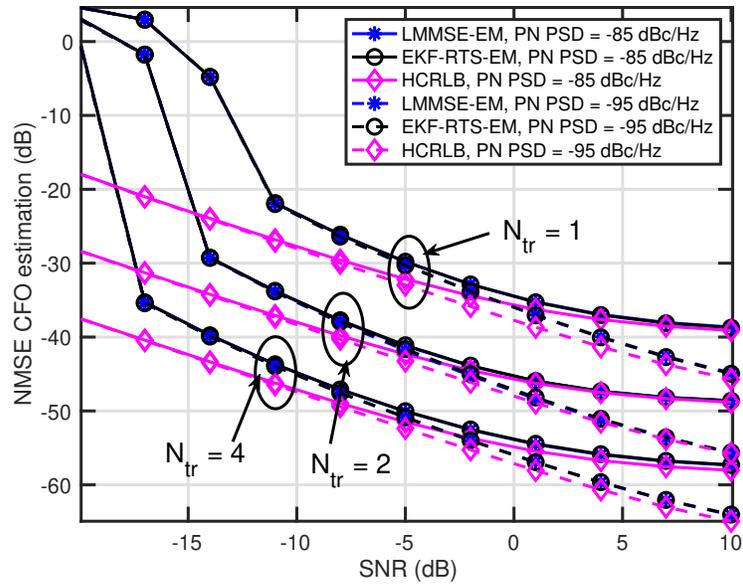
#### 4.7.1 Performance analysis of EM-based algorithms

In this subsection, I analyze the performance of the proposed LMMSE-EM and EKF-RTS-EM algorithms from Section 4.5, and compare their achievable performance to the derived hybrid CRLB from Section 4.4.

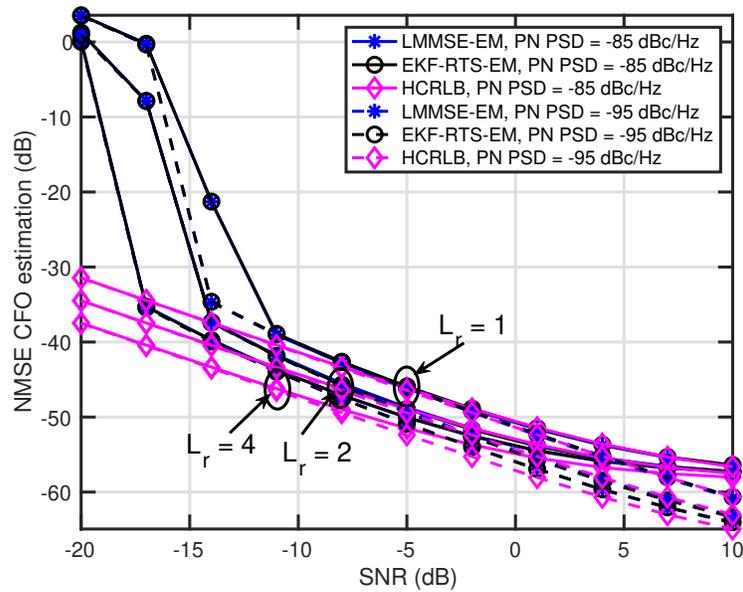
In Fig. 4.6, I show the evolution of the NMSE of the CFO estimates versus SNR for the proposed LMMSE-EM and EKF-RTS-EM algorithms. The hybrid CRLB is also provided as an estimation performance bound. I evaluate both algorithms using two different values for the PN variance, which are  $G_\theta = -85$  dBc/Hz and  $G_\theta = -95$  dBc/Hz. In Fig. 4.6 (a), I set the number of receive RF chains to  $L_r = 4$  and sweep  $N_{tr}$  within the range  $\{1, 2, 4\}$  OFDM training symbols. Conversely, in Fig. 4.6 (b), the number of OFDM training symbols is set to  $N_{tr} = 4$ , and the number of receive RF chains is swept within the range  $\{1, 2, 4\}$ .

Several observations can be made from Fig. 4.6:

- The proposed LMMSE-EM and EKF-RTS-EM algorithms exhibit very similar estimation performance, which suggests that the proposed EKF-



(a)



(b)

Figure 4.6: Evolution of the NMSE of the CFO estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound.

RTS-EM algorithm does not compromise estimation performance while dramatically reducing computational complexity during the measurement update in PN estimation.

- The estimation performance of the proposed algorithms exhibits a small gap with respect to the hybrid CRLB. At low SNR, the gap between the NMSE and the hybrid CRLB is more noticeable, but it shrinks as  $\text{SNR} \rightarrow \infty$ . It is also observed that the NMSE and the hybrid CRLB are monotonically decreasing proportionally to the SNR. There is, however, a certain SNR value beyond which both the performance of the proposed algorithms and the hybrid CRLB saturate and exhibit a plateau effect. This behavior sets the distinction between the noise-limited regime and the PN-limited regime, whereby estimation performance cannot longer improve even if  $\text{SNR} \rightarrow \infty$ . This behavior is shown in Fig. 4.7 for  $N_{\text{tr}} = 4$  and  $L_r = 4$ .
- The estimation performance in the low SNR regime does not depend on the PSD of the PN, which indicates that the synchronization performance is limited by the AWGN. Notice, however, that as  $\text{SNR} \rightarrow \infty$ ,
- The estimation performance in the low SNR regime does not depend on the PSD of the PN, which indicates that the synchronization performance is limited by the AWGN. Notice, however, that as  $\text{SNR} \rightarrow \infty$ , both the estimation performance of the proposed algorithms and the hybrid CRLB are different for the two values of the PSD  $G_\theta$  of the PN. In-

tuitively, as  $G_\theta$  increases, the information coupling between the PN and the CFO impairments increases, thereby reducing both the achievable CFO estimation performance of the proposed algorithms and the hybrid CRLB. Conversely, reducing  $G_\theta$  reduces this information coupling, which results in better CFO estimates and lower hybrid CRLB.

- When the SNR is very low, the NMSE of the proposed algorithms is high. However, when the SNR increases, a waterfall effect is observed, and the SNR at which this effect happens depends on both the number of RF chains  $L_r$  and the number of OFDM training symbols  $N_{tr}$ . More specifically, increasing  $N_{tr}$  or  $L_r$  shifts the minimum SNR at which this waterfall effect is observed. Thereby, increasing  $N_{tr}$  and  $L_r$  results in more accurate estimates of the CFO parameter, even for SNR  $< -10$  dB.
- Last, increasing the number of OFDM training symbols  $N_{tr}$  and the number of receive RF chains  $L_r$  have a different impact on both the CFO estimation performance and the hybrid CRLB. More specifically, doubling  $N_{tr}$  results in a performance gain of approximately 9 dB, which indicates that the estimation performance depends on  $N_{tr}^3$ , which is a similar result to the CFO estimation performance and CRLB from Chapter 3. Regarding the number of receive RF chains, doubling  $L_r$  enhances estimation performance by a factor of 3 dB, which indicates that the estimation approach averages the receive noise across multiple receive

RF chains, thereby exhibiting an NMSE estimation performance proportional to  $L_r^{-1}$ .

In Fig. 4.8 (a), I show the NMSE evolution of the equivalent beamformed channels versus SNR, for both the proposed LMMSE-EM and the EKF-RTS-EM algorithms. The number of receive RF chains is set to  $L_r = 4$ , and the number of OFDM training symbols  $N_{tr}$  is swept within  $\{1, 2, 4\}$ . A similar behavior to that in Fig. 4.6 is observed. For both proposed algorithms, the estimation performance is very close to the hybrid CRLB, although there is a more noticeable performance gap for  $\text{SNR} < 10$  dB. Similar to Fig. 4.6, it is observed that the reduction in computational complexity of the second proposed EKF-RTS-EM algorithm does not compromise estimation performance, thereby showing that synchronization in the low SNR regime can be successfully accomplished with reduced computational complexity. It is also observed that doubling the number of OFDM training symbols  $N_{tr}$  results in enhanced estimation performance by a factor of 3 dB, which is expected since it was observed that the Fisher information of the channel coefficients increases linearly with the number of training samples.

Furthermore, the performance of the proposed algorithms, as well as the hybrid CRLB, do not depend on the PSD of the PN in both the mid and low SNR regimes. Notice, however, that for  $\text{SNR} > 0$  dB, the estimation performance and hybrid CRLB depend on the PSD of the PN, similar to Fig. 4.6. This behavior sets the beginning of the PN-limited regime, which is

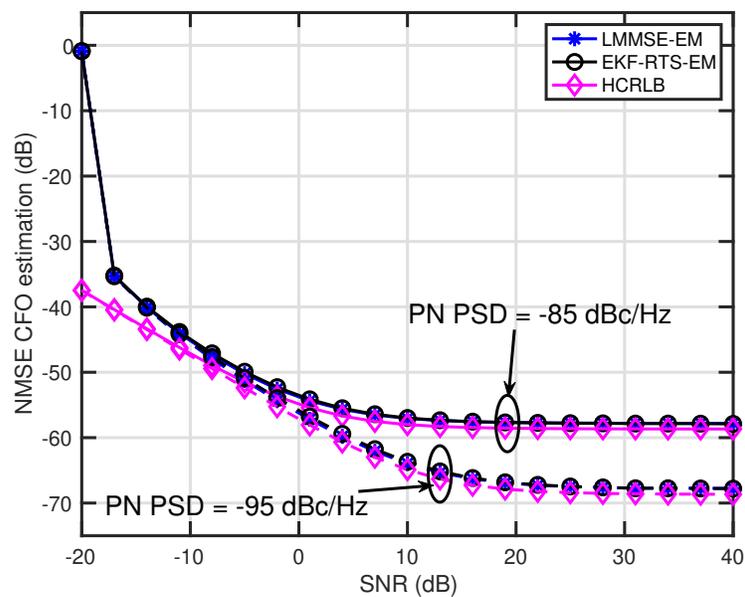
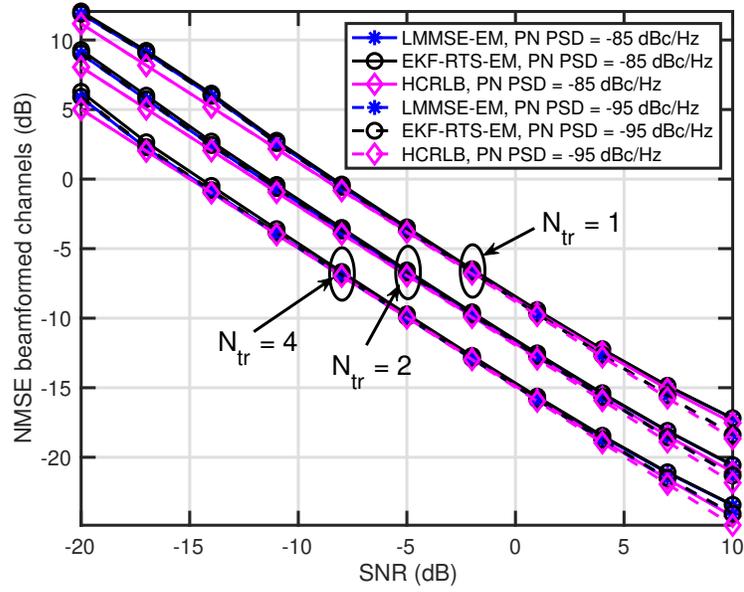
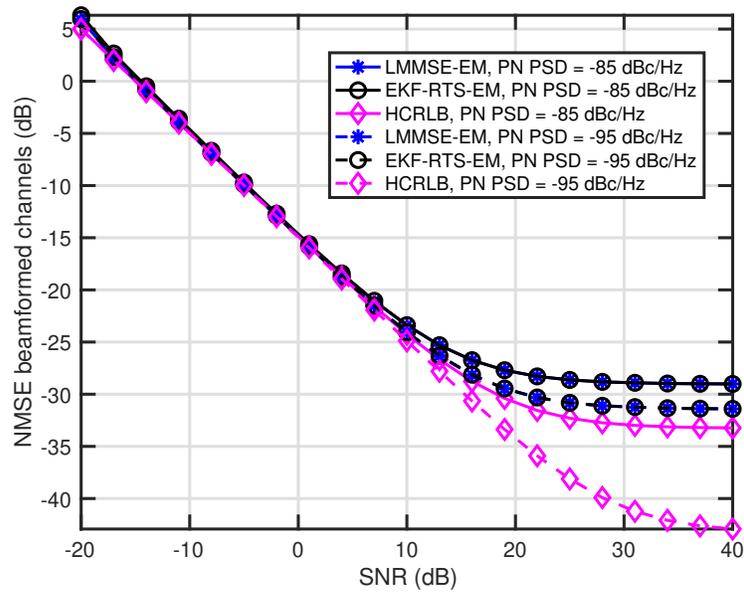


Figure 4.7: Asymptotic evolution of the NMSE of the CFO estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound.



(a)



(b)

Figure 4.8: Evolution of the NMSE of the beamformed channel estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound.

more pronounced as  $\text{SNR} \rightarrow \infty$ , as shown in Fig. 4.8 (b). In Fig. 4.8, the number of OFDM training symbols is set to  $N_{\text{tr}} = 4$ . It is observed that the estimation performance gap between the proposed algorithms and the hybrid CRLB increases as  $\text{SNR} \rightarrow \infty$ , and it is more pronounced for smaller values of the PSD of the PN. This behavior is due to two different factors: i) instead of using the MMSE estimator for the PN, the proposed algorithms attempt to approximate this estimator using statistical linearization (LMMSE), such that non-linearities are not dealt with, and ii) for smaller values of the PSD of the PN, the covariance matrix of the PN has smaller eigenvalues, thereby reducing the amount of prior information on this parameter. This second fact makes the covariance after propagation update be significantly larger than the AWGN impairment, thereby making the Kalman gain for the PN estimator rely more heavily on the measurement and less on the AWGN. Consequently, the PN impairment is more difficult to estimate, which affects the estimation of the equivalent beamformed channels.

Last, I show the estimation performance of the proposed LMMSE-EM and EKF-RTS-EM algorithms versus SNR in Fig. 4.9 and Fig. 4.10. In Fig. 4.9, the number of receive RF chains is swept within  $\{1, 2, 4\}$ , and it is set to  $L_r = 4$  in Fig. 4.10. The PSD of the PN is set to  $G_\theta = -85$  dB in Fig. 4.9, which corresponds to a stronger PN process. The number of OFDM training symbols is set to  $N_{\text{tr}} = 4$  in both Fig. 4.9 and Fig. 4.10.

The first observation from both Fig. 4.9 and Fig. 4.10 is that the PN estimation performance of the first proposed LMMSE-EM algorithm sig-

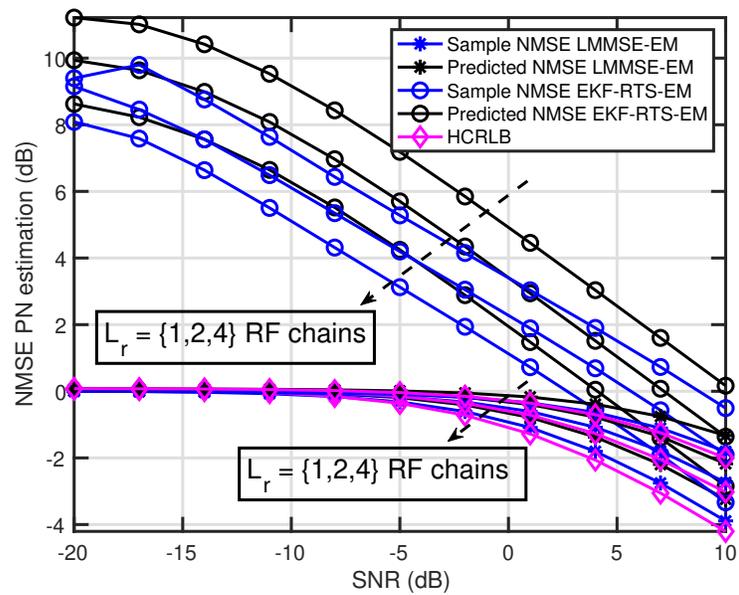


Figure 4.9: Evolution of the NMSE of the PN estimates obtained using the proposed algorithms versus SNR. The hybrid CRLB is also provided as a performance bound.

