UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE TELEINFORMÁTICA
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

DANIEL COSTA ARAÚJO

CHANNEL ESTIMATION TECHNIQUES APPLIED TO MASSIVE MIMO SYSTEMS USING SPARSITY AND STATISTICAL APPROACHES

FORTALEZA

2016

DANIEL COSTA ARAÚJO

CHANNEL ESTIMATION TECHNIQUES APPLIED TO MASSIVE MIMO SYSTEMS USING SPARSITY AND STATISTICAL APPROACHES

Tese de doutorado apresentada ao Curso de Engenharia de Teleinformática da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Doutor em Engenharia de Teleinformática.

Orientador: Prof. Dr.João César Moura Mota

Co-Orientador: Prof. Dr. André Lima Férrer de Almeida

FORTALEZA

2016

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A688c Araújo, Daniel Costa.
Channel Estimation Techniques Applied to Massive MIMO Systems Using Sparsity and Stastistical Approaches / Daniel Costa Araújo. – 2016.
124 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2016.
Orientação: Prof. Dr. João César Moura Mota.
Coorientação: Prof. Dr. André Lima Férrer de Almeida.

1. Massive MIMO. 2. Channel estimation. 3. Compressive Sensing. 4. Hybrid Beamforming. 5. Sparsity. I. Título.

CDD 621.38

DANIEL COSTA ARAÚJO

CHANNEL ESTIMATION TECHNIQUES APPLIED TO MASSIVE MIMO SYSTEMS USING SPARSITY AND STATISTICAL APPROACHES

Tese de doutorado apresentada ao Curso de Engenharia de Teleinformática da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Doutor em Engenharia de Teleinformática.

Aprovada em 29/10/2016

BANCA EXAMINADORA

Prof. Dr.João César Moura Mota (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. André Lima Férrer de Almeida
(Co-Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr.Charles Casemiro Cavalcante
Universidade Federal do Ceará (UFC)

Prof. Dr. Sérgio Lima Neto
Universidade Federal do Rio de Janeiro

Prof. Dr. Felix Anterich
German Aerospace Center

This thesis is dedicated to my parents, Moisés and Nágila, my brother, David, and my beloved fiancé, Nicole, for their endless love and support.

ACKNOWLEDGEMENTS

I am very thankful to everyone that has supported me in my Doctor of Philosophy (PhD) period. I could not have done this thesis alone.

First of all, my thanks are to my advisor and co-advisor Dr. João César Moura Mota and Dr. André Lima F. de Almeida, respectively. Prof. Dr. João César has a history in my academic career. He accepted me as his Master student in 2010, and since ever he has helped by providing guidance, trust and advises in my academic career. I am also grateful to my co-advisor Prof. Dr. André Lima for his knowledge, teachings, technical discussion and incentive in producing papers and participating in research projects. Both Prof. Dr. João César and Prof. Dr. André Lima have a large contribution to my academic career.

My gratitude is also extended to my colleagues in the Wireless Telecommunication Research Group (GTEL). The group gave me the opportunity of working in two Ericsson's projects. In such projects, the team has given me a lot of inputs in technical discussions and has also contributed a lot to build my understanding on telecommunication systems. The outcomes from those fruitful discussions are part of this thesis today and, for the people that I have had the pleasure of working with, I must say thank you for your contributions on this thesis. Besides this, I must also express my thanks to Prof. Dr. Francisco Rodrigo Porto Cavalcanti who has accepted me in the projects and gave me his trust by offering me the chance to be an intern at Ericsson. The internship was indeed a very important part of the process to achieve the PhD degree.

Ericsson and CAPES have provided all the conditions to financially supporting my internships at Ercisson Research in Sweden. As an intern, I had the opportunity of working with remarkable researchers in the Department of Radio Access Technology (RAT). The people in RAT contributed a lot with their knowledge, technical discussions and their simulators which helped me to generate many of the results in this thesis.

To the last, but not least, I have special thanks to my family, especially to my parents Moisés David Façanha Araújo and Nágila Costa Araújo. They gave me incentive since ever by supporting me whenever I needed. This thesis would not be possible without them. I am also thankful to my beloved fiancé and future wife, Nicole Marques Sousa, who has been with me during this period and has been a source of inspiration to me.

