Waveform Design for Sparse Delay-Doppler Channel Estimation

Dissertation

submitted to

Sorbonne Université

in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Author:

Wissal BENZINE

Scheduled for defense on the 28^{th} *April of* 2025 *before a committee composed of:*

Reviewers

Prof.	**Giuseppe Caire**	TU-Berlin, Germany
Prof.	**Didier Le Ruyet**	CNAM, France

Examiners

Prof.	**Lina Mroueh**	ISEP, France
Dr.	**Maxime Guillaud**	INRIA, France
Prof.	**Dirk Slock**	Eurecom, France

Thesis Advisor

Prof.	**Dirk Slock**	Eurecom, France

Co-supervisor of thesis

Dr.	**Nassar Ksairi**	Huawei R&D, France

Conception de Formes d'Onde pour l'Estimation de Canaux Parcimonieux en Délai-Doppler

Thèse

soumise à

Sorbonne Université

pour l'obtention du Grade de Docteur

présentée par:

Wissal BENZINE

Soutenance de thèse prévue le 28 Avril 2025 devant le jury composé de:

Rapporteur

Prof.	**Giuseppe Caire**	TU-Berlin, Allemagne
Prof.	**Didier Le Ruyet**	CNAM, France

Examinateur

Prof.	**Lina Mroueh**	ISEP, France
Dr.	**Maxime Guillaud**	INRIA, France
Prof.	**Dirk Slock**	Eurecom, France

Directeur de Thèse

Prof.	**Dirk Slock**	Eurecom, France

Co-encadrant de thèse

Dr.	**Nassar Ksairi**	Huawei R&D, France

Abstract

With the rapid evolution of wireless communication technologies, ensuring reliable and efficient data transmission in high-mobility scenarios has become a critical challenge. In particular, accurate channel estimation is essential to maintain communication quality in environments characterized by significant Doppler effects and dynamic propagation conditions. This thesis explores the design of waveforms for efficient sparse delay-Doppler channel estimation in high-mobility wireless communication systems. As future wireless networks demand robust and accurate channel estimation techniques, particularly in the presence of high Doppler shifts, the study focuses on both on-grid and off-grid approaches for doubly sparse linear time-varying (DS-LTV) channels.

In the first part, we investigate on-grid DS-LTV channel estimation and introduce three different sparsity models that characterize practical propagation environments. We propose an optimized estimation framework leveraging the minimum mean squared error (MMSE) criterion and basis expansion models (BEMs). Through theoretical analysis and simulations, we demonstrate that Affine Frequency Division Multiplexing (AFDM) outperforms traditional waveforms such as Orthogonal Frequency Division Multiplexing (OFDM) and Orthogonal Time Frequency Space (OTFS) in terms of pilot overhead reduction and estimation accuracy.

In the second part, we extend our study to off-grid DS-LTV channel estimation, addressing the issue of mismatches between actual Doppler shifts and predefined grid points. By employing novel off-grid approximation techniques based on multiple shifted

elementary BEMs, we enhance estimation robustness and improve channel prediction capabilities. Our findings confirm AFDM's efficiency in handling off-grid Doppler shifts and its potential for adaptive transmission strategies.

Beyond channel estimation, we explore the broader implications of our research for radar and sensing applications, demonstrating the feasibility of sub-Nyquist radar techniques that optimize sampling rates while maintaining detection accuracy. This interdisciplinary approach highlights the impact of our work beyond wireless communication systems. The conclusions drawn from this research provide valuable insights for the development of next-generation communication technologies. Future work could explore adaptive sparsity-aware estimation techniques, machine learning-based approaches, and real-world experimental validations to further enhance the practical deployment of AFDM in high-mobility scenarios.

Résumé

Avec l'évolution rapide des technologies de communication sans fil, garantir une transmission de données fiable et efficace dans des scénarios de haute mobilité est devenu un défi crucial. En particulier, une estimation précise du canal est essentielle pour maintenir la qualité de communication dans des environnements caractérisés par des effets Doppler significatifs et des conditions de propagation dynamiques.

Cette thèse explore la conception de formes d'onde pour une estimation efficace des canaux parcimonieux en delai-Doppler dans les systèmes de communication sans fil à haute mobilité. Alors que les futurs réseaux sans fil exigent des techniques d'estimation de canal robustes et précises, en particulier en présence de forts décalages Doppler, cette étude se concentre sur les approches sur grille et hors grille pour les canaux linéaires temporellement variables (DS-LTV) à double parcimonie.

Dans la première partie, nous étudions l'estimation des canaux DS-LTV sur grille et introduisons trois modèles de parcimonie différents caractérisant les environnements de propagation pratiques. Nous proposons un cadre d'estimation optimisé exploitant le critère de l'erreur quadratique moyenne minimale (MMSE) et les modèles d'expansion de base (BEMs). Grâce à des analyses théoriques et des simulations, nous démontrons que le multiplexage par division de fréquence affine (AFDM) surpasse les formes d'onde traditionnelles telles que le multiplexage par division orthogonale de fréquence (OFDM) et l'espace temps-fréquence orthogonal (OTFS) en termes de réduction de la surcharge des pilotes et d'amélioration de l'exactitude de l'estimation.

Dans la seconde partie, nous étendons notre étude à l'estimation des canaux DS-LTV hors grille, en abordant le problème des écarts entre les décalages Doppler réels et les points de grille prédéfinis. En employant de nouvelles techniques d'approximation hors grille basées sur des modèles d'expansion de base élémentaires déplacées multiples, nous améliorons la robustesse de l'estimation et les capacités de prédiction des canaux. Nos résultats confirment l'efficacité de l'AFDM dans le traitement des décalages Doppler hors grille et son potentiel pour des stratégies de transmission adaptatives.

Au-delà de l'estimation de canal, nous explorons les implications plus larges de notre recherche pour les applications radar et de détection, démontrant la faisabilité des techniques radar sous-Nyquist qui optimisent les taux d'échantillonnage tout en maintenant la précision de détection. Cette approche interdisciplinaire met en évidence l'impact de notre travail au-delà des systèmes de communication sans fil.

Les conclusions tirées de cette recherche fournissent des perspectives précieuses pour le développement des technologies de communication de nouvelle génération. Les travaux futurs pourraient explorer des techniques d'estimation adaptatives tenant compte de la parcimonie, des approches basées sur l'apprentissage automatique et des validations expérimentales en conditions réelles afin de renforcer le déploiement pratique de l'AFDM dans les scénarios de haute mobilité.

Acknowledgements

This PhD was a remarkable journey, one that I could not have completed alone. I would like to express my deepest gratitude and sincere thanks to:

Dirk SLOCK, my supervisor. I am truly honored to have had the opportunity to work under his supervision and learn from his expertise; to **Nassar KSAIRI**, my co-supervisor, for his constant guidance, unwavering support, and belief in my abilities. His insightful advice, technical expertise, and continuous encouragement were instrumental in shaping the success of this work.

To the jury members, for their time, constructive feedback, and the thoughtful evaluation of my research.

To my managers at Huawei Technologies, **Jean-Claude BELFIORE** and **Lirong**, for providing me with an inspiring and supportive work environment and for their continued trust and leadership; to my wonderful colleagues and friends **Salma EL MIZ**, **Maha BOUHADIDA** , **Melissa ZEBBOUDJ**, **Thanina KOUDIA**, **Mustapha HAMAD**, **Rima KHOUJA**, **Ali BEMANI** and **Serge AFOUDA**, for the constructive discussions, good spirit, and memorable moments we shared. To my colleague and friend at EURECOM, **Zixuan HE**, for the good moments we spent together.

To my best friends, **Imene GHAMNIA** and **Chaymae EL QOURCHI**, for their unconditional support, kindness, and the many moments of joy and laughter that helped me stay grounded.

To my wonderful sisters **Salma** and **Arwa**, for their love, encouragement, and the positive energy they bring into my life. To my parents, for their infinite love, understanding, and moral support that have guided me through every step of my life. I am deeply grateful for their sacrifices and belief in me.

And finally, to my dear husband **Souhail**, for his patience, love, and unwavering support. His encouragement and belief in me have been my driving force throughout this journey.

Thank you all.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

The acronyms and abbreviations used throughout the manuscript are specified in the following. They are presented here in their singular form, and their plural forms are constructed by adding and s, e.g. BS (base stations) and BSs (Base stations). The meaning of an acronym is also indicated the first time that it is used.

AFDM	affine frequency division multiplexing
AFT	affine Fourier transform
BEM	basis expansion model
BER	Bit Error Rate
CCDF	complementary cumulative distribution function
CP	cyclic prefix
CPP	chirp-periodic prefix
CS	compressed sensing
DAFT	discrete affine Fourier transform
DD	delay-Doppler
DDP	delay-Doppler profile
DFT	discrete Fourier transform
DPSS	Discrete Prolate Spheroidal Sequences
DS	doubly selective
DS-LTV	Doubly sparse linear time-varying
HTP	hard thresholding pursuit
HiRIP	hierarchical restricted isometry property
ICI	inter-carrier interference
iid	independently and identically distributed
IDAFT	inverse discrete affine Fourier transform

ISAC	integrated sensing and communications
LTV	linear time-varying
LTI	linear time-invariant
MIMO	multiple-input multiple-output
MMSE	minimum mean squared error
MSE	mean square error
OFDM	orthogonal frequency division multiplexing
OTFS	orthogonal time frequency space
PDP	power delay profile
PROMP	parameter-refined orthogonal matching pursuit
QAM	quadrature amplitude modulation
RF	radio frequency
RIP	restricted isometry property
SBL	sparse Bayesian learning
SCM	single-carrier modulation
SNR	signal-to-noise ratio
TF	time-frequency
UWB	ultra-wideband
V2X	vehicle-to-everything

Notations

The next list describes an overview on the notation used throughout this manuscript. We use boldface uppercase letters ($\mathbf{A}$) for matrices, boldface lowercase letters for vectors ($\mathbf{a}$), and regular letters for scalars (a or A). Sets are represented by calligraphic uppercase letters ($\mathcal{A}$).

Bernoulli(p)	The Bernoulli distribution with probability p
B(n, p)	Binomial distribution with parameters (n, p)
$\mathcal{CN}\left(0, \sigma^2\right)$	Zero-mean Complex Gaussian distribution with variance σ^2
$X \sim F$	Random variable X has distribution F
$\mathbb{P}$	Probability measure associated with a random variable
$\mathbb{E}[X]$	Expected value of random variable X
$\lvert\mathcal{A}\rvert$	Cardinality of the set $\mathcal{A}$
$(x)_+$	Maximum between the real number x and 0, $\max(x, 0)$
$[\![l..m]\!]$	Integer interval between l and m (including l and m)
$\mathbf{A}^H$	Hermitian transpose of the matrix $\mathbf{A}$
$[\mathbf{A}]_c$	The c-th column of the matrix $\mathbf{A}$
$[\mathbf{A}]_{r,c}$	The r-th entry of the c-th column of the matrix $\mathbf{A}$
$(\cdot)_N$	Modulo N
$\lceil . \rceil$	Ceil function
$\lfloor . \rfloor$	Floor function
$\lVert\mathbf{a}\rVert$	Euclidian norm of the vector $\mathbf{a}$
$\lvert a\rvert$	Absolute value of the variable a
$\triangleq$	Used for definition
$\log(a)$	Logarithm of the positive number a

Chapter 1

Introduction

Supporting a wide range of services and applications is the stated aim of next-generation wireless systems (beyond 5G/6G). This involves maintaining dependable communications in settings characterized by high mobility, like diverse automotive communications including communications like V2X and high-speed railway systems. As wireless networks evolve to meet these demands, one key challenge remains: how to maintain robust, efficient, and reliable communications in highly dynamic environments where mobility introduces rapid time variations in the propagation channel, particularly in the case of linear time-varying (LTV) channels.

The current backbone of wireless communication, Orthogonal Frequency Division Multiplexing (OFDM), has proven effective in stable or slowly varying environments. OFDM is central to modern communication standards such as 4G and 5G due to its efficient spectrum utilization in typical terrestrial communication channels [1]. However, OFDM begins to show limitations in high-mobility contexts, particularly where Doppler shifts, resulting from relative motion between the transmitter and receiver, significantly affect the communication channel. In such cases, the frequency channel becomes time-varying, leading to interference, specifically Inter-Carrier Interference (ICI), which arises from the mismatch in frequency synchronization between transmitters and receivers.

This interference degrades system performance and limits the effectiveness of OFDM in high-mobility scenarios [2].

With the push towards higher frequency bands, such as millimeter-wave (mmWave) bands, which offer vast bandwidth but are more susceptible to Doppler shifts, the challenges for traditional OFDM systems are further exacerbated. The need for new, more robust communication methods to handle these high-mobility, high-frequency environments has never been more pressing [2]. However, existing methods often suffer from high computational complexity and pilot overhead, making them inefficient in practical deployments.

Traditional time-frequency (TF) domain-based methods for channel estimation struggle to cope with these rapid variations inherent in LTV channels. These methods typically assume slow or moderate channel dynamics, an assumption that does not hold in high-speed, high-frequency scenarios where channel conditions can change dramatically over short timescales. Furthermore, the need for frequent pilot signals in TF-based approaches leads to a substantial waste of bandwidth, which could otherwise be used for transmitting actual data. The delay-Doppler (DD) domain, on the other hand, offers a promising alternative for channel estimation in such scenarios. In the DD domain, LTV channels that exhibit rapid time variations appear almost stationary, simplifying the estimation process by extending the time period over which the channel can be considered constant [3]. However, despite this advantage, efficient channel estimation in the DD domain remains a challenge due to the high-dimensional nature of the channel parameters.

Additionally, sparsity is a key characteristic of wireless channels that can be exploited to reduce the amount of pilot overhead required for accurate channel estimation. In systems operating in sub-6 GHz bands, the channel sparsity is primarily observed in the delay domain [4, 5], where the number of significant delay taps is much smaller than the total possible delay spread. In high-mobility scenarios, this sparsity extends into the Doppler domain, where the number of significant Doppler shifts is also small, even for

relatively high transmitter-receiver velocities.

A question thus arises as to **which waveforms can offer for communications on LTV channels both good data transmission reliability and efficient channel estimation schemes that can take advantage of DD sparsity**. To deal with high-mobility communications, orthogonal time frequency space (OTFS) modulation has been recently proposed [6]. OTFS is a two-dimensional (2D) modulation technique that spreads the information symbols over the delay-Doppler domain [3]. It has been proved that OTFS outperforms OFDM in doubly selective channels [7]. However, channel estimation overhead in OTFS cannot be reduced when the channel exhibits more sparsity (or at best it can be reduced to a very limited extent) unless non-orthogonal pilot-data multiplexing is employed. Indeed, one way of exploiting delay-Doppler sparsity in OTFS is according to [8] by using a sparse superimposed pilot (SP-Sparse) scheme for channel estimation. SP-Sparse is a non-orthogonal scheme where pilot symbols are superimposed on top of data symbols in the transform domain of OTFS. When (the more practical) orthogonal pilot transmission is instead maintained, the performance gap between OTFS and OFDM narrows in favor of OFDM on channels with delay domain sparsity [9]. Another recently proposed waveform for communications in high-mobility scenarios is Affine Frequency Division Multiplexing (AFDM) [10, 11]. While in AFDM data and pilot symbols are not directly transmitted in the delay-Doppler domain, AFDM can still reconstruct a DD representation of the channel achieving full diversity on doubly dispersive channels. In the absence of DD sparsity, AFDM has a comparable bit error rate (BER) performance to that of OTFS but with the advantage of requiring less channel estimation overhead [10]. However, the study of AFDM reliability and channel estimation performance under the assumption of DD sparsity has not been investigated yet.

Therefore, in this thesis, we explore advanced waveform design strategies tailored for *sparse* delay-Doppler channel estimation. Specifically, we analyze how different waveforms—such as OTFS, OFDM, single-carrier modulation (SCM) and AFDM—perform

under varying sparsity conditions in high-mobility environments. Our objective is to determine which waveform best exploits delay-Doppler sparsity to minimize pilot overhead while maintaining accurate channel estimation.

Before addressing that problem, we first give in the next section some common background material that will be useful for the remaining chapters of the thesis. This includes a reminder on LTV channels and on AFDM—a waveform that has demonstrated robustness in high-mobility scenarios and provides insight into practical communication systems designed for such channels.

1.1 Background

1.1.1 Linear time-varying channels

This thesis deals mainly with waveform and pilot design solutions for wireless propagation links that can be modeled as LTV channels that are further characterized by double sparsity, in both their delays and Doppler components. In this section, we introduce the mathematical representations of LTV channels. To accurately model and estimate such channels, we adopt a probabilistic delay-Doppler representation, which provides a more structured and intuitive framework for analyzing channel sparsity.

A LTV channel is a model of multipath propagation that is characterized by changes in its impulse response over time, caused by Doppler frequency shifts. The received signal at the channel output corresponding to a signal $s(t)$ at its input is expressed as:

$$r(t) = \int_{\tau} s(t)h(t,\tau)d\tau + z(t), t \in \mathbb{R}, \tag{1.1}$$

where $z(t)$ is the additive white Gaussian noise process and

$$h(t,\tau) = \sum_{p=1}^{N_{\mathrm{p}}} g_p e^{\imath 2\pi\nu_p t\Delta f}\delta(\tau-\tau_p), \tag{1.2}$$

is the continuous-time impulse response of the channel. Here, $N_{\mathrm{p}} \geq 1$ is the number of paths (a random variable in general), $\delta(\cdot)$ is the Dirac delta function, and g_p, ν_p and τ_p are the (in general random) complex gain, Doppler shift (normalized with respect to the subcarrier spacing Δf i.e., the frequency resolution), and delay (normalized with respect to the sample period T_{s} i.e., the time resolution) associated with the p-th path, respectively. We define

$$\tau_p = l_p + \iota_p \,, \tag{1.3}$$

where $l_p \in [\![0..L-1]\!]$ is its integer part, while ι_p is the fractional part that satisfies $\frac{-1}{2} < \iota_p \leq \frac{1}{2}$. We also define

$$\nu_p = q_p + \kappa_p \,, \tag{1.4}$$

where $q_p \in [\![-Q..Q]\!]$ is its integer part, while κ_p is the fractional part satisfying $\frac{-1}{2} < \kappa_p \leq \frac{1}{2}$.

In practice, the transmitted signal $s(t)$ is the continuous-time version of a discrete-time signal $s_n \triangleq s\left(nT_{\mathrm{s}}\right)$ generated (assuming a sample rate T_{s}) from a vector $\mathbf{x}$ of N symbols (for some integer value $N > 0$). These symbols could be either data symbols, pilot symbols or a combination of both. Defining $r_n \triangleq r(nT_{\mathrm{s}})$ and $z_n \triangleq z(nT_{\mathrm{s}})$, the discrete-time version of the LTV channel model in (1.1) becomes

$$r_n = \sum_{p=1}^{N_{\mathrm{p}}} g_p e^{\imath 2\pi \nu_p n \Delta f T_{\mathrm{s}}} s(nT_{\mathrm{s}} - \tau_p) + z_n, \quad n \in \mathbb{Z} \,. \tag{1.5}$$

From now on, $z_n \sim \mathcal{CN}\left(0, \sigma_w^2\right)$ and the process $(z_n)_{n\in\mathbb{Z}}$ is modeled as independent identically distributed (i.i.d.).

The input-output relation in (1.5), while defining the LTV model, is not sufficient to give a rigorous sense to delay-Doppler sparsity. Such a rigorous definition of delay-Doppler sparsity is crucial to get the kind of mathematical model of doubly-sparse (DS)-LTV channels that can enable developing sparsity-aware waveform, pilot design, and channel

estimation solutions and for the thorough analysis of those solutions. Indeed, while it is tempting to associate sparsity with the value of N_{p} being small relative to some relevant measure, this does not correspond to many realistic wireless propagation scenarios which are characterized simultaneously by a large value of N_{p} and effective sparsity. This apparent contradiction disappears when we consider that many of the N_{p} channel paths have in real-world scenarios delay and Doppler frequency shift values that are very close with respect to the time and frequency resolutions of the wireless transmission. To remedy this limitation, a rigorous double-sparsity modeling is provided in Chapter 2 and extended in Chapter 3.

1.1.2 Affine frequency division multiplexing (AFDM)

A recently proposed waveform that can achieve robust communication performance in high mobility scenarios is Affine Frequency Division Multiplexing (AFDM) [10, 11]. AFDM employs multiple orthogonal chirps generated using the discrete affine Fourier transform (DAFT). With chirp parameters adapted to the channel characteristics, AFDM can reconstruct a delay-Doppler representation of the channel, achieving full diversity on doubly dispersive channels. In comparison with OTFS, AFDM has comparable bit error rate (BER) performance but with the advantage of requiring less channel estimation overhead [10].

In AFDM, modulation is achieved through the use of DAFT. DAFT is a discretized version [12] of the affine Fourier transform (AFT) [10, 13] with chirp $e^{-\imath 2\pi(c_2k^2+\frac{1}{N}kn+c_1n^2)}$ as its kernel(see Fig. 1.1).In this work, we focus on a single design parameter, c_1, which is adjusted to account for the level of sparsity in the channel, governed by its delay-Doppler structure. The parameter c_1 is defined as $c_1 = \frac{P}{2N}$ by appropriately tuning c_1, the effective delay spread in the DAFT domain can be made large enough to minimize overlap between channel taps. This facilitates better exploitation of the channel's sparsity and helps reduce inter-tap interference in the transformed domain.

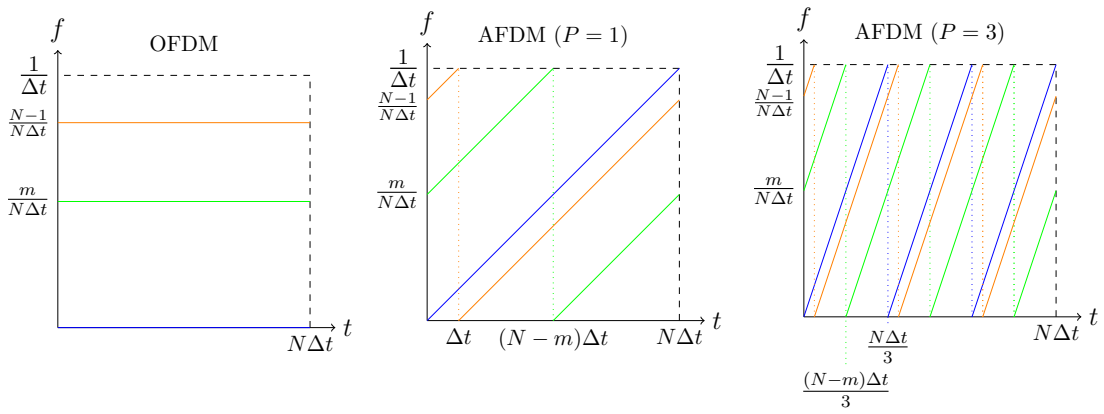


Figure 1.1: Time-frequency representation of three subcarriers of OFDM and AFDM ($c_1 = \frac{P}{2N}$). Each subcarrier is represented with a different colour.

Consider a set of quadrature amplitude modulation (QAM) symbols denoted $\{x_k\}_{k=0\cdots N-1}$. AFDM employs inverse DAFT (IDAFT) to map $\{x_k\}_{k=0\cdots N-1}$ to $\{s_n\}_{n=0\cdots N-1}$ as follows:

$$s_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x_k e^{\imath 2\pi (c_2 k^2 + \frac{1}{N} kn + c_1 n^2)}, n = 0 \cdots N-1 \tag{1.6}$$

with the following so-called *chirp-periodic prefix* (CPP)

$$s_n = s_{N+n} e^{-\imath 2\pi c_1 (N^2 + 2Nn)}, \quad n = -L_{\text{CPP}} \cdots -1 \tag{1.7}$$

where L_{CPP} denotes an integer that is greater than or equal to the number of samples required to represent the maximum delay of the wireless channel. The CPP simplifies to a cyclic prefix (CP) whenever $2c_1 N$ is integer and N is even, an assumption that will be considered to hold from now on.

When AFDM is used for transmission on LTV channels, it has the property that a channel path with a delay and a Doppler frequency shift equal in samples to respectively l and q, appears in the DAFT domain as a path with an effective delay equal to $-2Nc_1 l + q$. This particular way of mixing delays and Doppler shifts in the DAFT domain that depends on the AFDM parameters was shown in [10] to be the key to AFDM achieving the full

diversity of LTV channels. Note that throughout this thesis we show the relevance of the AFDM not only for achieving the full diversity of LTV channels but also for conceiving efficient channel estimation solutions that can take advantage of the delay-Doppler sparsity of doubly sparse linear time-varying (DS-LTV) channels.

1.2 Problem to solve

The problem addressed in this manuscript is the following.

1. To develop a comprehensive framework for efficient and accurate delay-Doppler sparse channel estimation, including rigorous modeling and closed-form performance analysis.

2. To use the developed framework to compare different wireless waveforms in terms of sparse delay-Doppler channel estimation performance and to propose efficient pilot schemes based on some of these waveforms capable of leveraging the delay-Doppler sparsity of the channel.

To achieve this, we propose a mathematical representation of the channel that will be formulated into two approximations: an on-grid and an off-grid model. The on-grid approximation leverages a discretized delay-Doppler domain, facilitating structured analysis and computational efficiency, while the off-grid model captures the continuous nature of the channel in the Doppler domain, addressing inaccuracies introduced by grid-based methods. Specifically, we propose a method for sparse channel estimation based on an off-grid channel representation using multiple elementary basis expansion models (BEMs). This channel representation will allow us to propose novel waveform designs, and their related pilot schemes, that minimize channel estimation pilot overhead while maintaining high estimation accuracy. These designs exploit the sparsity in both the delay and Doppler domains, and closed-form asymptotic results will be derived to quantify their associated minimal pilot overhead. Through both theoretical analysis and

numerical simulations, we will demonstrate the superiority of the proposed framework, particularly for practical system parameters, and highlight the potential benefits of AFDM in terms of pilot overhead reduction and estimation accuracy.

We show in the next section that while previous research has explored the potential of exploiting delay-Doppler sparsity for channel estimation, many of its approaches rely on restrictive assumptions or fail to account for the full complexity of real-world propagation environments. Specifically, existing methods either assume a grid-based discretization of the delay and Doppler dimensions or rely on simplistic models that do not capture the nuances of delay-Doppler sparsity. Furthermore, these approaches often fail to provide analytical insights into the optimal pilot overhead or the performance comparison of various waveforms under different sparsity levels.

1.3 State of the art

In this section, we provide a comprehensive overview of the background knowledge essential for understanding the key contributions of this manuscript, summarizing existing solutions and approaches related to the problem stated previously. We will go through the various methods used for channel estimation, focusing on the exploitation of sparsity in the delay and Doppler domains. This includes discussing grid-based approaches, compressive sensing (CS) techniques, and off-grid methods, as well as their strengths and limitations in high-mobility and high-frequency environments. By reviewing these existing strategies, we will lay the groundwork for the novel contributions presented in this manuscript.