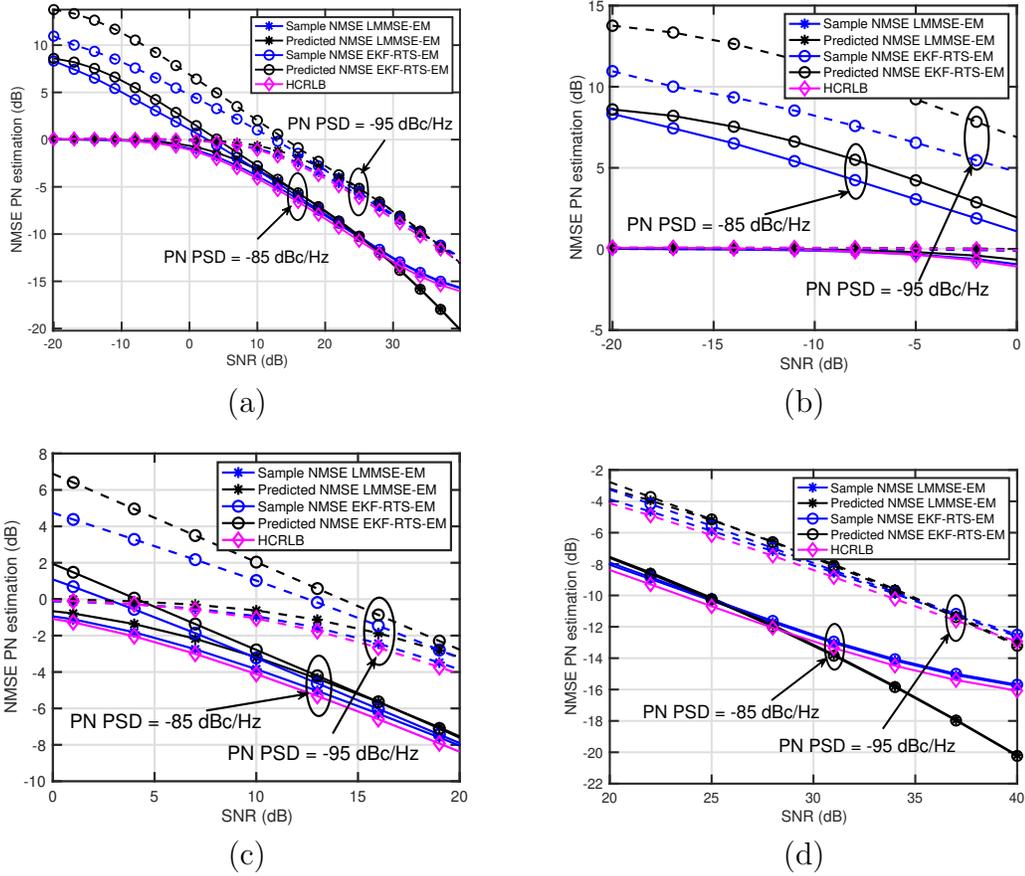


Figure 4.10: Asymptotic evolution of the NMSE of the PN estimates obtained using the proposed algorithms versus SNR. The curves in (a) show this asymptotic evolution for SNR values between  $-20$  and  $40$  dB. Magnified curves of the asymptotic evolution are also shown for  $\text{SNR} \in [-20, 0]$  dB in (b),  $\text{SNR} \in [0, 20]$  dB in (c), and  $\text{SNR} \in [20, 40]$  dB in (d). The hybrid CRLB is also provided as a performance bound.

nificantly outperforms that of the second proposed EKF-RTS-EM algorithm. This is not surprising, since the estimation approach in both cases comes from a linearization of the measurement signal through its Jacobian matrix, which is very sensitive to AWGN if a small number of measurements are processed. If a single  $L_r$ -dimensional measurement is processed, as in the EKF-RTS-EM algorithm, the Jacobian matrix of the measurement varies significantly depending on the AWGN variance and the PSD of the PN process, thereby making it harder for the second proposed algorithm to track the PN variations. If every measurement is processed in a larger-dimensional batch, as in the first proposed LMMSE-EM algorithm, the Kalman measurement update is performed taking the multiple measurements into account, thereby contributing to stabilizing the Jacobian matrix and making it easier to track the PN variations. This is a significant effect only in the low SNR regime, and the performances of both proposed algorithms exhibit convergence as SNR increases, as shown in Fig. 4.10.

It is also observed that increasing  $L_r$  results in a noise averaging effect, which results in an estimation performance improvement of approximately 3 dB. This is a similar effect to that in Fig. 4.8 as a function of  $N_{tr}$ , which indicates that the AWGN can be more effectively filtered out as  $L_r$  increases.

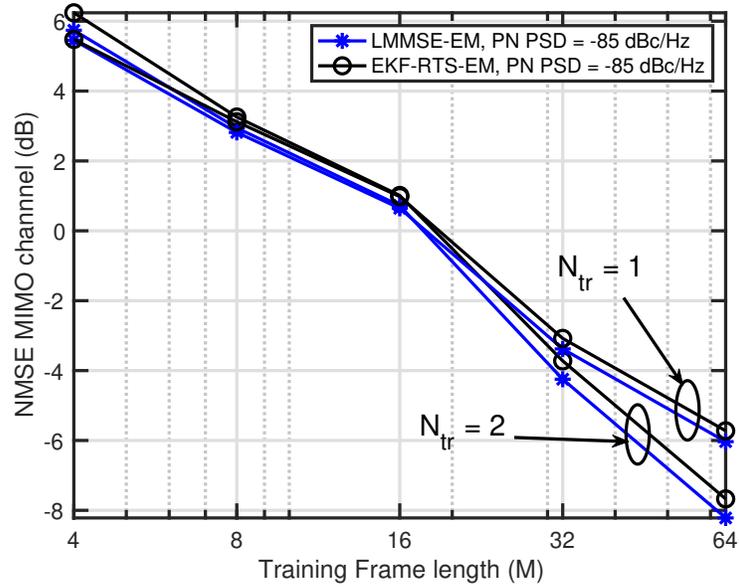
It is also observed that, for a wide range of SNR values, the first proposed LMMSE-EM algorithm exhibits estimation performance lying very close to the hybrid CRLB, and divergence from the bound is observed as  $\text{SNR} \rightarrow \infty$ , for similar reasons as with Fig. 4.8 (b). It is also observed that the NMSE

predicted from the first proposed LMMSE-EM algorithm is very close to the actual NMSE performance, while for the second proposed EKF-RTS-EM algorithm it is more difficult to predict the NMSE performance in the low SNR regime, which is expected due to the varying nature of the Jacobian matrix of the measurement when the received samples are sequentially processed, instead of performing simultaneous batch-processing, as in the first proposed LMMSE-EM algorithm.

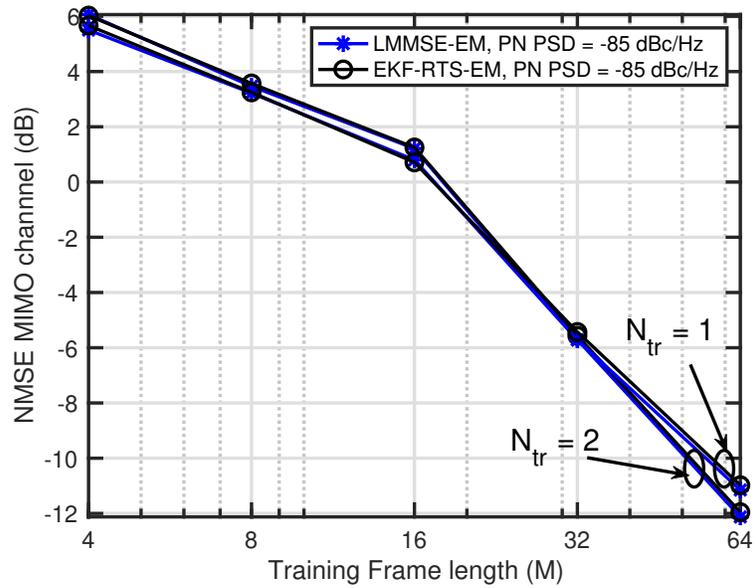
#### 4.7.2 NMSE of channel estimator

In this subsection, I analyze the performance of the proposed channel estimation algorithms in the absence of synchronization. I show in Fig. 4.11 the evolution of the NMSE of the high-dimensional MIMO channel estimates versus the training frame length  $M$ , for  $\text{SNR} = \{-10, 0\}$  dB and  $N_{\text{tr}} = \{1, 2\}$  OFDM training symbols.

First, it can be observed from Fig. 4.11 that the estimation performance of the proposed algorithms increases with  $M$ , as expected. Notice, however, that for  $\text{SNR} = -10$  dB, there is a small yet noticeable performance gap between the first proposed LMMSE-EM algorithm and the second proposed EKF-RTS-EM algorithm. This effect is due to the more effective PN estimation carried out by the batch-processing-based LMMSE-EM algorithm, which helps obtaining more accurate estimates of both the CFO and the frequency-selective beamformed channels. Furthermore, the stopping criterion of the proposed channel estimation algorithm depends on the hybrid CRLB for



(a)



(b)

Figure 4.11: Evolution of the NMSE of the mmWave MIMO channel estimates obtained using the proposed algorithms versus  $M$ , for SNR = -10 dB (a) and SNR = 0 dB (b).

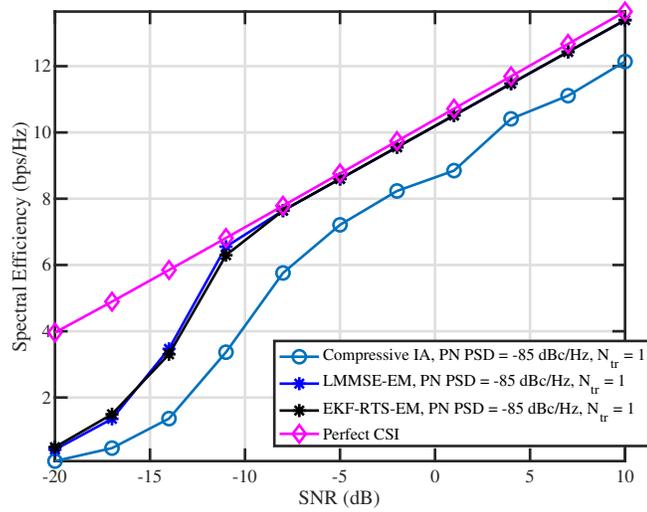
the estimation of the different parameters. Therefore, even though both synchronization algorithms exhibit similar performance as shown in the previous subsection, the first proposed LMMSE-EM algorithm produces slightly more accurate phase information estimates than the second proposed EKF-RTS-EM algorithm, thereby resulting in MIMO channel estimates having slightly higher quality. Notice, however, that using  $N_{\text{tr}} = 2$  results in the EKF-RTS-EM algorithm producing MIMO channel estimates that have slightly higher quality than the LMMSE-EM algorithm using only  $N_{\text{tr}} = 1$ . This suggests a trade-off between overhead and computational complexity between the LMMSE-EM and the EKF-RTS-EM algorithms. Notice that accurate estimation of the CFO parameter in the low SNR regime heavily depends on  $N_{\text{tr}}$ , such that using  $N_{\text{tr}} = 2$  is desirable to guarantee accurate synchronization in the practical SNR range that mmWave systems are expected to work.

Furthermore, using  $N_{\text{tr}} = 2$  OFDM training symbols increases the estimation performance gap between both algorithms at  $\text{SNR} = -10$  dB, while there is marginal improvement at  $\text{SNR} = 0$  dB. The reasoning for this effect is similar to the effect discussed in the previous paragraph. At  $\text{SNR} = -10$  dB, increasing  $N_{\text{tr}}$  significantly enhances the CFO estimator performance, while at  $\text{SNR} = 0$  dB such enhancement is more negligible. Thereby, it can be concluded that having  $N_{\text{tr}} > 1$  plays a pivotal role in obtaining accurate phase synchronization information in the low SNR regime, while using  $N_{\text{tr}} = 1$  in the mid SNR regime yields phase synchronization information having comparable quality to that obtained using  $N_{\text{tr}} = 2$ .

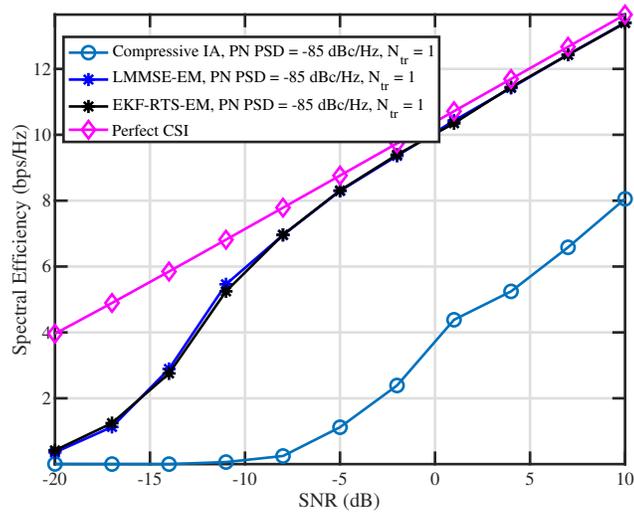
### 4.7.3 Spectral efficiency

In this subsection, I analyze the achievable ergodic spectral efficiency obtained with the proposed algorithms, and compare it to that of the state-of-the-art compressive initial access algorithm in [5]. In Fig. 4.12, I show the evolution of the spectral efficiency versus SNR using  $N_{\text{tr}} = 1$  OFDM symbol and  $L_{\text{r}} = 1$  RF chains, for both of the proposed algorithms as well as the algorithm in [5]. The number of training frames is set to  $M = 32$ . The channel model is set as the 3GPP UMi LOS channel model in Fig. 4.12 (a), and the 3GPP UMi NLOS channel model in Fig. 4.12 (b), with a Rician factor of 0 dB.

The first observation from Fig. 4.12 is that the algorithm in [5] is unable to properly estimate the AoD and AoA of the dominant multipath component, while the proposed algorithms manage to successfully establish synchronization for  $\text{SNR} \geq -11$  dB in both cases. This is due to: i) the TO and CFO synchronization algorithm in [5], which is based on the Moose algorithm [110], ii) the algorithm in [5] does not consider the effect of PN impairments on the received signal, and iii) the algorithm in [5] was designed specifically for LOS channels, so that the algorithm is ineffective under NLOS channel models. The Moose algorithm is very effective when both the transmit training sequences are periodic and the SNR is high enough, although it is also used when the transmit training sequences are aperiodic yet the channel has high enough Rician factor. Since the training sequences used in [5] are the Synchronization Signal (SS) Blocks used in 5G NR, these sequences are



(a)



(b)

Figure 4.12: Evolution of the ergodic spectral efficiency achieved by the proposed algorithms and the compressive beamforming algorithm in [5] under PN impairments versus SNR, for  $N_{\text{tr}} = 1$  OFDM training symbol. The curves in (a) are obtained using the mmMAGIC LOS channel model, while those in (b) are obtained using the mmMAGIC NLOS channel model with a Rician factor of  $-10$  dB.

periodic across SS Blocks comprising 4 OFDM symbols. Using only  $M = 32$  training frames, the algorithm in [5] estimates  $M - 1 = 31$  CFO estimates by comparing the phase differences of the received signal between consecutive SS Blocks. Thereby, the long periodicity of the SS Blocks reduces the CFO estimation performance. Further, the inability to estimate and keep track of the PN impairment hardens the CFO estimation task, especially in the low SNR regime. In Fig. 4.12 (b), the CFO estimation task is significantly more challenging under a NLOS channel, thereby limiting the application of the Moose algorithm for frequency synchronization. Last, regarding AoD and AoA estimation, the algorithm in [5] assumes the channel comprises a single pair of AoD and AoA, which is not true under NLOS channels. When several AoDs/AoAs are present in the channel, the lack of orthogonality between array steering vectors results in information coupling between different AoDs and AoAs, which must be exploited to correctly estimate the different angular parameters [145]. For this reason, the algorithm in [5] fails to estimate the AoD and AoA corresponding to the dominant multipath component, thereby significantly reducing the achievable spectral efficiency.

In view of the results in Fig. 4.12, it can be concluded that the proposed algorithms significantly outperform the beam-training-based strategy in [5] in terms of achievable spectral efficiency, thereby making the proposed algorithms attractive owing to their applicability to more challenging NLOS scenarios even in the low SNR regime.

I show in Fig. 4.13 the evolution of the ergodic spectral efficiency

achieved by the proposed algorithms versus SNR, for  $N_{\text{tr}} = 2$ ,  $M = 32$  training frames, and  $N_s = \{1, 2, 4\}$  data streams.

The first observation from Fig. 4.13 is that the proposed algorithms are capable of achieving near-optimum spectral efficiency values for  $N_s = \{1, 2\}$ , yet there is a more noticeable spectral efficiency performance gap between the proposed algorithms and the perfect CSI scenario. This is due to the fact that, under the 3GPP UMi NLOS channel model with Rician factor of  $-10$  dB, the mmWave MIMO channel comprises a large number of multipath components, and exploiting the benefits of using a larger number of spatial degrees of freedom requires a more accurate estimation of the MIMO channel. This issue can be circumvented by increasing the number of training frames  $M$ , as well as the dictionary sizes  $G_t$  and  $G_r$ , as also discussed in Section 2.4.1.

Furthermore, it is also observed that both the first proposed LMMSE-EM algorithm and second proposed EKF-RTS-EM algorithm perform similarly for the entire SNR range under study. The more complex PN estimator in the LMMSE-EM algorithm results in slightly more accurate estimation of the different unknown parameters than in the EKF-RTS-EM algorithm, which in turn results in slightly higher spectral efficiency values. Notice, however, that this marginal increase in ergodic spectral efficiency comes at the expense of significantly higher computational complexity, such that the second proposed EKF-RTS-EM algorithm is a more flexible alternative enabling synchronization with reduced computational complexity.