This thesis is for me and the people who have been with me. Thanks!

ABSTRACT

Massive MIMO has the potential of greatly increasing the system spectral efficiency by employing many individually steerable antenna elements at the base station (BS). This potential can only be achieved if the BS has sufficient channel state information (CSI) knowledge. The way of acquiring it depends on the duplexing mode employed by the communication system. Currently, frequency division duplexing (FDD) is the most used in the wireless communication system. However, the amount of overhead necessary to estimate the channel scales with the number of antennas which poses a big challenge in implementing massive MIMO systems with FDD protocol. To enable both operating together, this thesis tackles the channel estimation problem by proposing methods that exploit a compressed version of the massive MIMO channel. There are mainly two approaches used to achieve such a compression: sparsity and second order statistics. To derive sparsity-based techniques, this thesis uses a compressive sensing (CS) framework to extract a sparse-representation of the channel. This is investigated initially in a flat channel and afterwards in a frequency-selective one. In the former, we show that the Cramer-Rao lower bound (CRLB) for the problem is a function of pilot sequences that lead to a Grassmannian matrix. In the frequency-selective case, a novel estimator which combines CS and tensor analysis is derived. This new method uses the measurements obtained of the pilot subcarriers to estimate a sparse tensor channel representation. Assuming a Tucker3 model, the proposed solution maps the estimated sparse tensor to a full one which describes the spatial-frequency channel response. Furthermore, this thesis investigates the problem of updating the sparse basis that arises when the user is moving. In this study, an algorithm is proposed to track the arrival and departure directions using very few pilots. Besides the sparsity-based techniques, this thesis investigates the channel estimation performance using a statistical approach. In such a case, a new hybrid beamforming (HB) architecture is proposed to spatially multiplex the pilot sequences and to reduce the overhead. More specifically, the new solution creates a set of beams that is jointly calculated with the channel estimator and the pilot power allocation using the minimum mean square error (MMSE) criterion. We show that this provides enhanced performance for the estimation process in low signal-noise ratio (SNR) scenarios.

Keywords: Massive MIMO, channel estimation, sparsity, compressive sensing, tensor analysis, hybrid beamforming.

RESUMO

Pesquisas em sistemas MIMO massivo (do inglês multiple-input multiple-output) ganharam muita atenção da comunidade científica devido ao seu potencial em aumentar a eficiência espectral do sistema comunicações sem-fio utilizando centenas de elementos de antenas na estação de base (EB). Porém, tal potencial só poderá é obtido se a EB possuir suficiente informação do estado de canal. A maneira de adquiri-lo depende de como os recursos de comunicação tempo-frequência são empregados. Atualmente, a solução mais utilizada em sistemas de comunicação sem fio é a multiplexação por divisão na frequência (FDD) dos pilotos. Porém, o grande desafio em implementar esse tipo solução é porque a quantidade de tons pilotos exigidos para estimar o canal aumenta com o número de antenas. Isso resulta na perda do eficiência espectral prometido pelo sistema massivo. Esta tese apresenta métodos de estimação de canal que demandam uma quantidade de tons pilotos reduzida, mas mantendo alta precisão na estimação do canal. Esta redução de tons pilotos é obtida porque os estimadores propostos exploram a estrutura do canal para obter uma redução das dimensões do canal. Nesta tese, existem essencialmente duas abordagens utilizadas para alcançar tal redução de dimensionalidade: uma é através da esparsidade e a outra através das estatísticas de segunda ordem. Para derivar as soluções que exploram a esparsidade do canal, o estimador de canal é obtido usando a teoria de "compressive sensing" (CS) para extrair a representação esparsa do canal. A teoria é aplicada inicialmente ao problem de estimação de canais seletivos e não-seletivos em frequência. No primeiro caso, é mostrado que limitante de Cramer-Rao (CRLB) é definido como uma função das sequências pilotos que geram uma matriz Grassmaniana. No segundo caso, CS e a análise tensorial são combinado para derivar um novo algoritmo de estimatição baseado em decomposição tensorial esparsa para canais com seletividade em frequência. Usando o modelo Tucker3, a solução proposta mapeia o tensor esparso para um tensor cheio o qual descreve a resposta do canal no espaço e na frequência. Além disso, a tese investiga a otimização da base de representação esparsa propondo um método para estimar e corrigir as variações dos ângulos de chegada e de partida, causados pela mobilidade do usuário. Além das técnicas baseadas em esparsidade, esta tese investida aquelas que usam o conhecimento estatístico do canal. Neste caso, uma nova arquitetura de *beamforming* híbrido é proposta para realizar multiplexação das sequências pilotos. A nova solução consite em criar um conjunto de feixes, que são calculados conjuntamente com o estimator de canal e alocação de potência para os pilotos, usand o critério de minimização erro quadrático médio. É mostrado que esta solução reduz a sequencia pilot e mostra bom desempenho e cenários de baixa relação sinal ruído (SNR).