Sparsity plays a crucial role in wireless communication, particularly in the estimation of LTV channels. The concept of sparsity emerges due to the limited number of dominant scatterers contributing to signal transmission. By leveraging sparsity, advanced estimation methods can mitigate the challenges posed by high-mobility scenarios where channel variations are rapid and unpredictable. In wireless communication systems operating in sub-6GHz frequency bands, sparsity is often observed in the delay domain, where only a

few dominant delay taps carry most of the channel energy [4, 5]. Channel estimation methods leveraging this delay-domain sparsity have been proposed in various studies, such as [14], which uses grid-based discretization of the delay dimension to enable a compressive sensing framework. However, in high-mobility scenarios, such as vehicular or high-speed train communications, sparsity extends to the Doppler domain as well. Also, in high-frequency bands, Doppler domain sparsity appears even at moderate transmitter-receiver relative velocity values. Delay-Doppler sparsity was assumed in [15, 16] and leveraged to conceive enhanced channel estimation schemes for time-varying channels using the sparse Bayesian learning (SBL) framework. However, delay-Doppler sparsity was modeled as the sparsity of a one-dimensional array with no way to assign different sparsity levels to the delay and Doppler domains. The model in [17] also assumes a form of delay-Doppler sparsity, where one Doppler shift is considered per delay tap, though this assumption can be restrictive for real-world wireless propagation channels. Other studies have explored discretized delay-Doppler representations for improved channel estimation, highlighting the efficiency of compressive sensing techniques in capturing the sparse characteristics of wireless channels [3, 18]. The authors of [19] use the *virtual channel representation* i.e, multi-dimensional Fourier basis functions, to model multiple-input multiple-output (MIMO) time-varying channels and introduce sparsity in the delay, Doppler and spatial angle domains of that representation with no restriction on the number of Doppler shifts per delay tap. Classical on-grid models assume that delay and Doppler shifts are quantized to a predefined grid, allowing for structured sparse recovery methods. This quantization simplifies estimation but introduces grid mismatch errors [20], particularly in practical scenarios where the actual Doppler shifts do not align perfectly with the assumed grid. Such mismatches degrade estimation accuracy and necessitate advanced approaches capable of handling off-grid effects.

To overcome the limitations of strict grid-based models, off-grid channel estimation techniques have been introduced. These methods relax the assumption of discretized

Doppler shifts, enabling more accurate modeling of real-world wireless channels. Grid refinement techniques have been explored to mitigate the performance degradation caused by grid mismatches. [21] presents a gridless approach for channel estimation, it makes use of off-grid super-resolution techniques to effectively identify and reconstruct doubly dispersive channels. Similar challenges arise in other fields, such as in microwave imaging [22]. In a similar vein, other papers introduce the joint recovery of signals with a continuous double domain formulation [23, 24]. The method in [23], based on concatenated atomic norm minimization, is particularly applicable to delay-Doppler sparsity in systems such as radar and communication networks, where signals are sparse in both the delay and Doppler dimensions. The proposed approach significantly improves the recovery of signals in the presence of off-the-grid frequency components, offering advantages over traditional separate recovery methods. [25] highlights how delay and Doppler sparsity are critical in radar imaging, particularly for rotating targets with off-grid scatterers. The authors in [25] propose parameter-refined orthogonal matching pursuit (PROMP), which enhances grid refinement using a nonlinear least-squares optimization, thus improving scatterer position and reflectivity estimation. Off-grid sparse Bayesian learning (SBL) [26] is another solution to handle sparsity in off-grid delay-Doppler models. This method improves the accuracy of channel estimation by partially mitigating the discretization errors associated with grid-based methods. However, the disadvantage of this approach lies in its computational complexity, as it involves solving optimization problems iteratively, which can be resource-intensive and may limit its applicability in real-time or large-scale systems. Grid refinement methods improve frequency estimation by increasing grid resolution, but they still face limitations. Despite finer grids, they cannot fully eliminate basis mismatch between real-world signals and the grid, leading to residual errors in cases with off-the-grid or continuous frequency components. This is particularly the case, as is shown in Chapter 3, when the propagation link is characterized by a large number of "physical" channel paths contribute to each of the (refined) grid

points. A scenario that matches most of the real-world propagation environments.

Basis expansion models (BEM) can in principle offer an alternative to grid refinement. Based on the BEM approach in [27, 28], the off-grid channel model demonstrates sparsity in the delay domain, but this sparsity does not extend to the Doppler domain. The approach in [27, 29] employs one BEM for each delay tap to capture channel variations, without making any assumptions about sparsity in the Doppler domain. Similar to this, [30, 31] introduces channel estimation scheme for OFDM systems over a doubly selective channel leveraging compressive sensing and BEM and exploiting only delay domain sparsity. These approaches fails to exploit both delay and Doppler domain sparsity simultaneously, leading to inefficiencies in modeling complex, dynamic channels. Resorting to BEM schemes aligns with the work in [32], which explores off-grid channel estimation for OTFS systems with fractional delay and Doppler shifts. While the paper addresses channel estimation in the delay-Doppler domain, it does not fully exploit DD sparsity to achieve pilot overhead reduction.

As for the issue of waveform performance comparison under sparsity assumptions, the work in [9] compares OFDM and OTFS under delay-Doppler sparsity in terms of the *pragmatic capacity* i.e., the mutual information of the virtual channel having at its input the constellation symbols *excluding the pilot and guard symbols* and at its output the detector soft-outputs. While this overhead-aware comparison constitutes a step forward, the restrictive sparsity model does not allow to do the comparison under realistic propagation conditions nor to devise pilot patterns with adjustable time and frequency densities for different delay-Doppler sparsity levels. Nonetheless, such works point towards the fact that some waveforms are more suited to take advantage of delay-Doppler sparsity than others. For instance, channel estimation overhead in OTFS cannot be significantly reduced when the channel exhibits more sparsity unless non-orthogonal pilot-data multiplexing is used as in [8]. For such a scheme, sparsity in the channel delay-Doppler response lessens inter-pilot and pilot-data interference. However, the use

of iterative detection methods becomes necessary which not only makes this approach require high computational complexity but also makes it prone to error propagation. We thus restrict our work in this thesis to the case of orthogonal resources for pilot and data symbols.

1.4 Contribution

This thesis begins with a comprehensive overview in Chapter 1, establishing the foundational knowledge necessary to understand the research contributions presented in this manuscript. Following this, the key contributions of this work are outlined, with each subsequent chapter expanding upon these advancements.

1. **Chapter 2: Estimation of doubly sparse linear time-varying channels with the on-grid approximation.** This chapter introduces the concept of DS-LTV channels, focusing on the on-grid modeling of delay-Doppler sparsity. The on-grid approach assumes strict alignment with a predefined grid, simplifying analysis and practical implementation. Three distinct delay-Doppler sparsity models—Type-1, Type-2, and Type-3—are presented, capturing key characteristics of high-mobility and high-frequency wireless communication environments. Furthermore, this chapter derives statistical properties of the minimal pilot overhead required for reliable channel estimation. Using these statistical findings, closed-form asymptotic results are obtained for the average minimal pilot overhead and the mean squared error (MSE) of four different waveforms. The results analytically demonstrate the superiority of Affine Frequency Division Multiplexing (AFDM) in minimizing pilot overhead while maintaining accurate channel estimation. Numerical simulations validate these theoretical findings, confirming their relevance for practical transmission settings.

2. **Chapter 3: Estimation and extrapolation of doubly sparse linear time-varying channels with off-grid Doppler shifts.** This chapter extends the analysis to off-grid

models, where Doppler shifts do not align perfectly with a predefined grid. A novel off-grid approximation method is proposed, leveraging basis expansion models (BEMs) to enhance channel estimation accuracy. The impact of off-grid Doppler shifts on estimation performance is studied, and techniques for extrapolating sparse delay-Doppler channels are introduced. The findings demonstrate significant improvements in estimation robustness and predictive capabilities, which are critical for high-mobility communication scenarios.

3. **Chapter 4: Further applications.** This chapter explores the broader applications of delay-Doppler sparsity beyond wireless communications. It examines its potential benefits in radar and sensing applications, particularly in sub-Nyquist radar systems, where it enables reduced sampling rates and hardware complexity while enhancing target detection and estimation.

The relation between the chapters and the publications is as identified in Table 1.1.

Table 1.1: Publications and their contributions to the chapters.

Type	Reference	Ch. 2	Ch. 3	Ch. 4
C	W. Benzine, A. Bemani, N. Ksairi and D. Slock, "Affine Frequency Division Multiplexing For Communications on Sparse Time-Varying Channels,"*GLOBECOM 2023 - 2023 IEEE Global Communications Conference*, Kuala Lumpur, Malaysia, 2023, pp. 4921-4926	✓		
C	W. Benzine, A. Bemani, N. Ksairi and D. Slock, "Affine Frequency Division Multiplexing for Compressed Sensing of Time-Varying Channels," *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Lucca, Italy, 2024, pp. 916-920	✓		✓
J	W. Benzine, A. Bemani, N. Ksairi and D. Slock, "Models, Methods and Waveforms for Estimation and Prediction of Delay-Doppler Sparse Time-Varying Channels," *IEEE transactions on wireless communications (2025)*	✓	✓	

C = Conference, J = Journal

Chapter 2

Estimation of Doubly Sparse Linear Time-Varying Channels with the On-grid Approximation.

This chapter begins by discussing on-grid delay-Doppler double sparsity, where channel components align with a predefined grid, simplifying DS-LTV channel modeling. The chapter then explores estimation techniques for known delay-Doppler profiles, comparing SCM, OFDM, and AFDM in terms of minimal pilot overhead required by different waveforms to achieve identifiability or a target error performance. Next, it addresses unknown delay-Doppler profile (DDP) using compressed sensing, introducing hierarchical sparsity and recovery algorithm HiHTP. Finally, numerical results highlight AFDM's efficiency in minimizing pilot overhead while maintaining estimation accuracy, demonstrating its advantages over other waveforms in DS-LTV channel estimation.

2.1 On-grid doubly sparse linear time-varying channels

In a first approximation, we assume that both ι_p and κ_p are zero and we define $L \triangleq \max_{p=1\cdots N_\mathrm{p}} \frac{\tau_p}{T_\mathrm{s}} + 1$. Then the discrete-time LTV input-output model in (1.5) becomes

$$r_n = \sum_{l=0}^{L-1} s_{n-l} h_{l,n} + z_n, \quad n \in \mathbb{Z}\,. \tag{2.1}$$

The input-output relation in (2.1) defines an on-grid LTV channel with a $L-1$ maximum delay shift with the complex gain $h_{l,n}$ of the l-th path varying with time index n as

$$h_{l,n} = \sum_{p=1}^{N_\mathrm{p}} g_p e^{\imath 2\pi \frac{qn}{N}} \delta_{l-l_p}, l = 0 \cdots L-1\,. \tag{2.2}$$

Under this assumption, we define g_p as:

$$g_p = \alpha_{l_p,q_p} I_{l_p,q_p}\,, \tag{2.3}$$

where $I_{l,q}$ is given by:

$$I_{l,q} = \begin{cases} 1 & \text{if } \exists p \text{ such that } (l,q) = (l_p, q_p), \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The number of paths N_p, as defined in equation (1.2), can now be expressed as:

$$N_\mathrm{p} = \sum_{l=0}^{L-1} \sum_{q=-Q}^{Q} I_{l,q} \quad \text{(total \# paths in the on-grid model)} \tag{2.5}$$

Here, $I_{l,q}$ for any l and q is a binary random variable that, when non-zero, indicates that a channel path with delay l, Doppler shift q and complex gain $\alpha_{l,q}$ is active and contributes to the channel output. Note that the distribution of the random variables

$\{I_{l,q}\}_{l,q}$ controls the kind of sparsity the LTV channel might have. The complex gain is assumed to satisfy $\alpha_{l,q} \sim \mathcal{CN}\left(0, \sigma_\alpha^2\right)$ with σ_α^2 satisfying the channel power normalization

$$\sum_{l=0}^{L-1} \sum_{q=-Q}^{Q} \mathbb{E}\left[|\alpha_{l,q}|^2 I_{l,q}\right] = 1\,. \tag{2.6}$$

Definition 2.1.1 (On-grid Delay-Doppler double sparsity, [33])**.** The complex gain $h_{l,n}$ of the l-th path of the LTV channel varies with time as

$$h_{l,n} = \sum_{q=-Q}^{Q} \alpha_{l,q} I_{l,q} e^{\imath 2\pi \frac{nq}{N}}, \quad l = 0, \ldots, L-1, \tag{2.7}$$

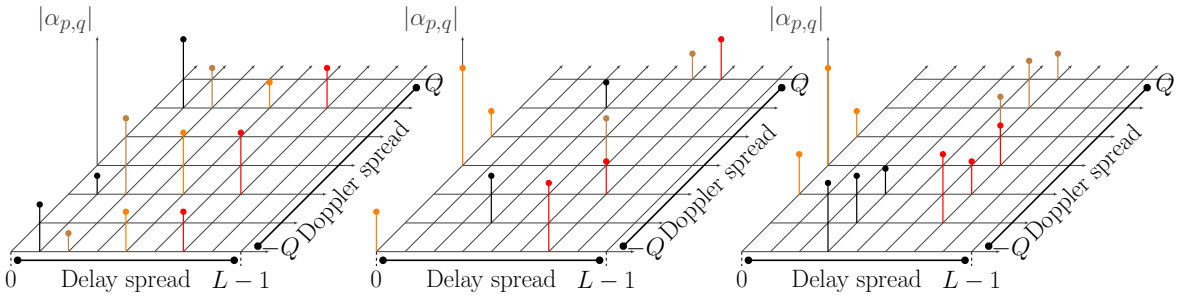
and there exist $0 < p_d, p_D < 1$ such that

$$I_{l,q} = I_l I_q^{(l)}, \quad \forall (l,q) \in [\![0..L-1]\!] \times [\![-Q..Q]\!], \tag{2.8}$$

where $I_l \sim \text{Bernoulli}(p_d)$ and $I_q^{(l)} \sim \text{Bernoulli}(p_D)$. Moreover, $I_{l,q}$ and $\alpha_{l,q}$ are independent.

Fig. 2.1 illustrates three different delay-Doppler sparsity models, fully described in [33] and dubbed Type-1, Type-2 and Type-3, that all fall under the scope of Definition 2.1.1 each with an additional different assumption on I_l and $I_q^{(l)}$. Here, we just point out that the difference between Type-2 and Type-3 of Figures 2.1b and 2.1c, respectively, is that in the latter the active Doppler bins per delay tap appear in clusters of random positions but of deterministic length as opposed to the absence of clusters in the former. The case where the delay taps have all the same (random) sparsity (as in Type-1 models of Fig. 2.1a) also falls under Definition 2.1.1 by setting $I_q^{(l)} = I_q^{(0)}, \forall l$.

The above models are not exhaustive. For instance, block sparsity can be extended to the delay domain. Furthermore, each model can be extended by removing the on-grid approximation. In that case, $I_{l,q}$ will only represent the closest grid point in the delay-Doppler domain to a channel path instead of representing the path itself. Nonetheless,



(a) Type-1 sparsity (b) Type-2 sparsity (c) Type-3 sparsity

Figure 2.1: Examples of channels satisfying the 3 types of delay-Doppler sparsity

the three models capture important features of wireless channels in high-frequency bands that are subject to user mobility.

Note that under Definition 2.1.1, $s_{\mathrm{d}} \triangleq \mathbb{E}\left[\sum_l I_l\right] = p_d L$ is the mean number of active delay taps in the delay-Doppler profile of the channel and can be thought of as the *delay domain sparsity level* while $s_{\mathrm{D}} \triangleq \mathbb{E}\left[\sum_q I_q^{(l)}\right] = p_D(2Q+1)$ is the mean number of active Doppler bins per delay tap and can be thought of as the *Doppler domain sparsity level*.

To ensure sparsity in a stronger sense i.e., with high probability (as $L, Q, Lp_{\mathrm{d}}, (2Q+1)p_{\mathrm{D}}$ grow), we require that the following assumption hold.

Assumption 2.1.1. *$\{I_l\}_{l=0\cdots L-1}$ are mutually independent. Also, random variables $\left\{I_{q_t}^{(l_t)}\right\}_{t=1\cdots T}$ are mutually independent for any $(l_1, q_1), \ldots, (l_T, q_T)$ ($T \leq \min(l, 2Q+1)$) such that $l_s \neq l_t$ and $q_s \neq q_t$ whenever $s \neq t$. Moreover, the complementary cumulative distribution function (CCDF) $\overline{F}_{S_{\mathrm{D},l}}(m)$ of the random variable $S_{\mathrm{D},l} \triangleq \sum_{q=-Q}^{Q} I_q^{(l)}$ for any $l \in [\![0..L-1]\!]$ is upper-bounded for any integer $m > (2Q+1)p_{\mathrm{D}}$ by the CCDF of $\mathrm{B}(2Q+1, p_{\mathrm{D}})$.*

The CCDF upper bound condition in Assumption 2.1.1 is not arbitrary. Indeed, Type-1 and 2 models easily satisfy it simply by requiring that $\{I_q^{(0)}\}_q$ in the first and $\{I_q^{(l)}\}_q$ for any l in the second to be mutually independent (and to thus satisfy $\overline{F}_{S_{\mathrm{D},l}}(m) = \overline{F}_{\mathrm{B}(2Q+1,p_{\mathrm{D}})}(m), \forall m$). For Type-3 models, $S_{\mathrm{D},l}$ is deterministic and hence its CCDF is trivially upper-bounded. As the following lemma rigorously shows, the mutual independence of $\{I_l\}_{l=0\cdots L-1}$ in Assumption 2.1.1 guarantees strong delay domain

sparsity while Doppler sparsity is guaranteed in a more explicit manner by the CCDF upper bound.

2.2 Waveform definitions

In this section, we define the waveforms used for estimation in subsequent sections. For that sake, let $\mathbf{s}$ be the N-long vector of samples s_n at the input of the channel. Define $\mathbf{x}$ as the vector of data symbols and embedded channel estimation pilots, which modulate the waveform in use to produce the time-domain samples vector $\mathbf{s}$. For all the considered waveforms, we can write the dependence of $\mathbf{s}$ on $\mathbf{x}$ using a modulation matrix $\mathbf{\Phi_{tx}}$

$$\mathbf{s} = \mathbf{\Phi}_{\mathrm{tx}}\mathbf{x}\,. \tag{2.9}$$

For single-carrier modulation (SCM), $\mathbf{\Phi}_{\mathrm{tx}} = \mathbf{I}_N$. For AFDM, $\mathbf{\Phi}_{\mathrm{tx}} = \mathbf{\Lambda}_{c2}\mathbf{F}_N\mathbf{\Lambda}_{c1}$, with $\mathbf{\Lambda}_c = \mathrm{diag}(e^{-\imath 2\pi c n^2}, n = 0, \ldots, N-1)$ and $\mathbf{F}_N$ being the N-order discrete Fourier transform (DFT) matrix. For an OFDM grid as the one shown in Figure 2.3 composed of N_{s} symbols each having N_{fft} sub-carriers and a cyclic prefix of length N_{cp}, $\mathbf{\Phi}_{\mathrm{tx}} = \mathrm{blkdiag}\left(\mathbf{T}_{\mathrm{cp}}\mathbf{F}^{\mathrm{H}}_{N_{\mathrm{fft}}}, \ldots, \mathbf{T}_{\mathrm{cp}}\mathbf{F}^{\mathrm{H}}_{N_{\mathrm{fft}}}\right)$ with $\mathbf{T}_{\mathrm{cp}} = \begin{bmatrix} \mathbf{0}, \mathbf{I}_{N_{\mathrm{cp}}} \\ \mathbf{I}_{N_{\mathrm{fft}}, N_{\mathrm{fft}}} \end{bmatrix}$ being the matrix for cyclic prefix insertion. For OTFS, $\mathbf{\Phi}_{\mathrm{tx}} = \mathbf{F}^{\mathrm{H}}_{M_{\mathrm{otfs}}} \otimes \mathbf{I}_{N_{\mathrm{otfs}}}$.

At the receiver, let $\mathbf{r}$ be the N-long vector of the channel output samples r_n defined in (2.1). Depending on the waveform employed, the receiver applies a demodulation matrix $\mathbf{\Phi}_{\mathrm{rx}}$ to produce the vector of transform domain samples $\mathbf{y}$. For both SCM and AFDM, $\mathbf{\Phi}_{\mathrm{rx}} = \mathbf{\Phi}^{\mathrm{H}}_{\mathrm{tx}}$. For the OFDM grid of Figure 2.3, $\mathbf{\Phi}_{\mathrm{rx}} = \mathrm{blkdiag}\left(\mathbf{F}_{N_{\mathrm{fft}}}\mathbf{R}_{\mathrm{cp}}, \ldots, \mathbf{F}_{N_{\mathrm{fft}}}\mathbf{R}_{\mathrm{cp}}\right)$ and $\mathbf{R}_{\mathrm{cp}} = [\mathbf{0}, \mathbf{I}_{N_{\mathrm{fft}}}]$ is the matrix for cyclic prefix removal. For OTFS, $\mathbf{\Phi}_{\mathrm{rx}} = \mathbf{\Phi}^{\mathrm{H}}_{\mathrm{tx}}$. Let $\mathcal{P} \subset [\![0..N-1]\!]$ designate the indexes of the received samples associated with N_{p} transform domain pilots, of values $\{\mathrm{p}_p\}_{p=1\cdots N_{\mathrm{p}}}$ inserted at indexes $\{m_p\}_{p=1\cdots N_{\mathrm{p}}}$ within the vector $\mathbf{x}$ and each surrounded with a waveform-dependent number of zero guard

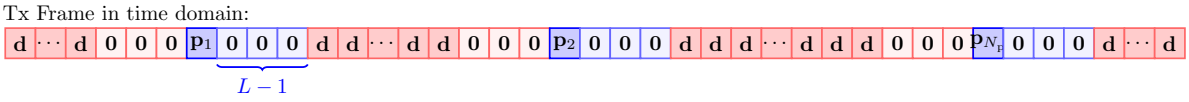


Figure 2.2: An example of a SCM frame composed of data samples and pilot symbols, each of the latter surrounded by $2L-1$ guard samples.

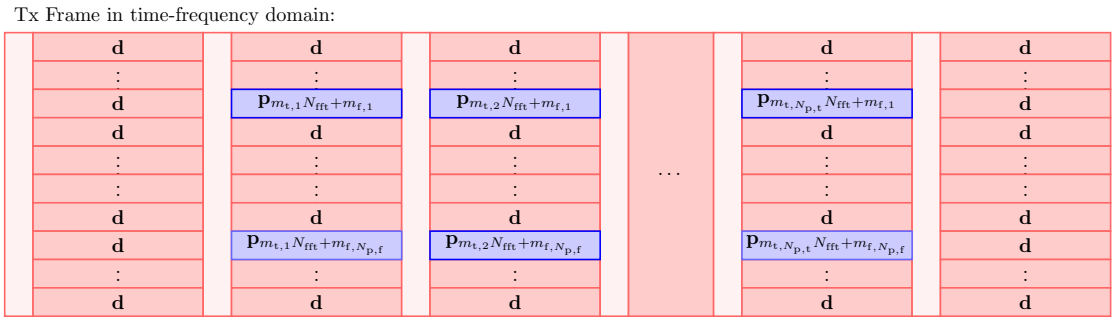


Figure 2.3: OFDM frame with pilot (blue), guard (light blue and red) and data (red) subcarriers. Each symbol is preceded by $L-1$ CP samples (light red)

samples (see Figures 2.2, 2.3 and 2.5).

For SCM,

$$\mathcal{P} = \bigcup_{p=1}^{N_\mathrm{p}} [\![m_p .. m_p + L - 1]\!]. \quad \text{(SCW)} \tag{2.10}$$

As for OFDM, let $m_{p_\mathrm{t}} \in [\![0..N_\mathrm{s}-1]\!]$ be the time domain position of the p-th pilot and $m_{p_\mathrm{f}} \in [\![0..N_\mathrm{fft}-1]\!]$ be its frequency domain position. Then

$$\mathcal{P} = \{ m_p = m_{\mathrm{t},p_\mathrm{t}} N_\mathrm{fft} + m_{\mathrm{f},,p_\mathrm{f}} \}_{\substack{p_\mathrm{t}=1\cdots N_{\mathrm{p,t}}, \\ p_\mathrm{f}=1\cdots N_{\mathrm{p,f}}}}, \quad N_\mathrm{p} = N_{\mathrm{p,t}} N_{\mathrm{p,f}}, \quad \text{(OFDM)} \tag{2.11}$$

For OTFS, the set $\mathcal{P}$ is the vectorized form of the Frame shown in Figure 2.4.

While for AFDM, it holds [34] that

$$\begin{aligned} y_k &= \sum_{l=0}^{L-1} \sum_{q=-Q}^{Q} \alpha_{l,q} I_{l,q} e^{\imath 2\pi (c_1 l^2 - \frac{ml}{N} + c_2(m^2-k^2))} x_k + w_k \,, \\ m &\triangleq (k - q + 2Nc_1 l)_N \,. \end{aligned} \tag{2.12}$$

The samples related to the p-th pilot symbol thus occupy $2N|c_1|(L-1)+2Q+1$

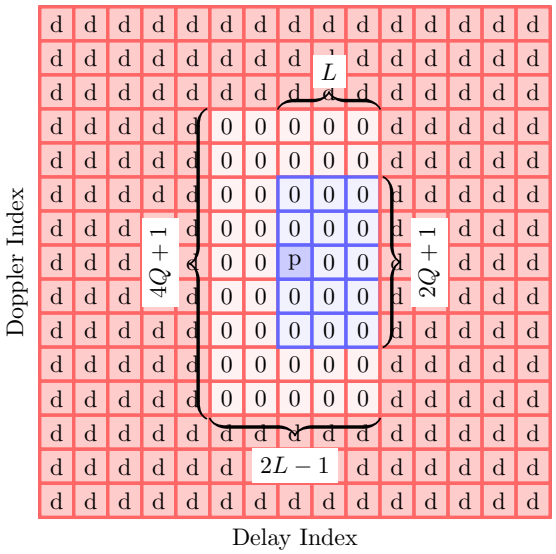


Figure 2.4: An OTFS symbol composed in the Zak domain of data samples (red), a pilot sample (blue) and guard samples (light blue and red)

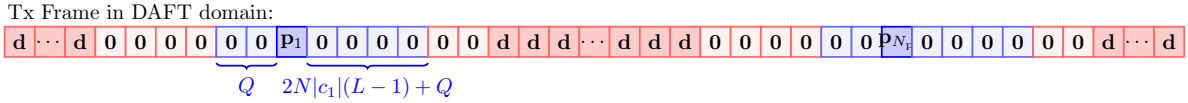


Figure 2.5: AFDM symbol composed of data samples, N_{p} pilot symbols and their guard samples.