Last, I show in Fig. 4.14 the evolution of the ergodic spectral efficiency

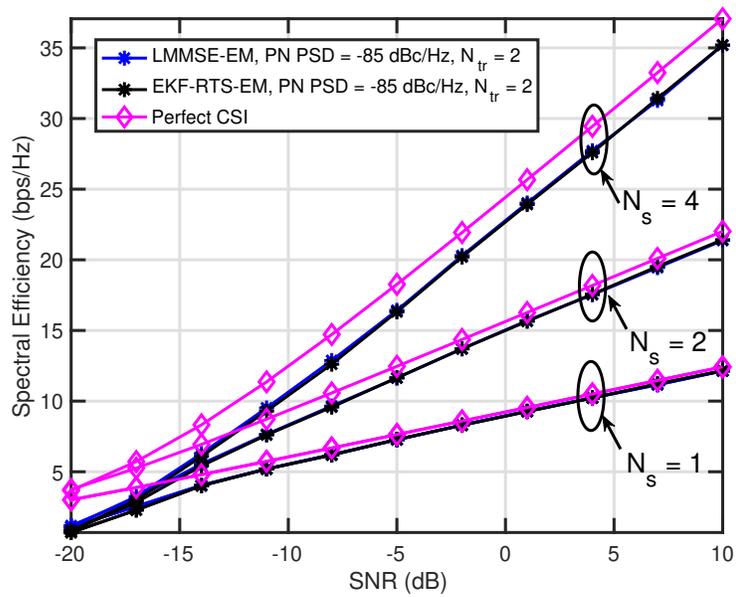
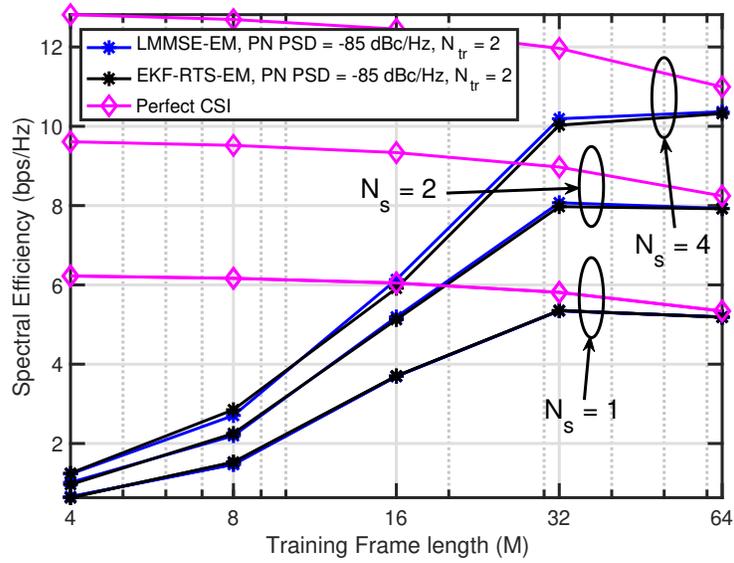


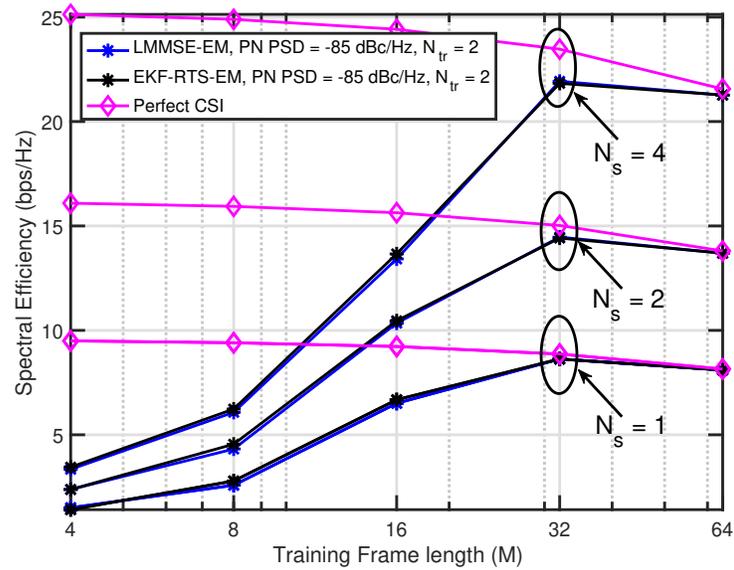
Figure 4.13: Evolution of the ergodic spectral efficiency achieved by the proposed algorithms versus SNR, for  $N_{\text{tr}} = 2$  OFDM training symbol. The curves are obtained using the 3GPP UMi NLOS channel model [6].

versus the number of training frames  $M$ , for SNR =  $-10$  dB (a) and SNR =  $0$  dB.

The first observation from Fig. 4.14 is that increasing the number of training frames  $M$  increases the ergodic spectral efficiency until a certain value of  $M$ , beyond which additional training results in an inadequate allocation of resources for channel estimation and data transmission, as expected. It is also observed that, similar to Fig. 4.13, near-optimum spectral efficiency values can be attained for  $N_s = \{1, 2\}$ , yet the performance gap with respect to perfect CSI increases for  $N_s = 4$ . The reasoning for this effect is similar to the one provided for Fig. 4.13. It is also observed that, for  $M = 64$  training frames, the proposed algorithms exhibit near-optimum spectral efficiency even for  $N_s = 4$ . This suggests that using more training pilots allows more accurate channel estimation since more compressed measurements are available to estimate the channel matrices. Notice, however, that the achievable spectral efficiency for  $M = 64$  at SNR =  $0$  dB is lower than its  $M = 32$  counterpart, which indicates that over-training the channel results in a penalty for the achievable spectral efficiency that can be achieved. Furthermore, for  $M < 32$ , there is a huge performance difference between the achievable spectral efficiency and the perfect CSI scenario. Therefore, in view of Fig. 4.14, it can be concluded that using  $M = 32$  comprises a reasonable trade-off between overhead and spectral efficiency performance.



(a)



(b)

Figure 4.14: Evolution of the ergodic spectral efficiency achieved by the proposed algorithms versus the number of training frames  $M$ , for  $\text{SNR} = -10$  dB (a) and  $\text{SNR} = 0$  dB (b), and  $N_{\text{tr}} = 2$  OFDM training symbols. The curves are obtained using the 3GPP UMi NLOS channel model [6].

#### 4.7.4 Computational complexity analysis

In this subsection, I analyze the computational complexity of the proposed LMMSE-EM and EKF-RTS-EM algorithms, as well as the proposed data detection and PN mitigation algorithm. For the proposed LMMSE-EM and EKF-RTS-EM algorithms, the variable  $N_{\Delta f}$  denotes the number of discrete frequencies used to perform the initial exhaustive-search-based CFO estimation, and the variable  $N_{\text{SIC}}$  denotes the number of iterative SIC-based optimization steps performed by the SS-SW-OMP+Th algorithm presented in Chapter 2. The online computational complexity of the proposed algorithms is summarized in Table 4.2. The computational complexity of the compressive beam-training-based approach in [5] is also included in Table 4.3, in which I include the notation used therein. Likewise, I include the number of measured FLOPs required by the proposed algorithms in Table 4.4 and Table 4.5, for the two following sets of system parameters:

- System Parameters 1:  $K = 256$ ,  $L_r = 4$ ,  $N_{\text{tr}} = 2$ ,  $N_{\Delta f} = 129$ , and  $N_{\text{DMRS}} = 2$ .
- System Parameters 2:  $K = 1024$ ,  $L_r = 4$ ,  $N_{\text{tr}} = 2$ ,  $N_{\Delta f} = 129$ , and  $N_{\text{DMRS}} = 2$ .

The number of measured FLOPs required by the proposed algorithms has been obtained using the MATLAB library in [153].

The main observation from Table 4.2 is that the proposed LMMSE-EM and EKF-RTS-EM algorithms exhibit higher online computational complexity

Table 4.2: Online computational complexity of proposed synchronization and channel estimation algorithms

LMMSE-EM	
Operation	Complexity
Initialize $\widehat{\Delta f}^{(m,0)}$	$\mathcal{O}(N_{\Delta f} L_r N_{tr} K \log_2(K))$
$M \times$ Initialize $\hat{\mathbf{g}}_i^{(m,0)}$	$\mathcal{O}(M L_r N_{tr} K \log_2(K))$
$M \times$ <b>Update PN estimate</b>	$\mathcal{O}(M N_{iter} (L_r K N_{tr})^3)$
$M \times$ Update $\widehat{\Delta f}_{ML}^{(m,n)}$	$\mathcal{O}(M N_{iter} L_r N_{tr} K \log_2(K))$
$M \times$ Update $\hat{\mathbf{g}}_{i,ML}^{(m,n)}$	$\mathcal{O}(M N_{iter} L_r N_{tr} K \log_2(K))$
$M \times$ Update LF difference	$\mathcal{O}(M N_{iter} L_r N_{tr} K)$
Overall synchronization	$\mathcal{O}(L_r K N_{tr} (\log_2(K) N_{\Delta f} + M N_{iter} (L_r K N_{tr})^2))$
Compute hybrid CRLB	$\mathcal{O}(L_r^2 L_c K \log_2(K))$
Channel estimation	$\mathcal{O}(N_{SIC} L_r M K_p G_r G_t)$
EKF-RTS-EM	
Operation	Complexity
Initialize $\widehat{\Delta f}^{(m,0)}$	$\mathcal{O}(N_{\Delta f} L_r N_{tr} K \log_2(K))$
$M \times$ Initialize $\hat{\mathbf{g}}_i^{(m,0)}$	$\mathcal{O}(M L_r N_{tr} K \log_2(K))$
$M \times$ <b>Update PN estimate</b>	$\mathcal{O}(M N_{iter} K N_{tr} L_r^3)$
$M \times$ Update $\widehat{\Delta f}_{ML}^{(m,n)}$	$\mathcal{O}(M N_{iter} L_r N_{tr} K \log_2(K))$
$M \times$ Update $\hat{\mathbf{g}}_{i,ML}^{(m,n)}$	$\mathcal{O}(M N_{iter} L_r N_{tr} K \log_2(K))$
$M \times$ Update LF difference	$\mathcal{O}(M N_{iter} L_r N_{tr} K)$
Overall synchronization	$\mathcal{O}(L_r K N_{tr} (\log_2(K) (N_{\Delta f} + M N_{iter}) + M N_{iter} L_r^2))$
Compute hybrid CRLB	$\mathcal{O}(L_r^2 L_c K \log_2(K))$
Channel estimation	$\mathcal{O}(N_{SIC} L_r M K_p G_r G_t)$
PN mitigation and data detection	
Operation	Complexity
Estimate $\hat{\mathbf{g}}_{i,ML}$	$\mathcal{O}(L_r N_{DMRS} K \log_2(K))$
Initialize $\hat{\mathbf{s}}_{t,MAP}^{(0)}$	$\mathcal{O}(L_r K \log_2(K))$
<b>Update PN estimate</b>	$\mathcal{O}(N_{iter} N_{data} K L_r^3)$
Update $\hat{\mathbf{s}}_{t,MAP}^{(n)}$	$\mathcal{O}(N_{iter} N_{data} L_r K \log_2(K))$
Overall	$\mathcal{O}(L_r K (N_{DMRS} \log_2(K) + N_{iter} N_{data} (L_r^2 + \log_2(K))))$

Table 4.3: Online computational complexity of previously proposed algorithm

Compressive beam training from [5]	
Operation	Complexity
PSS FIR corr.	$\mathcal{O}(PN_B)$
Detection and time sync.	$\mathcal{O}(N_B)$
Excess. delay est.	$\mathcal{O}(PG_D)$
AoA/AoD est.	$\mathcal{O}(MG_T G_R)$
CFO est.	$\mathcal{O}(MG_T G_R)$
Alternative updates	$\mathcal{O}(K_{\text{iter}}N_T N_R MP)$
Error norm evaluation	$\mathcal{O}(K_{\text{iter}}MP)$
Overall	$\mathcal{O}(MG_T G_R + P(K_{\text{iter}}MN_T N_R + M + G_D + N_B))$

than prior work in [5]. This fact is due to the algorithm in [5] being targeted at LOS channels, thereby only estimating a single multipath component in the frequency-selective mmWave MIMO channel. Consequently, the AoA/AoD estimation requires single-shot estimation and additional gradient-based refinement steps. Single-path estimation thereby requires a single compressive estimation step, without performing SIC-based multipath estimation as in the proposed SS-SW-OMP+Th algorithm in Chapter 2.

Notice, however, that the computational complexity of the approach in [5] increases linearly with the number of subcarriers (denoted by  $P$ ), the number of gradient iterations (denoted by  $K_{\text{iter}}$ ), and the number of transmit and receive antennas (denoted by  $N_T$  and  $N_R$ , respectively). The number of gradient iterations is set to  $K_{\text{iter}} = 1000$  for the algorithm in [5], which indicates this algorithm exhibits slow convergence for the desired estimation accuracy. For large numbers of transmit and receive antennas, the algorithm

Table 4.4: Asymptotic complexity from Tables 4.2 and 4.3 and number of measured FLOPs required by the proposed LMMSE-EM, EKF-RTS-EM, and PN mitigation and data detection algorithms for the set of system parameters in Table 4.1 of  $K = 256$ ,  $L_r = 4$ , and  $N_{tr} = 2$ , and algorithm parameters  $N_{\Delta f} = 129$ , and  $N_{DMRS} = 2$ .

LMMSE-EM		
Operation	Complexity	FLOPs
Initialize $\widehat{\Delta f}^{(m,0)}$	$2.1135 \cdot 10^6$	$5.0328 \cdot 10^6$
$M \times$ Initialize $\hat{\mathbf{g}}_i^{(m,0)}$	$16384M$	$13360M$
$M \times$ <b>Update PN estimate</b>	$8.5899 \cdot 10^9 MN_{iter}$	$10.7694 \cdot 10^9 MN_{iter}$
$M \times$ Update $\widehat{\Delta f}_{ML}^{(m,n)}$	$16384MN_{iter}$	$13817MN_{iter}$
$M \times$ Update $\hat{\mathbf{g}}_{i,ML}^{(m,n)}$	$16384MN_{iter}$	$13360MN_{iter}$
$M \times$ Update LF difference	$2048MN_{iter}$	$4108MN_{iter}$
Compute hybrid CRLB	$1.048 \cdot 10^6$	$3.984 \cdot 10^6$
Channel estimation	$1.342 \cdot 10^8 N_{SIC}$	$4.871 \cdot 10^8 N_{SIC}$
EKF-RTS-EM		
Operation	Complexity	FLOPs
Initialize $\widehat{\Delta f}^{(m,0)}$	$2.1135 \cdot 10^6$	$5.0328 \cdot 10^6$
$M \times$ Initialize $\hat{\mathbf{g}}_i^{(m,0)}$	$16384M$	$13360M$
$M \times$ <b>Update PN estimate</b>	$32768MN_{iter}$	$77815MN_{iter}$
$M \times$ Update $\widehat{\Delta f}_{ML}^{(m,n)}$	$16384MN_{iter}$	$13817MN_{iter}$
$M \times$ Update $\hat{\mathbf{g}}_{i,ML}^{(m,n)}$	$16384MN_{iter}$	$13360MN_{iter}$
$M \times$ Update LF difference	$2048MN_{iter}$	$4108MN_{iter}$
Compute hybrid CRLB	$1.048 \cdot 10^6$	$3.984 \cdot 10^6$
Channel estimation	$1.342 \cdot 10^8 N_{SIC}$	$4.871 \cdot 10^8 N_{SIC}$
PN mitigation and data detection		
Operation	Complexity	FLOPs
Estimate $\hat{\mathbf{g}}_{i,ML}$	$16384$	$13360$
Initialize $\hat{\mathbf{s}}_{t,MAP}^{(0)}$	$8192N_{iter}N_{data}$	$6696N_{iter}N_{data}$
<b>Update PN estimate</b>	$16384N_{iter}N_{data}$	$38904N_{iter}N_{data}$
Update $\hat{\mathbf{s}}_{t,MAP}^{(n)}$	$8192N_{iter}N_{data}$	$6696N_{iter}N_{data}$

Table 4.5: Asymptotic complexity and number of measured FLOPs required by the proposed LMMSE-EM, EKF-RTS-EM, and PN mitigation and data detection algorithms for the set of System Parameters 2:  $K = 1024$ ,  $L_r = 4$ ,  $N_{\text{tr}} = 2$ ,  $N_{\Delta f} = 129$ , and  $N_{\text{DMRS}} = 2$ .

LMMSE-EM		
Operation	Complexity	FLOPs
Initialize $\widehat{\Delta f}^{(m,0)}$	$1.056 \cdot 10^7$	$2.0091 \cdot 10^7$
$M \times$ Initialize $\hat{\mathbf{g}}_i^{(m,0)}$	$81920M$	$53296M$
$M \times$ <b>Update PN estimate</b>	$5.497 \cdot 10^{11} MN_{\text{iter}}$	$6.3602 \cdot 10^{11} MN_{\text{iter}}$
$M \times$ Update $\widehat{\Delta f}_{\text{ML}}^{(m,n)}$	$81920MN_{\text{iter}}$	$58573MN_{\text{iter}}$
$M \times$ Update $\hat{\mathbf{g}}_{i,\text{ML}}^{(m,n)}$	$81920MN_{\text{iter}}$	$53296MN_{\text{iter}}$
$M \times$ Update LF difference	$8192MN_{\text{iter}}$	$16396MN_{\text{iter}}$
Compute hybrid CRLB	$5.242 \cdot 10^6$	$18.350 \cdot 10^6$
Channel estimation	$1.342 \cdot 10^8 N_{\text{SIC}}$	$7.895 \cdot 10^8 N_{\text{SIC}}$
EKF-RTS-EM		
Operation	Complexity	FLOPs
Initialize $\widehat{\Delta f}^{(m,0)}$	$1.056 \cdot 10^7$	$2.0091 \cdot 10^7$
$M \times$ Initialize $\hat{\mathbf{g}}_i^{(m,0)}$	$81920M$	$53296M$
$M \times$ <b>Update PN estimate</b>	$131072MN_{\text{iter}}$	$305124MN_{\text{iter}}$
$M \times$ Update $\widehat{\Delta f}_{\text{ML}}^{(m,n)}$	$81920MN_{\text{iter}}$	$58573MN_{\text{iter}}$
$M \times$ Update $\hat{\mathbf{g}}_{i,\text{ML}}^{(m,n)}$	$81920MN_{\text{iter}}$	$53296MN_{\text{iter}}$
$M \times$ Update LF difference	$8192MN_{\text{iter}}$	$16396MN_{\text{iter}}$
Compute hybrid CRLB	$5.242 \cdot 10^6$	$18.350 \cdot 10^6$
Channel estimation	$1.342 \cdot 10^8 N_{\text{SIC}}$	$7.895 \cdot 10^8 N_{\text{SIC}}$
PN mitigation and data detection		
Operation	Complexity	FLOPs
Estimate $\hat{\mathbf{g}}_{i,\text{ML}}$	$81920$	$53296$
Initialize $\hat{\mathbf{s}}_{t,\text{MAP}}^{(0)}$	$40960N_{\text{data}}$	$26704N_{\text{data}}$
<b>Update PN estimate</b>	$65536N_{\text{iter}}N_{\text{data}}$	$152558N_{\text{iter}}N_{\text{data}}$
Update $\hat{\mathbf{s}}_{t,\text{MAP}}^{(n)}$	$40960N_{\text{data}}$	$26704N_{\text{data}}$

in [5] exhibits high computational complexity to estimate a single path in the mmWave MIMO channel. In simulation results, the number of EM iterations the proposed algorithms require to reach convergence is upper bounded by  $N_{\text{iter}} = 4$ , thereby indicating that the proposed synchronization strategies manage to obtain highly reliable estimates significantly faster than the algorithm in [5].