Palavras-chave: MIMO massivo, estimação de canal, *compressive sensing* ,análise tensorial, *beamforming* híbrido.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ADC	analog-digital converter
ADMM	alternate direction method of multipliers
ALSP	alternating least-square with projection
AN	accesses node
AoA	angle of arrival
BS	base station
CRLB	Cramer-Rao lower bound
CS	compressive sensing
CSI	channel state information
D2D	device-to-device
DAC	digital to analog converter
DB	digital beamforming
DFT	discrete Fourier transmform
DL	downlink
DoA	direction of arrival
DoD	direction of departure
DoF	degree of freedom
FDD	frequency division duplexing
FP	frequency dependent precoder
HB	hybrid beamforming
HetNet	heterogeneous network
KKT	Karush Kuhn Tucker
LOS	line-of-sight
LS	least square
LTE	Long Term Evolution
MIMO	multiple-input multiple-output
mm	milimeter
MMSE	minimum mean square error
MRC	maximum ratio combiner
MRT	maximum ratio transmission
MSE	mean square error
MVU	minimum variance and unbiased
NLOS	non-line-of-sight

NMSE	Normalized Mean Square Error
OFDM	Orthogonal Frequency Division Multiplexing
OMP	orthogonal matching-pursuit
PARAFAC	Parallel Factors
PCWP	phase-constrained wideband precoder
PDF	probability density function
RF	radio frequency
SC	small cells
SNR	signal-noise ratio
SVD	singular value decomposition
TDD	time-division duplexing
UE	user equipment
UL	uplink
WP	wideband precoder