DAFT domain indexes. More precisely, and in case c_1 is negative,

$$\mathcal{P} = \bigcup_{p=1}^{N_\mathrm{p}} [\![m_p - Q \,..\, m_p + 2N|c_1|(L-1) + Q]\!]. \quad \text{(AFDM)} \tag{2.13}$$

2.3 Channel estimation with known delay-Doppler profile

This section focuses on on-grid channel estimation in time-varying communication systems, where Doppler shifts align with the predefined discrete grid. In this scenario, the channel is modeled under the assumption that Doppler shifts are exactly represented by the grid points, which simplifies the estimation process. The section presents techniques for estimating channel coefficients by leveraging pilot symbols and applying minimum mean squared error (MMSE) estimation considering that the knowledge of the delay-Doppler profile (DDP) i.e., of $\{I_{l,q}\}_{l=0\cdots L-1, q=-Q\cdots Q}$, at the receiver side is assumed. This bears similarities with the knowledge of the power delay profile (PDP) for linear time-invariant (LTI) channel estimation [35]. The performance of on-grid channel estimation is also analyzed, with numerical results demonstrating its effectiveness in achieving accurate and reliable estimates, while maintaining lower computational complexity due to the grid alignment.

2.3.1 Estimation of DS-LTV channels with known DDP using different waveforms

We now turn our attention to the estimation of the DS-LTV channel using practical waveforms and embedded pilots.

Let $\boldsymbol{\alpha}_\mathrm{ddp} \triangleq [\alpha_{l,q}]_{(l,q)\mathrm{s.t.}I_{l,q}=1}$ designate the vectorized form of the unknown channel gains associated with active delay-Doppler components such that $\boldsymbol{\alpha} = \mathbf{A}_\alpha \boldsymbol{\alpha}_\mathrm{ddp}$ ($\boldsymbol{\alpha}_\mathrm{ddp}$ being the restriction of the vector $\boldsymbol{\alpha}$ to the DDP support) . For any of the above waveforms, let $\mathbf{y}_\mathrm{p} \triangleq [y_k]_{k\in\mathcal{P}}$ be the vector of received pilot samples. Inserting (2.7) and

(2.9) into (2.1) gives the following signal model for recovery the vector $\boldsymbol{\alpha}_{\mathrm{ddp}}$

$$\mathbf{y}_{\mathrm{p}} = \underbrace{\mathbf{A}_{\mathcal{P}}\mathbf{M}\mathbf{A}_{\alpha}}_{\triangleq \mathbf{M}_{\mathrm{p}}} \boldsymbol{\alpha}_{\mathrm{ddp}} + \mathbf{w}_{\mathrm{p}} \tag{2.14}$$

where $\mathbf{x}_{\mathrm{p}}$ is a vector of the same length as $\mathbf{x}$ with entries equal to $\mathrm{p}_1, \ldots, \mathrm{p}_{N_{\mathrm{p}}}$ at indexes $\{m_p\}_{p=1\cdots N_{\mathrm{p}}}$ and to zero elsewhere, $\mathbf{w}_{\mathrm{p}}$ are the corresponding noise samples and the matrix $\mathbf{M}$ is a $N \times L(2Q+1)$ partitioned matrix expressed as

$$\mathbf{M} = \Big[[\boldsymbol{\gamma}_{0,-Q} \quad \cdots \quad \boldsymbol{\gamma}_{0,Q}] \quad \cdots \quad [\boldsymbol{\gamma}_{L-1,-Q} \quad \cdots \quad \boldsymbol{\gamma}_{L-1,Q}] \Big] \tag{2.15}$$

with the columns of $\mathbf{M}$ being given as

$$\boldsymbol{\gamma}_{l,q} = \sum_{i=1}^{M} p_i \left[\boldsymbol{\Phi}_{\mathrm{rx}} \boldsymbol{\Delta}_q \boldsymbol{\Pi}^l \boldsymbol{\Phi}_{\mathrm{tx}}^{\mathrm{H}} \right]_{k_{p_i}} . \tag{2.16}$$

Here, $\mathbf{A}_{\mathcal{P}}$ is the $|\mathcal{P}| \times N$ matrix that chooses from the transform domain received vector the entries corresponding to $\mathcal{P}$. $\mathbf{A}_{\alpha}$ is the matrix that augments $\boldsymbol{\alpha}_{\mathrm{ddp}}$ with zeros corresponding to $I_{l,q} = 0$ resulting in a $L(2Q+1)$-long vector $\mathbf{A}_{\alpha}\boldsymbol{\alpha}_{\mathrm{ddp}}$. $\boldsymbol{\Delta}_q = \mathrm{diag}(e^{\imath 2\pi qn}, n = 0 \cdots N-1)$, $\boldsymbol{\Pi}$ is the N-order permutation matrix.

The minimum mean squared error (MMSE) estimate, $\hat{\boldsymbol{\alpha}}_{\mathrm{ddp}}$, of $\boldsymbol{\alpha}_{\mathrm{ddp}}$ based on $\mathbf{y}_{\mathrm{p}}$ is given by [36]

$$\hat{\boldsymbol{\alpha}}_{\mathrm{ddp}} = \sigma_{\alpha}^2 (\sigma_{\alpha}^2 \mathbf{M}_{\mathrm{p}}^{\mathrm{H}} \mathbf{M}_{\mathrm{p}} + \sigma_w^2 \mathbf{I})^{-1} \mathbf{M}_{\mathrm{p}}^{\mathrm{H}} \mathbf{y}_{\mathrm{p}} . \tag{2.17}$$

For any (l,q) satisfying $I_{l,q} = 1$, define $\hat{\alpha}_{l,q}$ as the corresponding entry of vector $\hat{\boldsymbol{\alpha}}_{\mathrm{ddp}}$. For any $(l,q) \in [\![0..L-1]\!] \times [\![-Q..Q]\!]$ such that $I_{l,q} = 0$, set $\hat{\alpha}_{l,q} = 0$. Finally, define

$$\hat{h}_{l,n} \triangleq \textstyle\sum_{q=-Q}^{Q} \hat{\alpha}_{l,q} I_{l,q} e^{\imath 2\pi \frac{nq}{N}}, \quad n = 0, \ldots, N-1 . \tag{2.18}$$

as the resulting MMSE estimate of $h_{l,n}$. In what follows, we give indications on how to set N_{p} and c_1 based on the delay-Doppler sparsity level of the channel so that the

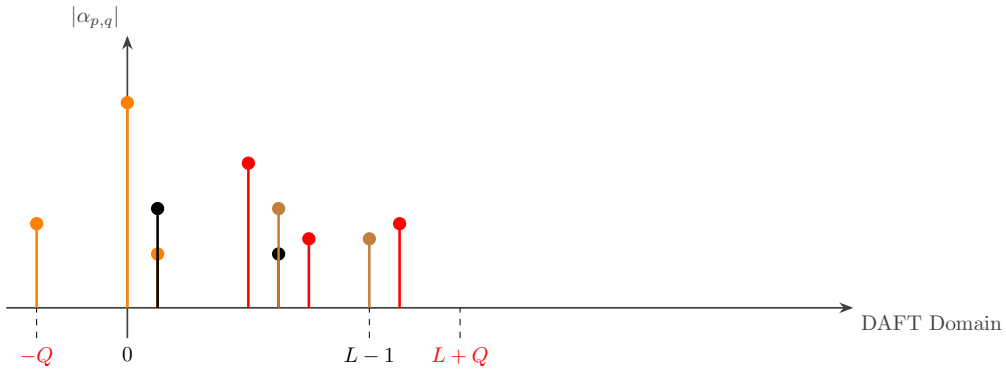
minimal pilot overhead needed to guarantee a vanishing (with respect to an increasing signal-to-noise ratio (SNR)) mean squared error (MSE) $\mathbb{E}[\sum_{l=0}^{L-1}\frac{1}{N}\sum_{n=0}^{N-1}|h_{l,n}-\hat{h}_{l,n}|^2] = \mathbb{E}[\|\hat{\boldsymbol{\alpha}}_{\mathrm{ddp}}-\boldsymbol{\alpha}_{\mathrm{ddp}}\|^2]$.

2.3.2 AFDM parameters setting for transmission over DS-LTV channels

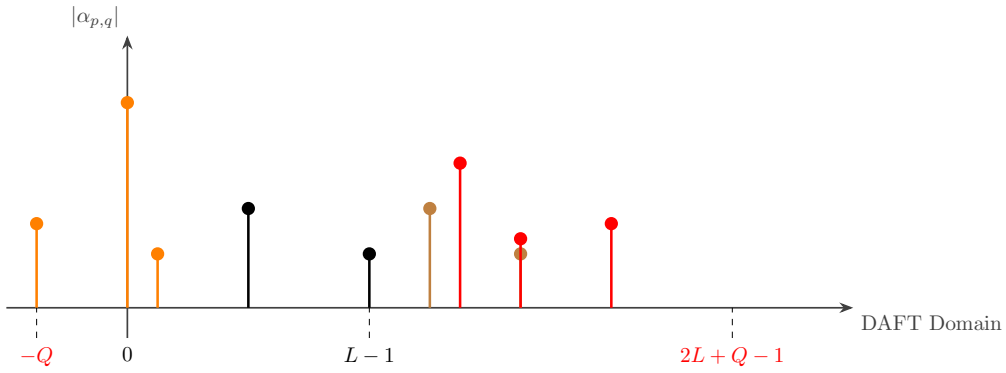
In the case of AFDM, thanks to these zero guard samples, the amplitude of the pilot symbols can be boosted as follows without violating the (time domain) transmit power constraint:

$$|p_i| = \sqrt{2|c_1|N(L-1)+2Q+1}, \quad i = 1\cdots M\,. \tag{2.19}$$

Let $u_m^{(l,q)}$ ($m\in\mathbb{Z}$) be the individual DAFT domain impulse response of the part of the channel associated with delay-Doppler component $\alpha_{l,q}$. Since $[\![0..L-1]\!]\times[\![-Q..Q]\!]$ in the delay-Doppler domain maps to an interval in the DAFT domain that is either $[\![-Q..2|c_1|N(L-1)+Q]\!]$ if c_1 is negative or $[\![-Q-2|c_1|N(L-1)..Q]\!]$ if c_1 is positive, the latter interval is the support of $u_m^{(l,q)}$. We designate by *DAFT domain representation of the channel* the collection $\{u_m^{(l,q)}\}_{(l,q)\in[\![0..L-1]\!]\times[\![-Q..Q]\!]}$ of all individual DAFT domain impulse responses. Figure 2.6a shows the DAFT domain representation of a channel in the case $c_1=\frac{-1}{2N}$ while Figure 2.6b shows that representation when $c_1=\frac{-2}{2N}$. In what follows we restrict c_1 to be negative without loss of generality. Define the random variable $X_k \triangleq |\{(l,q)\in[\![0..L-1]\!]\times[\![-Q..Q]\!], I_{l,q}=1, q-2c_1Nl=k\}|$ for any $k\in[\![-Q..2|c_1|N(L-1)+Q+1]\!]$, i.e., X_k is the number of non-zero components $\alpha_{l,q}$ appearing at index k in the DAFT domain representation. It is also the number of terms in the mixture of complex sinusoids that constitute the sample y_k in (2.12) and is thus closely related to pilot overhead and channel estimation performance. For instance, under a given channel realization, the minimal number of DAFT domain pilots needed for full identifiability i.e., for the measurement matrix $\mathbf{M}_{\mathrm{p}}$ in (2.22) to have full column rank, should be at least equal to $\max_{k\in[\![-Q..2|c_1|N(L-1)+Q+1]\!]} X_k$. We will also show that the distribution of X_k affects directly the MSE of $\hat{\boldsymbol{\alpha}}_{\mathrm{ddp}}$. This is why we examine in what follows that probability



(a) $c_1 = -1/2N$



(b) $c_1 = -2/2N$

Figure 2.6: DAFT domain representation of the channel realization of Figure 2.1b for different values of c_1

distribution. First, by referring to the signal relation in (2.12) we can see that with "enough" sparsity i.e., if the number of nonzero channel components is sufficiently smaller than the support $[\![-Q..2|c_1|N(L-1)+Q+1]\!]$ of the channel DAFT domain representation, it is unlikely that X_k takes large values and hence it is unlikely that a large number of DAFT domain pilots would be needed to get a target estimation error performance. This can be seen from Figure 2.6 where $\max_k X_k = 3$ when $|c_1| = \frac{1}{2N}$ and $\max_k X_k = 1$ when $|c_1| = \frac{2}{2N}$. The following lemma and the ensuing theorem give a rigorous confirmation of the above intuition.

Lemma 2.3.1. *For $c_1 = -\frac{P}{2N}$ ($P \in \mathbb{N}^*$) and any $k \in [\![-Q..2|c_1|N(L-1)+Q]\!]$ the complementary cumulative distribution function (CCDF) of X_k under Assumption 2.1.1*

is upper bounded by the CCDF of $\mathrm{B}((2\lceil\frac{Q}{P}\rceil+1, p_{\mathrm{d}}p_{\mathrm{D}})$.

Proof. The proof of Lemma 2.3.1 is provided in Appendix A.1. □

For tractability and more insights, the asymptotic regime for N, L, Q defined by the following assumption will be helpful. Note that the numerical results given in Section 2.3.3 are not asymptotic but are obtained with finite values of N, L, Q.

Assumption 2.3.1. *$L = O(K)$, $Q = O(K)$, $s_{\mathrm{d}} = O\left(K^{\kappa_{\mathrm{d}}}\right)$ and $s_{\mathrm{D}} = O\left(K^{\kappa_{\mathrm{D}}}\right)$ for some $\kappa_{\mathrm{d}}, \kappa_{\mathrm{D}} \in [0, 1)$.*

Remark 1. *Assuming $L = O(K)$ and $Q = O(K)$ implies that $N = O\left(K^2\right)$. Indeed, assuming that the maximum delay L increases to infinity as K implies that the transmission bandwidth increases at the same rate. Also, assuming that the maximum Doppler shift Q increases as K implies that the transmission duration increases at the same rate. Therefore, the frame size in samples i.e., N, increases as K^2.*

Theorem 2.3.1. *Under Assumption 2.3.1 and Assumption 2.1.1, if we set P s.t. $(L-1)P+2Q+1 = O(s_{\mathrm{d}}s_{\mathrm{D}})$ then DAFT domain pilot overhead needed for the MSE $\mathbb{E}[\|\hat{\boldsymbol{\alpha}}_{\mathrm{ddp}}-\boldsymbol{\alpha}_{\mathrm{ddp}}\|^2]$ to tend to zero as $K \to \infty$ and $\sigma_w^2 \to 0$ is $O\left(K^{\kappa_{\mathrm{d}}+\kappa_{\mathrm{D}}} \log K\right)$.*

Proof. A sketch of the proof is given in Appendix A.2. □

Remark 2. *Theorem 2.3.1 implies that AFDM is order-optimal in terms of channel estimation overhead for DS-LTV channels since the total overhead needed for vanishing channel estimation MSE has the same asymptotic order (up to a logarithmic term) as the smallest possible overhead which is equal to the average number of unknowns and thus to*

$$\mathbb{E}\left[|\{(l,q) \in [\![0..L-1]\!] \times [\![-Q..Q]\!] \text{s.t.} I_{l,q} = 1\}|\right] = s_{\mathrm{d}}s_{\mathrm{D}} = O\left(K^{\kappa_{\mathrm{d}}+\kappa_{\mathrm{D}}}\right). \tag{2.20}$$

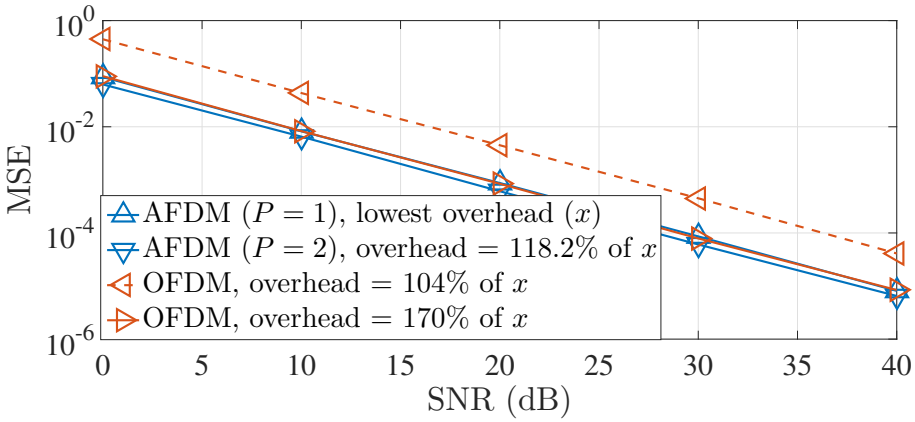
This optimality of AFDM is confirmed by the comparison done in the following section of its channel estimation overhead and channel estimation error performance to that required by OFDM, OTFS and SCM.

2.3.3 Numerical results

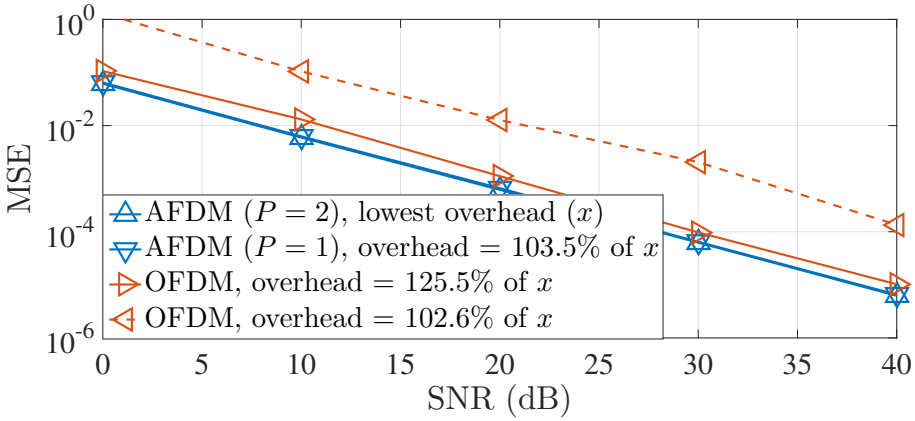
In this section, we compare the channel estimation performance of AFDM and OFDM for DS-LTV channels, focusing on both the MSE and pilot overhead. We present results that highlight the advantage of AFDM in terms of reduced overhead, particularly when channel sparsity is high, and further discuss its optimal performance in terms of diversity order and error performance compared to OFDM, SCM, and OTFS.

In Figure 2.7, the channel estimation MSE of AFDM and OFDM are compared. The simulations used 100 realizations of channels with a type-1 delay-Doppler sparsity, with parameters set to $N = 8192, L = 60, Q = 15$ (corresponding to a 12 MHz transmission at a 70 GHz carrier frequency, a relative moving speed of 340 km/h, and a delay spread of 5 μs) and $p_{\mathrm{d}} = 0.2$. In the solid lines, the number of AFDM pilots N_{p} and OFDM pilots $N_{\mathrm{p,t}}$ were chosen for each channel realization above $N_{\mathrm{p\,min}}$ and $N_{\mathrm{p,t\,min}}$, respectively, to achieve a 10^{-3} MSE at SNR = 20 dB. The dashed line represents the MSE of OFDM when $N_{\mathrm{p,t}}$ is reduced to match the pilot overhead of AFDM. As stated in Theorem 2.3.1, AFDM with $P = 1$ exhibits the lowest overhead, with $\mathbb{E}[N_{\mathrm{p}}] = 7$, as shown in Figure 2.7a where $p_{\mathrm{d}} = 0.2$. Similarly, AFDM with $P = 2$ and $\mathbb{E}[N_{\mathrm{p}}] = 7$ is observed in Figure 2.7b for $p_{\mathrm{d}} = 0.3$. This confirms the results stated in Remark 2, which implies that AFDM is order-optimal in terms of channel estimation overhead for DS-LTV channels.

In Figure 2.8, the average pilot overhead needed to achieve the target MSE is plotted for different values of p_{d} while $p_{\mathrm{D}} = 0.2$. As expected, the gain with respect to OFDM, SCM, and OTFS is the largest when sparsity is the highest. When there is no sparsity (p_{d} close to 1), performance measures other than pilot overhead can be used, such as diversity order or channel delay-Doppler components separability. AFDM has been shown to achieve the optimal diversity order of LTV channels [10] in the general case, irrespective of sparsity.



(a) $p_{\mathrm{D}} = 0.2$



(b) $p_{\mathrm{D}} = 0.3$

Figure 2.7: MSE performance for $N = 8192, L = 60, Q = 15, p = 0.2$.

2.4 Channel estimation with unknown delay-Doppler profile

In this section, the estimation of DS-LTV channels is analyzed under the assumption of an unknown delay-Doppler profile and an on-grid model. A hierarchical sparsity framework is used to apply compressed sensing techniques, with theoretical guarantees provided via the hierarchical restricted isometry property (HiRIP). Different waveforms, including OTFS, OFDM, and AFDM, are compared, demonstrating AFDM's superiority in reducing pilot overhead while maintaining estimation accuracy.

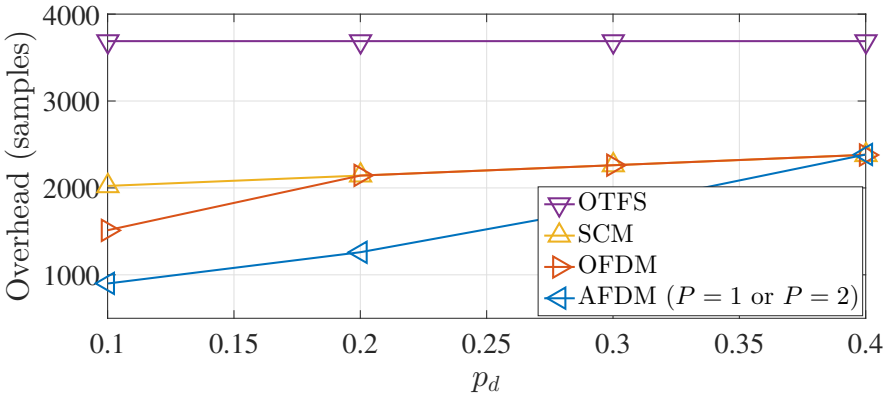


Figure 2.8: Channel estimation overhead for a target MSE $= 10^{-3}$ at SNR $= 20$ dB for $N = 8192, L = 60, Q = 15, p_{\mathrm{D}} = 0.2$

2.4.1 Relation to hierarchical sparsity

Definition 2.4.1 (Hierarchical sparsity, [37])**.** A vector $\mathbf{x} \in \mathbb{C}^{NM}$ is (s_N, s_M)-sparse if it consists of N blocks of size M, s_N of which at most are non-zero, and each non-zero block is s_M-sparse.

To analyze hierarchically sparse recovery schemes, a modified version of the restricted isometry property (RIP) called the hierarchical RIP (HiRIP) was proposed in the literature.

Definition 2.4.2 (HiRIP, [37])**.** The HiRIP constant δ_{s_N,s_M} of a matrix $\mathbf{A}$ is the smallest $\delta \geq 0$ such that for all (s_N, s_M)-sparse vectors $\mathbf{x} \in \mathbb{C}^{NM}$

$$(1-\delta)\left\|\mathbf{x}\right\|^2 \leq \left\|\mathbf{A}\mathbf{x}\right\|^2 \leq (1+\delta)\left\|\mathbf{x}\right\|^2 \tag{2.21}$$

DS-LTV sparsity is probabilistic while hierarchical sparsity of Definition 2.4.1 is deterministic. The two models are nonetheless related under Assumption 2.1.1.

Lemma 2.4.1. *With probability* $1 - e^{-\Omega(\min((2Q+1)p_{\mathrm{D}}, Lp_{\mathrm{d}}))}$, $\boldsymbol{\alpha}$ *is* $(s_{\mathrm{d}}, s_{\mathrm{D}})$*-sparse under Assumption 2.1.1.*

Proof. The proof of Lemma 2.4.1 is provided in Appendix A.3 □

Algorithm 2.4.1 HiHTP for compressive sensing of DS-LTV channels

Input: $\mathbf{M}_{\mathrm{p}}$, $\mathbf{y}_{\mathrm{p}}$, maximum number of iterations $k_{\max}$, s_{d}, s_{D}
$\hat{\boldsymbol{\alpha}}^{(0)} = 0$, $k = 0$
repeat
$\Omega^{(k+1)} = L_{s_{\mathrm{d}},s_{\mathrm{D}}}\left(\boldsymbol{\alpha}^{(k)} + \mathbf{M}_{\mathrm{p}}^{\mathrm{H}}\left(\mathbf{y}_{\mathrm{p}} - \mathbf{M}_{\mathrm{p}}\boldsymbol{\alpha}^{(k)}\right)\right)$
$\boldsymbol{\alpha}^{(k+1)} = \arg\min\left\{\left\|\mathbf{y}_{\mathrm{p}} - \mathbf{M}_{\mathrm{p}}\mathbf{z}\right\|, \sup\left(\mathbf{z}\right) \subset \Omega^{(k+1)}\right\}$
$k = k + 1$
until $k = k_{\max}$ or $\Omega^{(k+1)} = \Omega^{(k)}$ (whichever earlier)
Output: $(s_{\mathrm{d}}, s_{\mathrm{D}})$-sparse $\hat{\boldsymbol{\alpha}}^{(k)}$.

2.4.2 Compressed-sensing estimation of DS-LTV channels using different waveforms

For any of the above waveforms (see Figures 2.2, 2.3, 2.4, 2.5), let $\mathbf{y}_{\mathrm{p}} \triangleq [y_k]_{k\in\mathcal{P}}$ be the vector of received pilot samples. Inserting (2.7) and (2.9) into (2.1) gives the following signal model for recovery of the hierarchically sparse (per Lemma 2.4.1) vector $\boldsymbol{\alpha}$

$$\mathbf{y}_{\mathrm{p}} = \underbrace{\mathbf{A}_{\mathcal{P}}\mathbf{M}}_{\triangleq\mathbf{M}_{\mathrm{p}}}\boldsymbol{\alpha} + \mathbf{w}_{\mathrm{p}} \tag{2.22}$$

where $\mathbf{M}$ and $\mathbf{A}_{\mathcal{P}}$ are defined in (2.14). Hierarchical hard thresholding pursuit (HiHTP) [37] has been suggested in the literature for solving hierarchically-sparse recovery problems. When applied to Problem (2.22) it gives Algorithm 2.4.1.

HiHTP is a modification of the classical hard thresholding pursuit (HTP) [38] by replacing the thresholding operator employed at each iteration of HTP with a *hierarchically* sparse version $L_{s_{\mathrm{d}},s_{\mathrm{D}}}$. To compute $L_{s_{\mathrm{d}},s_{\mathrm{D}}}(\mathbf{x})$ for a vector $\mathbf{x} \in \mathbb{C}^{L(2Q+1)}$ first a s_{D}-sparse approximation is applied to each one of the L blocks of $\mathbf{x}$ by keeping in each of them the largest s_{D} entries while setting the remaining ones to zero. A s_{d}-sparse approximation is next applied to the result by identifying the s_{d} blocks with the largest l_2-norm. The following theorems give the conditions guaranteeing the convergence of Algorithm 2.4.1 and the recovery of $\boldsymbol{\alpha}$.