The proposed LMMSE-EM and EKF-RTS-EM algorithms exhibit higher computational complexity, which comes from the facts that: i) the proposed algorithms perform ML EM-based estimation of both the CFO and the frequency-selective beamformed channels, instead of using ad-hoc approaches as in [5], and ii) unlike the algorithm in [5], the proposed algorithms do perform statistical-linearization-based MMSE (LMMSE) estimation of the PN impairment. From a computational complexity standpoint, the first proposed LMMSE-EM algorithm differs from the second proposed EKF-RTS-EM algorithm in that the former performs LMMSE estimation of the PN impairment by processing the entire frame of  $N_{\text{tr}}$  OFDM training symbols, while the latter algorithm processes each time-domain received symbol in a sequential manner, thereby making the E-step significantly more efficient than in the former algorithm. More specifically, the E-step in the first proposed algorithm has complexity growing proportionally to  $\mathcal{O}((L_r K N_{\text{tr}})^3)$ , while the corresponding E-step in the second proposed algorithm exhibits linear complexity on the number of subcarriers,  $K$ , and the number of OFDM training symbols,  $N_{\text{tr}}$ , and exhibits cubic complexity in the number of RF chains  $L_r$ . Since the number of RF

chains is usually a small number at mmWave frequencies (i.e.  $L_r \leq 16$ ), the second proposed algorithm exhibits significantly lower complexity than the first proposed algorithm, and without significantly compromising estimation performance, as shown in the numerical results from the previous sections.

#### 4.7.5 Bit Error Rate and Modulation Error Ratio

In this last section, I present numerical results on the achievable communication performance of the proposed joint PN mitigation and data detection algorithm, both in terms of BER and MER. I consider the PDSCH in 5G NR with 16-QAM, 64-QAM, and 256-QAM modulations with LDPC channel coding using a target code rate of  $R_{\text{LDPC}} = 4/5$ . The number of transmitted data streams is set to  $N_s = 2$ . The MER is defined as

$$\text{MER} = -10 \log_{10} \left( \frac{1}{N_{\text{data}}} \sum_{t=1}^{N_{\text{data}}} \frac{1}{|\mathcal{P}_{t,\text{data}}|} \|\mathbf{s}_t\|_{\mathcal{P}_{t,\text{data}}} - [\hat{\mathbf{s}}_{t,\text{MAP}}]_{\mathcal{P}_{t,\text{data}}} \|_2^2 \right), \quad (4.99)$$

where  $\hat{\mathbf{s}}_{t,\text{MAP}}$ ,  $\mathbf{s}_t$  denote the MAP estimates of the detected QAM symbols and the original transmitted constellation symbols, respectively, and  $|\mathcal{P}_{t,\text{data}}|$  is the number of data subcarriers in an OFDM symbol. The average constellation energy is set to 1 (i.e. normalized constellation).

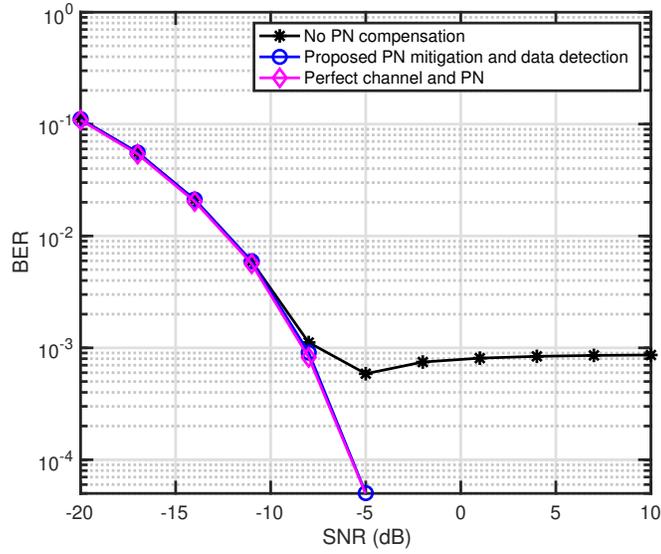
I show in Fig. 4.15, Fig. 4.16, and Fig. 4.17 the average BER (a) and MER (b) for 16-QAM, 64-QAM, and 256-QAM constellations, respectively.

The first observation from Figs. 4.15-4.17 is that the proposed PN mitigation and data detection algorithm exhibits a very small performance gap with respect to the case in which both the channel and the PN impairments are

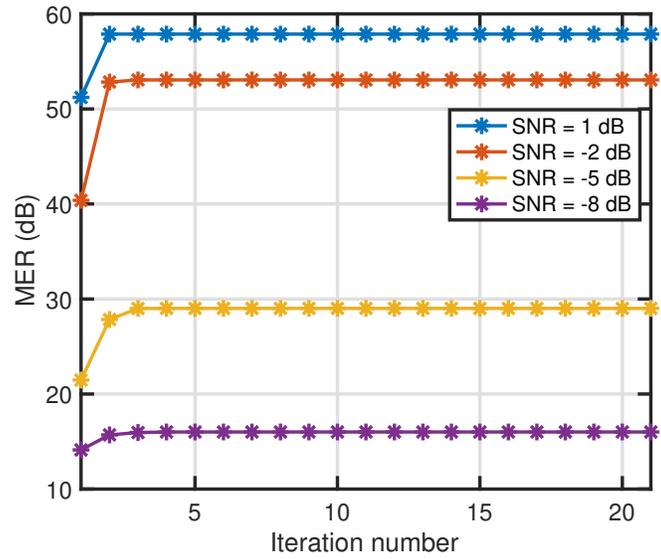
perfectly known, which is labeled as 'Perfect channel and PN'. Moreover, not compensating for the PN impairment results in similar BER as both the proposed algorithm and the perfect channel and PN case in the low SNR regime. In the mid-high SNR regimes, not compensating for the PN impairment results in a plateau effect in the BER performance, which is due to the PN having a significantly more degrading effect on the received signal than the AWGN impairment, which also makes it harder to improve the MER across iterations. For this reason, demodulation and equalization introduce CPE in the received OFDM data symbols, which makes it harder to estimate the LLRs and decode the receive data. This effect becomes less significant as the constellation order increases, which comes from the fact that, as the constellation size increases, noise amplification brought by equalization significantly degrades the quality of the received constellation. Hence, the effect of the PN impairment is significantly harder to remove. This is also the reason why, for 64-QAM and 256-QAM constellations, the BER performance of the proposed algorithm exhibits a larger performance gap with respect to the perfect channel and PN case, especially in the high SNR regime.

## 4.8 Conclusions

In this chapter, I proposed a framework for time-frequency synchronization that is robust under PN impairments and suitable for channel estimation in frequency-selective mmWave MIMO systems with hybrid architectures. I also developed a joint PN mitigation and data detection algorithm for

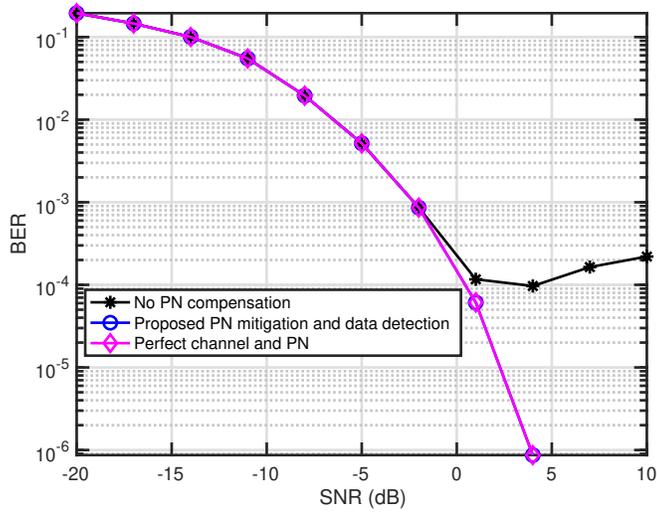


(a)

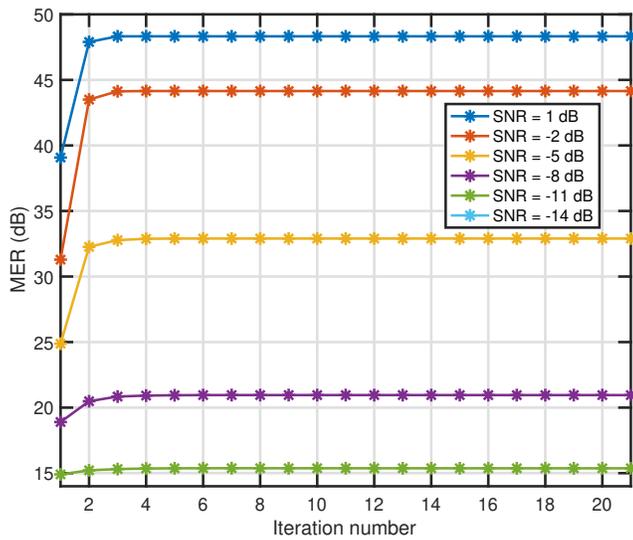


(b)

Figure 4.15: Evolution of the BER versus SNR (a) and MER versus number of iterations in the proposed PN mitigation and data detection algorithm (b), for  $\text{SNR} = \{-14, -11, -8, -5, -2\}$  dB before beamforming, using a 16-QAM constellation and  $N_s = 2$  data streams.

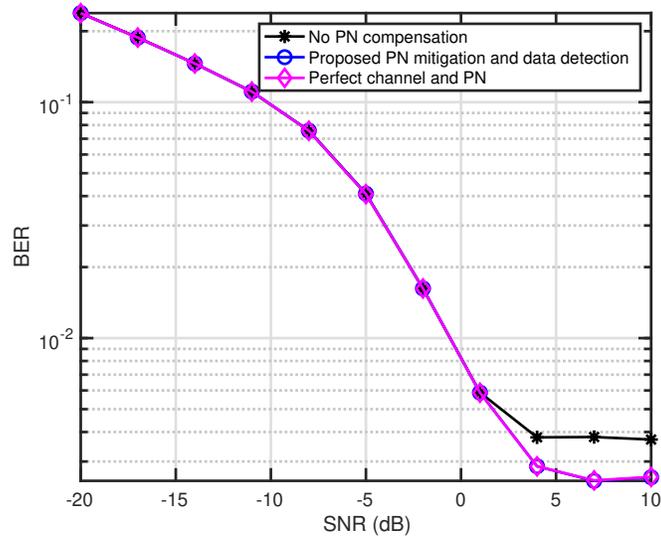


(a)

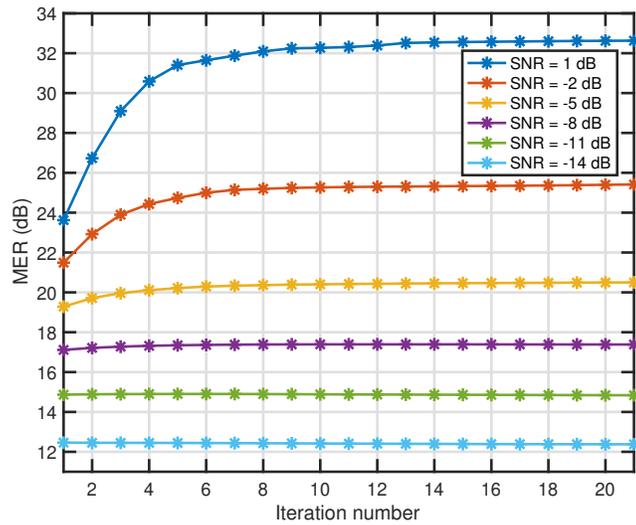


(b)

Figure 4.16: Evolution of the BER versus SNR (a) and MER versus number of iterations in the proposed PN mitigation and data detection algorithm (b), for  $\text{SNR} = \{-14, -11, -8, -5, -2, 1\}$  dB before beamforming, using a 64-QAM constellation and  $N_s = 2$  data streams.



(a)



(b)

Figure 4.17: Evolution of the BER versus SNR (a) and MER versus number of iterations in the proposed PN mitigation and data detection algorithm (b), for  $\text{SNR} = \{-14, -11, -8, -5, -2, 1\}$  dB before beamforming, using a 256-QAM constellation and  $N_s = 2$  data streams.

data transmission under the 5G NR frame structure for the PDSCH. I formulated the problem of jointly estimating the TO, CFO, PN impairments, and frequency-selective beamformed channels. Then, I theoretically analyzed the fundamental limits of the estimation problem using the theory of CRLB, and provided closed-form expressions for the hybrid CRLB when estimating the CFO, PN, and equivalent beamformed channels once time synchronization has already been performed. Then, I proposed two novel algorithms based on the EM approach and, using both the estimates obtained from these algorithms as well as their hybrid CRLB, I proposed a sparse reconstruction algorithm based on the solutions provided in Chapter 2 to retrieve the high-dimensional broadband mmWave MIMO channel. Finally, I proposed a joint PN mitigation and data detection algorithm using also the EM approach to detect the transmitted data under a variety of different constellation formats considered in 5G NR. Simulation results showed that synchronization can take place efficiently even in the low SNR regime, and near-optimum values of spectral efficiency can be obtained as well enabling spatial multiplexing. Further, the proposed algorithms were shown to significantly outperform prior work in terms of spectral efficiency, and without incurring higher overhead and keeping computational complexity low. Last, simulations of the joint proposed synchronization, compressive channel estimation, and data detection framework showed that unprecedented communication performance can be attained even when the mmWave link is configured in the low SNR regime, with communication performance very close to the perfect CSI case.

## Chapter 5

### Conclusions and Future Work

This chapter concludes the dissertation with a summary of contributions in Section 5.1 and potential future research directions in Section 5.2.

#### 5.1 Summary

Unlike sub-6 GHz wireless cellular communication systems such as the Long Term Evolution (LTE) family, in which link configuration can be performed in the medium/high SNR regimes using quasi-omnidirectional antenna arrays, enabling broadband mmWave MIMO systems in practice requires addressing the challenges of TO and CFO synchronization, and PN tracking and compensation prior to undergoing Channel State Information (CSI) acquisition that kicks off link configuration. A popular solution to solve this problem is joint beam training and synchronization, which enables synchronization in the high SNR regime once the SINR-maximizing beam pair is probed, and it has been incorporated in the 5G NR wireless cellular standard owing to its simplicity and robustness. Beam training techniques, however, are difficult to extend to multi-stream and multi-user settings, and both their complexity and overhead dramatically increases when more than a single communication path

is to be discovered.

In my dissertation, I develop efficient synchronization and channel estimation techniques by leveraging both the spatially common sparsity present in the mmWave MIMO channel and advanced estimation-theoretic techniques to obtain high quality CSI and enable data transmission without compromising the resulting overhead and computational complexity. Further, it is also important that these solutions satisfy the hardware constraints imposed by hybrid MIMO architectures, as well as being robust at low SNR. Robustness at low SNR is important since the path loss in the propagation environment increases quadratically with the carrier frequency (thereby significantly increasing when moving from sub-6 GHz to mmWave bands), the use of large bandwidth, and the high directionality of antenna arrays. I use different estimation, information-theoretic, and communication performance metrics (i.e. variance, Mean Squared Error (MSE), NMSE, CRLB, spectral efficiency, BER, MER, overhead, and complexity) throughout this dissertation to show that advanced hybrid analog-digital signal processing techniques can enable unprecedented communication performance while keeping training overhead low, even in the practical scenario of link configuration in the low SNR regime.

I first developed two CS-based techniques to estimate frequency-selective mmWave MIMO channels. The first algorithm I proposed leverages the spatially common sparsity present in mmWave MIMO channels, while fully exploiting the multiple received signals at different OFDM subcarriers. The second proposed algorithm, however, exploits a reduced number of subcar-

rier signals to estimate the sparse channel support, thereby reducing computational complexity with respect to the first algorithm, and incorporates a thresholding-based pruning procedure to discard residual multipath components having small energy that are not contained in the actual mmWave channel. I provided theoretical convergence guarantees for both algorithms, and showed that they offer excellent estimation and communication performance in comparison with prior solutions.

Thereafter, I focused on the joint CFO and narrowband channel estimation problem in mmWave systems using hybrid MIMO architectures in order to obtain further insight into the fundamental limits of both channel estimation and communication performance when the received signal is impaired by CFO synchronization uncertainties. I formulated the problem of estimating the unknown synchronization and baseband equivalent channel parameters, and theoretically obtained closed-form expressions regarding the CRLB for the estimation of these parameters. Then, I proposed a multi-stage solution to estimate these parameters as well as the high-dimensional mmWave MIMO channel, relying on both the estimators for the synchronization and baseband equivalent channel parameters and their CRLB. I used several numerical examples to show that the proposed strategies provide excellent estimation and communication performance, even for practical channels having a significant number of clusters with multiple rays per cluster. Last, I also showed that the proposed multi-stage solution exhibits a marginal increase in computational complexity with respect to the solutions proposed in the first part of my Ph.D.

research.