CONTENTS

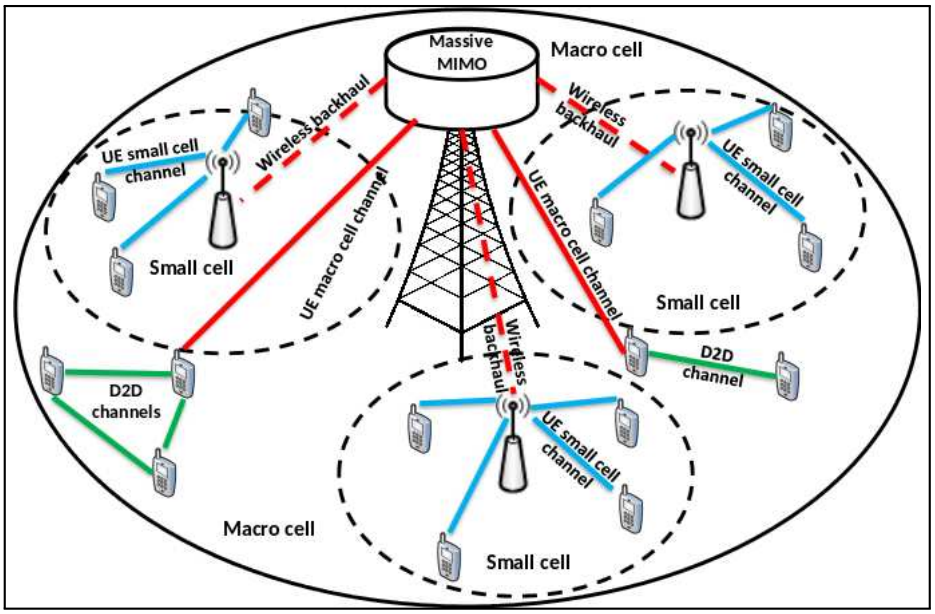
1 INTRODUCTION TO MASSIVE MIMO

Mobile networks are currently facing rapid traffic growth from both smartphones and tablets. Sequential improvements in service quality set the new challenge of increasing wireless network capacity about a thousand times within the next decade [1, 2], but no current wireless access technique can provide a significant improvement in capacity. One of the main candidates to deal with this traffic demand is using the massive MIMO systems. The underlying idea is to scale up the number of antennas at the base station (BS) by at least two orders of magnitude. With such large arrays, the author in [1] shows that the effects of small fading and additive noise are reduced. In a multiuser scenario, massive MIMO opens the possibility to steer many spatial streams to dozens of UE in the same cell, at the same frequency, and at the same time.

Massive MIMO will play a key technological role to create new spectral and energy-efficient networks [3]. This technology has a great potential to meet the requirements of a next-generation wireless system. In [4], the energy efficiency of massive MIMO and small cells (SC) has been studied. The authors proved that massive MIMO has better energy efficiency when the number of SC is low, while the latter offers better performance when the number of cells is high. However, a globally optimal trade-off between massive MIMO and SC efficiency are hard to achieve due to dynamic network behaviour. A viable solution could be found by converging massive MIMO, SC, device-to-device (D2D) communications into a single cloud-controlled heterogeneous network (HetNet), as shown in Fig. 1.1.

To achieve the promised benefits of massive MIMO, some questions must be answered regarding practical aspects and scenarios. An important issue is channel state information (CSI) knowledge because the BS uses it to design transmit beamformers and then provide high spatial multiplexing gains to the system. The CSI estimation is performed thanks to the use of pilot sequences that are used by the receiver as input of channel estimator. This process has impact on the channel spectral efficiency because some communication resources are used to transmit pilots instead of transmitting data. The number of resources used for the CSI acquisition process depends on the type of pilots: uplink (UL) or downlink (DL). In the first type, the UE sends the sequences using the UL channel. The amount of resources used for channel estimation increases proportionally with the number antennas at the UE. In the second type, the BS sends the pilots using the DL channel. In such a case, the amount of communication resources allocated for pilots increases proportionally with the number of antennas at the BS. This vanishes the spectral efficiency and creates big issues for systems that employ DL pilots challenges.

Figura 1.1 – The architecture of a heterogeneous network.



Source provided by the author.

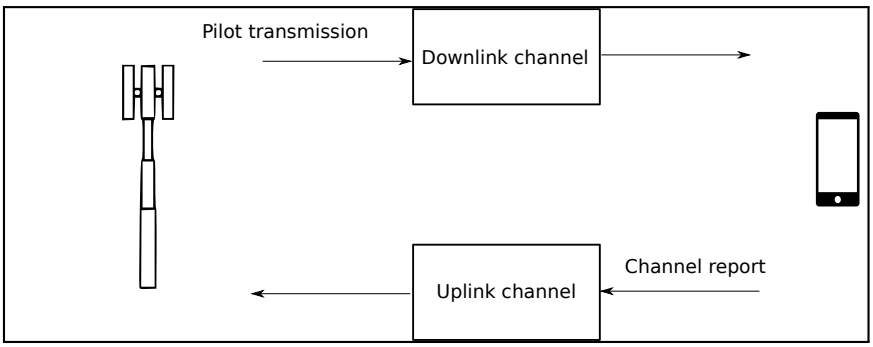
1.1 Channel Estimation Using Reciprocity