Theorem 2.4.1 (HiRIP for SCM and OFDM based measurements)**.** *Under Assump-*

tion 2.1.1 and for sufficiently large L, Q, sufficiently small δ_{t}, sufficiently small δ_{f}, $N_{\mathrm{p,t}} > O\left(\frac{1}{\delta_{\mathrm{t}}^2}\log^2\frac{1}{\delta_{\mathrm{t}}}\log\frac{s_{\mathrm{D}}}{\delta_{\mathrm{t}}}s_{\mathrm{D}}\log(2Q+1)\right)$ and $N_{\mathrm{p,f}} > O\left(\frac{1}{\delta_{\mathrm{f}}^2}\log^2\frac{1}{\delta_{\mathrm{f}}}\log\frac{s_{\mathrm{d}}}{\delta_{\mathrm{f}}}s_{\mathrm{d}}\log L\right)$, if $\mathbf{\Phi}_{\mathrm{tx}} = \mathbf{\Phi}_{\mathrm{rx}} = \mathbf{I}_N$ and $\mathcal{P} = \mathcal{P}(N_{\mathrm{p,t}})$ is defined by (2.10) then the HiRIP constant $\delta_{s_{\mathrm{d}},s_{\mathrm{D}}}$ of matrix $\mathbf{M}_{\mathrm{p}}$ satisfies $\delta_{s_{\mathrm{d}},s_{\mathrm{D}}} \leq \delta_{\mathrm{t}}$ with probability $1-e^{-\Omega\left(\log L\log\frac{s_{\mathrm{D}}}{\delta_{\mathrm{t}}}\right)}$. If $\mathbf{\Phi}_{\mathrm{tx}} = \mathrm{blkdiag}\left(\mathbf{T}_{\mathrm{cp}}\mathbf{F}_{N_{\mathrm{fft}}}^{\mathrm{H}},\ldots,\mathbf{T}_{\mathrm{cp}}\mathbf{F}_{N_{\mathrm{fft}}}^{\mathrm{H}}\right)$, $\mathbf{\Phi}_{\mathrm{rx}} = \mathrm{blkdiag}\left(\mathbf{F}_{N_{\mathrm{fft}}}\mathbf{R}_{\mathrm{cp}},\ldots,\mathbf{F}_{N_{\mathrm{fft}}}\mathbf{R}_{\mathrm{cp}}\right)$ and $\mathcal{P} = \mathcal{P}(N_{\mathrm{p,t}}, N_{\mathrm{p,f}})$ is defined by (2.11) then the HiRIP constant $\delta_{s_{\mathrm{d}},s_{\mathrm{D}}}$ of matrix $\mathbf{M}_{\mathrm{p}}$ satisfies $\delta_{s_{\mathrm{d}},s_{\mathrm{D}}} \leq \delta_{\mathrm{t}}+\delta_{\mathrm{f}}+\delta_{\mathrm{t}}\delta_{\mathrm{f}}$ with probability $1-e^{-\Omega\left(\min\left\{\log L\log\frac{s_{\mathrm{d}}}{\delta_{\mathrm{t}}},\log(2Q+1)\log\frac{s_{\mathrm{D}}}{\delta_{\mathrm{f}}}\right\}\right)}$.

Proof. The proof of the theorem is given in Appendix A.4. □

Theorem 2.4.2 (HiRIP for AFDM based measurements)**.** *Assume $\mathbf{\Phi}_{\mathrm{tx}} = \mathbf{\Lambda}_{c2}\mathbf{F}_N\mathbf{\Lambda}_{c1} = \mathbf{\Phi}_{\mathrm{rx}}^{\mathrm{H}}$, $|c_1| = \frac{P}{2N}$ and let P be set as the* smallest *integer satisfying $(L-1)P+2Q+1 \geq s_{\mathrm{d}}s_{\mathrm{D}}$ and $\mathcal{P} = \mathcal{P}(N_{\mathrm{p}})$ be defined by (2.13). Then under Assumption 2.1.1 and for sufficiently large L, Q, sufficiently small δ, and $N_{\mathrm{p}} > O\left(\frac{1}{\delta^2}\log^2\frac{1}{\delta}\log\frac{\log(LP)}{\delta}\log(LP)\log\frac{Q}{P}\right)$, the HiRIP constant $\delta_{s_{\mathrm{d}},s_{\mathrm{D}}}$ of matrix $\mathbf{M}_{\mathrm{p}}$ defined in (3.14) satisfies $\delta_{s_{\mathrm{d}},s_{\mathrm{D}}} \leq \delta$ with probability $1-e^{-\Omega\left(\log\left(2\lceil\frac{Q}{P}\rceil+1\right)\log\frac{\log(LP)}{\delta}\right)}$.*

Proof. The proof of the theorem is given in Appendix A.5. □

When $P \triangleq 2N|c_1|$ is set to $2Q+1$, AFDM achieves full diversity [10] and the measurements are non-compressive. The setting $P=1$ on the other hand is the most compressive. By choosing for P a value between these two extremes as in the statement of the theorem, each pilot instance gives in its $(L-1)P+2Q+1$-long guard interval a number of measurements close with high probability to the number $s_{\mathrm{d}}s_{\mathrm{D}}$ of unknowns. Of course, a number $N_{\mathrm{p}} > 1$ of pilot instances is still required as the sparsity support needs to be estimated. But, asymptotically, this number has only a logarithmic growth with respect to *both* delay and Doppler spreads. This property is to be contrasted with the SCM and OFDM HiRIP result showing first-degree polynomial dependence of N_{p} on s_{d} or s_{D} as stated by Theorem 2.4.1.

Corollary 2.4.1 (Recovery guarantee for compressive sensing of DS-LTV channels)**.** *If* $\mathbf{M}_{\mathrm{p}}$ *satisfies the conditions of Theorems 2.4.1 or 2.4.2, the sequence* $\hat{\boldsymbol{\alpha}}^{(k)}$ *defined by Algorithm 2.4.1 satisfies* $\|\hat{\boldsymbol{\alpha}}^{(k)} - \boldsymbol{\alpha}\| \leq \rho^k \|\boldsymbol{\alpha}^{(0)} - \boldsymbol{\alpha}\| + \tau \|\mathbf{w}_{\mathrm{p}}\|$ *where* $\rho < 1$ *and* τ *are constants defined in [37, Theorem 1].*

Proof. Thanks to Theorems 2.4.1 and 2.4.2, matrix $\mathbf{M}_{\mathrm{p}}$ with large enough L, Q, N_{p} can be made to have a HiRIP constant that satisfies $\delta_{3s_{\mathrm{d}},2s_{\mathrm{D}}} < \frac{1}{\sqrt{3}}$. The conditions of [37, Theorem 1] are thus satisfied, and the corollary follows from that theorem. □

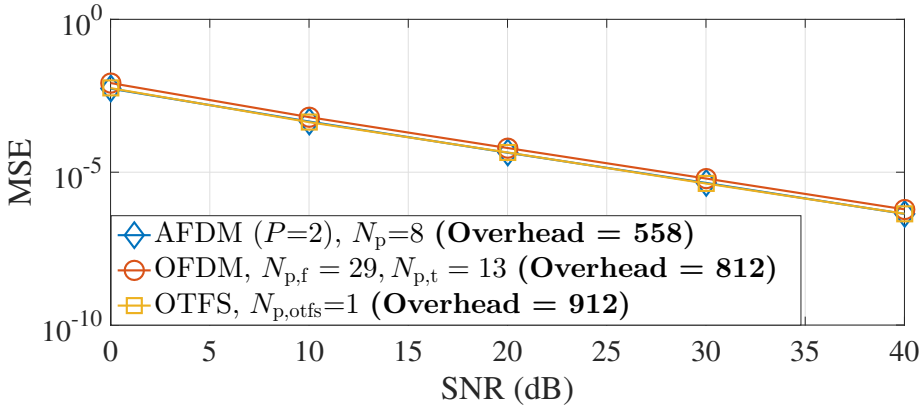
2.4.3 Numerical results

In this section, we compare the sparse recovery performance of AFDM with that of OFDM and OTFS, focusing on pilot overhead and MSE. The simulations assess the effectiveness of each waveform in achieving target MSE performance while minimizing the required pilot resources.

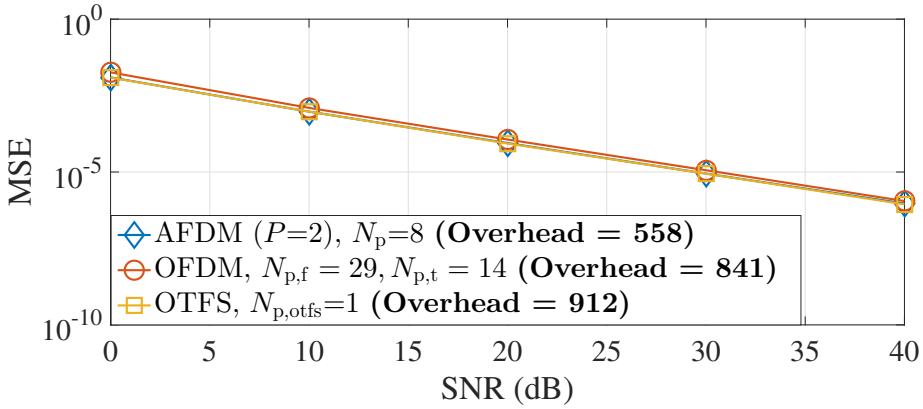
AFDM sparse recovery performance is now compared to that of OFDM and OTFS. For OFDM, transmission is organized in N-long frames, each constructed from $N_{\mathrm{s}} \approx 2Q + 1$ OFDM symbols each of which costing $L - 1$ in CP overhead. Within each frame, $N_{\mathrm{p,f}}$ subcarriers within $N_{\mathrm{p,t}}$ OFDM symbols are set as pilots [33]. As for OTFS, subcarriers are in the delay-Doppler domain forming a $M_{\mathrm{otfs}} \times N_{\mathrm{otfs}}$ grid (with $M_{\mathrm{otfs}} N_{\mathrm{otfs}} = N$). OTFS with orthogonal data-pilot resources [39] requires at least $N_{\mathrm{p,otfs}} = 1$ pilot symbols with $\min(4Q + 1, N_{\mathrm{otfs}}) \min(2L - 1, M_{\mathrm{otfs}})$ guard samples.

We used 100 realizations of channels having a Type-1 delay-Doppler sparsity with $p_{\mathrm{d}} = 0.2$, $p_{\mathrm{D}} \in \{0.2, 0.4\}$ and $N = 4096, L = 30, Q = 7$ (corresponding to a 30 MHz transmission at a 70 GHz carrier frequency, a maximum target moving speed of 396 km/h and a maximum target range of 300 meters). For both AFDM and OFDM, sparse recovery of $\boldsymbol{\alpha}$ is done using HiHTP (Algorithm 2.4.1). For OTFS, since sensing is done without compression, non-compressive estimation algorithms can be used [10]. For each waveform, the number of pilots was set in such a way that the mean squared error

MSE $\triangleq \mathbb{E}[\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2]$ is approximately 10^{-4} at SNR = 20 dB. Fig. 2.9 shows an advantage of AFDM in terms of pilot overhead i.e., the number of samples in each frame needed as pilots and guards to achieve the target MSE performance.



(a) $p_{\mathrm{D}} = 0.2$



(b) $p_{\mathrm{D}} = 0.4$

Figure 2.9: MSE and pilot overhead for $N = 4096, L = 30, Q = 7, p_{\mathrm{d}} = 0.2, N_{\mathrm{s}} = 16, N_{\mathrm{otfs}} = 16, M_{\mathrm{otfs}} = 256$.

2.5 Conclusion

This chapter explored channel estimation for doubly dispersive wireless links that are sparse in both the delay and the Doppler domains. A special focus was given to

the minimal pilot overhead required by different waveforms to achieve a target error performance while solving that problem. AFDM was shown to be optimal with respect to that performance measure when compared to OFDM and OTFS using both mathematical analysis and numerical results. In the case of an unknown DDP, DS channel has been recovered by linking delay-Doppler sparsity to the paradigm of hierarchically-sparse recovery. Numerical results confirmed AFDM's superiority in reducing pilot overhead while maintaining estimation accuracy, making it a promising approach for DS-LTV channel estimation in next-generation wireless systems.

Chapter 3

Estimation and Extrapolation of Doubly Sparse Linear Time-Varying Channels with Off-grid Doppler Shifts

While the on-grid approximation of Definition 2.1.1 is useful for the conception of LTV channel estimation and sensing schemes and for the analysis of their performance as we argued in Chapter 2, it lacks support for the finer Doppler resolution needed for channel prediction or mitigation of channel aging. Indeed, most channel prediction paradigms [40, 41, 42, 43] involve, explicitly of implicitly, the estimation of Doppler frequency shifts to within an error margin smaller than the frequency resolution characteristic of the duration of the channel observation interval. For that sake, we now present a second approximation for LTV channels that allows, in contrast to the first approximation in Definition 2.1.1, for fractional-valued Doppler frequency shifts.

3.1 Off-grid doubly sparse linear time-varying channels

To make the sparse channel model closer to the original model in (1.2), we allow for fractional-valued Doppler frequency shifts. In this case, only ι_p is zero, while κ_p may take non-zero values with κ_p sampled from a uniform distribution i.e., $\kappa_p \sim \mathcal{U}([-\frac{1}{2}, \frac{1}{2}])$. Define $N_{\mathrm{D},l,q}$ as the number of (sub-)paths with the same delays and Doppler shifts integer part and which only differ in their Doppler shifts fractional part defined as:

$$N_{\mathrm{D},l,q} \triangleq |\{p|l_p = l, q_p = q\}|, \quad p = 1 \dots N_{\mathrm{p}} \tag{3.1}$$

In this configuration, the (random) number of paths N_{p} in (1.2) is expressed as follows:

$$N_{\mathrm{p}} = \sum_{l=0}^{L-1} \sum_{q=-Q}^{Q} N_{\mathrm{D},l,q} I_{l,q} \quad \text{(total \# paths in the off-grid model)} \tag{3.2}$$

Depending on the scenario, it is in principle possible to model $N_{\mathrm{D},l,q}$ either as a fixed value or as a random variable, for example, a uniform random variable drawn over the range $[\![1..N_{\mathrm{D}}]\!]$. For simplicity, we opt for the first option, that is, $N_{\mathrm{D},l,q} = N_{\mathrm{D}}, \forall l, q$ for some value N_{D} and we impose that the following channel power normalization should be satisfied.

$$\sum_{l=0}^{L-1} \sum_{q=-Q}^{Q} \sum_{i=1}^{N_{\mathrm{D}}} \mathbb{E}\left[|\alpha_{l,q,i}|^2 I_{l,q} \right] = 1 \,. \tag{3.3}$$

Definition 3.1.1 (off-grid Delay-Doppler double sparsity)**.** The complex gain $h_{l,n}$ of the l-th path of the LTV channel varies with time as

$$h_{l,n} = \sum_{q=-Q}^{Q} I_{l,q} \sum_{i=1}^{N_{\mathrm{D}}} \alpha_{l,q,i} e^{\imath 2\pi \frac{n(q+\kappa_i)}{N}}, \quad l = 0 \cdots L-1 \tag{3.4}$$

for some value N_{D} and where $I_{l,q}$ retains the same description provided in Definition 2.1.1 and $I_{l,q}, \{\alpha_{l,q,i}, \kappa_i\}_{i=1\cdots N_{\mathrm{D}}}$ are statistically mutually independent. The complex gain

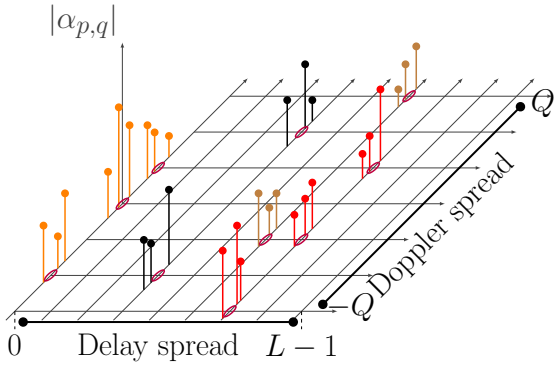


Figure 3.1: An instance of the Delay-Doppler domain response of a doubly sparse time-varying channel satisfying Definition 3.1.1

is assumed to satisfy $\alpha_{l,q,i} \sim \mathcal{CN}\left(0, \sigma_\alpha^2\right)$ with σ_α^2 chosen so that (3.3) is respected.

Figure 3.1 illustrates the delay-Doppler sparsity structure under the Type-2 model in an off-grid scenario, where active components deviate from a predefined grid. This figure effectively demonstrates the irregular yet structured sparsity inherent to the Type-2 model as defined in Definition 2. Note that the above model is an off-grid approximation of a time-varying channel.

Remark 3. *Maintaining the delay taps as integer values is not to the detriment of the core message of this work since the same approach we develop to deal with off-grid Doppler shifts can be extended to off-grid delays. Moreover, the effect of fractional delay values can still be made to fall under the current model by increasing the value of the delay domain sparsity parameter p_{d} of the model sufficiently to account for leakage due to off-grid delay shifts.*

Estimating the unknown parameters of the off-grid channel model in (3.4) remains a challenging task. Indeed, the N_{D} Doppler frequency shifts $\{q+\kappa_i\}_{i=1\cdots N_{\mathrm{D}}}$ associated with a grid point (l, q) are only different by their fractional part and hence render the problem of estimating the corresponding complex coefficients $\{\alpha_{l,q,i}\}_{i=1\cdots N_{\mathrm{D}}}$ ill-conditioned. Moreover, the number of unknowns in the model i.e., N_{p}, could be prohibitively large. For both these reasons, we propose a new model based on multiple "elementary" basis

expansion models (BEM), each set with a bandwidth equal to the frequency resolution and shifted in frequency to be centered at one of the active grid points. Thanks to its optimality in terms of time-frequency localization [44], we opt for BEM based on discrete prolate spheroidal sequences (DPSS).

3.2 Background: DPSS basis expansion model

We begin with an overview of DPSS BEM modeling of a baseband signal h_n that occupies a bandwidth $(-W, W)$ in digital frequencies with $W \in \left(0, \frac{1}{2}\right)$. DPSS basis vectors $u_{b,n}^{(N,W)}$ $(b = 1, \ldots, N)$ are the eigenvectors of the prolate matrix [30]:

$$\sum_{k=0}^{N-1} C_{k,n}^{(N,W)} u_{b,k}^{(N,W)} = \lambda_b^{(N,W)} u_{b,n}^{(N,W)}, b = 1 \cdots N, n = 0 \cdots N-1 \tag{3.5}$$

where $C_{k,n}^{(N,W)}$ is the $(k-n)$-th entry of the prolate matrix:

$$C_{k,n}^{(N,W)} = \frac{\sin\left(2\pi W(k-n)\right)}{\pi(k-n)}\,. \tag{3.6}$$

The eigenvectors are normalized so that $\sum_{n=1}^{N} \left(u_{b,n}^{(N,W)}\right)^2 = 1$. The eigenvalues are ordered according to their values starting with the largest one: $1 \geq \lambda_0^{(N,W)} \geq \ldots \geq \lambda_{N-1}^{(N,W)} \geq 0$. The eigenvalues $\lambda_b^{(N,W)}$ (representing energy concentration) are clustered near 1 for $b \leq 2WN$, and rapidly drop to zero for $b > 2WN$ [45].

In this study we employ multiple DPSS BEMs, each of which is used to represent the channel signal component related to the fractional part of the N_{D} Doppler shifts around one of the grid points (l, q) i.e., $\sum_{i=1}^{N_{\mathrm{D}}} \alpha_{l,q,i} e^{\imath 2\pi \frac{n\kappa_i}{N}}$, and not the channel signal associated with the whole Doppler spread i.e., $h_{l,n}$. Each of these BEMs is defined using (3.5) and (3.6) with $W = \frac{1}{2N}$.

In the interest of simplicity, all parameters with the superscript (N, W) in their notation, such as $u_{b,n}^{(N,W)}$, will be replaced by their simplified forms e.g., $u_{b,n}$.

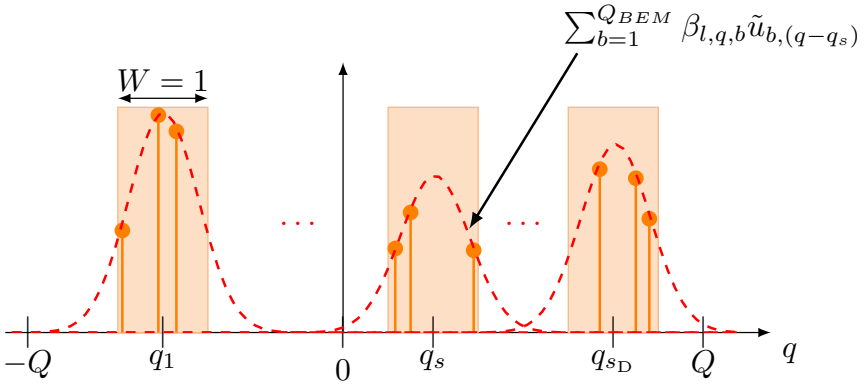


Figure 3.2: An illustration of how the leakage due to S_{D} clusters of off-grid Doppler shifts $\{q_s+\kappa_i\}_{\substack{s=1\cdots S_{\mathrm{D}},\\ i=1\cdots N_{\mathrm{D}}}}$ can be captured by S_{D} frequency shifted copies of a DPSS BEM $\{u_{b,n}\}_{b=1\cdots Q_{\mathrm{BEM}}}$ ($\tilde{u}_{b,k}$ in the figure being the DFT of $u_{b,n}$)

3.3 Off-grid approximation using multiple shifted elementary BEMs

The following term in (3.4) represents a signal occupying a frequency band centered at digital frequency $\frac{q}{N}$ and of a bandwidth equal to $\frac{1}{N}$

$$h_{l,q,n} \triangleq \sum_{i=1}^{N_{\mathrm{D}}} \alpha_{l,q,i} e^{\imath 2\pi \frac{n(q+\kappa_i)}{N}} \tag{3.7}$$

Its baseband version defined as $e^{-\imath 2\pi \frac{nq}{N}} h_{l,q,n}$ can thus be modeled (see Figure 3.2) using a DPSS BEM of an order equal to Q_{BEM} (chosen large enough depending on the required modeling precision) and a bandwidth $(-W, W)$ with $W = \frac{1}{2N}$ by defining

$$h^{\mathrm{BEM}}_{l,q,n} \triangleq e^{\imath 2\pi \frac{nq}{N}} \sum_{b=1}^{Q_{\mathrm{BEM}}} \beta_{l,q,b} u_{b,n}\,. \tag{3.8}$$

In vector form

$$\boldsymbol{\beta}_{l,q} = \mathbf{U}^{\mathrm{H}}_{Q_{\mathrm{BEM}}} \mathbf{E}^{\mathrm{H}}_{\frac{q}{N}} \mathbf{h}_{l,q} \tag{3.9}$$

$$\mathbf{h}^{\mathrm{BEM}}_{l,q} = \mathbf{E}_{\frac{q}{N}} \mathbf{P}^{\mathrm{BEM}} \mathbf{E}^{\mathrm{H}}_{\frac{q}{N}} \mathbf{h}_{l,q} \tag{3.10}$$

with $\mathbf{U}_{Q_{\mathrm{BEM}}} \triangleq [\mathbf{u}_1 \dots \mathbf{u}_{Q_{\mathrm{BEM}}}]$ and $\mathbf{P}^{\mathrm{BEM}} = \mathbf{U}_{Q_{\mathrm{BEM}}} \mathbf{U}_{Q_{\mathrm{BEM}}}^{\mathrm{H}}$ being the orthogonal projection matrix associated with the DPSS BEM representation and $\mathbf{E}_f \triangleq \mathrm{diag}\left(e^{\imath 2\pi f 0} \dots e^{\imath 2\pi f(N-1)}\right)$.

Inserting the elementary BEMs defined in (3.8) for each Doppler grid point into (3.4) gives

$$\begin{aligned} h_{l,n}^{\mathrm{BEM}} &= \sum_{q=-Q}^{Q} I_{l,q} e^{\imath 2\pi \frac{nq}{N}} \sum_{b=1}^{Q_{\mathrm{BEM}}} \beta_{l,q,b} u_{b,n}, \\ &= \sum_{q=-Q}^{Q} I_{l,q} h_{l,q,n}^{\mathrm{BEM}},\ l = 0 \cdots L-1 \quad (\text{“multiple shifted elementary BEMs” model}) \end{aligned} \tag{3.11}$$

The precision of the above representation in relation to the order of the elementary BEMs in the model (3.11) is provided by the following theorem.

Theorem 3.3.1. *For any $\epsilon > 0$, if we set $Q_{\mathrm{BEM}} \geq C \log \frac{1}{\epsilon}$ for some constant C then the representation error of the model in* (3.11) *satisfies $\sum_{l=0}^{L-1} \mathbb{E}\left[\left|h_{l,n} - h_{l,n}^{\mathrm{BEM}}\right|^2\right] < \epsilon$ for a sufficiently large N.*

Proof. The proof of Theorem 3.3.1 is given in Appendix B.1. □

Theorem 3.3.1 states that the number of DPSS basis functions needed to represent the channel component associated with a single delay-Doppler grid point grows only logarithmically with the inverse of the target precision. The following figure (as well as the numerical results of Section 3.6) shows that as few as 4 DPSS basis functions ($Q_{\mathrm{BEM}} = 4$) are sufficient to get good enough precision. Note that this value is unrelated to the Doppler spread value i.e., $2Q+1$: a larger Doppler spread does not imply the need for a larger value of Q_{BEM}.