Last, I focused on the joint problem of synchronization and broadband channel estimation, encompassing the estimation of TO, CFO, PN, and the high-dimensional frequency-selective mmWave MIMO channel. I formulated the problem of estimating the different synchronization impairments as a prior step to MIMO channel estimation, and theoretically analyzed the hybrid CRLB for the estimation of the TO, CFO, PN, and frequency-selective complex baseband equivalent channels. Then, I extended the multi-stage solution proposed in the second part of my Ph.D. research to estimate the unknown parameters. More specifically, I proposed two synchronization algorithms to estimate the CFO, PN, and frequency-selective baseband equivalent channels. The first proposed synchronization algorithm processes the complete received signal to estimate the PN impairments using a batch-processing-based EKF, while the second algorithm sequentially processes the received time-domain samples using a dual forward-and-backward-filtering-based EKF and RTS smoothing. Then, the estimates of the different synchronization impairments and their hybrid CRLB are exploited to estimate the mmWave MIMO channel using the frequency-selective channel estimation algorithms developed in the first part of my Ph.D. research. I showed that the proposed multi-stage solution offers excellent estimation and communication performance, and that the frequency-selective mmWave MIMO channel can be accurately reconstructed despite lack of synchronization, even in the low SNR regime. A qualitative summary of the proposed broadband synchronization and channel estimation algorithms

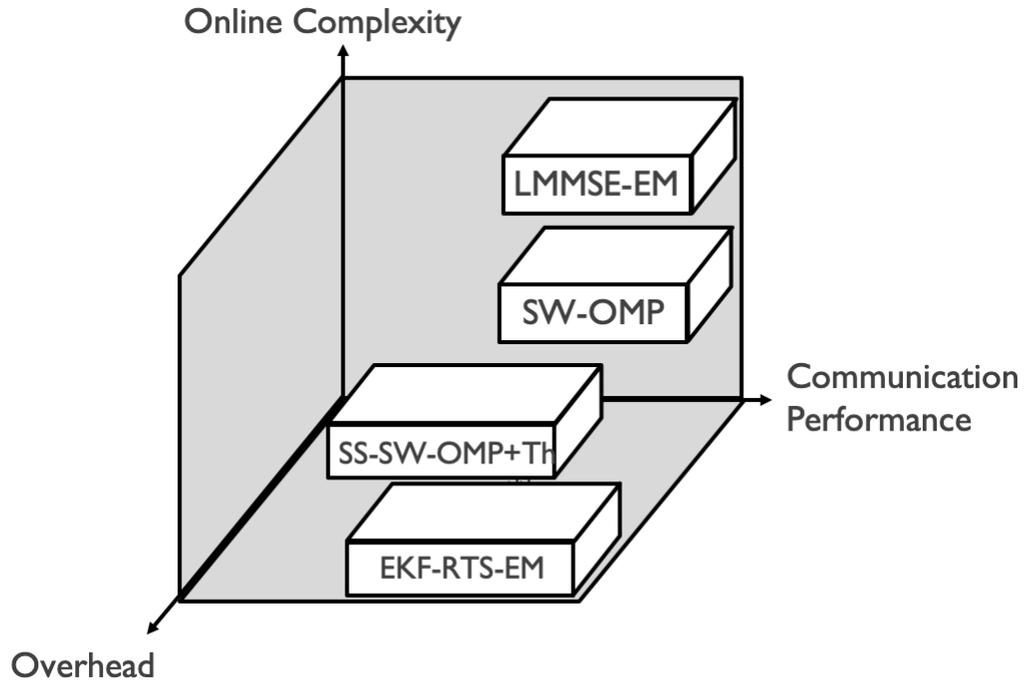


Figure 5.1: Illustration of qualitative performance of proposed broadband channel estimation and synchronization algorithms.

introduced in this dissertation is included in Fig. 5.1.

## 5.2 Future Work

In this dissertation, I have addressed some of the critical issues to enable hybrid-architecture-based link configuration of mmWave MIMO systems. There are still issues left to be addressed to successfully deploy these ideas into mmWave communication systems in the field. Hereafter, I present promising future research directions related to the topics in this dissertation.

- **Channel estimation without assuming knowledge on the antenna array geometry:** The broadband channel estimation algorithms presented in Chapter 2 achieved near-optimal spectral efficiency with reasonable computational complexity and low overhead. An important assumption in those algorithm is that the antenna array geometry is fully known, which is necessary to find the sparsifying basis of the mmWave MIMO channel. Then, the next question would be how to estimate the channel when only partial or no knowledge on the array geometry is known. Based on the linear model relating the received signal to the vectorized mmWave MIMO channel, dictionary learning techniques can be used to solve this problem. Dictionary learning is a branch of compressive sampling that aims to finding the optimum sparsifying basis in linear models that enable retrieval of sparse vectors lying on an unknown vector basis. As the dimensionality of the antenna array increases, it is expected that the channel estimation problem becomes more challenging, since the number of degrees of freedom in the unknown vector basis increases. Likewise, when partial antenna geometry knowledge is available, the problem is expected to be easier owing to the reduction in dimensionality of the unknown dictionary. Therefore, it is necessary to thoroughly investigate channel estimation techniques when knowledge of antenna array geometry cannot be assumed, as well as investigating the suitability of sparsifying basis such as the DFT basis or the one given by the Karhunen-Loeve Transform (KLT).

- **Channel estimation using prior information on the AoD/AoA:**

The channel estimation algorithms presented in Chapter 2 assume no prior information on the AoD/AoA is available beforehand. While this choice ensures robustness of the proposed algorithms when no information on the channel is available, the training overhead can be further reduced if statistical priors on the angular parameters are exploited, and consequently computational complexity can be reduced as well. Therefore, the questions here are how to obtain and exploit statistical priors on the small-scale parameters. The first question can be answered through the use of ray-tracing software to predict/analyze statistical deviations of the angular parameters for different snapshots, as well as prior information on the user's location using Global Positioning System (GPS) information. Once statistical priors are obtained, non-linear estimation techniques can be used to more accurately estimate the angular parameters and the high-dimensional MIMO channels themselves, such as Gaussian-Sum Filtering (GSF) for beam alignment and Marginalized Particle Filtering (MPF) for channel estimation.

- **Multi-user channel estimation in the presence of CFO uncertainties:**

The joint synchronization and channel estimation framework presented in Chapter 3 assumes a single-user scenario in which the received signal is impaired by CFO synchronization errors. While this scenario is interesting to shed light on the achievable performance of a hybrid mmWave MIMO system, mmWave transceivers are part of a

multi-user network, which results in the corresponding BS measuring a superposition of signals corresponding to each different user in the network. Furthermore, each user may be moving at a different speed, thereby resulting in a per-user CFO term that must be inferred and compensated before the channel can be estimated in practice. The questions here are how to obtain estimates of the multi-user channel and how to compensate for the potentially different CFO terms. An answer to the first question can be given by using either of the proposed broadband channel estimation algorithms from Chapter 2, SW-OMP or SS-SW-OMP+Th, which have already been shown to offer excellent performance in a multi-user setting under a perfect synchronization assumption [26, 154, 155]. The second question, however, is more challenging to answer. Assuming perfect TO synchronization and no PN impairments, the problem of estimating the multiple CFOs corresponding to the different users translates into a subspace alignment problem, similar to the interpretation in [141]. Using an ML formulation, and assuming  $U$  different active users communicating with the BS, the problem of estimating the  $U$  CFO terms consists of finding the  $U$  scalars that parameterize the vector subspace in which the received multi-user signal belongs, thus being a non-convex matched-filtering problem. In this context, the SNR is defined as the ratio between the energy of the projected signal onto the aforementioned subspace and the energy of the projected signal onto the corresponding orthogonal subspace. The most

straightforward yet complex approach to find such  $U$  scalars is through performing an exhaustive search over the  $U$ -dimensional hypercube defined by such scalars. Note that, similar to the EM-based algorithms developed in Chapter 4, an iterative quadratic optimization algorithm can be used to relax this non convex problem and thereby reduce computational complexity.

- **Joint PN mitigation and data detection using Hybrid Automatic Repeat Request (HARQ):** The joint PN mitigation and data detection algorithm developed in Chapter 4 assumes transmission of a single OFDM-modulated data stream to the receiver. In practice, however, one or more retransmissions are usually enabled, in which the estimated LLRs of the received bits are jointly processed using the concept of soft-combining. The question here is whether enabling HARQ along with soft-combining can further reduce the receive BER in comparison with single OFDM symbol transmission. For small constellation sizes (i.e. 4-QAM), the use of HARQ may not bring about significant performance gains. For especially dense constellations, such as 64-QAM or even denser constellations, the use of HARQ enables the availability of several observations at the receiver, which can be processed to improve the quality of the received LLRs, thereby enabling the use of such dense constellations for highly frequency-selective scenarios.
- **Joint synchronization, PN mitigation and data detection under non-linear power amplifiers:** The joint PN mitigation and data detec-

tion algorithm developed in Chapter 4 assumes a linear structure relating the time-domain mmWave MIMO channel and the training pilots/data symbols. In practical systems, however, hybrid MIMO transceivers are equipped with power amplifiers, which introduce non-linear distortions in both the amplitude and phase of the input signal. The severity of such non-linear distortions greatly depends on the dynamic range of the input signal. For OFDM-based signaling, the effect of non-linearities is more pronounced as either/both the number of subcarriers or the size of the constellation grows, so that properly accounting for such non-linearities is crucial to ensure that both synchronization is successful and the transmitted data can be successfully decoded. The questions here are how to efficiently synchronize the received signal and detect the transmitted data under the effect of these non-linearities. The first question can be answered using the EM algorithm, for instance. The E-Step in the EM algorithm can be used to find the approximate MMSE estimator of the PN impairment, while the M-Step would be used to estimate both the CFO and the low-dimensional equivalent beamformed channels, after (potentially) a Taylor-series-based quadratic approximation of the transmitted signal. The effect of the equivalent beamformed channels on the received signals is linear, which in turns enables the application of LS-like estimators to retrieve the beamformed channels. The second question is slightly more involved than the first. Mitigating the PN impairment while simultaneously estimating the transmitted constel-

lation symbols is not straightforward. However, the availability of prior statistical information on both the PN impairment and the transmitted symbols enables the application of Kalman Filter (KF)-like stochastic estimation techniques that might prove useful in solving this problem. Motivated by the typically small-amplitude of the PN impairment, a statistical linearization approximation might suffice to retrieve the MMSE estimate the PN impairment. Owing to the non-linearities introduced by the transmit power amplifiers, estimation of the transmitted data symbols, however, cannot be dealt with through simple linearization. Different stochastic estimation approaches can be used to solve this problem, such as the GSF with backward smoothing or the MPF with backward smoothing.

## Appendices

## Appendix A

### Derivation of the Cramér-Rao Lower Bound

Let us consider the block diagonal FIM  $\mathbf{I}(\boldsymbol{\xi}^{(m)})$  in (3.23). The minimum variance of any unbiased estimator of  $\alpha_i^{(m)}$  is given by the  $i$ -th diagonal elements of  $\mathbf{I}_1^{-1}(\boldsymbol{\xi}^{(m)})$ ,  $i = 1 \dots, L_r$  in (3.24) which is simply given by

$$\text{var} \left\{ \hat{\alpha}_i^{(m)} \right\} \geq \frac{\sigma^2}{2N}. \quad (\text{A.1})$$

Likewise, the corresponding minimum variance for the estimation of  $\sigma^2$  is given by the  $(L_r + 1)$ -th diagonal element in  $\mathbf{I}(\boldsymbol{\xi}^{(m)})$ , given by

$$\text{var} \left\{ \hat{\sigma}^2 \right\} \geq \frac{\sigma^4}{L_r N}. \quad (\text{A.2})$$

Now, the CRLB for the estimation of  $\Delta f^{(m)}$  and  $\beta_i^{(m)}$ ,  $1 \leq i \leq L_r$ , is given by the diagonal elements in  $\mathbf{I}_2^{-1}(\boldsymbol{\xi}^{(m)})$ . Let the  $\{i, j\}$ -th minor of  $\mathbf{I}_2(\boldsymbol{\xi}^{(m)})$  be defined as  $\mathbf{A}_{i,j} = [\mathbf{I}_2(\boldsymbol{\xi}^{(m)})]_{\mathcal{S}_i, \mathcal{S}_j}$ , with  $\mathcal{S}_i = \{1, \dots, i-1\} \cup \{i+1, \dots, L_r+1\}$  and  $\mathcal{S}_j = \{1, \dots, j-1\} \cup \{j+1, \dots, L_r+1\}$ . The element corresponding to  $\Delta f^{(m)}$  can be found as

$$\text{var} \left\{ \hat{\Delta f}^{(m)} \right\} \geq \frac{\det \{ \mathbf{A}_{1,1} \}}{\det \{ \mathbf{I}_2(\boldsymbol{\xi}^{(m)}) \}}, \quad (\text{A.3})$$

whereas the elements corresponding to  $\beta_i^{(m)}$  are given by

$$\text{var} \left\{ \hat{\beta}_i^{(m)} \right\} \geq \frac{\det \{ \mathbf{A}_{i+1, i+1} \}}{\det \{ \mathbf{I}_2(\boldsymbol{\xi}^{(m)}) \}}. \quad (\text{A.4})$$

Hereafter, I introduce  $S_1 = \sum_{n=0}^{N-1} 2\pi n$  and  $S_2 = \sum_{n=0}^{N-1} (2\pi n)^2$  to simplify the notation. The determinant of  $\mathbf{I}_2(\boldsymbol{\xi}^{(m)})$  can be found as

$$\det \{ \mathbf{I}_2(\boldsymbol{\xi}^{(m)}) \} = \left( \frac{2}{\sigma^2} \right)^{L_r+1} \left| \underbrace{\begin{bmatrix} \text{trace} \{ \mathbf{P}^{(m)} \} S_2 & S_1 \mathbf{1}_{L_r}^T \mathbf{P}^{(m)} \\ S_1 \mathbf{P}^{(m)} \mathbf{1}_{L_r} & N \mathbf{P}^{(m)} \end{bmatrix}}_{\mathbf{B}} \right|. \quad (\text{A.5})$$

The determinant of  $\mathbf{B}$  can be found by developing the determinant along its first column as

$$\begin{aligned} \det \{ \mathbf{B} \} &= \text{trace} \{ \mathbf{P}^{(m)} \} S_2 N^{L_r} \det \{ \mathbf{P}^{(m)} \} \\ &\quad - S_1^2 \alpha_1^{(m)2} N^{L_r-1} \det \{ \mathbf{P}^{(m)} \} - \\ &\quad - S_1^2 \alpha_2^{(m)2} N^{L_r-1} \det \{ \mathbf{P}^{(m)} \} - \dots \\ &\quad - S_1^2 \alpha_{L_r}^{(m)2} N^{L_r-1} \det \{ \mathbf{P}^{(m)} \} \\ &= \text{trace} \{ \mathbf{P}^{(m)} \} S_2 N^{L_r} \det \{ \mathbf{P}^{(m)} \} \\ &\quad - S_1^2 \sum_{i=1}^{L_r} \alpha_i^{(m)2} N^{L_r-1} \det \{ \mathbf{P}^{(m)} \}. \end{aligned} \quad (\text{A.6})$$

The numerator in (A.3) is found to be given by

$$\det \{ \mathbf{A}_{1,1} \} = N^{L_r} \det \{ \mathbf{P}^{(m)} \}. \quad (\text{A.7})$$

The values of  $S_1 = (2\pi)N(N+1)/2$  and  $S_2 = (2\pi)^2 N(N-1)(2N-1)/6$  can be plugged in (A.6) and thereafter substituted in (A.5) and (A.3), which yields the bound in (3.28).

As to the bound in (A.4), the only term left to compute is the deter-

minant of  $\mathbf{A}_{i+1,i+1}$ . Let us define  $\mathbf{P}_{-(i)}^{(m)} \in \mathbb{R}^{L_r-1 \times L_r-1}$  as

$$\mathbf{P}_{-(i)}^{(m)} = \text{diag}\{\mathbf{e}_1^T \mathbf{P}^{(m)} \mathbf{e}_1, \dots, \mathbf{e}_{i-1}^T \mathbf{P}^{(m)} \mathbf{e}_{i-1}, \mathbf{e}_{i+1}^T \mathbf{P}^{(m)} \mathbf{e}_{i+1}, \dots, \mathbf{e}_{L_r}^T \mathbf{P}^{(m)} \mathbf{e}_{L_r}\}. \quad (\text{A.8})$$

Then, the numerator in (A.4) is given by

$$\det \{\mathbf{A}_{i+1,i+1}\} = \left(\frac{2}{\sigma^2}\right)^{L_r} \left| \underbrace{\begin{bmatrix} \text{trace}\{\mathbf{P}^{(m)}\} S_2 & S_1 \mathbf{1}_{L_r-1}^T \mathbf{P}_{-(i)}^{(m)} \\ S_1 \mathbf{P}_{-(i)}^{(m)} \mathbf{1}_{L_r-1} & N \mathbf{P}_{-(i)}^{(m)} \end{bmatrix}}_{\mathbf{D}} \right|. \quad (\text{A.9})$$

The determinant of  $\mathbf{D}$  can be obtained by again developing the determinant of the minors along its first column as

$$\begin{aligned} \det\{\mathbf{D}\} &= \text{trace}\{\mathbf{P}^{(m)}\} S_2 N^{L_r-1} \frac{\det\{\mathbf{P}^{(m)}\}}{\alpha_i^{(m)2}} \\ &\quad - S_1^2 N^{L_r-2} \alpha_1^{(m)2} \frac{\det\{\mathbf{P}^{(m)}\}}{\alpha_i^{(m)2}} - \dots \\ &\quad \dots - S_1^2 N^{L_r-2} \alpha_{i-1}^{(m)2} \frac{\det\{\mathbf{P}^{(m)}\}}{\alpha_i^{(m)2}} \\ &\quad - S_1^2 N^{L_r-2} \alpha_{i+1}^{(m)2} \frac{\det\{\mathbf{P}^{(m)}\}}{\alpha_i^{(m)2}} - \\ &\quad - S_1^2 N^{L_r-2} \alpha_{L_r}^{(m)2} \frac{\det\{\mathbf{P}^{(m)}\}}{\alpha_i^{(m)2}} \\ &= \text{trace}\{\mathbf{P}^{(m)}\} S_2 N^{L_r-1} \frac{\det\{\mathbf{P}^{(m)}\}}{\alpha_i^{(m)2}} \\ &\quad - S_1^2 N^{L_r-2} \frac{\det\{\mathbf{P}^{(m)}\}}{\alpha_i^{(m)2}} \text{trace}\{\mathbf{P}_{-(i)}^{(m)}\}. \end{aligned} \quad (\text{A.10})$$

Finally, the values of  $S_1$  and  $S_2$  can be plugged in (A.10) and (A.9), such that (A.4) yields the bound in (3.29).