One of the big advantages of TDD systems lays on the possibility of estimating the downlink channel based on the assumption of reciprocity. This means that the UL channel matrix within the channel coherence time is a transposed version of the DL one. In this case, the UE sends out the pilots and the BS estimates the UL channel. Then, the BS is able to steer narrow beams towards the UE with a transposed version of the estimated UL channel matrix. The overall process has to be performed within the duration of the channel coherence time, which may not hold in scenarios with moderate or high mobility. This is still true even though, in reality, each UE has RF mixers, filters, and analog-digital converter (ADC)s that make up the equivalent channel model and can potentially break the assumptions on uplink/downlink channel reciprocity [5]. It is important to highlight to the reader that this issue is not uniquely related to massive MIMO, but it has also been studied over the past few years in the context of traditional MIMO systems [5]. However, with massive MIMO systems, a lack of understanding about hardware impairments and a lack of proper modelling persists, and it is still not clear whether they challenge reciprocity or how to mitigate any associated negative effects to keep transmissions reliable. For example, mutual coupling among the BS antennas may not be neglected and may limit massive MIMO performance [6].

1.2 Problem Definition: Channel Estimation Without Reciprocity

Even though channel reciprocity is a key aspect to acquire the channel at the BS, there are systems where such property does not hold. For instance, many current wireless systems

Figura 1.2 – The FDD system is composed of two channels: uplink and downlink. The UE estimates the channel by using the pilot sequence sent in the downlink channel, then reports channel estimation in the feedback channel (uplink channel).



Source provided by the author.

operate in FDD. In this case, the BS needs to spend communication resources on the DL channel to transmit pilots and needs the UL one to acquire the channel. The CSI is estimated by the UE and then reported to the BS via feedback control channel.

The benefits of using conventional DL channel estimation techniques in traditional MIMO systems that assume small or moderate antenna arrays are well known. However, the amount of time-frequency resources used by the pilot sequence grows with the number of transmit antennas. In the absence of a specific strategy to reduce the sequence length and keep a reliable CSI acquisition, the gains in terms of spectral efficiency of massive MIMO systems tend to disappear [3]. To tackle the DL channel estimation problem, effective pilot transmission methodologies that accurately estimates the CSI with minimal signalling overhead play a crucial role in the success of massive MIMO for FDD systems.

Although traditional MIMO channel estimation techniques require the length of the pilot sequence to be on the same order as the number of BS antennas, optimality is related to the channel covariance matrix [7]. Specifically, by considering the link between BS and UE, the training sequence for each antenna element should be orthogonal. The optimal length corresponds to the number of antennas at the transmitter in i.i.d./Rayleigh channels. However, under spatially correlated channels, and from the assumption that the BS knows the channel correlation matrix perfectly, the authors in [7] and [8] affirm that the optimal training sequence length is equal to the number of dominant eigenvalues. For massive MIMO, it is likely true that the number of eigenvalues is smaller than the number of antennas which means this can be exploited to potentially reduce the pilot overhead.

Another challenge to make massive MIMO systems a reality is related to the feedback channel. Recall that the channel estimation in an FDD-based massive MIMO system is performed by the UE, which is responsible for reporting CSI to the BS. However, due to the huge number of channel coefficients, the amount of information that goes through channel feedback can easily take a large portion of the available bandwidth. The feedback overhead increases linearly as the

number of antennas increases. Because of this issue, channel feedback has gained much attention in the massive MIMO research community. Current strategies to tackle the feedback overhead problem try to compress the channel matrix so that the UE sends fewer information to the BS and it can still reconstruct the channel. For instance, non-coherent trellis-coded quantization has been proposed by authors in [9] and the techniques has high compression rates, but at the cost of moderate computation complexity using Viterbi algorithm.

Another perspective are solutions that use sparsity-aware channel estimation techniques. In this case, the UE estimates a sparse channel representation and feeds the non-zero entries back to the BS. Another example of using sparsity is to identify the most representative "channel directions" (transmission angles), which can be fed back to the BS using a few bits [10, 11].

Statistical approaches are also interesting to reduce the system overhead. Using the spatial covariance matrix, the UE and BS can use this information to obtain a reduced version of the massive MIMO channel matrix [12]. This approach will be explained in more detail in this thesis.