3.4 DS-LTV off-grid channel estimation

Let $\boldsymbol{\beta}$ be the vectorized form of the BEM coefficients $\{\beta_{l,q,b}\}_{\substack{b=1\cdots Q_{\mathrm{BEM}} \\ l,q,I_{l,q}\neq 0}}$ associated with the active delay-Doppler grid points. Define $\mathbf{B}$ as the following block-diagonal matrix

$$\mathbf{B} = \mathrm{blkdiag}\left(\widetilde{\mathbf{B}}, \ldots, \widetilde{\mathbf{B}}\right) \tag{3.12}$$

with $\widetilde{\mathbf{B}} \triangleq \left[\widetilde{\mathbf{B}}_{-Q} \ldots \widetilde{\mathbf{B}}_{Q}\right]$ being a $N \times (2Q+1)Q_{\mathrm{BEM}}$ matrix satisfying $\left[\widetilde{\mathbf{B}}_q\right]_{n,b} = \frac{1}{\sqrt{N}}\tilde{U}_{b,(n-q)}$ and $\tilde{U}_{b,n} = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} u_{b,k} e^{-\imath 2\pi \frac{nk}{N}}$ being the DFT of the DPSS basis vector $u_{b,n}$. Let $\mathbf{A}_\beta$ be the matrix that places the blocks of $\boldsymbol{\beta}$, each of size Q_{BEM}, within the positions corresponding to $I_{l,q} \neq 0$ in a larger vector of length $LQ_{\mathrm{BEM}}(2Q+1)$ resulting in a $LQ_{\mathrm{BEM}}(2Q+1)$-long vector $\mathbf{A}_\beta\boldsymbol{\beta}$ that is block sparse.Now, define

$$\boldsymbol{\alpha} \triangleq \mathbf{B}\mathbf{A}_\beta\boldsymbol{\beta}\,. \tag{3.13}$$

Inserting (3.13) into (2.22), we can write the received pilot samples vector $\mathbf{y}_{\mathrm{p}}$ as

$$\mathbf{y}_{\mathrm{p}} = \underbrace{\mathbf{A}_{\mathcal{P}}\mathbf{M}\mathbf{B}\mathbf{A}_\beta}_{\triangleq \mathbf{M}_{\mathrm{p}}}\boldsymbol{\beta} + \mathbf{w}_{\mathrm{p}}\,. \tag{3.14}$$

where $\mathbf{M}$ and $\mathbf{A}_{\mathcal{P}}$ are defined in (2.14). The minimum mean squared error (MMSE) estimate, $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$ based on $\mathbf{y}_{\mathrm{p}}$ is given by [36].

$$\hat{\boldsymbol{\beta}} = \mathbf{B}\left(\sigma_\alpha^2\mathbf{M}_{\mathrm{p}}^H\mathbf{M}_{\mathrm{p}} + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{M}_{\mathrm{p}}^H\mathbf{y}_{\mathrm{p}}\,. \tag{3.15}$$

Note that the knowledge of the DDP at the receiver side is here assumed. Finally, define

$$\hat{h}_{l,n}^{\mathrm{BEM}} = \sum_{q=-Q}^{Q} I_{l,q} e^{\imath 2\pi \frac{nq}{N}} \sum_{b=1}^{Q_{\mathrm{BEM}}} \hat{\beta}_{l,q,b} u_{b,n}, \quad n = 0, \ldots, N-1\,. \tag{3.16}$$

as the resulting MMSE estimate of $h_{l,n}^{\mathrm{BEM}}$. In vector form

$$\hat{\mathbf{h}}_l^{\mathrm{BEM}} = \sum_{q=-Q}^{Q} I_{l,q} \underbrace{\mathbf{E}_{\frac{q}{N}} \mathbf{U}_{Q_{\mathrm{BEM}}} \hat{\boldsymbol{\beta}}}_{\triangleq \hat{\mathbf{h}}_{l,q}^{\mathrm{BEM}}} . \tag{3.17}$$

The mean squared error (MSE) conditioned on a given $I_{l,q}$ realization is then written as

$$\begin{aligned} \sum_{l=0}^{L-1} \frac{1}{N} \mathbb{E}\left[\left\| \mathbf{h}_l^{\mathrm{BEM}} - \hat{\mathbf{h}}_l^{\mathrm{BEM}} \right\|^2 \right] &\leq \frac{1}{N} \sum_{l=0}^{L-1} \sum_{q=-Q}^{Q} I_{l,q} \mathbb{E}\left[\left\| \mathbf{h}_{l.q}^{\mathrm{BEM}} - \hat{\mathbf{h}}_{l,q}^{\mathrm{BEM}} \right\|^2 \right] \\ &= \frac{1}{N N_{\mathrm{D}} \sigma_\alpha^2} \mathbb{E}\left[\left\| \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \right\|^2 \right] \end{aligned} \tag{3.18}$$

where the inequality is due to the triangle inequality and where the equality is due to the definition of $\hat{\mathbf{h}}_{l,q}^{\mathrm{BEM}}$ in (3.17), to the fact that $\sum_{l=0}^{L-1} \sum_{q=-Q}^{Q} I_{l,q} = \frac{1}{N_{\mathrm{D}} \sigma_\alpha^2}$ per (3.3) and to the fact that $\mathbf{E}_{\frac{q}{N}}^{\mathrm{H}} \mathbf{E}_{\frac{q}{N}} = \mathbf{I}_N$ and $\mathbf{U}_{Q_{\mathrm{BEM}}}^{\mathrm{H}} \mathbf{U}_{Q_{\mathrm{BEM}}} = \mathbf{I}_{Q_{\mathrm{BEM}}}$.

Assumption 3.4.1. *The number N_{p} of pilots is sufficiently large for the MSE associated with estimating $\boldsymbol{\beta}$ (and hence $\mathbf{h}_l^{\mathrm{BEM}}$ per (3.18)) to converge to zero as $\sigma_w^2 \to 0$.*

Remark 4 (Impact of the sensing waveform)**.** *The value of N_{p} needed for Assumption 3.4.1 to hold and its associated pilot overhead both depend on the particular waveform in use. The numerical results given in Section 3.6 show that the relative advantage of AFDM over OFDM and SCM, which has been analytically established by Theorems 2.4.1 and 2.4.2 under the on-grid channel model, is still valid in the off-grid Doppler shifts case.*

Assumption 3.4.1 is about the MSE of estimating the BEM representation of the channel. The following corollary to Theorem 3.3.1 provides an analysis of the estimation MSE when computed with respect to the actual channel defined by (3.1.1) instead of its BEM representation.

Corollary 3.4.1. *Under Assumption 3.4.1 and Definition 3.1.1, provided that N is large*

enough and $Q_{\mathrm{BEM}} \geq C \log \frac{1}{\epsilon}$ *for any* $\epsilon > 0$ *and some constant* C*, then*

$$\lim_{\sigma_w^2 \to 0} \mathbb{E}\left[\sum_{l=0}^{L-1} \frac{1}{N} \left\| \mathbf{h}_l - \hat{\mathbf{h}}_l^{\mathrm{BEM}} \right\|^2\right] \leq \epsilon \tag{3.19}$$

Proof. Apply the triangle inequality to $\mathbf{h}_l - \hat{\mathbf{h}}_l^{\mathrm{BEM}} = \left(\mathbf{h}_l - \mathbf{h}_l^{\mathrm{BEM}}\right) + \left(\mathbf{h}_l^{\mathrm{BEM}} - \hat{\mathbf{h}}_l^{\mathrm{BEM}}\right)$ followed by applying Assumption 3.4.1 to the first term and Theorem 3.3.1 to the second. □

3.5 Application to channel extrapolation and prediction

We know from the literature on the Slepian basis [43] that there is a "natural" way to extend the finite sequences $u_{b,n}$ from the smaller interval $[\![0..N-1]\!]$ to the larger one $[\![-N_{\mathrm{ext}}..N+N_{\mathrm{ext}}]\!]$, where N_{ext} denotes the additional channel samples to be extrapolated. This is achieved by letting the index n in (3.5) be defined over $\mathbb{Z}$ instead of being confined to $[\![0..N-1]\!]$ leading to the following infinite discrete-time signal

$$u_{b,n}^{\mathrm{ext}} \triangleq \frac{1}{\lambda_b^{(N,W)}} \sum_{k=0}^{N-1} C_{k,n}^{(N,W)} u_{b,k}, n \in \mathbb{Z}\,. \tag{3.20}$$

Signal $(u_{b,n}^{\mathrm{ext}})_{n\in\mathbb{Z}}$ ("ext" stands for "extrapolation") has a discrete-time Fourier transform (DTFT) that is zero outside $(-W, W)$ and is the signal that has the least energy outside the time interval $[\![0..N-1]\!]$ from among all the discrete-time signals band-limited to $(-W, W)$ [43]. Once we have estimated the multiple-BEM representation of the off-grid DS-LTV channel on the interval $[\![0..N-1]\!]$ as in (3.16) and once we have calculated the infinite-time version of the DPSS basis function as in (3.20), the channel can thus be extrapolated as follows

$$h_{l,n}^{\mathrm{ext}} \triangleq \sum_{q=-Q}^{Q} I_{l,q} \underbrace{e^{\imath 2\pi \frac{nq}{N}} \sum_{b=1}^{Q_{\mathrm{BEM}}} \hat{\beta}_{l,q,b} u_{b,n}^{\mathrm{ext}}}_{\triangleq h_{l,q,n}^{\mathrm{ext}}},\ l = 0 \cdots L-1\,,\ n \in \mathbb{Z}\,. \tag{3.21}$$

Here, we defined

$$h_{l,q,n}^{\text{ext}} \triangleq e^{\imath 2\pi \frac{nq}{N}} \sum_{b=1}^{Q_{\text{BEM}}} \hat{\beta}_{l,q,b} u_{b,n}^{\text{ext}} \,. \tag{3.22}$$

Defining $\mathbf{u}_n^{\text{ext}} \triangleq \begin{bmatrix} u_{1,n}^{\text{ext}} & \cdots & u_{Q_{\text{BEM}},n}^{\text{ext}} \end{bmatrix}^{\text{T}}$ and referring to (3.17) gives

$$h_{l,q,n}^{\text{ext}} = e^{\imath 2\pi \frac{nq}{N}} \left(\mathbf{u}_n^{\text{ext}}\right)^{\text{T}} \mathbf{U}_{Q_{\text{BEM}}}^{\text{H}} \mathbf{E}_{\frac{q}{N}}^{\text{H}} \hat{\mathbf{h}}_{l,q}^{\text{BEM}} \tag{3.23}$$

Theorem 3.5.1. *The predictor $h_{l,n}^{\text{ext}}$ defined by (3.21) for any $n > N-1$ based on extrapolating multiple estimated BEMs of order Q_{BEM} converges in the squared-mean sense in the limit of a vanishing noise variance to a reduced-rank (with a reduced rank equal to Q_{BEM}) MMSE estimator of sample $h_{l,n}$ under the assumptions of Definition 3.1.1 and Assumption 3.4.1.*

Proof. The proof of Theorem 3.5.1 is given in Appendix B.2. □

Remark 5 (Setting the value of Q_{BEM} for prediction purposes)**.** *While Theorem 3.5.1 establishes the optimality in a certain sense of the predictor $h_{l,n}^{\text{ext}}$ by relating it to a reduced-rank MMSE estimator, the theorem tells nothing about the optimal value of the BEM order Q_{BEM} to be used while deriving $h_{l,n}^{\text{ext}}$ (or equivalently about the optimal rank of the related reduced-rank estimator). The properties of the DPSS basis can give us some insight. Indeed, it is known [44] that the b-th infinite-length DPSS has $1-\lambda_b^{(N,W)}$ (respectively $\lambda_b^{(N,W)}$) of its energy outside (respectively inside) the interval $[\![0..N-1]\!]$ as illustrated in Figure 3.3.*

To get a non-vanishing value $h_{l,q,n}^{\text{ext}} = e^{\imath 2\pi \frac{nq}{N}} \sum_{b=1}^{Q_{\text{BEM}}} \beta_{l,q,b} u_{b,n}^{\text{ext}}$ for $n > N-1$, the sum should include terms with large-enough samples $u_{b,n}^{\text{ext}}$ i.e., with small-enough $\lambda_b^{(N,W)}$. While this constrains Q_{BEM} to be sufficiently large, it simultaneously constrains the estimation SNR to be high enough to guarantee a precise estimation of coefficients $\beta_{l,q,b}$ associated with basis functions that have little contribution to the signal received inside the observation interval $[\![0..N-1]\!]$.

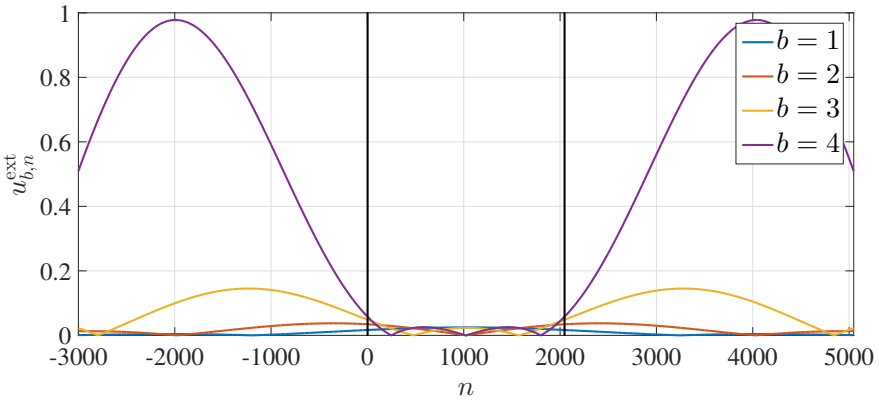


Figure 3.3: Extrapolated versions of the first four DPSS ($N = 2048, W = \frac{1}{2N}$)

Remark 6 (Relation to other DPSS extrapolation prediction methods)**.** *Our approach applies multiple elementary DPSS-based BEMs each modeling the narrow-band channel component* $e^{-\imath 2\pi \frac{ng}{N}} h_{l,q,n}$ *associated with one of the delay-Doppler grid points. The DPSS-based BEM modeling in [43], on the other hand, uses a single BEM to represent the multi-band signal* $h_{l,n} = \sum_{q=-Q}^{Q} I_{l,q} \sum_{i=1}^{N_{\mathrm{D}}} \alpha_{l,q,i} e^{\imath 2\pi \frac{n(q+\kappa_i)}{N}}$ *of each channel delay tap to give*

$$\check{h}_{l,n}^{\mathrm{BEM}} = \sum_{b=1}^{\check{Q}_{\mathrm{BEM}}} \check{\beta}_{l,q,b} \check{u}_{b,n}, \quad n = 0, \ldots, N-1\,. \tag{3.24}$$

where $\check{u}_{b,k}^{(N,W)}$ $(b = 1, \ldots, \check{Q}_{\mathrm{BEM}})$ *are the basis vectors of the single multi-band BEM defined as*

$$\sum_{k=0}^{N-1} \check{C}_{k,n}^{(N,W)} \check{u}_{b,k} = \check{\lambda}_{b}^{(N,W)} \check{u}_{b,n}, b = 1 \cdots N, n = 0 \cdots N-1 \tag{3.25}$$

where $\check{C}_{k,n}^{(N,W)}$ *is the* $(k-n)$*-th entry of the multi-band prolate matrix:*

$$\check{C}_{k,n}^{(N,W)} \triangleq \sum_{\substack{q=-Q \\ I_{l,q}=1}}^{Q} e^{\imath 2\pi \frac{q(k-n)}{N}} \frac{\sin\left(2\pi W(k-n)\right)}{\pi(k-n)}\,. \tag{3.26}$$

The associated channel predictor is

$$\check{h}_{l,n}^{\text{ext}} \triangleq \sum_{b=1}^{\check{Q}_{\text{BEM}}} \check{\beta}_{l,q,b} \check{u}_{b,n}^{\text{ext}},\ l = 0 \cdots L-1\ ,\ n \in \mathbb{Z}\,, \tag{3.27}$$

where the associated extrapolated basis vector is given by

$$\check{u}_{b,n}^{\text{ext}} \triangleq \frac{1}{\check{\lambda}_b^{(N,W)}} \sum_{k=0}^{N-1} \check{C}_{k,n}^{(N,W)} \check{u}_{b,k}, n \in \mathbb{Z}\,. \tag{3.28}$$

The main issue with the multi-band DPSS approach is the size of the codebook that needs to be computed or stored at the network device performing the channel estimation. As can be seen from (3.14)*, the codebook size in our approach is the number of columns of the matrix* $\mathbf{MB}$ *which equals* $Q_{\text{BEM}} L(2Q+1)$*. In the case of the multi-band BEM approach, the codebook size would be* $L \sum_{k=1}^{2Q+1} \binom{2Q+1}{k} \check{Q}_{\text{BEM}}(k) \gg Q_{\text{BEM}} L(2Q+1)$ *as every different combination of* k *active Doppler grid points (*$k \in [\![1..2Q+1]\!]$*) would result in a different multi-band prolate matrix* $\check{C}_{k,n}^{(N,W)}$ (3.26) *and thus in a different DPSS basis* (3.25)*. Note that, for the same reason, the multi-band DPSS BEM approach is not suited for compressed sensing applications with Doppler domain sparsity. Another advantage of the multiple shifted BEMs scheme is the fact that we can analytically quantify its precision, as we did in Theorem 3.3.1 and Corollary 3.4.1.*

3.6 Numerical results

Since it has been theoretically proven that OFDM has a smaller overhead than SCM, the numerical results will focus on comparing AFDM sparse recovery performance to that of OFDM. We used 100 realizations of channels having a Type-1 delay-Doppler sparsity with $p_{\text{d}} = 0.2$, $p_{\text{D}} = 0.2$ and $N = 2048, L = 20, Q = 7$ (corresponding to a 15 MHz transmission at a 70 GHz carrier frequency, a maximum target moving speed of 396 km/h and a maximum target range of 400 meters).

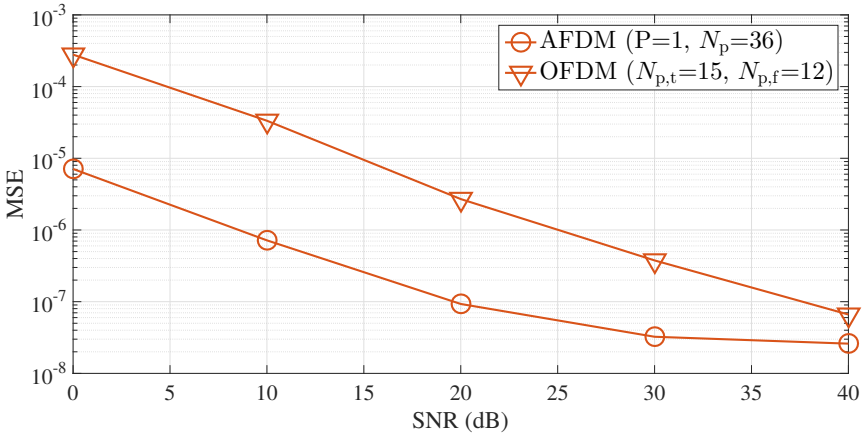


Figure 3.4: Channel estimation MSE for $N = 2048, L = 20, Q = 7, p_{\mathrm{d}} = 0.2$, $p_{\mathrm{D}} = 0.2, N_{\mathrm{s}} = 32, N_{\mathrm{fft}} = 64$

OFDM transmission is organized in N-long frames, each constructed from N_{s} OFDM symbols each of which costing $L-1$ in CP overhead. Within each frame, $N_{\mathrm{p,f}}$ subcarriers within $N_{\mathrm{p,t}}$ OFDM symbols are set as pilots as it is shown by Figure 2.3. As highlighted previously in remark 4, Figure 3.4 illustrates that AFDM waveform maintains its superiority particularly in handling off-grid Doppler shifts. This is particularly evident when comparing AFDM and OFDM, both of which have nearly identical overheads, with $overhead_{\mathrm{AFDM}} = 767$ and $overhead_{\mathrm{OFDM}} = 769$. Despite the minimal difference in overhead, AFDM shows superior performance in channel estimation. Similarly, Figure 3.5 shows in the case of additional channel samples to be extrapolated $N_{\mathrm{ext}} = 500$ that AFDM has a superior performance in terms of channel prediction compared to OFDM .

Figure 3.6 compares the MSE performance of AFDM using different approaches, including our proposed method, a single BEM approach, and grid refinement with a refinement factor of $O = 4$ [46], for channel estimation and Figure 3.7 for channel prediction with $N_{\mathrm{ext}} = 500$ (represented by the blue lines) and $N_{\mathrm{ext}} = 1000$ (represented by the yellow lines). All methods are evaluated under the same overhead. For channel estimation, the multiple shifted BEMs approach and the single BEM method achieve similar performance, both outperforming the grid refinement technique. However, for

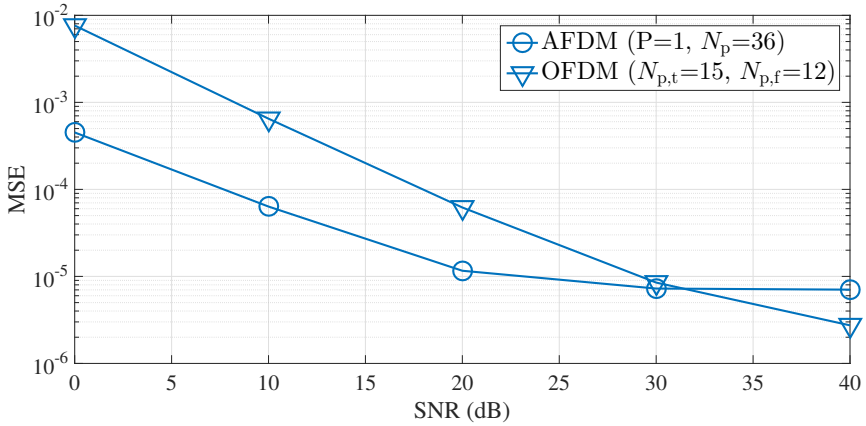


Figure 3.5: Channel prediction MSE for $N = 2048, L = 20, Q = 7, p_{\mathrm{d}} = 0.2$, $p_{\mathrm{D}} = 0.2, N_{\mathrm{s}} = 32, N_{\mathrm{fft}} = 64, N_{\mathrm{ext}} = 500$

Table 3.1: Size of the codebook for each method

Method	**Codebook Size**
One BEM	2,621,360
Multiple BEMs	1,200

channel prediction, our approach exhibits a slight performance degradation compared to the single BEM method. Nevertheless, as highlighted in Remark 6, our method requires a significantly smaller codebook size than the single BEM approach shown by the Table 3.1 (for the same settings as the previous simulations), making our approach more efficient in practical implementations.

3.7 Conclusion

In this chapter, we introduced our approach for estimating and extrapolating doubly sparse linear time-varying (DS-LTV) channels with off-grid Doppler shifts. We demonstrated the limitations of traditional on-grid approximations and proposed an off-grid model that accounts for fractional Doppler shifts. To address the challenges of estimating off-grid DS-LTV channels, we leveraged multiple shifted elementary BEMs based on DPSS. This approach allowed us to efficiently capture the sparse and structured nature of delay-

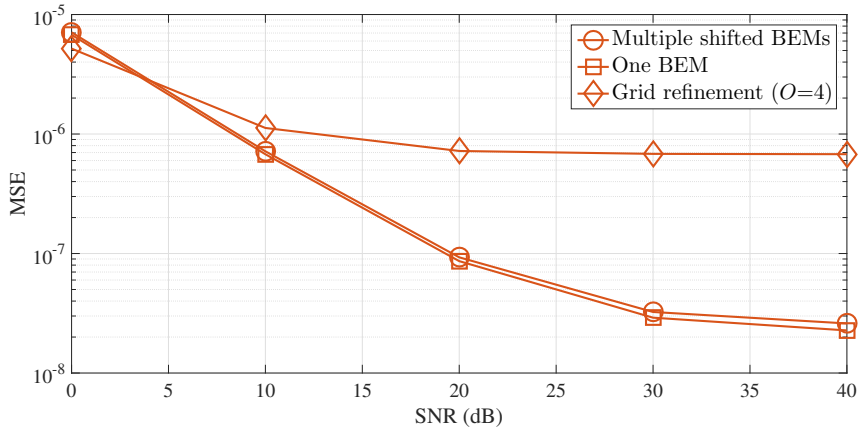


Figure 3.6: Channel estimation MSE performance of AFDM with different approaches, all using the same overhead for $N = 2048, L = 20, Q = 7, p_{\mathrm{d}} = 0.2, p_{\mathrm{D}} = 0.2$

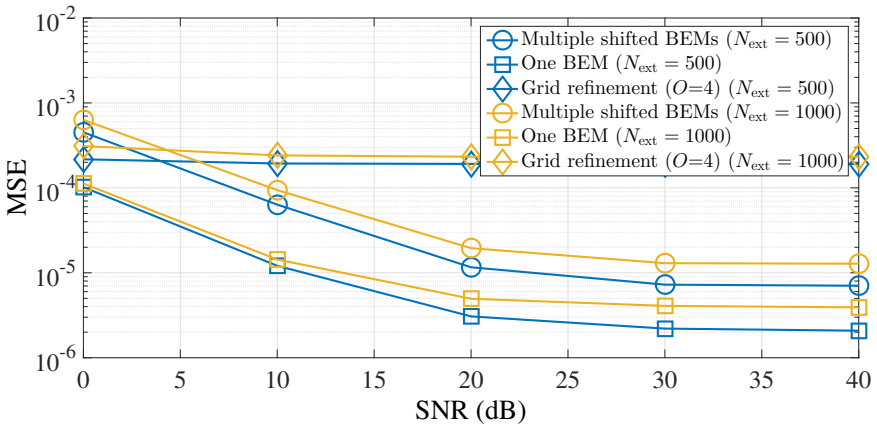


Figure 3.7: channel prediction MSE performance of AFDM with different approaches, all using the same overhead, for $N = 2048, L = 20, Q = 7, p_{\mathrm{d}} = 0.2, p_{\mathrm{D}} = 0.2, N_{\mathrm{ext}} \in \{500, 1000\}$

Doppler domain responses while ensuring minimal codebook size compared to the single BEM approach. Theoretical analysis established the accuracy of our proposed model, showing that the estimation error diminishes as the number of DPSS basis functions increases logarithmically. We also extended the proposed model to channel extrapolation and prediction, utilizing DPSS properties to extend the estimated BEM coefficients beyond the observation interval. Theoretical proofs confirmed that our method achieves a reduced-rank MMSE estimator in the low-noise regime. Through numerical simulations, we validated the effectiveness of our approach, demonstrating superior channel estimation and prediction performance compared to conventional OFDM. Our multiple shifted BEMs approach has been compared to the single BEM method and grid-refinement technique. Our method outperforms grid refinement for channel estimation but shows a slight performance degradation in channel prediction compared to the single BEM approach. However, it remains more efficient in practical implementations due to its significantly smaller codebook size. Overall, the proposed framework offers a robust and computationally efficient solution for DS-LTV channel estimation and extrapolation, making it highly suitable for next-generation wireless communication systems, particularly in high-mobility environments where precise Doppler estimation is crucial.

Chapter 4

Further Applications

In this chapter, further applications of delay-Doppler sparsity are discussed. More precisely, we first point to the close relation between Integrated Sensing and Communication (ISAC) and the problem of DS-LTV channel estimation and how the results we obtained pertaining to the latter paradigm hence extend to the former. Focusing on the application of Affine Frequency Division Multiplexing (AFDM) in ISAC, we further demonstrate its potential for efficient delay-Doppler estimation in sub-Nyquist settings.

4.1 Integrated sensing and communications

The results obtained in the thesis pertaining to DS-LTV channel estimation extend to target detection and target parameter estimation in ISAC applications. Indeed, the LTV channel model originally given in (1.2) and reproduced below can be seen as the echo signal from N_{p} point targets

$$h(t,\tau) = \sum_{p=1}^{N_{\mathrm{p}}} g_p e^{\imath 2\pi \nu_p t \Delta f} \delta(\tau - \tau_p) \,. \tag{4.1}$$

In such a case, the delays and the Doppler frequency shifts are related to the target parameters as follows. Let c, f_{c}, v_p, r_p denote the speed of light (m/s), carrier frequency

(Hz), relative velocity (m/s) and range (m) associated with the p-th point target, respectively. In a mono-static setting (in which the ISAC signal transmitter and receiver are co-located), the range and relative velocity of the p-th target are respectively

$$r_p = c \cdot \frac{\tau_p}{2}\,, \quad v_p = c \cdot \frac{\nu_p \Delta f}{2 f_c}\,. \tag{4.2}$$

4.1.1 Application of the DS-LTV model to ISAC

When the number N_{p} of point targets is relatively small, it is thus clear from (4.1) and (4.2) that the problem of target detection and target range and relative velocity estimation is closely related to the problem of estimation of doubly sparse LTV channels. Indeed, under the on-grid approximation, estimating the delay-Doppler sparsity support of the DS-LTV channel (see Definition 2.1.1) associated with the echo signal, that is the set $\{(l,q), I_{l,q} \neq 0\}$, gives an estimate of the targets' range and relative velocity.