## Appendix B

### Fundamentals of Compressive Sensing

In this appendix, I provide background on compressed sensing (CS) theory, which is used to develop the solutions proposed in this dissertation.

Let us imagine a continuous-time signal which we wish to sample. According to the sampling theorem [156], the sampling frequency has to be at least twice the maximum frequency present in the signal (or twice the signal bandwidth, if the continuous-time signal is a bandpass signal) in order to guarantee lossless reconstruction. This is a deterministic focus, in the sense that if the signal is sampled at a rate that satisfies the requirements of the sampling theorem, then the probability of perfect reconstruction is 1, i.e., it always holds. This fact can be also analyzed from the linear algebra point of view.

Let  $\mathbf{y} \in \mathbb{C}^M$ ,  $\mathbf{A} \in \mathbb{C}^{M \times N}$ , and  $\mathbf{x} \in \mathbb{C}^N$ , with  $M \leq N$ , and consider the problem of finding the vector  $\mathbf{x}$  satisfying  $\mathbf{A}\mathbf{x} = \mathbf{y}$ . To recover  $\mathbf{x}$  from  $\mathbf{y}$ , then it is compulsory to store at least as many measurements in  $\mathbf{y}$  as the dimension of  $\mathbf{x}$ , i.e,  $N$ . This is a deterministic principle of digital processing systems, in which under  $M = N$ , and  $\mathbf{A}$  having linearly independent columns, the system is determined and the solution is unique. If  $M < N$ , however, linear algebra

tells us that, in principle, the linear system of equations  $\mathbf{y} = \mathbf{A}\mathbf{x}$  has no unique solution. Despite this, there are some cases in which the solution is unique, as long as two conditions are met: sparsity, which pertains to the vector of interest  $\mathbf{x}$ , and incoherence, which refers to the measurement process undergone through the matrix  $\mathbf{A}$ . CS theory asserts that certain vectors/signals can be recovered from a much smaller number of samples or measurements than traditional methods use, as long as these conditions are met [142].

In discrete-time systems, sparsity expresses the idea that a discrete signal depends on a number of degrees of freedom which is comparably much smaller than its length. More precisely, many natural signals are sparse or compressible in the sense that they have a concise representation when expressed in an appropriate basis  $\Psi$ . For instance, a sampled cosine signal is generally dense in the time domain. When expressed in the frequency domain, however, most of the signal content is concentrated around the discrete tone. Thus, this signal is compressible in the DFT basis. Incoherence extends the duality between time and frequency and conveys the intuition that signals having a sparse representation in  $\Psi$  must be spread out in the domain in which they are acquired. In our example, the DFT representation of a cosine is (approximately) a spike in the frequency domain, while it is spread out in the time domain.

Intuitively, if a vector  $\mathbf{x} \in \mathbb{C}^N$  can be expressed in a basis  $\Psi \in \mathbb{C}^{N \times L}$  as  $\mathbf{x} = \Psi\mathbf{z}$ , with  $\mathbf{z} \in \mathbb{C}^L$  being  $K$ -sparse (i.e., only  $K$  entries in  $\mathbf{z}$  are non-zero), then it is not necessary to store  $N$  samples in  $\mathbf{y}$  such that  $\mathbf{x}$  can be

retrieved through these measurements. Only a number  $M > K$  measurements are needed to reconstruct  $\mathbf{x}$  from  $\mathbf{y}$ . Then, under the assumption that  $\mathbf{x}$  is  $K$ -sparse, the vector  $\mathbf{x}$  can be reconstructed from  $\mathbf{y}$  when the number of measurements  $M$  obeys  $K < M \ll N$ , because the sparse vector has much lower information entropy than a non-sparse one. Consequently, the amount of samples needed to represent it is much lower. To undergo this reconstruction, the matrix  $\mathbf{A}$  must fulfill the Restricted Isometry Property (RIP), which can be stated as follows [142, 157]:

**Definition:** For each integer  $K = 1, 2, \dots$ , the isometry constant  $\delta_K$  of a matrix  $\mathbf{A}$  is the smallest number such that

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2 \quad (\text{B.1})$$

holds for all  $K$ -sparse vectors  $\mathbf{x}$ .

A matrix  $\mathbf{A}$  is loosely said to fulfill the RIP of order  $K$  if  $\delta_K$  is not too close to one. Intuitively, if  $\delta_K$  is small, and  $\mathbf{x}$  is any  $K$ -sparse vector, then (B.1) establishes that any set of  $K$  columns of  $\mathbf{A}$  approximately behaves like an orthonormal system. Therefore, when  $\mathbf{A}$  fulfills this property, then  $\mathbf{A}$  approximately preserves the Euclidean length of  $S$ -sparse signals, which implies that  $S$ -sparse vectors cannot be in the null space of  $\mathbf{A}$  (which is useful, as otherwise these vectors could not be reconstructed). The connection between the RIP and CS can be made as follows. Let us imagine that we wish to acquire any  $K$ -sparse vector using  $\mathbf{A}$ , and assume that  $\delta_{2K}$  is sufficiently smaller than

one. The RIP then translates into

$$(1 - \delta_{2K})\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \leq \|\mathbf{A}\mathbf{x}_1 - \mathbf{A}\mathbf{x}_2\|_2^2 \leq (1 + \delta_{2K})\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2, \quad (\text{B.2})$$

in which  $\mathbf{x}_1, \mathbf{x}_2$  are  $K$ -sparse vectors, and hence  $\mathbf{x}_1 - \mathbf{x}_2$  is at most  $2K$ -sparse. The result in (B.2) implies that all pairwise distances between any two  $K$ -sparse signals must be well preserved in the measurement space. The result in (B.2) guarantees the existence of efficient and robust algorithms for discriminating  $K$ -sparse signals based on their compressive measurements [142]. Examples of matrices that fulfill (B.2) with high probability are matrices formed by independent and identically distributed samples taken from a normal, a symmetric Bernoulli, or a sub-Gaussian distribution [142]. With overwhelming probability, all these matrices obey the RIP providing that [142]

$$M \geq CK \log(N/K), \quad (\text{B.3})$$

where  $C$  is some constant. Therefore, using these matrices makes it possible to retrieve  $K$ -sparse vectors using a number of measurements  $M \ll N$ . The main difficulty of applying CS theory into real-world problems stems from the fact that natural signals are not exactly  $K$ -sparse, but approximately  $K$ -sparse, in which  $K$  is generally unknown. This means that most of the content of these signals is concentrated on  $K$  components, but no knowledge on the value of  $K$  is available a priori.

Let us imagine that we wish to recover a  $K$ -sparse signal  $\mathbf{x}$  from  $\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}$ , where  $\mathbf{w} \in \mathbb{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ . Likewise, let us assume that  $\mathbf{x}$  can be

(approximately) sparsely represented in the vector basis  $\Psi$  as  $\mathbf{x} = \Psi\mathbf{z}$ . The conditional density of the measurement is given by

$$p(\mathbf{y}|\mathbf{z}) = \underbrace{n_c(\Phi\Psi\mathbf{z}, \sigma^2\mathbf{I})}_{\text{Complex Gaussian density}}. \quad (\text{B.4})$$

Intuitively, to retrieve a  $K$ -sparse vector  $\mathbf{z}$ , it would be desirable to use the  $\ell_0$ -norm of  $\mathbf{z}$  to find the sparsest solution, which is unique if the RIP holds. Unfortunately, the  $\ell_0$ -norm is not mathematically tractable to solve an optimization problem. In the Bayesian framework, a prior Laplacian PDF on  $\mathbf{z}$  is usually chosen since it has high kurtosis

$$p(\mathbf{z}) = \frac{\lambda}{2} e^{-\lambda\|\mathbf{z}\|_1}, \quad (\text{B.5})$$

which depends on  $\lambda$ , a parameter that controls the spikeness of  $p(\mathbf{z})$ . The larger the value of  $\lambda$ , the more concentrated the probability density of  $\mathbf{z}$  around zero. Using this prior PDF, the MAP estimator of  $\mathbf{z}$ ,  $\hat{\mathbf{z}}_{\text{MAP}}$ , can be found as

$$\begin{aligned} \hat{\mathbf{z}} &= \arg \max_{\mathbf{z}} p(\mathbf{y}|\mathbf{z})p(\mathbf{z}) \\ &= \arg \min_{\mathbf{z}} \|\mathbf{y} - \Phi\Psi\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_1. \end{aligned} \quad (\text{B.6})$$

The optimization problem in (B.6) is called Basis Pursuit De-Noising (BPDN) [158], and it can be interpreted as follows. The first term in (B.6) is the usual LS error term characterizing estimation problems with conditionally Gaussian densities, while the second term is a penalty factor that depends on both  $\lambda$  and the sparsity on  $\mathbf{z}$ . The larger the value of  $\lambda$ , the higher the penalty that the second term in (B.6) introduces for having a non-zero element in  $\mathbf{z}$ .

Intuitively, this means that higher values of  $\lambda$  will result in sparser estimates  $\mathbf{z}$ . The problem in (B.6) is often expressed as [142]

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} \|\mathbf{z}\|_1, \quad \text{subject to} \quad \|\mathbf{y} - \Phi\Psi\mathbf{z}\|_2^2 \leq \epsilon, \quad (\text{B.7})$$

where  $\epsilon$  bounds the amount of noise in the measurement  $\mathbf{y}$ . The problem in (B.7) is often called Least Absolute Shrinkage and Selection Operator (LASSO) [159–161]. It is a Second-Order Cone Program (SOCP) (hence convex) which can be solved efficiently. The problem in (B.7) is different from the one in (B.6) in that (B.7) allows setting a maximum allowable LS error in  $\mathbf{z}$  to control how well the sparse estimate  $\mathbf{z}$  fits the data, and the sparsity in  $\mathbf{z}$  is optimized within an  $\epsilon$ -error interval. In general, the formulation in (B.7) is preferred when prior information on the measurement noise is available, since it allows finding sparse estimates of  $\mathbf{z}$  while controlling how well  $\mathbf{z}$  explains the measurement  $\mathbf{y}$ . There is a variety of algorithms to solve the optimization problem in (B.7), some of which include the Compressive Sampling Matching Pursuit (CoSaMP) [162], and the OMP [163] algorithms. Notice that the problems in (B.6)-(B.7) can be seen as static sparse estimation problems. Extensions to scenarios with time-varying dynamic sparse signals have also been investigated [158], and algorithms to efficiently retrieve dynamic sparse signals have also been proposed, some of which exhibit convergence guarantees.

## Appendix C

### Fundamentals of Estimation Theory

In this appendix, I provide background on estimation theory, which is used to develop the solutions proposed in this dissertation.

Let us consider two vectors  $\boldsymbol{\xi} \in \mathbb{C}^P$  and  $\mathbf{y} \in \mathbb{C}^N$ , and let us assume that they are related as  $\mathbf{y} = \mathbf{h}(\boldsymbol{\xi}) + \mathbf{w}$ , with  $\mathbf{h} : \mathbb{C}^P \rightarrow \mathbb{C}^N$  being, in general, a non-linear function, and  $\mathbf{w} \in \mathbb{C}^N$  being additive noise with a certain PDF  $p_{\mathbf{w}}(\mathbf{w})$ . Estimation theory deals with the problem of how to optimally infer  $\boldsymbol{\xi}$  from  $\mathbf{y}$ , i.e. how to find optimal functions  $\mathbf{g}(\mathbf{y})$  to approximate  $\boldsymbol{\xi}$ . This process makes sense only if  $\mathbf{y}$  and  $\boldsymbol{\xi}$  depend on each other or, equivalently, if  $\mathbf{y}$  contains some information on  $\boldsymbol{\xi}$ . There are two main flavors of estimation theory, which are usually called classical estimation and Bayesian estimation. The main difference between them is that the former deals with optimal estimation of deterministic parameters/vectors from noisy observations, while in the latter the vectors to be estimated are random, hence belonging to a certain distribution with prior PDF  $p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$ .

There are two main questions that can be formulated from an estimation problem. The first one is related to how well the vector  $\boldsymbol{\xi}$  can be estimated from  $\mathbf{y}$ , and the second one relates to how to find functions  $\mathbf{g}(\mathbf{y})$  such that

$\hat{\boldsymbol{\xi}} = \mathbf{g}(\mathbf{y})$  is as close as possible to  $\boldsymbol{\xi}$ . The most common metric to assess the quality of an estimator is the MSE of  $\mathbf{e} = \boldsymbol{\xi} - \mathbf{g}(\mathbf{y})$ , which is denoted as  $\mathbf{m}_{\mathbf{e}\mathbf{e}} = \mathbb{E}_{\mathbf{e}}\{\mathbf{e}\mathbf{e}^*\}$ . In the particular event that  $\mathbb{E}_{\mathbf{e}}\{\mathbf{g}(\mathbf{y})\} = \mathbb{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}\} = \mathbf{m}_{\boldsymbol{\xi}}$ , the estimator  $\hat{\boldsymbol{\xi}} = \mathbf{g}(\mathbf{y})$  is said to be unbiased, and the MSE boils down to the covariance matrix of the estimation error  $\mathbf{e}$ . It turns out that in many scenarios there is a lower bound for the MSE any unbiased estimator can attain, which is called the CRLB.

## C.1 Fisher Information and Cramér-Rao Lower Bound

The CRLB is a lower bound on the MSE any unbiased estimator can attain, hence a lower bound on the covariance matrix of any unbiased estimator. Let us consider the measurement model introduced earlier

$$\mathbf{y} = \mathbf{h}(\boldsymbol{\xi}) + \mathbf{w}, \quad (\text{C.1})$$

where  $\mathbf{w}$  is fully described by its PDF  $p_{\mathbf{w}}(\mathbf{w})$ . Let us first consider that  $\boldsymbol{\xi}$  is a deterministic parameter. The Fisher's score function is defined as the partial derivative of the LLF of  $\mathbf{y}$ ,  $\partial \log p(\mathbf{y}; \boldsymbol{\xi}) / \partial \boldsymbol{\xi}$ , which plays a central role in both the existence and derivation of the CRLB. A condition for the CRLB to exist is that the Fisher's score function is zero mean

$$\mathbb{E} \left\{ \frac{\partial \log p_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right\} = \mathbf{0}, \quad (\text{C.2})$$

which generally holds unless the support of  $p_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\xi})$  depends on  $\boldsymbol{\xi}$ . The condition in (C.2) is called the regularity condition [135]. The amount of information

that  $\mathbf{y}$  provides about  $\boldsymbol{\xi}$  is collected in the FIM, usually denoted by  $\mathbf{I}(\boldsymbol{\xi})$ , which can be computed as the covariance matrix of Fisher's score [135]

$$\mathbf{I}(\boldsymbol{\xi}) = -\mathbb{E} \left\{ \frac{\partial \log p(\mathbf{y}; \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \left( \frac{\partial \log p(\mathbf{y}; \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right)^* \right\}. \quad (\text{C.3})$$

The result in (C.3) is general and independent of the distribution of  $\mathbf{w}$ . In the event that  $\mathbf{y}$  follows a Gaussian distribution,  $\mathbf{y} \sim \mathcal{CN}(\mathbf{m}_y(\boldsymbol{\xi}), \mathbf{C}_{yy}(\boldsymbol{\xi}))$ , then the FIM in (C.3) can be computed using the Slepian-Bangs formula [135]

$$[\mathbf{I}(\boldsymbol{\xi})]_{i,j} = 2 \operatorname{Re} \left\{ \frac{\partial \mathbf{m}_y^*(\boldsymbol{\xi})}{\partial \xi_i} \mathbf{C}_{yy}^{-1}(\boldsymbol{\xi}) \frac{\partial \mathbf{m}_y(\boldsymbol{\xi})}{\partial \xi_j} \right\} + \operatorname{trace} \left\{ \mathbf{C}_{yy}^{-1}(\boldsymbol{\xi}) \frac{\partial \mathbf{C}_{yy}(\boldsymbol{\xi})}{\partial \xi_i} \mathbf{C}_{yy}^{-1}(\boldsymbol{\xi}) \frac{\partial \mathbf{C}_{yy}(\boldsymbol{\xi})}{\partial \xi_j} \right\}. \quad (\text{C.4})$$

Let  $\hat{\boldsymbol{\xi}} = \mathbf{g}(\mathbf{y})$  be any estimator of  $\boldsymbol{\xi}$ . Then, the CRLB inequality states that the covariance matrix of the estimation error  $\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}$  is lower bounded as [135]

$$\underbrace{\mathbb{E}_{\mathbf{y}} \{ (\boldsymbol{\xi} - \mathbf{g}(\mathbf{y})) (\boldsymbol{\xi} - \mathbf{g}(\mathbf{y}))^* \}}_{\mathbf{C}_{\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}}} \geq \mathbf{I}^{-1}(\boldsymbol{\xi}), \quad (\text{C.5})$$

where  $\mathbf{A} \geq \mathbf{B}$  indicates that  $\mathbf{A} - \mathbf{B}$  is a positive semidefinite matrix. Furthermore, an unbiased estimator may be found that attains the bound in that  $\mathbf{C}_{\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}} = \mathbf{I}^{-1}(\boldsymbol{\xi})$  if and only if

$$\frac{\partial \log p(\mathbf{y}; \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \mathbf{I}(\boldsymbol{\xi}) (\mathbf{g}(\mathbf{y}) - \boldsymbol{\xi}), \quad (\text{C.6})$$

for some function  $\mathbf{g}(\mathbf{y})$ . Such estimator is  $\hat{\boldsymbol{\xi}} = \mathbf{g}(\mathbf{y})$ , and its covariance matrix is  $\mathbf{I}^{-1}(\boldsymbol{\xi})$ . For a more detailed discussion about the CRLB, the reader is referred to [135].