This thesis focuses on deriving channel estimators that support applications where the reciprocity does not hold. In this context, the techniques developed in this work relies on two approaches: sparsity and statistics. Such strategies are applied mainly for the pilot and feedback overhead problems. Figure 1.2 summarizes how the system works with DL and UL channels.

1.3 Sparsity and Statistical Approaches to Deal With Overhead

The pilot overhead demanded by channel estimators depends on the number of parameters to infer. In the massive MIMO case, the channel matrix is naturally large which implies the need for many pilots. To reduce the overhead, channel estimators must exploit other representations that lead to a compressed version of the channel.

Usually, the channel matrix is represented with respect to the antenna domain, but such a large matrix is likely to be low-rank. The number of scatters in most environments is usually much less than the number of antennas and generates fewer dominant paths. In this case, the channel can be represented in a compressed version by using a given basis whose choice determines the rate of compression. The optimum basis varies according to the environment characteristics, however, this information is not available in advance for most applications. In such a case, the system imposes a fixed basis up to an error in the estimation process. The approaches considered in this thesis use two frameworks to achieve channel compression: i) compressive sensing (CS) [13, 14] and ii) statistical channel representation [15].

The CS theory is very attractive in problems where the signal of interest has a sparse representation [13, 16]. This subject has been under intensive study and several works have exploited this concept in different areas, such as sampling theory [17], ultrawideband channel estimation [18, 19] and radar [20], just to mention a few. The use of this theory holds for

signals which are sparse in the canonical coordinate basis or sparse with respect to some other orthonormal basis. In the context of this thesis, the massive MIMO channel matrix itself is not sparse, but it admits a sparse decomposition.

Second order statistics can also be exploited to obtain channel compression. Environments with only a few scatters lead to a low-rank spatial channel covariance matrix. This means that to accurately represent the instantaneous channel, the estimator only needs a subspace defined by the eigenvectors. Knowing this subspace, the BS smartly designs the pilot sequences [21] and also their length.

In this thesis, each chapter covers a specific aspect of the estimation process between BS and UE, as detailed below

- Chapter 2 (*Compressive Sensing for Estimating Narrowband Channels*): In this chapter, we present the fundamental concepts of CS and use them to derive a channel estimator that exploits the flat channel sparsity. Two CS-based estimators are derived, one using a greedy algorithm, called OMP, and one using the ADMM.
- Chapter 3 (*Tensor Compressive Sensing for Estimating Frequency Selective Channels*): In this chapter, we investigate channel estimation applied to frequency-selective channels. Such a case involves estimating a collection of matrices, each one corresponding to a given a channel frequency response. To do this, we combine the CS and tensor analysis [22, 23] to jointly estimate the space-frequency channel.
- Chapter 4 (*Sparse Channel Estimation for mm-wave Massive MIMO Systems*): In this chapter, two problems are investigated: i) one is to optimize the sparse representation of the channel leading to a better channel compression; ii) the other one consist of tracking the angle variations due to the UE mobility.
- Chapter 5 (*Statistical Hybrid Beamforming for massive MIMO Channel Estimation*): In this chapter, we tackle the problem of estimating the channel using the spatial channel covariance matrix. To exploit such an information, a beamforming technique and a channel estimator are jointly designed to acquire the channel.