The results given in Chapters 2 and 3 pertaining to the pilot overhead needed when using different waveforms to estimate the DS-LTV channel with an *unknown* delay-Doppler profile are also relevant for comparing the pilot overhead needed by different waveforms when these waveforms are used for sensing purposes. All that is needed is to set the sparsity parameters of the DS-LTV model, namely p_{d} and p_{D}, in accordance with the characteristics of the wireless propagation environment related to the targets to be detected.

4.2 Sub-Nyquist radar

Traditional sub-Nyquist radar techniques exploit channel sparsity to reduce sampling rates, but many existing methods rely on impractical random sampling strategies or complex analog-domain processing, which introduce hardware challenges and noise susceptibility [47, 48, 49]. These limitations hinder their applicability in real-world scenarios, necessitating more efficient solutions. As an alternative, we show in this section

that the optimality of AFDM in terms of pilot overhead under delay-Doppler sparsity along with the chirp nature of its waveform make it relevant for sub-Nyquist radar applications.

Indeed, in sensing and radar applications, the sub-Nyquist radar paradigm [47] leverages wireless channel sparsity to allow for sub-Nyquist receivers. For example, some of the solutions presented in [47] consider the use of random sub-Nyquist sampling to achieve a low coherence for the sensing matrix, a property that translates into good compressed sensing performance. However, random sub-Nyquist sampling is technically impossible in many applications [48]. Other solutions in [47] don't rely on random sampling but require dedicated analog-domain processing, more precisely multi-channel processing with multiple mixers and integrators for direct Fourier coefficient extraction. Similarly, the sub-Nyquist sampling method in [49] relies on dedicated analog-domain components needed to perform continuous-time signal differentiation. This requirement complicates hardware design and introduces noise susceptibility, making the approach less robust. Also, the approach in [48] relies on specialized hardware components, including radio frequency (RF) signal splitters and multiple analog delay modules, which not only increase the complexity of implementation but also reduces signal-to-noise ratio due to analog domain splitting.

4.2.1 AFDM-based sub-Nyquist radar

We now consider the case where the AFDM signal is destined for a sensing receiver either co-located with the transmitter (the mono-static setting) or in a remote device (the bi-static setting). In any of these settings, the non-zero complex gains $\alpha_{l,q}$ in (2.7) will represent a point target with a delay l (related to the to-be-estimated range) and a Doppler frequency shift q (related to the to-be-estimated velocity).

Instead of applying DAFT to the received AFDM signal after sampling as in basic AFDM operation [10] (which would require a sampling rate at least equal to the signal

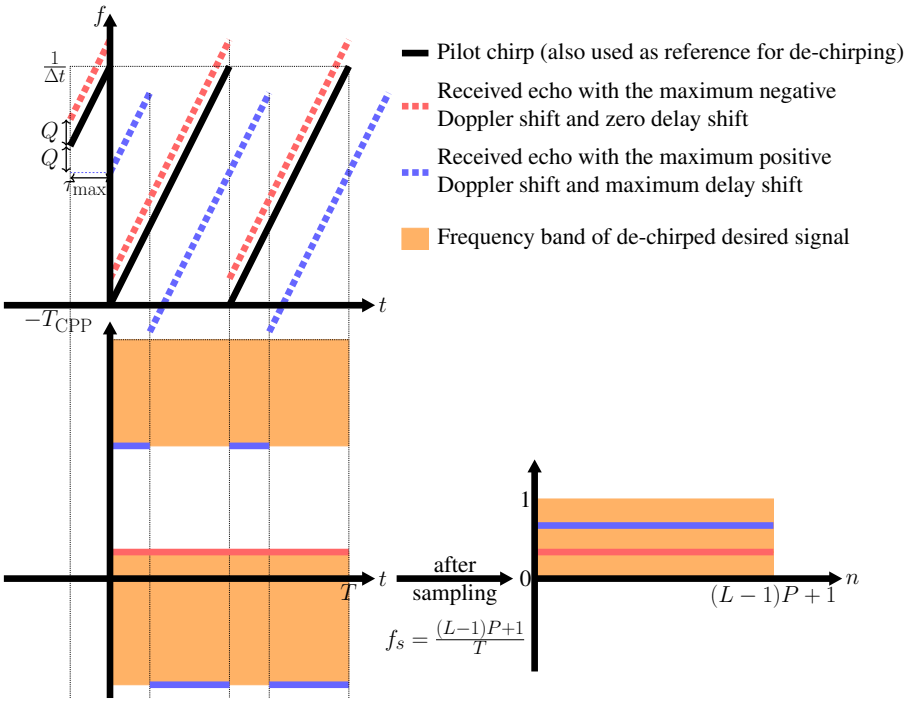


Figure 4.1: Time-frequency content of one AFDM pilot and its echoes, before and after analog de-chirping and sampling

bandwidth), an alternative consists in first de-chirping the received signal in the analog domain with a continuous-time version [50] of a DAFT chirp carrier e.g., of the 0-th chirp $\left(e^{\imath 2\pi(c_2 0^2+\frac{1}{N}0n+c_1 n^2)}\right)_n$. The result is a multi-tone signal (as shown in Fig. 4.1 in the case of $N_\mathrm{p}=1$ and $P=2$) with discontinuities due to the frequency wrapping characterizing AFDM chirp carriers. In this figure, the de-chirped signal occupies two disjoint frequency bands that get merged into one (without discontinuities) thanks to spectrum folding after sampling at rate $f_\mathrm{s}=\frac{(L-1)P+1}{T}$. In the general case of $N_\mathrm{p}\geq 1$ pilots, if we restrict the total subset $\mathcal{P}$ of pilot guard indexes to be an interval, then sampling after de-chirping can be done at rate $f_\mathrm{s}=\frac{N_\mathrm{p}((L-1)P+1)}{T}$ to yield the vector $\mathbf{y}_\mathrm{p}$ used for target estimation. In most practical configurations $\frac{N_\mathrm{p}((L-1)P+1)}{T}\ll\frac{N}{T}=\frac{1}{\Delta t}$, and hence the sampling rate needed for AFDM sensing is significantly smaller than what is needed in sensing based on OFDM or OTFS waveforms.

The gain that can be achieved in terms of sampling rate reduction when AFDM is

employed for sub-Nyquist sensing. This gain is illustrated (for the same setting as Fig. 2.7b) by Table 4.1.

Table 4.1: Minimal sampling rate at sensing receiver

Waveform	OFDM	OTFS	AFDM
Sampling rate f_{s} (MHz)	$30 = \mathrm{BW}$	$30 = \mathrm{BW}$	$3.45 = \frac{N_{\mathrm{P}}((L-1)P+1)}{T}$

4.3 Conclusion

In this chapter, we proposed a sensing solution with simplified analog-domain processing, specifically chirp carrier mixing, for the case where the ISAC signal was generated using AFDM waveform. AFDM, based on the Discrete Affine Fourier Transform (DAFT), had been shown to achieve full diversity over doubly dispersive channels, making it a strong candidate for efficient sensing applications. To further reduce the sampling rate requirements, we described a sensing receiver architecture that took advantage of both the multi-chirp nature of AFDM and the optimization of the DAFT chirp rate c_1 (through parameter P) to minimize the number of required pilots. This approach enabled accurate delay-Doppler estimation while significantly lowering hardware complexity. The proposed framework applied to both mono-static and bi-static sensing scenarios, where the receiver could be co-located with the AFDM transmitter or positioned remotely, such as in a base station-to-terminal configuration. By integrating AFDM's unique waveform properties with optimized sensing strategies, our method provided a practical and efficient solution for next-generation radar and wireless sensing applications.

Chapter 5

Conclusion

In this thesis, we have investigated the problem of waveform design for sparse delay-Doppler channel estimation in high-mobility wireless communication systems. Our study focused on developing and analyzing efficient estimation techniques for DS-LTV channels, leveraging both on-grid and off-grid approximations. The goal was to enhance channel estimation accuracy while minimizing pilot overhead, crucial for next-generation wireless technologies. In the first part, we explored DS-LTV channel estimation under an on-grid assumption, where Doppler shifts align with a predefined grid. We introduced three different sparsity models and analyzed their impact on estimation performance. By leveraging proposed framework, we demonstrated that AFDM provides superior performance compared to conventional waveforms such as OFDM and OTFS. The key conclusions from this part are:

- When the delay-Doppler sparsity support is known at the receiver side, the asymptotic analysis of the MSE associated with sparse channel estimation can be leveraged to compare the minimal pilot overhead requirement of different wireless waveforms. Using this approach, we demonstrated that waveforms differ in their capacity to exploit delay-Doppler sparsity, with AFDM offering an advantage over OTFS, OFDM and SCM.

- When the delay-Doppler sparsity support is not known at the receiver side, the DS-LTV channel estimation problem can be effectively transformed into a structured compressed sensing problem, enabling significant pilot overhead reduction. In this context, pilot overhead comparison of the different waveforms can be analyzed through its effect on performance criteria relevant to compressed sensing problems. Again, it turns out that the AFDM-based approach exploits delay-Doppler sparsity more efficiently than traditional waveforms, leading to improved estimation accuracy.

In the second part, we extended our study to off-grid DS-LTV channel estimation, addressing the challenge of mismatches between actual Doppler shifts and predefined grid points. We proposed novel off-grid approximation techniques using multiple shifted elementary DPSS BEMs, leading to enhanced estimation robustness. Additionally, we investigated the application of AFDM for sparse channel extrapolation and prediction. The key conclusions from this part include:

- Off-grid channel estimation significantly improves accuracy in practical scenarios where Doppler shifts do not perfectly align with predefined grid points. moreover, the particular off-grid paradigm we propose has the advantage, thanks to its reliance on an elementary DPSS basis, of offering channel representation precision guarantees with a dependence on the basis size that can be analytically established.

- Thanks to the inherent extrapolation capability of the DPSS basis, the proposed extrapolation techniques enable efficient channel prediction, crucial for adaptive transmission in high-mobility networks.

- AFDM continues to outperform existing methods in handling off-grid Doppler shifts, confirming its suitability for next-generation communication systems.

Beyond channel estimation, we explored further applications of our findings in radar and sensing systems. By leveraging delay-Doppler sparsity, we demonstrated the feasibility of AFDM based sub-Nyquist radar techniques that reduce sampling rates while preserving

detection accuracy. This highlights the broader impact of our research beyond traditional communication scenarios.

In summary, this thesis provides a comprehensive framework for efficient DS-LTV channel estimation, contributing to the development of robust, low-overhead communication techniques for high-mobility environments. Our findings pave the way for future advancements in adaptive waveform design, sparse channel estimation, and integrated communication-sensing applications.

Appendices

Appendix A

Chapter 2 Appendices

A.1 Proof of Lemma 2.3.1

We only consider $P = 1$. The proof for $P > 1$ follows the same arguments. For any $k \in [\![-Q..L-1+Q]\!]$, define

$$\begin{aligned} \mathcal{Q}_k &\triangleq \{l \in [\![0..L-1]\!] \text{s.t.} \exists q \in [\![-Q..Q]\!], q+l=k\} \\ &= [\![k-Q..k+Q]\!] \cap [\![0..L-1]\!] \triangleq [\![l_{k,\max}..l_{k,\min}]\!] \end{aligned} \tag{A.1}$$

For any $N_\mathrm{p} \in [\![0..2Q+1]\!]$ define $\mathcal{L}_{k,N_\mathrm{p}} \triangleq \begin{pmatrix} \mathcal{Q}_k \\ N_\mathrm{p} \end{pmatrix}$ as the set of all N_p-size subsets of $\mathcal{Q}_k$. Then $\left|\mathcal{L}_{k,N_\mathrm{p}}\right| = \begin{pmatrix} |\mathcal{Q}_k| \\ N_\mathrm{p} \end{pmatrix}$ and

$$\begin{aligned}
&\mathbb{P}[X_k = N_\mathrm{p}] = \\
&\mathbb{P}[\bigcup_{\substack{(l_1,\ldots,l_{N_\mathrm{p}}) \\ \in \mathcal{L}_{k,N_\mathrm{p}}}} \{\bigcap_{\substack{l \in \\ \{l_1,\ldots,l_{N_\mathrm{p}}\}}} \mathcal{I}_{l,l-k} \cap \bigcap_{\substack{l \in \mathcal{Q}_k \setminus \\ \{l_1,\ldots,l_{N_\mathrm{p}}\}}} \overline{\mathcal{I}}_{l,l-k}\}] \\
&= \sum_{\substack{(l_1,\ldots,l_{N_\mathrm{p}}) \\ \in \mathcal{L}_{k,N_\mathrm{p}}}} \mathbb{P}[\bigcap_{\substack{l \in \\ \{l_1,\ldots,l_{N_\mathrm{p}}\}}} \mathcal{I}_{l,l-k} \cap \bigcap_{\substack{l \in \mathcal{Q}_k \setminus \\ \{l_1,\ldots,l_{N_\mathrm{p}}\}}} \overline{\mathcal{I}}_{l,l-k}] \\
&= \sum_{\substack{(l_1,\ldots,l_{N_\mathrm{p}}) \\ \in \mathcal{L}_{k,N_\mathrm{p}}}} \prod_{\substack{l \in \\ \{l_1,\ldots,l_{N_\mathrm{p}}\}}} \mathbb{P}[\mathcal{I}_{l,l-k}] \prod_{\substack{l \in \mathcal{Q}_k \setminus \\ \{l_1,\ldots,l_{N_\mathrm{p}}\}}} \mathbb{P}[\overline{\mathcal{I}}_{l,l-k}]
\end{aligned} \tag{A.2}$$

where the second equality follows because the terms of the union are all disjoint events and where the third equality is due to the independence property established by Assumption 2.1.1 (in each term of the sum in the right-hand side of the second equality in (A.2), each pair of events is either $(\mathcal{I}_{l_1,q_1}, \mathcal{I}_{l_2,q_2})$, $\left(\mathcal{I}_{l_1,q_1}, \overline{\mathcal{I}}_{l_2,q_2}\right)$ or $\left(\overline{\mathcal{I}}_{l_1,q_1}, \overline{\mathcal{I}}_{l_2,q_2}\right)$ with $l_1 \neq l_2$ and $q_1 \neq q_2$).

If $k \in [\![Q..L-1-Q]\!]$, $l_{k,\min} = k - Q$ and $l_{k,\max} = k + Q$ i.e., $|\mathcal{Q}_k| = 2Q+1$ and $\left|\mathcal{Q}_k \setminus \{l_1, \ldots, l_{N_\mathrm{p}}\}\right| = 2Q + 1 - N_\mathrm{p}$ as shown in Figure A.1a.

Since $\mathbb{P}[\mathcal{I}_{l,l-k}] = p_\mathrm{d} p_\mathrm{D}$ and $\mathbb{P}[\overline{\mathcal{I}}_{l,l-k}] = 1 - p_\mathrm{d} p_\mathrm{D}$ due to Definition 2.1.1, we get

$$\mathbb{P}\left[X_k = N_\mathrm{p}\right] = \begin{pmatrix} 2Q+1 \\ N_\mathrm{p} \end{pmatrix} \left(p_\mathrm{d} p_\mathrm{D}\right)^{N_\mathrm{p}} \left(1 - p_\mathrm{d} p_\mathrm{D}\right)^{2Q+1-N_\mathrm{p}} . \tag{A.3}$$

Thus, $\forall k \in [\![Q..L-Q]\!], X_k \sim \mathrm{B}(2Q+1, p_\mathrm{d} p_\mathrm{D})$. If $F(n,k,p)$ is the cumulative distribution

function (CDF) of $\mathrm{B}(n,p)$ then

$$\mathbb{P}\left[X_k > N_\mathrm{p}\right] = 1 - F(2Q+1, N_\mathrm{p}, p_\mathrm{d}p_\mathrm{D}), \quad \forall k \in [\![Q..L-Q]\!]\,. \tag{A.4}$$

If $k \in [\![-Q..Q-1]\!] \cup [\![L-Q..L-1+Q]\!]$ then $l_{k,\min} = k-Q$ and $l_{k,\max} = k+Q$ cannot be both satisfied and $|\mathcal{Q}_k| < 2Q+1$ as shown in Figure A.1b. Define $\rho_k \triangleq |\mathcal{Q}_k|$. Either $\rho_k < N_\mathrm{p}$, in which case $\mathcal{L}_{k,N_\mathrm{p}} = \emptyset$ and $\mathbb{P}\left[X_k = N_\mathrm{p}\right] = 0$ or $M \leq \rho_k < 2Q+1$, in which case $\left|\mathcal{Q}_k \setminus \{l_1, \ldots, l_{N_\mathrm{p}}\}\right| = \rho_k - N_\mathrm{p}$ and

$$\mathbb{P}\left[X_k = N_\mathrm{p}\right] = \binom{\rho_k}{N_\mathrm{p}} (p_\mathrm{d}p_\mathrm{D})^{N_\mathrm{p}} (1-p_\mathrm{d}p_\mathrm{D})^{\rho_k - N_\mathrm{p}}\,. \tag{A.5}$$

We thus have $X_k \sim \mathrm{B}(\rho_k, p_\mathrm{d}p_\mathrm{D})$ leading to

$$\begin{aligned} \mathbb{P}\left[X_k > N_\mathrm{p}\right] &= 1 - F(\rho_k, N_\mathrm{p}, p_\mathrm{d}p_\mathrm{D}) \\ &\leq 1 - F(2Q+1, N_\mathrm{p}, p_\mathrm{d}p_\mathrm{D}) \\ \forall k \in [\![-Q..Q-1]\!] &\cup [\![L-Q..L-1+Q]\!]\,. \end{aligned} \tag{A.6}$$

The inequality in (A.6) follows from the decreasing monotonicity property in n of the CDF of the binomial distribution $\mathrm{B}(n,k,p)$. Combining (A.4) and (A.6) gives us a uniform upper bound on the CCDF of X_k for any $k \in [\![-Q..L-1+Q]\!]$.

A.2 Proof of Theorem 2.3.1

First, write

$$\begin{aligned} \mathbb{P}\left[X_k > N_\mathrm{p}\right] \leq& \mathbb{P}\left[B > N_\mathrm{p}\right] \\ \leq& \left(\frac{p_d p_D}{N_\mathrm{p}/(2\lceil \frac{Q}{P} \rceil + 1)}\right)^{N_\mathrm{p}} \times \quad \left(\frac{1-p_d p_D}{1-N_\mathrm{p}/(2\lceil \frac{Q}{P} \rceil + 1)}\right)^{2\lceil \frac{Q}{P} \rceil + 1 - N_\mathrm{p}} \end{aligned} \tag{A.7}$$

where the first inequality is due to Lemma 2.3.1 and the third is due to the Chernoff's bound applied to $\mathrm{B}\left(2\lceil\frac{Q}{P}\rceil+1, p_d p_D\right)$. Inserting $N_\mathrm{p} = O(\log K)$ in (A.7), it can be shown after some manipulations that the right-hand side of (A.7) is $O\left(\frac{1}{K}\right)$ and hence $\mathbb{P}\left[X_k > N_\mathrm{p}\right] = O\left(\frac{1}{K}\right)$.

Next, define $\Delta_{l,q} \triangleq |\alpha_{l,q} I_{l,q} - \hat{\alpha}_{l,q} I_{l,q}|$, $k_{l,q} \triangleq q - 2c_1 N l$ and note that the MSE writes now as $\sum_{l=0}^{L-1}\sum_{q=-Q}^{Q}\mathbb{E}[\Delta_{l,q}^2]$ with

$$\begin{aligned}
&\mathbb{E}[\Delta_{l,q}^2] = \mathbb{E}[\Delta_{l,q}^2 | I_{l,q}=0]\mathbb{P}[I_{l,q}=0] + \\
&\mathbb{E}[\Delta_{l,q}^2 | I_{l,q}=1, 0 < X_{k_{l,q}} \leq N_\mathrm{p}]\mathbb{P}[I_{l,q}=1, 0 < X_{k_{l,q}} \leq N_\mathrm{p}] + \\
&\mathbb{E}[\Delta_{l,q}^2 | I_{l,q}=1, X_{k_{l,q}} > N_\mathrm{p}]\mathbb{P}[I_{l,q}=1, X_{k_{l,q}} > N_\mathrm{p}] \,.
\end{aligned} \tag{A.8}$$

The first term in the right-hand side of (A.8) is straightforwardly zero. As for the second term, let $\boldsymbol{\alpha}_{k_{l,q}}$ be the $X_{k_{l,q}}$-long vector of complex gains $\alpha_{\tilde{l},\tilde{q}}$ satisfying $\tilde{q} - 2Nc_1\tilde{l} = k_{l,q}$ of which $\alpha_{l,q}$ occupies the $i_{l,q}$-th entry. Recalling the pilot pattern definition in (2.13) and the signal input-output relation in (2.12), the pilot samples related to $\alpha_{\tilde{l},\tilde{q}}$ are received at DAFT domain positions $\{(m_p + k_{l,q})_N\}_{p=1\cdots N_\mathrm{p}}$ where m_p is the index of the p-th DAFT domain pilot symbol. Let $\mathbf{y}_{k_{l,q}}$ be the N_p-long vector formed by these samples. It relates to $\boldsymbol{\alpha}_{k_{l,q}}$

$$\mathbf{y}_{k_{l,q}} = \mathbf{M}_{k_{l,q}}^{(i)} \boldsymbol{\alpha}_{k_{l,q}} + \mathbf{w}_{k_{l,q}} \tag{A.9}$$

where is the vector composed of the entries associated with the received samples $\mathbf{y}_{k_{l,q}}$ of the noise vector $\mathbf{w}_\mathrm{p}$ from (2.14) and where matrix $\mathbf{M}_{k_{l,q}}^{(i)}$ is the $N_\mathrm{p} \times X_{k_{l,q}}$ measurement matrix associated with configuration i and formed by N_p rows and $X_{k_{l,q}}$ columns of the $|\mathcal{P}| \times L(2Q+1)$ matrix $\mathbf{A}_\mathcal{P}\mathbf{M}$ (see (2.14)) where, we recall, $|\mathcal{P}| = N_\mathrm{p}(2|c_1|N(L-1) + 2Q + 1) = N_\mathrm{p}(P(L-1) + 2Q + 1)$. Let $\hat{\boldsymbol{\alpha}}_{k_{l,q}}$ be the MMSE estimate of $\boldsymbol{\alpha}_{k_{l,q}}$ based on $\mathbf{y}_{k_{l,q}}$. Using the law of total expectation and the expression of the error covariance of the

MMSE estimator [36, (12.29)] we obtain

$$\begin{aligned}&\mathbb{E}\left[\left\|\boldsymbol{\alpha}_{k_{l,q}}-\hat{\boldsymbol{\alpha}}_{k_{l,q}}\right\|^2 |I_{l,q}=1, 0<X_{k_{l,q}}\leq N_{\mathrm{p}}\right]\\&=\sum_{i=1}^{C_{N_{\mathrm{p}}}}\left(\frac{1}{\sigma_w^2}\left(\mathbf{M}_{k_{l,q}}^{(i)}\right)^{\mathrm{H}}\mathbf{M}_{k_{l,q}}^{(i)}+\frac{1}{\sigma_\alpha^2}\mathbf{I}\right)^{-1}p^{(i)}\end{aligned}\tag{A.10}$$

where $C_{N_{\mathrm{p}}}\triangleq\sum_{X_{k_{l,q}}=1}^{N_{\mathrm{p}}}\begin{pmatrix}|\mathcal{Q}_{k_{l,q}}|-1\\X_{k_{l,q}}-1\end{pmatrix}\leq\sum_{X_{k_{l,q}}=1}^{N_{\mathrm{p}}}\begin{pmatrix}2\lceil\frac{Q}{P}\rceil\\X_{k_{l,q}}-1\end{pmatrix}$ is the number of possible delay-Doppler configurations leading to $0<X_{k_{l,q}}\leq N_{\mathrm{p}}$ and $I_{l,q}=1$ and $p^{(i)}$ is the probability of the i-th configuration among them so that $\sum_{i=1}^{C_{N_{\mathrm{p}}}}p^{(i)}=O(1)$. We thus have

$$\begin{aligned}&\mathbb{E}\left[\Delta_{l,q}^2|I_{l,q}=1, 0<X_{k_{l,q}}\leq N_{\mathrm{p}}\right]\\&=\sum_{i=1}^{C_{N_{\mathrm{p}}}}\sigma_w^2\left[\left(\left(\mathbf{M}_{k_{l,q}}^{(i)}\right)^{\mathrm{H}}\mathbf{M}_{k_{l,q}}^{(i)}+\frac{\sigma_w^2}{\sigma_\alpha^2}\mathbf{I}\right)^{-1}\right]_{(i_{l,q},i_{l,q})}p^{(i)}\end{aligned}\tag{A.11}$$

From the definition of matrix $\mathbf{M}$ in (2.15), (2.16) and the formula for pilot symbols amplitude in (2.19), it can be shown that $\frac{1}{\sqrt{P(L-1)+2Q+1}}\mathbf{M}_{k_{l,q}}^{(i)}$ has full rank in the case $X_{k_{l,q}}\leq N_{\mathrm{p}}$ with singular values that are bounded from below by a non-zero constant[1] uniformly in (l,q), i and K. This, along with the assumption $P=O\left(K^{(\kappa_d+\kappa_D-1)_+}\right)$ made in the theorem statement, leads to $\left(\mathbf{M}_{k_{l,q}}^{(i)}\right)^{\mathrm{H}}\mathbf{M}_{k_{l,q}}^{(i)}$ having eigenvalues that are $O\left(P(L-1)+2Q+1\right)=O(K^{\kappa_d+\kappa_D})\ \forall i$. The asymptotic order of the right-hand side of (A.11) is thus $\frac{\sigma_\alpha^2\sigma_w^2}{K^{\kappa_d+\kappa_D}\sigma_\alpha^2+\sigma_w^2}$. Combining this fact with (2.6) we conclude that there exists $C_{1,1}$ such that for K large enough and σ_w^2 small enough the following holds uniformly in

[1] Assuming for simplicity that N is an integer multiple of N_{p} and that the pilots indexes are dispersed uniformly throughout the AFDM symbol with a minimal inter-pilot interval of size $\frac{N}{N_{\mathrm{p}}}$, it is straightforward to show that $\frac{1}{\sqrt{P(L-1)+2Q+1}}\mathbf{M}_{k_{l,q}}^{(i)}$ is, up to a complex scalar multiplication, a sub-matrix of the N_{p}-point DFT matrix composed of $X_{k_{l,q}}$ of its columns

l, q

$$\mathbb{E}\left[\Delta_{l,q}^2 | I_{l,q} = 1, 0 < X_{k_{l,q}} \leq N_{\mathrm{p}}\right] \leq \frac{C_{1,1}\sigma_w^2}{K^{\kappa_d+\kappa_D}} \tag{A.12}$$

Now note that since $\mathbb{P}\left[I_{l,q} = 1, 0 < X_{k_{l,q}} \leq N_{\mathrm{p}}\right] \leq \mathbb{P}\left[I_{l,q} = 1\right] = p_d p_D$ then there exists $C_{1,2}$ such that for K large enough the following holds $\forall (l, q)$

$$\mathbb{P}\left[I_{l,q} = 1, 0 < X_{k_{l,q}} \leq N_{\mathrm{p}}\right] \leq C_{1,2} K^{\kappa_d+\kappa_D-2}\,. \tag{A.13}$$