Now, let us turn our attention into the scenario in which some prior information on  $\boldsymbol{\xi}$  is available. In statistical estimation, prior information means

that a prior PDF  $p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$  is available. The extension of the FIM for the estimation of random parameters is called the HIM, which is usually denoted by  $\mathbf{H}(\boldsymbol{\xi})$ . This matrix is defined as [148]

$$\mathbf{H}(\boldsymbol{\xi}) \triangleq \mathbf{I}_D(\boldsymbol{\xi}) + \mathbf{I}_P(\boldsymbol{\xi}), \quad (\text{C.7})$$

where  $\mathbf{I}_D(\boldsymbol{\xi})$  is the expected value of the FIM  $\mathbf{I}(\boldsymbol{\xi})$  in (C.3) with respect to  $\boldsymbol{\xi}$

$$\mathbf{I}_D(\boldsymbol{\xi}) \triangleq \mathbb{E}_{\boldsymbol{\xi}} \{ \mathbf{I}(\boldsymbol{\xi}) \}, \quad (\text{C.8})$$

and  $\mathbf{I}_P(\boldsymbol{\xi})$  is the prior information matrix given by

$$\mathbf{I}_P(\boldsymbol{\xi}) \triangleq -\mathbb{E}_{\boldsymbol{\xi}^{(m)}} \left\{ \frac{\partial \log p(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \left( \frac{\partial \log p(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right)^* \right\}. \quad (\text{C.9})$$

Then, for any unbiased estimator  $\hat{\boldsymbol{\xi}} = \mathbf{g}(\mathbf{y})$ , the hybrid or Bayesian CRLB inequality states that the covariance matrix of the estimation error  $\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}$  is lower bounded as [148]

$$\underbrace{\mathbb{E}_{\boldsymbol{\xi}, \mathbf{y}} \{ (\boldsymbol{\xi} - \mathbf{g}(\mathbf{y})) (\boldsymbol{\xi} - \mathbf{g}(\mathbf{y}))^* \}}_{\mathbf{C}_{\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}}} \geq \mathbf{I}^{-1}(\boldsymbol{\xi}), \quad (\text{C.10})$$

which is recognized as an extension of (C.5). For a more detailed discussion about the hybrid CRLB and other types of Bayesian bounds, the reader is referred to [148].

## C.2 Finding Optimal Estimators

In the previous section, an introduction to the theory of CRLB was provided, both for estimation of deterministic and random parameters. Let

us now turn our attention into the problem of finding unbiased estimators  $\hat{\boldsymbol{\xi}} = \mathbf{g}(\mathbf{y})$  whose performance index, i.e. covariance matrix, is as close as possible to the CRLB. In the context of classical estimation of deterministic parameters, the most common estimators are the MVUE, the BLUE, and the ML estimator. The MVUE  $\hat{\boldsymbol{\xi}}_{\text{MVUE}}$  is known to be an efficient estimator, meaning that it is unbiased and its covariance attains the CRLB with equality. Usually, it is found through (C.6) when the Fisher's score can be appropriately factorized. The BLUE estimator is a linear estimator that has the property of exhibiting the minimum variance among all the unbiased estimators that are linear on the data. In general, the BLUE is not an optimal estimator, but it is known to be optimal when the measurement model falls under the GLM [135]

$$\mathbf{y} = \mathbf{H}\boldsymbol{\xi} + \mathbf{w}, \quad (\text{C.11})$$

with  $\mathbf{w}$  having covariance matrix  $\mathbf{C}_{\mathbf{w}\mathbf{w}}$ . In this case, the BLUE is given by  $\hat{\boldsymbol{\xi}}_{\text{BLUE}} = (\mathbf{H}^* \mathbf{C}_{\mathbf{w}\mathbf{w}}^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{C}_{\mathbf{w}\mathbf{w}}^{-1} \mathbf{y}$ , as long as  $\mathbf{H}$  is full column rank. The ML estimator is the maximizer of the likelihood function  $p(\mathbf{y}; \boldsymbol{\xi})$  parameterized by  $\boldsymbol{\xi}$ . This estimator is not optimal in general, but under certain conditions on the PDF [135], the ML estimator is asymptotically efficient for large data records or as  $N \rightarrow \infty$ , where we recall that  $N$  is the dimension of  $\mathbf{y}$ . Hence, asymptotically it is the MVUE. The performance of the ML estimator for finite number of samples  $N$  depends on the PDF. Asymptotically, however, the ML estimator is distributed as

$$\hat{\boldsymbol{\xi}}_{\text{ML}} \stackrel{a}{\sim} \mathcal{CN}(\boldsymbol{\xi}, \mathbf{I}^{-1}(\boldsymbol{\xi})). \quad (\text{C.12})$$

In the particular case that the measurement model is linear and driven by Gaussian noise, then the MVUE, BLUE, and ML estimators coincide.

Now, let us focus on the Bayesian scenario, in which an unbiased estimator  $\hat{\boldsymbol{\xi}} = \mathbf{g}(\mathbf{y})$  is to be found, and prior knowledge on the distribution of  $\boldsymbol{\xi}$ , denoted by  $p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$ , is known. Usually, a Bayesian estimator is found when optimizing a performance index depending on the posterior PDF of the random vector we wish to estimate. In this context, the most common estimators are the MAP and the MMSE estimators. Since our main interest is on finding unbiased estimators of  $\boldsymbol{\xi}$ , only the MMSE estimator will be discussed here. Let  $\mathcal{G}$  denote a set of functions closed under addition (i.e.  $\forall \mathbf{g}_1(\mathbf{y}), \mathbf{g}_2(\mathbf{y}) \in \mathcal{G}, \mathbf{g}_1(\mathbf{y}) + \mathbf{g}_2(\mathbf{y}) \in \mathcal{G}$ ) and scalar multiplication (i.e.  $\forall \alpha \in \mathbb{C}, \mathbf{g}(\mathbf{y}) \in \mathcal{G}, \alpha \mathbf{g}(\mathbf{y}) \in \mathcal{G}$ ). The MMSE estimator  $\hat{\boldsymbol{\xi}}_{\text{MMSE}}$  is the function  $\mathbf{g}^*(\mathbf{y}) \in \mathcal{G}$  that satisfies

$$\hat{\boldsymbol{\xi}}_{\text{MMSE}} \triangleq \arg \min_{\mathbf{g}(\mathbf{y}) \in \mathcal{G}} \mathbb{E}_{\boldsymbol{\xi}, \mathbf{y}} \{(\boldsymbol{\xi} - \mathbf{g}(\mathbf{y}))^* (\boldsymbol{\xi} - \mathbf{g}(\mathbf{y}))\}. \quad (\text{C.13})$$

By virtue of the orthogonality principle, the statement in (C.13) is equivalent to

$$\mathbb{E}_{\boldsymbol{\xi}, \mathbf{y}} \{(\boldsymbol{\xi} - \mathbf{g}^*(\mathbf{y})) \mathbf{g}(\mathbf{y})\} = \mathbf{0}, \quad \forall \mathbf{g}(\mathbf{y}) \in \mathcal{G}, \quad (\text{C.14})$$

which must hold for any  $\mathbf{g}(\mathbf{y}) \in \mathcal{G}$ . Therefore, it must hold for  $\mathbf{g}(\mathbf{y}) = \mathbf{e}_i$ , for instance, where  $\mathbf{e}_i$  is the  $i$ -th vector forming the canonical basis. This leads to

$$\mathbf{g}^*(\mathbf{y}) = \mathbb{E}_{\boldsymbol{\xi} | \mathbf{y}} \{\boldsymbol{\xi}\}, \quad (\text{C.15})$$

whereby the optimal MMSE estimator is given by the conditional mean. The main difficulty about finding this estimator is that it is distribution-dependent,

and it requires, in general, knowledge about the PDF of every random vector involved in the estimation problem. For this reason, a linearization approach is usually followed in order to find the best possible estimator within the class of affine functions. This leads to the LMMSE estimator, which is known to be the statistical linearization approximation of the MMSE. For the general non-linear model in (C.1), the LMMSE estimator reads as

$$\hat{\boldsymbol{\xi}}_{\text{LMMSE}} = \mathbb{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}\} + \mathbf{C}_{\boldsymbol{\xi}\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}(\mathbf{y} - \mathbf{h}(\mathbb{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}\})). \quad (\text{C.16})$$

The orthogonality principle in (C.14) can be used to derive the LMMSE estimator, but it can also be applied to classes of non-linear functions, such as polynomial functions. For instance, the Quadratic Minimum Mean Square Error (QMMSE) would be the quadratic function of  $\mathbf{y}$  that minimizes the MSE among any possible quadratic function, and it is known to be the statistical quadratic approximation of the MMSE estimator in this case. A more detailed treatment of Bayesian estimators can be found in [135], [164]. When applied to dynamic scenarios in which random vectors vary with time, hence being random processes, Bayesian estimators are referred to as Bayesian filters and Bayesian smoothers. Bayesian filters are time-varying estimators that filter the measurement data forward in time, while Bayesian smoothers perform both forward and backward filtering. A detailed treatment of both Bayesian filters and smoothers can be found in [164] and [149].

## Bibliography

- [1] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, “Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [2] J. Rodríguez-Fernández and N. González-Prelcic, “Channel estimation for hybrid mmwave MIMO systems with CFO uncertainties,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4636–4652, Oct 2019.
- [3] J. Rodriguez-Fernandez and N. Gonzalez-Prelcic, “Joint synchronization, phase noise and compressive channel estimation in hybrid frequency-selective mmWave MIMO systems,” *available at arXiv*, 2019.
- [4] S. Sun, G. R. MacCartney, and T. S. Rappaport, “A novel millimeter-wave channel simulator and applications for 5g wireless communications,” in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.
- [5] H. Yan and D. Cabric, “Compressive initial access and beamforming training for millimeter-wave cellular systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 1151–1166, Sep. 2019.

- [6] 3GPP, “Study on channel model for frequencies from 0.5 to 100 GHz (release 14.3.0),” *Technical Report*, Dec 2017.
- [7] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, June 2011.
- [8] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, April 2016.
- [9] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.
- [10] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5G be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [11] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.
- [12] X. Zhang, A. F. Molisch, and S.-Y. Kung, “Variable-phase-shift-based rf-baseband codesign for MIMO antenna selection,” *IEEE Trans. Signal*

*Process.*, vol. 53, no. 11, pp. 4091–4103, Nov 2005.

- [13] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, “MIMO precoding and combining solutions for millimeter-wave systems,” *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, December 2014.
- [14] R. M. Rial, C. Rusu, A. Alkhateeb, N. G. Prelcic, and R. W. Heath Jr., “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?” *IEEE Access*, no. 99, Jan 2016.
- [15] A. Alkhateeb, G. Leus, and R. W. Heath, “Limited feedback hybrid precoding for multi-user millimeter wave systems,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, Nov 2015.
- [16] R. Méndez-Rial, C. Rusu, N. González-Prelcic, and R. W. Heath, “Dictionary-free hybrid precoders and combiners for mmwave MIMO systems,” in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 151–155.
- [17] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct 2014.
- [18] R. López-Valcarce, N. González-Prelcic, C. Rusu, and R. W. Heath, “Hybrid precoders and combiners for mmwave MIMO systems with per-

- antenna power constraints,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.
- [19] P. L. Cao, T. J. Oechtering, and M. Skoglund, “Precoding design for massive MIMO systems with sub-connected architecture and per-antenna power constraints,” in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, March 2018, pp. 1–6.
- [20] R. Peng and Y. Tian, “Wideband hybrid precoder design in MU-MIMO based on channel angular information,” in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–5.
- [21] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, “Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, April 2016.
- [22] S. Park, A. Alkhateeb, and R. W. Heath, “Dynamic subarrays for hybrid precoding in wideband mmwave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2907–2920, May 2017.
- [23] J. P. González-Coma, N. González-Prelcic, L. Castedo, and R. W. Heath, “Frequency selective multiuser hybrid precoding for mmwave systems with imperfect channel knowledge,” in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 291–295.

- [24] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. W. Heath, “Channel estimation and hybrid precoding for frequency selective multiuser mmwave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 353–367, May 2018.
- [25] X. Yu, J. Zhang, and K. B. Letaief, “A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 282–297, May 2018.
- [26] J. Rodriguez-Fernandez and N. Gonzalez-Prelcic, “Low-complexity multiuser hybrid precoding and combining for frequency selective millimeter wave systems,” *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (accepted)*, 2018.
- [27] J. Rodriguez-Fernandez, R. Lopez-Valcarce, and N. Gonzalez-Prelcic, “Frequency-selective hybrid precoding and combining for mmwave MIMO systems with per-antenna power constraints,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4794–4798.
- [28] X. Yu, J. Zhang, and K. B. Letaief, “A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems,” *IEEE Jour-*

- nal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 282–297, May 2018.
- [29] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, 2018.
- [30] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. USA: Cambridge University Press, 2005.
- [31] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, “QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials,” *IEEE Trans. on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, June 2014.
- [32] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, “QuaDRiGa: QuaDRiGa - Quasi Deterministic Ratio Channel Generator, User Manual and Documentation,” *Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2.0.0*, 2017.
- [33] P. Schniter and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: The sparse way,” in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov 2014, pp. 273–277.
- [34] K. Venugopal, A. Alkhateeb, N. González Prelicic, and R. W. Heath, “Channel estimation for hybrid architecture-based wideband millimeter wave systems,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.

- [35] J. Ko, Y. Cho, S. Hur, T. Kim, J. Park, A. F. Molisch, K. Haneda, M. Peter, D. Park, and D. Cho, "Millimeter-wave channel measurements and analysis for statistical spatial channel model in in-building and urban environments at 28 GHz," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5853–5868, Sep. 2017.
- [36] E. Perahia *et al*, "IEEE 802.11ad: Defining the next generation multi-Gbps Wi-Fi," *IEEE Consumer Commun. and Networking Conf.*, pp. 1–5, Jan. 2010.
- [37] 3GPP, "Physical channels and modulation (release 15)," *Tech. Rep. v15.1.0*, 2017.
- [38] J. Wang, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, October 2009.
- [39] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *2011 6th International ICST Conference on Communications and Networking in China (CHINACOM)*, Aug 2011, pp. 361–365.
- [40] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, October 2013.

- [41] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, T. Korakis, S. S. Panwar, and T. S. Rappaport, “Directional cell discovery in millimeter wave cellular networks,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6664–6678, Dec 2015.
- [42] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, “A tutorial on beam management for 3GPP NR at mmwave frequencies,” *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 173–196, Firstquarter 2019.
- [43] Y. Li, J. Luo, M. H. Castañeda Garcia, R. Böhnke, R. A. Stirling-Gallacher, W. Xu, and G. Caire, “On the beamformed broadcasting for millimeter wave cell discovery: Performance analysis and design insight,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7620–7634, Nov 2018.
- [44] Y. Li, F. Baccelli, J. G. Andrews, and J. C. Zhang, “Directional cell search delay analysis for cellular networks with static users,” *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 4318–4332, Sep. 2018.
- [45] Y. Yang, H. S. Ghadikolaei, C. Fischione, M. Petrova, and K. W. Sung, “Reducing initial cell-search latency in mmwave networks,” in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018, pp. 686–691.

- [46] A. Alkhateeb, Y. Nam, M. S. Rahman, J. Zhang, and R. W. Heath, “Initial beam association in millimeter wave cellular systems: Analysis and design insights,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2807–2821, May 2017.
- [47] X. Meng, X. Gao, and X. Xia, “Omnidirectional precoding and combining based synchronization for millimeter wave massive MIMO systems,” *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1013–1026, March 2018.
- [48] M. Giordani, M. Mezzavilla, and M. Zorzi, “Initial access in 5G mmwave cellular networks,” *IEEE Communications Magazine*, vol. 54, no. 11, pp. 40–47, November 2016.
- [49] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, “Millimeter wave beam alignment: Large deviations analysis and design insights,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1619–1631, July 2017.
- [50] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, “Coding the beams: Improving beamforming training in mmwave communication system,” in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec 2011, pp. 1–6.
- [51] A. Alkhateeb, G. Leus, and R. W. Heath, “Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?”

- in *Proc. IEEE Int. Conf. Acoustics, Speech and Sig. Process. (ICASSP)*, April 2015, pp. 2909–2913.
- [52] Y. Han and J. Lee, “Two-stage compressed sensing for millimeter wave channel estimation,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 860–864.
- [53] J. Lee, G. T. Gil, and Y. H. Lee, “Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications,” in *Proc. of 2014 IEEE Global Communications Conference*, Dec 2014, pp. 3326–3331.
- [54] M. Iwen and A. Tewfik, “Adaptive strategies for target detection and localization in noisy environments,” *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2344–2353, 2012.
- [55] M. L. Malloy and R. D. Nowak, “Near-optimal compressive binary search,” *arXiv preprint arXiv:1306.6239*, 2012.
- [56] D. Ramasamy, S. Venkateswaran, and U. Madhow, “Compressive tracking with 1000-element arrays: A framework for multi-Gbps mm wave cellular downlinks,” in *Proc. Allerton Conf. Commun., Control, Computing*, Oct 2012, pp. 690–697.
- [57] D. E. Berraki, S. M. D. Armour, and A. R. Nix, “Application of compressive sensing in sparse spatial channel recovery for beamforming in

- mmWave outdoor systems,” in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2014, pp. 887–892.
- [58] Z. Marzi, D. Ramasamy, and U. Madhow, “Compressive channel estimation and tracking for large arrays in mm-wave picocells,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 514–527, April 2016.
- [59] I. Chafaa and M. Djeddou, “Improved channel estimation in mmwave communication system,” in *2017 Seminar on Detection Systems Architectures and Technologies (DAT)*, Feb 2017, pp. 1–5.
- [60] Z. Xiao, P. Xia, and X. G. 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.
- [61] J. Sung and B. L. Evans, “Hybrid beamformer codebook design and ordering for compressive mmwave channel estimation,” in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 914–919.
- [62] —, “Versatile compressive mmwave hybrid beamformer codebook design framework,” in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 1052–1057.
- [63] C. Rusu and N. Gonz´alez-Prelcic, “Designing incoherent frames through

- convex techniques for optimized compressed sensing,” *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2334–2344, 2016.
- [64] C. Rusu, N. González-Prelcic, and R. W. Heath, “The use of unit norm tight measurement matrices for one-bit compressed sensing,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4044–4048.
- [65] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, “Subspace estimation and decomposition for large millimeter-wave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, April 2016.
- [66] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, “Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation,” *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 601–616, Feb 2017.
- [67] M. Kokshoorn, P. Wang, Y. Li, and B. Vucetic, “Fast channel estimation for millimetre wave wireless systems using overlapped beam patterns,” in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1304–1309.
- [68] D. Zhu, J. Choi, and R. W. Heath, “Auxiliary beam pair enabled aod and aoa estimation in closed-loop large-scale millimeter-wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4770–4785, July 2017.