In the third term of the right-hand side of (A.8),

$$\begin{aligned}
\mathbb{E}\left[\Delta_{l,q}^2 | I_{l,q} = 1, X_{k_{l,q}} > N_{\mathrm{p}}\right] &= \mathbb{E}\left[|\alpha_{l,q} I_{l,q}|^2 \, | I_{l,q} = 1, X_{k_{l,q}} > N_{\mathrm{p}}\right] \\
&= \mathbb{E}\left[|\alpha_{l,q}|^2\right] \\
&= \frac{1}{p_d L p_D (2Q+1)} \\
&\leq \frac{C_{2,1}}{K^{\kappa_d+\kappa_D}}
\end{aligned} \tag{A.14}$$

where the first equality holds because whenever the problem of estimating the $X_{k_{l,q}}$ unknown complex gains appearing at position $k_{l,q}$ is infeasible (due to a number N_{p} of pilots smaller than the number of unknowns $X_{k_{l,q}}$ and hence to rank-deficient $\mathbf{A}_{k_{l,q}}^{(i)}$) we can set $\hat{\alpha}_{l,q} = 0$, the second equality holds because $\alpha_{l,q}$ is independent from $\left\{I_{\tilde{l},\tilde{q}}\right\}_{(\tilde{l},\tilde{q}) \neq (l,q)}$, the third equality is due to (2.6) and the inequality on the last line holds for sufficiently large K uniformly in l, q due to Assumption 2.3.1. Also, there exists $C_{2,2}$ such that for K large enough, the following holds uniformly in (l, q)

$$\begin{aligned}
\mathbb{P}\left[I_{l,q} = 1, X_{k_{l,q}} > N_{\mathrm{p}}\right] &= \mathbb{P}\left[X_{k_{l,q}} > N_{\mathrm{p}} | I_{l,q} = 1\right] \mathbb{P}\left[I_{l,q} = 1\right] \\
&\leq C_{2,2} \frac{1}{K} K^{\kappa_d+\kappa_D-2} = \frac{C_{2,2} K^{\kappa_d+\kappa_D}}{K^3}
\end{aligned} \tag{A.15}$$

This is the case because on the one hand and for any l, q, $\mathbb{P}\left[I_{l,q} = 1\right] = p_d p_D =$

$O\left(K^{\kappa_d+\kappa_D-2}\right)$. On the other hand, $\mathbb{P}\left[X_{k_{l,q}} > N_{\mathrm{p}} | I_{l,q}=1\right] = \mathbb{P}\left[\tilde{X}_{k_{l,q}} > N_{\mathrm{p}}-1\right]$ where $\tilde{X}_{k_{l,q}}$ is a random variable with a CCDF that is upper-bounded by the CCDF of a $\mathrm{B}\left(2\lceil\frac{Q}{P}\rceil, p_d p_D\right)$ distribution (instead of $\mathrm{B}\left(2\lceil\frac{Q}{P}\rceil+1, p_d p_D\right)$ for $X_{k_{l,q}}$), a result that can be proved using the same steps as the proof of Lemma 2.3.1 while replacing $\mathcal{Q}_k$ with $\mathcal{Q}_{k_{l,q}} \setminus \{l\}$ and $\mathcal{L}_k$ with $\begin{pmatrix} \mathcal{Q}_{k_{l,q}} \setminus \{l\} \\ N_{\mathrm{p}}-1 \end{pmatrix}$. It thus holds that $\mathbb{P}\left[\tilde{X}_{k_{l,q}} > N_{\mathrm{p}}-1\right] = O\left(\frac{1}{K}\right)$. Putting all pieces together, we get $\mathbb{E}\left[\Delta_{l,q}^2\right] \leq C_1 \frac{\sigma_w^2}{K^2} + C_2 \frac{1}{K^3}$ where $C_1 \triangleq C_{1,1}C_{1,2}$ and $C_2 \triangleq C_{2,1}C_{2,2}$. Recalling that $L, Q = O(K)$ gives that there exists a constant C such that for K large enough the following holds $\forall \sigma_w^2$

$$\begin{aligned}\sum_{l=0}^{L-1}\sum_{q=-Q}^{Q}\mathbb{E}\left[\Delta_{l,q}^2\right] &\leq \sum_{l=0}^{L-1}\sum_{q=-Q}^{Q}\left(\frac{C_1\sigma_w^2}{K^2}+\frac{C_2}{K^3}\right) \\ &\leq C\left(\sigma_w^2+\frac{1}{K}\right). \end{aligned} \tag{A.16}$$

Letting σ_w^2 tend to zero, we get

$$\begin{aligned}&\lim_{\sigma_w^2\to 0}\lim_{K\to\infty}\mathbb{E}\left[\sum_{l=0}^{L-1}\frac{1}{N}\sum_{n=0}^{N-1}\left|h_{l,n}-\hat{h}_{l,n}\right|^2\right] = \\ &\lim_{\sigma_w^2\to 0}\lim_{K\to\infty}\mathbb{E}\left[\sum_{l=0}^{L-1}\sum_{q=-Q}^{Q}\Delta_{l,q}^2\right] \leq \lim_{\sigma_w^2\to 0} C\sigma_w^2 = 0\end{aligned} \tag{A.17}$$

This proves that the MSE tends to zero when the number of pilots $N_{\mathrm{p\,min}} = O(\log K)$. This number of pilots, each costing $(L-1)P+2Q+1$ samples, results since $P = O(K^{(\kappa_{\mathrm{d}}+\kappa_{\mathrm{D}}-1)_+})$ in a total overhead $N_{\mathrm{p\,min}}\left((L-1)P+2Q+1\right) = O\left(K^{\kappa_{\mathrm{d}}+\kappa_{\mathrm{D}}}\log K\right)$. This completes the proof of Theorem 2.3.1.

A.3 Proof of Lemma 2.4.1

Let $S_{\mathrm{d}} \triangleq \sum_{l=0}^{L-1} I_l$ be the random variable representing the number of active delay taps. From Definition 2.1.1 and Assumption 2.1.1, $S_{\mathrm{d}} \sim \mathrm{B}\left(L, p_{\mathrm{d}}\right)$ since it is the sum of

independent Bernoulli random variables $I_l \sim \text{Bernoulli}\,(p_\text{d})$. Applying Chernoff's bound to S_d evaluated at $s_\text{d} = (1+\epsilon)Lp_\text{d}$ (with an $\epsilon > 0$ that can be set as small as needed) gives after some algebraic manipulations

$$\mathbb{P}\left[S_\text{d} > s_\text{d}\right] \leq \left(\frac{p_\text{d}}{\frac{s_\text{d}}{L}}\right)^{s_\text{d}} \left(\frac{1-p_\text{d}}{1-\frac{s_\text{d}}{L}}\right)^{L-s_\text{d}} = e^{-\Omega(Lp_\text{d})}. \tag{A.18}$$

As for $S_{\text{D},l}$, the *joint* sparsity of $\{I_q^{(l)}\}_{l=0\cdots L-1}$ follows from writing

$$\begin{aligned}
\mathbb{P}\left[\exists l, I_l = 1, S_{\text{D},l} > s_\text{D}\right] &\leq \sum_{l=0}^{L-1} \mathbb{P}\left[I_l = 1, S_{\text{D},l} > s_\text{D}\right] \\
&= \sum_{l=0}^{L-1} \mathbb{P}\left[S_{\text{D},l} > s_\text{D} | I_l = 1\right] \mathbb{P}\left[I_l = 1\right] \\
&= \sum_{l=0}^{L-1} \mathbb{P}\left[S_{\text{D},l} > s_\text{D} | I_l = 1\right] p_\text{d} \\
&\leq Lp_\text{d} \bar{F}_{\text{B}(2Q+1,p_\text{D})}\left(s_\text{D}\right) \\
&\leq \left(\frac{p_\text{D}}{\frac{s_\text{D}}{2Q+1}}\right)^{s_\text{D}} \left(\frac{1-p_\text{D}}{1-\frac{s_\text{D}}{2Q+1}}\right)^{2Q+1-s_\text{D}} \\
&= e^{-\Omega((2Q+1)p_\text{D})},
\end{aligned} \tag{A.19}$$

where the first inequality is due to the union bound, the second inequality is due to the fact that Assumption 2.1.1 upper-bounds its CCDF by that of a $\text{B}\left(2Q+1, p_\text{D}\right)$ distribution and the third inequality and the last equality follow from applying the Chernoff's bound to the latter evaluated at $s_\text{D} = (1+\epsilon)(2Q+1)p_\text{D}$.

Combining (A.18) and (A.19) completes the proof of the lemma.

A.4 Proof of Theorem 2.4.1

In the case of SCW, defining $\tilde{\alpha}_{l,q} \triangleq \alpha_{l,q} e^{\imath 2\pi \frac{lq}{N}}$, rearranging $\mathbf{y}_\text{p}$ in 2.22 into $\tilde{\mathbf{y}}_\text{p}$ (composed of L successive blocks, with the l-th blocks composed of the l-th sample in each of the

N_{p} pilot intervals) and assuming $2Q+1$ divides N and that the p-th pilot position (for any $p \in [\![1..N_{\mathrm{p}}]\!]$) satisfies $m_p = q_p \frac{N}{2Q+1}$ for some $q_p \in [\![0..2Q]\!]$, we can write $\tilde{\mathbf{y}}_{\mathrm{p}} = \widetilde{\mathbf{M}}_{\mathrm{p}}^{\mathrm{scw}} \tilde{\boldsymbol{\alpha}}$ where $\tilde{\boldsymbol{\alpha}}$ is the vectorized form of $\tilde{\alpha}_{l,q}$ and $\widetilde{\mathbf{M}}_{\mathrm{p}}^{\mathrm{scw}} \triangleq \mathbf{I}_L \otimes \left(\mathrm{diag}\left(\mathrm{p}_1, \ldots, \mathrm{p}_{N_{\mathrm{p}}}\right) \overline{\mathbf{F}}_{2Q+1,N_{\mathrm{p}}}\right)$. Here, $\overline{\mathbf{F}}_{2Q+1,N_{\mathrm{p}}}$ is the partial inverse Fourier measurement matrix formed from N_{p} rows of the $(2Q+1)$-point inverse DFT matrix. The HiRIP of $\widetilde{\mathbf{M}}_{\mathrm{p}}^{\mathrm{scw}}$ can thus be derived and proven to be equal to the value given in the theorem statement by using the known RIP of partial inverse Fourier measurement matrices followed by applying [37, Theorem 4] pertaining to the HiRIP of hierarchical measurement matrices having the Kronecker property. This completes the part of the proof related to SCM.

As for OFDM, it can be shown that the estimation problem has a measurement matrix $\widetilde{\mathbf{M}}_{\mathrm{p}}^{\mathrm{ofdm}} \triangleq \left(\mathrm{diag}\left(\mathrm{p}_1, \ldots, \mathrm{p}_{N_{\mathrm{p,f}}}\right) \mathbf{F}_{L,N_{\mathrm{p,f}}}\right) \otimes \overline{\mathbf{F}}_{2Q+1,N_{\mathrm{p,t}}}$ where $\mathbf{F}_{L,N_{\mathrm{p,f}}}$ is the partial Fourier measurement matrix formed from $N_{\mathrm{p,f}}$ rows of the L-point inverse DFT matrix. The value of the HiRIP of $\widetilde{\mathbf{M}}_{\mathrm{p}}^{\mathrm{ofdm}}$ given in the statement of the theorem follows thus from the RIP of the partial Fourier measurement matrix and the HiRIP result pertaining to Kronecker hierarchical measurements.

A.5 Proof of Theorem 2.4.2

First, out of the pilot samples set $\mathcal{P}$, consider the subset $\mathcal{P}_p$ associated with the p-th pilot symbol transmitted at the DAFT index m_p (Fig. 2.5). To homogenize the sensing signal model associated with edge samples and inner samples of $\mathcal{P}_p$, we apply two overlap-add operations: adding the samples received within the index interval $[\![m_p - Q..m_p - 1]\!]$ to those received within $[\![m_p + (L-1)P - Q..m_p + (L-1)P - 1]\!]$ and the samples received within $[\![m_p + (L-1)P + 1..m_p + (L-1)P + Q]\!]$ to those received within $[\![m_p + 1..m_p + Q]\!]$. Now, define

$$\mathcal{D}_l \triangleq \left\{(\tilde{l}, q) \text{s.t.} (q + P\tilde{l})_{(L-1)P+1} = l\right\} \tag{A.20}$$

as the set of delay-Doppler grid points that potentially contribute to the pilot sample received at DAFT domain index $l \in [\![m_p..m_p + (L-1)P]\!]$ (Fig. A.2) after the two overlap-add operations described above. Note that $\mathcal{D}_l$ does not depend on the pilot symbol index p and that it has a cardinality that does not change with l and which satisfies $|\mathcal{D}_l| \leq 2\lceil \frac{Q}{P} \rceil + 1$. Next, define $\boldsymbol{\alpha}_{\mathcal{D}_l} \triangleq [\alpha_{l,q}]_{(l,q)\in\mathcal{D}_l}$ and $\widetilde{\boldsymbol{\alpha}} \triangleq \begin{bmatrix} \boldsymbol{\alpha}_{\mathcal{D}_0}^{\mathrm{T}} & \cdots & \boldsymbol{\alpha}_{\mathcal{D}_{(L-1)P}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$. The entries of $\widetilde{\boldsymbol{\alpha}}$ are just a permutation of the entries of $\boldsymbol{\alpha}$ and estimating one of these vectors directly gives an estimate of the other. Now, it can be shown that when we set P as in the theorem and $\epsilon > 0$ as small as needed, then $\widetilde{\boldsymbol{\alpha}}$ is $(\tilde{s}_{\mathrm{d}}, \tilde{s}_{\mathrm{D}})$-hierarchically sparse with high probability.

$$\tilde{s}_{\mathrm{d}} = (L-1)P + 1, \quad \tilde{s}_{\mathrm{D}} = (1+\epsilon)\log(LP)\,. \tag{A.21}$$

Indeed, the first level (of size $(L-1)P+1$) of $\widetilde{\boldsymbol{\alpha}}$ is sensed without compression with a number of measurements equal to $(L-1)P+1$ while $\tilde{s}_{\mathrm{D}}$ can be determined thanks to Definition 2.1.1 and Assumption 2.1.1. Indeed, the latter assumption implies that random variables $\{I_q^{(l_1)}\}_{q=-Q\cdots Q}$ are independent from $\{I_q^{(l_2)}\}_{q=-Q\cdots Q}$ for any $l_1 \neq l_2$ and hence that $\tilde{S}_{\mathrm{D},l} \triangleq \sum_{(\tilde{l},q)\in\mathcal{D}_l} I_{\tilde{l},q}$ has a binomial distribution since it is the sum of mutually independent Bernoulli random variables. The same approach as in the proof of Lemma 2.4.1 can thus be applied to $\tilde{S}_{\mathrm{D},l}$. Now, we can write the signal model of sensing $\widetilde{\boldsymbol{\alpha}}$ as

$$\widetilde{\mathbf{y}}_{\mathrm{p}} = \widetilde{\mathbf{M}}_{\mathrm{p}}\widetilde{\boldsymbol{\alpha}} + \widetilde{\mathbf{w}}_{\mathrm{p}}, \tag{A.22}$$

where $\widetilde{\mathbf{y}}_{\mathrm{p}} = [\widetilde{\mathbf{y}}_{\mathrm{p},0}^{\mathrm{T}} \quad \cdots \quad \widetilde{\mathbf{y}}_{\mathrm{p},(L-1)P}^{\mathrm{T}}]^{\mathrm{T}}$. For each l, $\widetilde{\mathbf{y}}_{\mathrm{p},l}$ is a $N_{\mathrm{p}} \times 1$ vector composed of the pilot samples received at DAFT domain positions $\{m_p + l\}_{p=1\cdots N_{\mathrm{p}}}$. Note that by this definition $\tilde{\mathbf{y}}_{\mathrm{p}}$ is obtained by permuting $\mathbf{y}_{\mathrm{p}}$ in (2.22) in accordance with the permutation that gives $\tilde{\boldsymbol{\alpha}}$ from $\boldsymbol{\alpha}$. Next, by setting for each $p \in [\![1..N_{\mathrm{p}}]\!]$ $m_p = q_p \frac{N}{2\lceil \frac{Q}{P} \rceil + 1}$ for some

integer q_p it follows from (2.12) and (A.20) that $\widetilde{\mathbf{M}}_{\mathrm{p}}$ has the following Kronecker structure

$$\widetilde{\mathbf{M}}_{\mathrm{p}} = \mathbf{I}_{(L-1)P+1} \otimes \widetilde{\mathbf{M}}_{\mathcal{D}}, \tag{A.23}$$

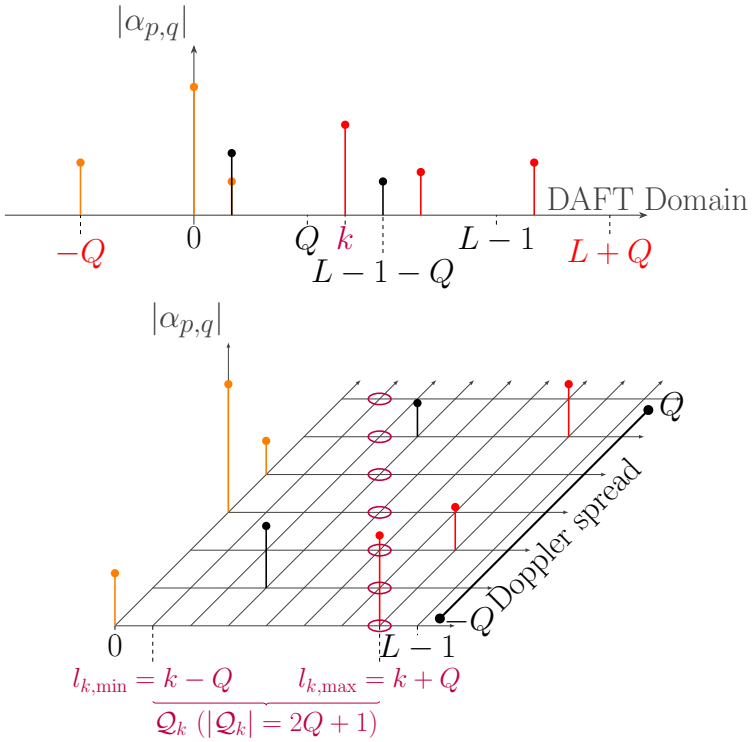
with $\widetilde{\mathbf{M}}_{\mathcal{D}} = \mathrm{diag}(\mathrm{p}_1 \cdots \mathrm{p}_{N_{\mathrm{p}}})\mathbf{F}_{2\lceil \frac{Q}{P} \rceil + 1, N_{\mathrm{p}}} \mathbf{\Psi}$, $\mathbf{F}_{2\lceil \frac{Q}{P} \rceil + 1, N_{\mathrm{p}}}$ is a $\left(2\lceil \frac{Q}{P} \rceil + 1\right) \times N_{\mathrm{p}}$ partial Fourier measurement matrix and $\mathbf{\Psi}$ is a diagonal matrix with unit-modulus entries. We can thus use [51, Theorem 4.5] pertaining to sub-sampled Fourier matrices to get that for sufficiently large L, Q, sufficiently small δ, and

$$N_{\mathrm{p}} > O\left(\frac{1}{\delta^2} \log^2 \frac{1}{\delta} \log \frac{\log(LP)}{\delta} \log(LP) \log \frac{Q}{P}\right) \tag{A.24}$$

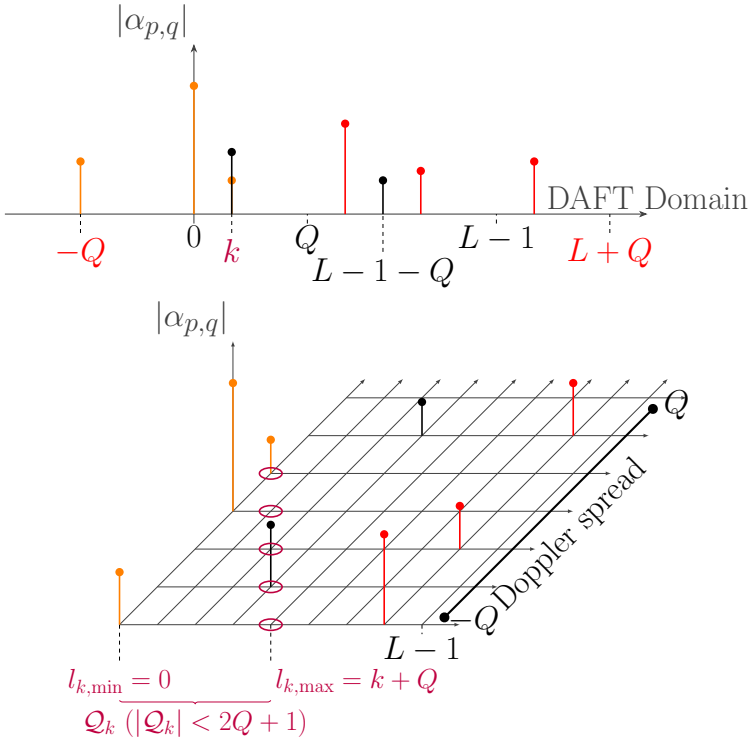
the RIP constant $\delta_{\tilde{s}_{\mathrm{D}}}$ of $\widetilde{\mathbf{M}}_{\mathcal{D}}$ satisfies $\delta_{\tilde{s}_{\mathrm{D}}} \leq \delta$ with probability $1 - e^{-\Omega\left(\log \frac{Q}{P} \log \frac{1}{\delta}\right)}$. The RIP of $\mathbf{I}_{(L-1)P+1}$ trivially satisfies $\delta_{\tilde{s}_{\mathrm{d}}} = 0$. As for the HiRIP of $\widetilde{\mathbf{M}}_{\mathrm{p}}$, we can apply [37, Theorem 4] to (A.23) thanks to its Kronecker structure to get that, f N_{p} and δ are as in (A.24), then

$$\delta_{s_{\mathrm{d}}, s_{\mathrm{D}}} \leq \delta_{\tilde{s}_{\mathrm{d}}} + \delta_{\tilde{s}_{\mathrm{D}}} + \delta_{\tilde{s}_{\mathrm{d}}} \delta_{\tilde{s}_{\mathrm{D}}} \leq \delta\,. \tag{A.25}$$

This completes the proof of the theorem.

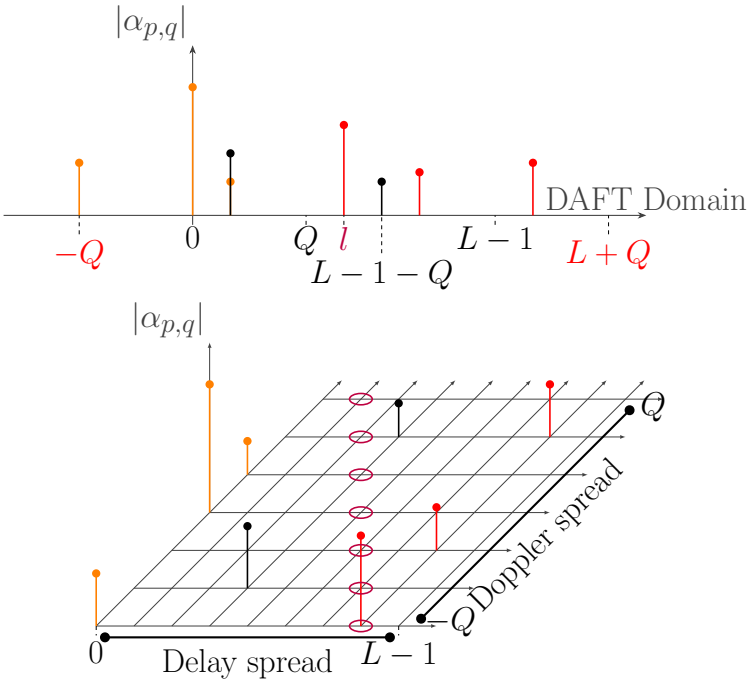


(a) $\rho_k \triangleq |\mathcal{Q}_k| = 2Q+1$

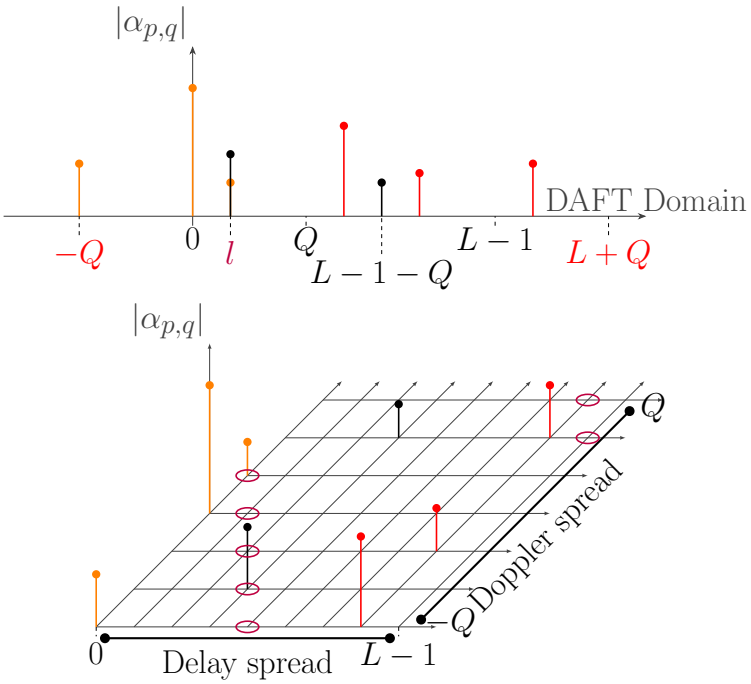


(b) $\rho_k < 2Q+1$

Figure A.1: Examples of interval $\mathcal{Q}_k$. Grid points surrounded by circles represent potential delay-Doppler taps that may appear at the k-th position in the DAFT domain.



(a) for an l resulting in a whole diagonal



(b) for an l resulting in a wrapped diagonal

Figure A.2: Two examples of the set $\mathcal{D}_l$. In each one of the two examples, the grid points forming $\mathcal{D}_l$ are shown surrounded by red rings. ($P = 1$, $m_p = 0$)

Appendix B

Chapter 3 Appendices

B.1 Proof of Theorem 3.3.1

We first need the following lemma taken from [52, Theorem 2.4].

Lemma B.1.1. *[52, Theorem 2.4] Let $h(t)$ be a continuous-time zero-mean wide-sense stationary random process with power spectrum $P_h(f) = \frac{1}{B}\mathrm{rect}_{[F_\mathrm{c}-B,F_\mathrm{c}+B]}(f)$.*

Denote by $\mathbf{h} = [h(0T_\mathrm{s}) \ \dots \ h((N-1)T_s)]^\mathrm{T}$ a vector of samples acquired from $h(t)$ with a sampling period $T_s \leq \frac{1}{2F_\mathrm{c}+B}$. Let $W = \frac{BT_s}{2}$, $\mathbf{U}_k$ the matrix form of the k first (N,W)-DPSS vectors, $\mathbf{E}_f \triangleq \mathrm{diag}\left(e^{\imath 2\pi f 0} \dots e^{\imath 2\pi f(N-1)}\right)$ and $\mathbf{P}_k \triangleq \mathbf{E}_{F_\mathrm{c}T_\mathrm{s}}\mathbf{U}_k\mathbf{U}_k^\mathrm{H}\mathbf{E}_{F_\mathrm{c}T_\mathrm{s}}^\mathrm{H}$. Then $\mathbb{E}\left[\|\mathbf{h}-\mathbf{P}_k\mathbf{h}\|_2^2\right] = \frac{1}{2W}\sum_{l=k}^{N-1}\lambda_l^{(N,W)}$.