- [69] Z. Gao, L. Dai, and Z. Wang, “Channel estimation for mmwave massive MIMO based access and backhaul in ultra-dense network,” in *Proc. IEEE Int. Conf. on Commun. (ICC)*, May 2016, pp. 1–6.
- [70] Z. Gao, C. Hu, L. Dai, and Z. Wang, “Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels,” *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, June 2016.
- [71] S. Liu, F. Yang, W. Ding, X. Wang, and J. Song, “Two-dimensional structured-compressed-sensing-based NBI cancelation exploiting spatial and temporal correlations in MIMO systems,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9020–9028, Nov 2016.
- [72] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014.
- [73] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, ser. Springer series in statistics. Springer, 2009. [Online]. Available: <https://books.google.com/books?id=eBSgoAEACAAJ>
- [74] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, “Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions,” *IEEE Access*, vol. 7, pp. 137 184–137 206, 2019.

- [75] B. Bojovic, E. Meshkova, N. Baldo, J. Riihijärvi, and M. Petrova, “Machine learning-based dynamic frequency and bandwidth allocation in self-organized LTE dense small cell deployments ,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 183, Aug. 2016. [Online]. Available: DOI:10.1186/s13638-016-0679-0
- [76] R. D. A. Timoteo, D. Cunha, and G. Cavalcanti, “A proposal for path loss prediction in urban environments using support vector regression,” *Proc. Adv. Intl. Conf. Telecommun.*, vol. 10, pp. 119–124, 2014.
- [77] J. Liu, R. Deng, S. Zhou, and Z. Niu, “Seeing the unobservable: Channel learning for wireless communication networks,” in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.
- [78] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, “Deep learning coordinated beamforming for highly-mobile millimeter wave systems,” *IEEE Access*, vol. 6, pp. 37 328–37 348, 2018.
- [79] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, “Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8549–8560, 2018.
- [80] G. C. Sobabe, Y. Song, X. Bai, and B. Guo, “A cooperative spectrum sensing algorithm based on unsupervised learning,” in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–6.

- [81] J. Gazda, E. SLapak, G. Bugár, D. Horváth, T. Maksymyuk, and M. Jo, “Unsupervised learning algorithm for intelligent coverage planning and performance optimization of multitier heterogeneous network,” *IEEE Access*, vol. 6, pp. 39 807–39 819, 2018.
- [82] L. Wang and S. Cheng, “Data-driven resource management for ultra-dense small cells: An affinity propagation clustering approach,” *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 267–279, 2019.
- [83] T. Kudo and T. Ohtsuki, “Cell range expansion using distributed q-learning in heterogeneous networks,” in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, 2013, pp. 1–5.
- [84] M. Simsek, M. Bennis, and Guvenç, “Learning based frequency- and time-domain inter-cell interference coordination in hetnets,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4589–4602, 2015.
- [85] E. Stevens-Navarro, Y. Lin, and V. W. S. Wong, “An mdp-based vertical handoff decision algorithm for heterogeneous wireless networks,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1243–1254, 2008.
- [86] Y. Wei, F. R. Yu, M. Song, and Z. Han, “User scheduling and resource allocation in hetnets with hybrid energy supply: An actor-critic reinforcement learning approach,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 680–692, 2018.

- [87] F. B. Mismar, B. L. Evans, and A. Alkhateeb, “Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1581–1592, 2020.
- [88] Y. Wang, A. Klautau, M. Ribero, A. C. K. Soong, and R. W. Heath, “Mmwave vehicular beam selection with situational awareness using machine learning,” *IEEE Access*, vol. 7, pp. 87 479–87 493, 2019.
- [89] W. Ma, C. Qi, and G. Y. Li, “Machine learning for beam alignment in millimeter wave massive MIMO,” *IEEE Wireless Communications Letters*, pp. 1–1, 2020.
- [90] Y. Heng and J. G. Andrews, “Machine learning-assisted beam alignment for mmWave systems,” in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.
- [91] X. Li and A. Alkhateeb, “Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems,” in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 800–805.
- [92] W. Ma, C. Qi, Z. Zhang, and J. Cheng, “Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO,” *IEEE Transactions on Communications*, pp. 1–1, 2020.
- [93] Y. Guo, Z. Wang, M. Li, and Q. Liu, “Machine learning based mmWave channel tracking in vehicular scenario,” in *2019 IEEE International Con-*

- ference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6.
- [94] Y. Jin, J. Zhang, S. Jin, and B. Ai, “Channel estimation for cell-free mmWave massive MIMO through deep learning,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 325–10 329, 2019.
- [95] F. B. Mismar and B. L. Evans, “Deep learning in downlink coordinated multipoint in new radio heterogeneous networks,” *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1040–1043, 2019.
- [96] S. Bay, C. Herzet, J. Brossier, J. Barbot, and B. Geller, “Analytic and asymptotic analysis of Bayesian Cramer–Rao bound for dynamical phase offset estimation,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 61–70, Jan 2008.
- [97] D. D. Lin, R. A. Pacheco, T. J. Lim, and D. Hatzinakos, “Joint estimation of channel response, frequency offset, and phase noise in OFDM,” *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3542–3554, Sep. 2006.
- [98] O. H. Salim, A. A. Nasir, H. Mehrpouyan, W. Xiang, S. Durrani, and R. A. Kennedy, “Channel, phase noise, and frequency offset in OFDM systems: Joint estimation, data detection, and hybrid Cramér-Rao lower bound,” *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3311–3325, Sep. 2014.

- [99] E. P. Simon, L. Ros, H. Hijazi, and M. Ghogho, "Joint carrier frequency offset and channel estimation for OFDM systems via the EM algorithm in the presence of very high mobility," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 754–765, Feb 2012.
- [100] R. Corvaja and A. G. Armada, "SINR degradation in MIMO-OFDM systems with channel estimation errors and partial phase noise compensation," *IEEE Transactions on Communications*, vol. 58, no. 8, pp. 2199–2203, August 2010.
- [101] Y. Zhang and H. Liu, "MIMO-OFDM systems in the presence of phase noise and doubly selective fading," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 2277–2285, July 2007.
- [102] Y. Fang, X. Li, and L. Qiu, "Asymptotic equivalent performance of uplink massive MIMO systems with phase noise," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.
- [103] A. Chorti and M. Brookes, "A spectral model for RF oscillators with power-law phase noise," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 9, pp. 1989–1999, Sep. 2006.
- [104] A. G. Armada and M. Calvo, "Phase noise and sub-carrier spacing effects on the performance of an OFDM communication system," *IEEE Communications Letters*, vol. 2, no. 1, pp. 11–13, Jan 1998.

- [105] A. Garcia Armada, “Understanding the effects of phase noise in orthogonal frequency division multiplexing (OFDM),” *IEEE Transactions on Broadcasting*, vol. 47, no. 2, pp. 153–159, June 2001.
- [106] D. Petrovic, W. Rave, and G. Fettweis, “Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation,” *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1607–1616, Aug 2007.
- [107] H. Mehrpouyan, M. R. Khanzadi, M. Matthaiou, A. M. Sayeed, R. Schober, and Y. Hua, “Improving bandwidth efficiency in E-band communication systems,” *IEEE Communications Magazine*, vol. 52, no. 3, pp. 121–128, March 2014.
- [108] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, “IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi,” *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, Dec 2017.
- [109] E. Onggosanusi, M. S. Rahman, L. Guo, Y. Kwak, H. Noh, Y. Kim, S. Faxer, M. Harrison, M. Frenne, S. Grant, R. Chen, R. Tamrakar, and a. Q. Gao, “Modular and high-resolution channel state information and beam management for 5G new radio,” *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 48–55, Mar. 2018.
- [110] P. H. Moose, “A technique for orthogonal frequency division multiplexing

- frequency offset correction,” *IEEE Transactions on Communications*, vol. 42, no. 10, pp. 2908–2914, Oct 1994.
- [111] T. M. Schmidl and D. C. Cox, “Robust frequency and timing synchronization for OFDM,” *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1613–1621, Dec 1997.
- [112] H. Yan and D. Cabria, “Compressive sensing based initial beamforming training for massive MIMO millimeter-wave systems,” in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2016, pp. 620–624.
- [113] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, “Noncoherent mmwave path tracking,” in *HotMobile '17*, 2017.
- [114] H. Hassanieh, O. Salehi-Abari, M. Rodriguez, M. A. Abdelghany, D. Katabi, and P. Indyk, “Fast millimeter wave beam alignment,” in *SIGCOMM '18*, 2018.
- [115] M. E. Rasekh and U. Madhow, “Noncoherent compressive channel estimation for mm-wave massive MIMO,” in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Oct 2018, pp. 889–894.
- [116] N. J. Myers and R. W. Heath, “A compressive channel estimation technique robust to synchronization impairments,” in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–5.

- [117] —, “Joint CFO and channel estimation in millimeter wave systems with one-bit ADCs,” in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2017, pp. 1–5.
- [118] C. Zhang, Z. Xiao, L. Su, L. Zeng, and D. Jin, “Iterative channel estimation and phase noise compensation for SC-FDE based mmwave systems,” in *Proc. of 2015 IEEE Int. Conf. on Communication Workshop (ICCW)*, June 2015, pp. 2133–2138.
- [119] M. Pajovic, P. Wang, T. Koike-Akino, and P. Orlik, “Estimation of frequency unsynchronized millimeter-wave channels,” in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2017, pp. 1205–1209.
- [120] N. J. Myers, A. Mezghani, and R. W. Heath, “Swift-link: A compressive beam alignment algorithm for practical mmwave radios,” *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 1104–1119, Feb 2019.
- [121] N. J. Myers and R. W. Heath, “Message passing-based joint CFO and channel estimation in mmwave systems with one-bit ADCs,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 3064–3077, June 2019.
- [122] J. T. Parker and P. Schniter, “Parametric bilinear generalized approximate message passing,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 795–808, June 2016.

- [123] S. Ling and T. Strohmer, “Self-calibration and biconvex compressive sensing,” *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.
- [124] J. P. Vila and P. Schniter, “Expectation-maximization Gaussian-mixture approximate message passing,” *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4658–4672, Oct 2013.
- [125] J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, K. Venugopal, and R. W. Heath, “Exploiting common sparsity for frequency-domain wideband channel estimation at mmwave,” in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–6.
- [126] Y. Liu, X. Fang, M. Xiao, and S. Mumtaz, “Decentralized beam pair selection in multi-beam millimeter-wave networks,” *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2722–2737, 2018.
- [127] Y. Ghasempour, M. K. Haider, C. Cordeiro, and E. W. Knightly, “Multi-user multi-stream mmwave w lans with efficient path discovery and beam steering,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 12, pp. 2744–2758, 2019.
- [128] D. Ramasamy, S. Venkateswaran, and U. Madhow, “Compressive adaptation of large steerable arrays,” in *Proc. of Inf. Theory and Applications Workshop (ITA)*, Feb 2012, pp. 234–239.
- [129] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath Jr., “A frequency-domain approach to wideband channel estima-

- tion in millimeter wave systems,” in *Proc. IEEE Intl. Conf. Commun. (ICC)*, 2017.
- [130] A. Alkhateeb and R. W. Heath Jr., “Frequency selective hybrid precoding for limited feedback millimeter wave systems,” *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [131] M. K. Samimi and T. S. Rappaport, “3-D millimeter-wave statistical channel model for 5G wireless system design,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2207–2225, July 2016.
- [132] S. G. Larew, T. A. Thomas, M. Cudak, and A. Ghosh, “Air interface design and ray tracing study for 5G millimeter wave communications,” in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec 2013.
- [133] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, “Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design,” *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.
- [134] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Simultaneous sparse approximation via greedy pursuit,” in *Proc. IEEE Conf. Acous., Speech, and Signal Processing (ICASSP)*, 2005.
- [135] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall PTR, 1993.

- [136] K. Venugopal, N. González-Prelcic, and R. W. Heath, “Optimality of frequency flat precoding in frequency selective millimeter wave channels,” *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 330–333, June 2017.
- [137] J. Rodríguez-Fernández, N. Gonzalez-Prelcic, and R. W. Heath, “Channel estimation for millimeter wave MIMO systems in the presence of CFO uncertainties,” in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.
- [138] J. Rodríguez-Fernández, N. González-Prelcic, and R. W. Heath, “Channel estimation for millimeter wave MIMO systems in the presence of CFO uncertainties,” in *in Proc. of 2018 IEEE Conference on Communications (ICC)*, 2018.
- [139] J. R. Hampton, *Introduction to MIMO Communications*. Cambridge University Press, 2014.
- [140] S. Kishimoto, K. Maruhashi, M. Ito, T. Morimoto, Y. Hamada, and K. Ohata, “A 60-GHz-band subharmonically injection locked VCO MMIC operating over wide temperature range,” in *IEEE MTT-S International Microwave Symposium Digest, 2005.*, June 2005, pp. 4 pp.–.
- [141] J. Rodríguez-Fernández and N. González-Prelcic, “Joint synchronization and compressive estimation for frequency-selective mmwave MIMO systems,” in *in Proc. of IEEE Asilomar Conference on Signals, Systems, and Computers*, 2018.

- [142] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.
- [143] E. J. Candes and Y. Plan, “A probabilistic and RIPless theory of compressed sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, Nov 2011.
- [144] R. Lopez-Valcarce, N. Gonzalez-Prelcic, C. Rusu, and R. W. Heath, “Hybrid precoders and combiners for mmwave MIMO systems with per-antenna power constraints,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.
- [145] J. Rodríguez-Fernández, N. González-Prelcic, and R. W. Heath, “A compressive sensing-maximum likelihood approach for off-grid wideband channel estimation at mmwave,” in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2017, pp. 1–5.
- [146] “IEEE draft standard for local and metropolitan area networks - specific requirements - part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications - amendment 3: Enhancements for very high throughput in the 60 ghz band,” *IEEE P802.11ad/D8.0, May 2012 (Draft Amendment based on IEEE 802.11-2012)*, pp. 1–667, June 2012.

- [147] T. A. Thomas, M. Cudak, and T. Kovarik, “Blind phase noise mitigation for a 72 GHz millimeter wave system,” in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1352–1357.
- [148] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.
- [149] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [150] R. W. Heath, *Introduction to Wireless Digital Communication: A Signal Processing Perspective*. Prentice Hall, 2017.
- [151] X. Yu, J. Zhang, and K. B. Letaief, “Partially-connected hybrid precoding in mm-wave systems with dynamic phase shifter networks,” in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017, pp. 1–5.
- [152] D. D. Lin and T. J. Lim, “The variational inference approach to joint data detection and phase noise estimation in OFDM,” *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1862–1874, May 2007.
- [153] *Counting the number of FLOPs, MATLAB library, available online at <https://www.mathworks.com/matlabcentral/fileexchange/50608-counting-the-floating-point-operations-flops>*, The Mathworks, Inc., Natick, Massachusetts, 2020.

- [154] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. W. Heath, “Channel estimation and hybrid precoding for frequency selective multiuser mmwave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 353–367, May 2018.
- [155] J. P. Gonzalez-Coma, J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, and L. Castedo, “Channel estimation and hybrid precoding/combining for frequency selective multiuser mmwave systems,” in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–6.
- [156] R. W. Heath Jr., *Introduction to Wireless Digital Communication: a Signal Processing Perspective*. Prentice Hall, 2016.
- [157] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.
- [158] A. S. Charles, A. Balavoine, and C. J. Rozell, “Dynamic filtering of time-varying sparse signals via  $\ell_1$  minimization,” *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5644–5656, Nov 2016.
- [159] R. Tibshirani, “Regression shrinkage and selection via the LASSO,” *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 58, pp. 267–288, 1994.
- [160] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, p. 129–159, Jan. 2001.

[Online]. Available: <https://doi.org/10.1137/S003614450037906X>

- [161] F. Santosa and W. W. Symes, “Linear inversion of band-limited reflection seismograms,” *SIAM J. Sci. Stat. Comput.*, vol. 7, no. 4, p. 1307–1330, Oct. 1986.
- [162] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Commun. ACM*, vol. 53, no. 12, p. 93–100, Dec. 2010. [Online]. Available: <https://doi.org/10.1145/1859204.1859229>
- [163] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.
- [164] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Ltd, 2002.

# Vita

Javier Rodriguez-Fernandez received the Bachelor and Master of Science degree in Telecommunication Engineering from the University of Vigo, Spain, in 2014 and 2016, respectively. He started the Ph.D. degree in Information and Communication Technologies in the University of Vigo in 2016, and transferred to The University of Texas at Austin in Spring 2018, where he is currently working towards the Ph.D. degree in Electrical and Computer Engineering.

Permanent address: 5413 Guadalupe Street  
Austin, Texas 78751

This dissertation was typeset with  $\text{\LaTeX}^\dagger$  by the author.

---

<sup>†</sup> $\text{\LaTeX}$  is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's  $\text{\TeX}$  Program.