In what follows, we use Lemma B.1.1 to upper bound $\mathbb{E}\left[\left|h_{l,q,n} - h_{l,q,n}^{\mathrm{BEM}}\right|^2\right]$. For that sake,

we rewrite $h_{l,q,n}$ as the sampled version of the continuous-time signal $h_{l,q}(t)$ defined as

$$h_{l,q}(t) \triangleq \sum_{i=1}^{N_\mathrm{D}} \alpha_{l,q,i} e^{\imath 2\pi f_i t}, \quad t \in \mathbb{R}. \tag{B.1}$$

with $f_i \triangleq \frac{q}{NT_\mathrm{s}} + \frac{\kappa_i}{NT_\mathrm{s}}$. To prove that the PSD of the random process $h_{l,q}(t)$ has the desired

property, we derive its autocorrelation function $R_{h_{l,q}}(\tau) \triangleq \mathbb{E}[h_{l,q}(t)h_{l,q}(t+\tau)^*]$ as

$$\begin{aligned} R_{h_{l,q}}(\tau) &= \sum_{i=1}^{N_D}\sum_{j=1}^{N_D} \mathbb{E}[\alpha_{l,q,i}\alpha_{l,q,j}^*]\,\mathbb{E}\left[e^{\imath 2\pi f_i t} e^{-\imath 2\pi f_j (t+\tau)}\right] \\ &= \sum_{i=0}^{N_D} \sigma_\alpha^2\,\mathbb{E}\left[e^{-\imath 2\pi f_i \tau}\right] \end{aligned} \tag{B.2}$$

where the first equality is due to (B.1) and the second to $f_i \sim \mathcal{U}\left(\left[\frac{q}{NT_\mathrm{s}} - \frac{1}{2NT_\mathrm{s}}, \frac{q}{NT_\mathrm{s}} + \frac{1}{2NT_\mathrm{s}}\right]\right)$, $\alpha_{l,q,i} \sim \mathcal{CN}(0, \sigma_\alpha^2)$ and the independence property of $\{\alpha_{l,q,i}\}_i$ as per Definition 3.1.1.

This gives:

$$R_{h_{l,q}}(\tau) = N_D\,\sigma_\alpha^2\, e^{-\imath 2\pi \frac{q}{NT_\mathrm{s}}\tau}\,\mathrm{sinc}\left(\frac{\tau}{NT_\mathrm{s}}\right). \tag{B.3}$$

The power spectral density (PSD) is thus

$$P_{h_{l,q}}(f) \triangleq \mathcal{F}\{R_{h_{l,q}}(\tau)\} = N_D\,\sigma_\alpha^2\,NT_\mathrm{s}\,\mathrm{rect}\left(\left(f - \frac{q}{NT_\mathrm{s}}\right)NT_\mathrm{s}\right). \tag{B.4}$$

The PSD of $h_{l,q}(t)$ thus satisfies the condition of Lemma B.1.1 with $B = \frac{1}{2NT_\mathrm{s}}$ and $F_\mathrm{c} = \frac{q}{NT_\mathrm{s}}$

giving

$$\mathbb{E}\left[\left|h_{l,q,n} - h_{l,q,n}^{\mathrm{BEM}}\right|^2\right] = \frac{1}{N}\mathbb{E}\left[\left\|\mathbf{h}_{l,q} - \mathbf{h}_{l,q}^{\mathrm{BEM}}\right\|^2\right] = \frac{N_\mathrm{D}\,\sigma_\alpha^2}{2WN}\sum_{b=Q_{\mathrm{BEM}}}^{N-1}\lambda_b^{(N,W)}. \tag{B.5}$$

Now define $\lambda_b^{(c)}$ as the b-th eigenvalue of the prolate spheroidal wave functions (PSWF) [53] with the bandwidth parameter $c \triangleq \pi NW$. This allows us to exploit existing results on the behavior of PSWF eigenvalues in the limit of $c \to \frac{\pi}{2}$ (or equivalently as W tends to zero at the rate $\frac{1}{2N}$) to upper bound the sum of DPSS eigenvalues $\lambda_b^{(N,W)}$ in (B.5). Indeed, due to [54, Theorem 2]

$$\lambda_b^{(N,W)} \leq A_W \lambda_b^{(c)}, \forall b = 1, \ldots, N\,. \tag{B.6}$$

Here, A_W is a function of W defined in [54, Eq. (45)] and whose image is fully included in the interval $\left[\frac{\pi^2}{8}, 2\right]$. Plugging (B.6) into (B.5) and noting that $2WN = 1$ lead to

$$\mathbb{E}\left[\left|h_{l,q,n} - h_{l,q,n}^{\mathrm{BEM}}\right|^2\right] \leq N_{\mathrm{D}}\,\sigma_\alpha^2\,A_W \sum_{b=Q_{\mathrm{BEM}}}^{N} \lambda_b^{(c)} \tag{B.7}$$

The right-hand side term in (B.7) can be upper bounded due to the fact that the PSWF eigenvalues decay at least exponentially[1] as b grows beyond $\frac{2\pi NW}{\pi} + O(\log(\pi NW)) = 1 + O(\log\frac{\pi}{2})$. More precisely, it follows from [56, Theorem 2.5] that $\lambda_b^{(c)} = O\left(e^{-\frac{\pi^2}{\log\frac{\pi}{2}}b}\right)$.

Plugging this into (B.7) results, for any $\epsilon > 0$ and $Q_{\mathrm{BEM}} > C\log\frac{2}{\epsilon}$ for sufficiently large C, in

$$\mathbb{E}\left[\left|h_{l,q,n} - h_{l,q,n}^{\mathrm{BEM}}\right|^2\right] < \frac{N_{\mathrm{D}}\,\sigma_\alpha^2\,A_W\epsilon}{2}\,. \tag{B.8}$$

Now, note that

$$\begin{aligned}\mathbb{E}\left[\sum_{l=0}^{L-1}\left|h_{l,n} - h_{l,n}^{\mathrm{BEM}}\right|^2\right] &= \mathbb{E}\left[\sum_{l=0}^{L-1}\left|\sum_{q=-Q}^{Q} I_{l,q}e^{\imath 2\pi\frac{nq}{N}}\sum_{b=1}^{Q_{\mathrm{BEM}}}\left(h_{l,q,n} - h_{l,q,n}^{\mathrm{BEM}}\right)\right|^2\right] \\ &= \sum_{l=0}^{L-1}\sum_{q=-Q}^{Q}\mathbb{E}\left[I_{l,q}\right]\mathbb{E}\left[\left|h_{l,q,n} - h_{l,q,n}^{\mathrm{BEM}}\right|^2\right]\end{aligned} \tag{B.9}$$

Plugging (B.8) into the right-hand side of (B.10) gives

$$\mathbb{E}\left[\sum_{l=0}^{L-1}\left|h_{l,n} - h_{l,n}^{\mathrm{BEM}}\right|^2\right] < \sum_{l=0}^{L-1}\sum_{q=-Q}^{Q}\mathbb{E}\left[I_{l,q}\right]\frac{N_{\mathrm{D}}\,\sigma_\alpha^2\,A_W\epsilon}{2} \leq \epsilon \tag{B.10}$$

where the second inequality is due to the fact that $A_W \leq 2$ and that $\sum_{l=0}^{L-1}\sum_{q=-Q}^{Q}\mathbb{E}\left[I_{l,q}\right] = \frac{1}{N_{\mathrm{D}}\,\sigma_\alpha^2}$ due to the power normalization condition in (3.3). This completes the proof of the theorem.

[1] Actually, even super-geometric decay can be proven [55]

B.2 Proof of Theorem 3.5.1

In what follows we use the notation $h^{\text{ext}}_{l,q,n}\left(\hat{\mathbf{h}}^{\text{BEM}}_{l,q}\right)$ to designate the DPSS extrapolation predictor defined in (3.23) to highlight its dependence on the estimated vector $\hat{\mathbf{h}}^{\text{BEM}}_{l,q}$. When $\hat{\mathbf{h}}^{\text{BEM}}_{l,q}$ in (3.23) is replaced with an arbitrary channel vector $\mathbf{h}$, the DPSS predictor generalizes to

$$h^{\text{ext}}_{l,q,n}\left(\mathbf{h}\right) \triangleq \underbrace{e^{\imath 2\pi \frac{nq}{N}} \left(\mathbf{u}^{\text{ext}}_{n}\right)^{\text{T}} \mathbf{U}^{\text{H}}_{Q_{\text{BEM}}} \mathbf{E}^{\text{H}}_{\frac{q}{N}}}_{\triangleq \left(\mathbf{f}^{\text{ext}}_{Q_{\text{BEM}}}\right)^{\text{H}}} \mathbf{h}\,. \tag{B.11}$$

For instance, $h^{\text{ext}}_{l,q,n}\left(\mathbf{h}^{\text{BEM}}_{l,q}\right)$ is the DPSS predictor based on the actual, not the estimated, vector $\mathbf{h}^{\text{BEM}}_{l,q}$. Next, we establish the link between the DPSS extrapolation predictor and MMSE prediction given the knowledge of the channel during the observation interval.

For that sake first note that for each l, q and any $n \in \mathbb{Z}$, the random variable $h_{l,q,n}$ as defined by (3.7) follows a complex symmetric Gaussian distribution $\mathcal{CN}\left(0, N_{\text{D}}\sigma^2_{\alpha}\right)$ under the conditions of Definition 3.1.1. Moreover, the random process $\left(h_{l,q,n}\right)_{n\in\mathbf{Z}}$ is stationary and has an auto-correlation $\mathbb{E}\left[h_{l,q,n}h^{*}_{l,q,m}\right] = \sigma^2_{\alpha}N_{\text{D}}e^{\imath 2\pi\frac{(n-m)q}{N}}\frac{N}{\pi(n-m)}\sin\left(\frac{\pi(n-m)}{N}\right)$ due to (3.7). Similarly, $\mathbf{h}_{l,q} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2_{\alpha}N_{\text{D}}N\mathbf{E}_{\frac{q}{N}}\mathbf{\Sigma}\mathbf{E}^{\text{H}}_{\frac{q}{N}}\right)$ and $\mathbf{h}^{\text{BEM}}_{l,q} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2_{\alpha}N_{\text{D}}N\mathbf{E}_{\frac{q}{N}}\mathbf{P}^{\text{BEM}}\mathbf{\Sigma}\mathbf{P}^{\text{BEM}}\mathbf{E}^{\text{H}}_{\frac{q}{N}}\right)$ with

$$\mathbf{\Sigma} \triangleq \left[\frac{1}{\pi(n-m)}\sin\frac{\pi(n-m)}{N}\right]_{\substack{n=0\ldots N-1\\ m=0\ldots N-1}} = \left[C^{(N,W)}_{n,m}\right]_{\substack{n=0\ldots N-1\\ m=0\ldots N-1}} \tag{B.12}$$

Therefore, the MMSE predictor of $h_{l,q,n}$ given the actual channel component $\mathbf{h}_{l,q}$ is

$$\begin{aligned}\hat{h}_{l,q,n}\left(\mathbf{h}_{l,q}\right) &= \mathbf{E}\left[h_{l,q,n}\mathbf{h}^{\text{H}}_{l,q}\right]\left(\mathbf{E}\left[\mathbf{h}_{l,q}\mathbf{h}^{\text{H}}_{l,q}\right]\right)^{-1}\mathbf{h}_{l,q}\\ &= e^{\imath 2\pi\frac{nq}{N}}\boldsymbol{\rho}\mathbf{E}_{\frac{q}{N}}\mathbf{\Sigma}^{-1}\mathbf{E}^{\text{H}}_{\frac{q}{N}}\mathbf{h}_{l,q}\,.\end{aligned} \tag{B.13}$$

Here $\boldsymbol{\rho} \triangleq \left[e^{-\imath 2\pi\frac{mq}{N}}C^{(N,W)}_{n,m}\right]^{N-1}_{m=0}$. It follows that the reduced-rank MMSE predictor $\hat{h}^{\text{RR}}_{l,q,n}$ of rank Q of $h_{l,q,n}$ given $\mathbf{h}_{l,q}$ (where notation 'RR' stands for "reduced rank") is [57]

$$\hat{h}^{\text{RR}}_{l,q,n}\left(\mathbf{h}_{l,q}\right) = \left(\mathbf{f}^{\text{RR}}_{Q}\right)^{\text{H}}\mathbf{h}_{l,q}\,, \tag{B.14}$$

$$\mathbf{f}_Q^{\mathrm{RR}} \triangleq \mathbf{E}_{\frac{q}{N}} \mathbf{U}_Q \mathrm{diag}\left(\frac{1}{\lambda_1^{(N,W)}}, \ldots, \frac{1}{\lambda_Q^{(N,W)}}\right) \mathbf{U}_Q^{\mathrm{H}} \mathbf{E}_{\frac{q}{N}}^{\mathrm{H}} \boldsymbol{\rho}^{\mathrm{H}}. \tag{B.15}$$

Next, we apply the triangle inequality to get

$$\begin{aligned}\mathbb{E}\left[\left|h_{l,q,n}^{\mathrm{ext}}\left(\hat{\mathbf{h}}_{l,q}^{\mathrm{BEM}}\right) - \hat{h}_{l,q,n}^{\mathrm{RR}}\left(\mathbf{h}_{l,q}\right)\right|^2\right] \leq \underbrace{\mathbb{E}\left[\left|h_{l,q,n}^{\mathrm{ext}}\left(\hat{\mathbf{h}}_{l,q}^{\mathrm{BEM}}\right) - h_{l,q,n}^{\mathrm{ext}}\left(\mathbf{h}_{l,q}^{\mathrm{BEM}}\right)\right|^2\right]}_{\triangleq E_1} + \\ \underbrace{\mathbb{E}\left[\left|h_{l,q,n}^{\mathrm{ext}}\left(\mathbf{h}_{l,q}^{\mathrm{BEM}}\right) - h_{l,q,n}^{\mathrm{ext}}\left(\mathbf{h}_{l,q}\right)\right|^2\right]}_{\triangleq E_2} + \underbrace{\mathbb{E}\left[\left|h_{l,q,n}^{\mathrm{ext}}\left(\mathbf{h}_{l,q}\right) - \hat{h}_{l,q,n}^{\mathrm{RR}}\left(\mathbf{h}_{l,q}\right)\right|^2\right]}_{\triangleq E_3}.\end{aligned} \tag{B.16}$$

Since $\lim_{\sigma_w^2 \to 0} \mathbb{E}\left[\left\|\hat{\mathbf{h}}_{l,q}^{\mathrm{BEM}} - \mathbf{h}_{l,q}^{\mathrm{BEM}}\right\|^2\right]$ due to Assumption 3.4.1, it follows from (B.11) by standard MSE derivations that $\lim_{\sigma_w^2 \to 0} E_1 = 0$. As for E_2, and since $\mathbf{h}_{l,q}^{\mathrm{BEM}} = \mathbf{E}_{\frac{q}{N}} \mathbf{P}^{\mathrm{BEM}} \mathbf{E}_{\frac{q}{N}}^{\mathrm{H}} \mathbf{h}_{l,q}$ per (3.10), we have that $\mathbf{U}_{Q_{\mathrm{BEM}}}^{\mathrm{H}} \mathbf{E}_{\frac{q}{N}}^{\mathrm{H}} \mathbf{h}_{l,q} = \mathbf{U}_{Q_{\mathrm{BEM}}}^{\mathrm{H}} \mathbf{E}_{\frac{q}{N}}^{\mathrm{H}} \mathbf{h}_{l,q}^{\mathrm{BEM}}$. It follows from (B.11) that $h_{l,q,n}^{\mathrm{ext}}\left(\mathbf{h}_{l,q}\right) = h_{l,q,n}^{\mathrm{ext}}\left(\mathbf{h}_{l,q}^{\mathrm{BEM}}\right)$ and hence that $E_2 = 0$. Finally, note by referring to (3.21), (B.11) and (B.15) that $\mathbf{f}_Q^{\mathrm{RR}} = \mathbf{f}_{Q_{\mathrm{BEM}}}^{\mathrm{ext}}$ if $Q = Q_{\mathrm{BEM}}$ leading to $E_3 = 0$.

Now, due to (3.21), $h_{l,n}^{\mathrm{ext}} = \sum_{l=0}^{L-1} I_{l,q} h_{l,q,n}^{\mathrm{ext}}$. Moreover, due to the independence conditions from Definition 3.1.1, $\hat{h}_{l,n}^{\mathrm{RR}} \triangleq \sum_{l=0}^{L-1} I_{l,q} \hat{h}_{l,q,n}^{\mathrm{RR}}$ is the reduced-rank MMSE estimate of $h_{l,n}$ given $\{\mathbf{h}_{l,q}\}_{q=-Q\cdots Q}$ and conditioned on a given realization of $I_{l,q}$. Putting all these pieces together, it follows that $\lim_{\sigma_w^2 \to 0} \left|h_{l,n}^{\mathrm{ext}} - \hat{h}_{l,n}^{\mathrm{RR}}\right|^2$. This completes the proof of the theorem.

Bibliography

[1] H.-F. Arraño-Scharager and C. Azurdia-Meza, "OFDM: Today and in the future of next generation wireless communications," 09 2016.

[2] Y. Zhou, H. Yin, J. Xiong, S. Song, J. Zhu, J. Du, H. Chen, and Y. Tang, "Overview and performance analysis of various waveforms in high mobility scenarios," in *2024 7th International Conference on Communication Engineering and Technology (ICCET)*, 2024, pp. 35–40.

[3] O. K. Rasheed, G. D. Surabhi, and A. Chockalingam, "Sparse delay-doppler channel estimation in rapidly time-varying channels for multiuser otfs on the uplink," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–5.

[4] A. J. G. F. Gómez-Cuba, "Compressed sensing channel estimation for OFDM with non-Gaussian multipath gains," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 47–61, 2020.

[5] A. A. K. Venugopal, J. R. W. Heath, and N. González-Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," *ICASSP*, p. 6493–6497, 2017.

[6] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[7] W. Anwar, A. Krause, A. Kumar, N. Franchi, and G. P. Fettweis, "Performance analysis of various waveforms and coding schemes in V2X communication scenarios," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–8.

[8] S. R. M. Fathima Jesbin and A. Chockalingam, "Sparse superimposed pilot based channel estimation in OTFS systems," in *WCNC*, 2023.

[9] L. Gaudio, G. Colavolpe, and G. Caire, "OTFS vs. OFDM in the presence of sparsity: A fair comparison," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4410–4423, 2022.

[10] A. Bemani, N. Ksairi, and M. Kountouris, "Affine frequency division multiplexing for next generation wireless communications," *IEEE Trans. Wireless Commun.*, 2023.

[11] ——, "AFDM: A full diversity next-generation waveform for high mobility communications," in *2021 IEEE Int. Conf. on Commun. Workshops (ICC Workshops)*, 2021, pp. 1–6.

[12] T. Erseghe, N. Laurenti, and V. Cellini, "A multicarrier architecture based upon the affine Fourier transform," *IEEE Trans. Commun.*, vol. 53, no. 5, pp. 853–862, May 2005.

[13] J. J. Healy, M. A. Kutay, H. M. Ozaktas, and J. T. Sheridan, *Linear canonical transforms: Theory and applications*. Springer, 2015.

[14] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 164–174, 2010.

[15] H. Groll *et al.*, "Sparsity in the delay-Doppler domain for measured 60 GHz vehicle-to-infrastructure communication channels," in *IEEE ICC Workshops*, 2019.

[16] C. K. Thomas and D. Slock, "BP-VB-EP based static and dynamic sparse Bayesian learning with kronecker structured dictionaries," in *IEEE ICASSP*, 2020, pp. 9095–9099.

[17] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *ICASSP*, 2008, pp. 2885–2888.

[18] V. Vahidi and E. Saberinia, "Compressed channel estimation methods for high mobility doubly selective channels in orthogonal frequency division multiplexing systems," *IET Communications*, vol. 13, no. 2, pp. 205–215, 2019. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-com.2018.5025

[19] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[20] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[21] R. Beinert, P. Jung, G. Steidl, and T. Szollmann, "Super-resolution for doubly-dispersive channel estimation," 2021. [Online]. Available: https://arxiv.org/abs/2101.11544

[22] C. Tian, J. Zhang, Q. Meng, and D. Wang, "Microwave staring correlated imaging using bomp based on adaptive meshing," *2016 CIE International Conference on Radar (RADAR)*, pp. 1–4, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:8167644

[23] Z. Lu, R. Ying, S. Jiang, Z. Zhang, P. Liu, and W. Yu, "Distributed compressed sensing off the grid," *IEEE Signal Processing Letters*, vol. 22, 07 2014.

[24] M. Eskandari and H. Bakhshi, "A gridless compressive sensing based channel estimation for millimeter wave massive mimo systems from 1-bit measurements," 2020. [Online]. Available: https://arxiv.org/abs/2003.10237

[25] N. H. Nguyen, K. Doğançay, H.-T. Tran, and P. E. Berry, "Parameter-refined omp for compressive radar imaging of rotating targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3561–3577, 2019.

[26] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Off-grid channel estimation with sparse bayesian learning for otfs systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7407–7426, 2022.

[27] A. Mohebbi, H. Abdzadeh-Ziabari, W. Zhu, and M. O. Ahmad, "A new channel estimation method for millimeter wave systems under high mobility," *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pp. 1–5, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:245015441

[28] Q. Qin, L. Gui, P. Cheng, and B. Gong, "Time-varying channel estimation for millimeter wave multiuser mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9435–9448, 2018.

[29] P. Cheng, Z. Chen, Y. Rui, Y. J. Guo, L. Gui, M. Tao, and Q. T. Zhang, "Channel estimation for OFDM systems over doubly selective channels: A distributed compressive sensing based approach," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4173–4185, 2013.

[30] ——, "Channel estimation for OFDM systems over doubly selective channels: A distributed compressive sensing based approach," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4173–4185, 2013.

[31] D. Fu, Y. Peng, and S. Zheng, "A compressive channel sensing method with optimal

thresholding for OFDM systems under fast fading channels," in *2016 IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–5.

[32] X. Li, X. He, Y. Liang, and Q. Wang, "Low complexity 2d off-grid otfs channel estimation in fractional delay-doppler scenarios," in *2023 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2023, pp. 1–5.

[33] W. Benzine, A. Bemani, N. Ksairi, and D. Slock, "Affine frequency division multiplexing for communications on sparse time-varying channels," in *IEEE GLOBECOM*, 2023.

[34] H. Shayanfar, W.-P. Zhu, and M. Swamy, "Compressed sensing based channel estimation for GFDM systems in high mobility scenario," in *IEEE SPAWC*, 2023, pp. 266–270.

[35] K.-C. Hung and D. W. Lin, "Pilot-based LMMSE channel estimation for OFDM systems with power–delay profile approximation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 150–159, January 2010.

[36] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory.* Prentice-Hall PTR, 2013.

[37] I. Roth, A. F. Martin Kliesch, G. Wunder, and J. Eisert, "Reliable recovery of hierarchically sparse signals for gaussian and kronecker product measurements," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4002–4016, 2020.

[38] S. FOUCART, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Num. An.*, vol. 49, p. 2543–2563, 2011.

[39] Y. Liu, Y. L. Guan, and D. G. G., "Near-optimal BEM OTFS receiver with low pilot overhead for high-mobility communications," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3392–3406, 2022.

[40] W. Li, H. Yin, Z. Qin, Y. Cao, and M. Debbah, "A multi-dimensional matrix pencil-based channel prediction method for massive mimo with mobility," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2215–2230, 2023.

[41] S. Uehashi, Y. Ogawa, T. Nishimura, and T. Ohgane, "Prediction of time-varying multi-user mimo channels based on doa estimation using compressed sensing," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 565–577, 2019.

[42] P. Chaki, J. Shikida, and K. Muraoka, "Exploiting sparse matrix solution to mitigate channel aging in sub-6 ghz 5g mu-mimo using beamspace-delay-doppler domain," in *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2024, pp. 475–480.

[43] T. Zemen, C. F. Mecklenbrauker, F. Kaltenberger, and B. H. Fleury, "Minimum-energy band-limited predictor with dynamic subspace selection for time-variant flat-fading channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4534–4548, 2007.

[44] S. Karnik, Z. Zhu, M. B. Wakin, J. Romberg, and M. A. Davenport, "The fast slepian transform," *Applied and Computational Harmonic Analysis*, vol. 46, no. 3, pp. 624–652, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1063520317300763

[45] D. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[46] S. Ganguly, I. Ghosh, R. Ranjan, J. Ghosh, P. K. Kumar, and M. Mukhopadhyay, "Compressive sensing based off-grid doa estimation using omp algorithm," in *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2019, pp. 772–775.

[47] D. Cohen and Y. C. Eldar, "Sub-Nyquist radar systems: Temporal, spectral, and spatial compression," *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 35–58, 2018.

[48] J. Cao, Z. Yang, and X. Chen, "Compressed line spectral estimation using covariance: A sparse reconstruction perspective," *IEEE Signal Processing Letters*, vol. 31, pp. 2540–2544, 2024.

[49] B. Liu, H. Zhang, W. Feng, Z. Liu, Z. Zhang, and Y. Liu, "Super-resolution generalized eigenvalue method with truly sub-nyquist sampling," 2024. [Online]. Available: https://arxiv.org/abs/2411.02700

[50] A. Bemani, N. Ksairi, and M. Kountouris, "Integrated sensing and communications with affine frequency division multiplexing," *IEEE Wireless Communications Letters*, 2024.

[51] I. Haviv and O. Regev, "The restricted isometry property of subsampled Fourier matrices," in *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*. Springer, 2017, pp. 163–179.

[52] Z. Zhu and M. B. Wakin, "Approximating sampled sinusoids and multiband signals using multiband modulated dpss dictionaries," *Journal of Fourier Analysis and Applications*, vol. 23, no. 6, pp. 1531–5851, December 2017.

[53] D. Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty — v: the discrete case," *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, 1978.

[54] M. Boulsane, N. Bourguiba, and A. Karoui, "Discrete prolate spheroidal wave functions: Further spectral analysis and some related applications," *Journal of Scientific Computing*, vol. 82, no. 3, February 2020.

[55] L.-L. WANG, "Analysis of spectral approximations using prolate spheroidal wave functions," *Mathematics of Computation*, vol. 79, no. 270, pp. 807–827, 2010. [Online]. Available: http://www.jstor.org/stable/40590432

[56] H. Xiao, V. Rokhlin, and N. Yarvin, "Prolate spheroidal wavefunctions, quadrature and interpolation," *Inverse Problems*, vol. 17, no. 4, p. 805, August 2001.

[57] F. Rubio and X. Mestre, "Consistent reduced-rank LMMSE estimation with a limited number of samples per observation dimension," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 2889–2902, 2009.