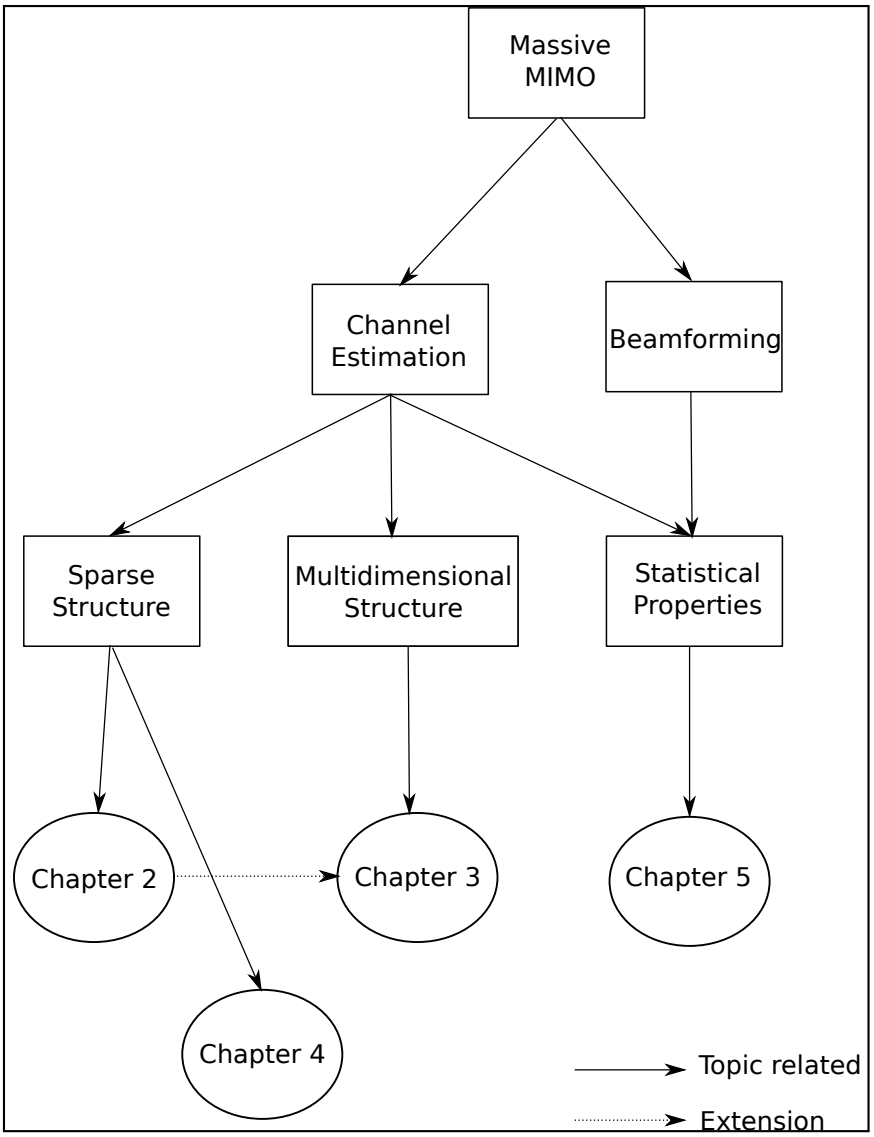
Fig. 1.3 summarizes in a block diagram how the thesis is organized.

1.4 List of Contributions

The contributions of this thesis are:

- The development of a Cramer-Rao lower bound (CRLB) for estimating sparse flat MIMO channels. The lowest bound can only be achieved by using pilot sequences that define Grasmanian matrices.

Figura 1.3 – Diagram of the thesis organization.



Source provided by the author.

- A low-complexity channel estimator that jointly estimates a space-frequency massive MIMO channel. This estimator is obtained from the combination between CS and tensor frameworks. To the best of author's knowledge, there is no estimator in the literature that exploits jointly sparse representation of the channel and still achieves a low-complexity estimator.

- A novel method that includes a new pilot placement on the time-frequency frame to estimate milimeter (mm)-wave channels and deal with the user mobility issues in indoor scenarios. The contribution of the method lays on the case where the UE has multiple antennas and the system uses an OFDM modulation .

- A new HB architecture is designed jointly with the channel estimator using the MMSE criterion. The optimum beams for transmitting pilot sequences are derived.

In terms of scientific production, this thesis has

- Journal
  - ARAÚJO, D.C., MAKSYMYUK, T., de ALMEIDA, A.L.F, MACIEL, T.F ; MOTA, J.C.M; "Massive MIMO: Survey and Future Research Topics", IET Communications, to appear, 2016.
- Conferences
  - ARAUJO, D. C. ; de ALMEIDA, A. L. F. ; MOTA, João C M ; HUI, D. "Estimation of Very Large MIMO Channels Using Compressed Sensing". In: Simpósio Brasileiro de Telecomunicações, 2013, Fortaleza. Anais do Simpósio Brasileiro de Telecomunicações, 2013
  - ARAUJO, D. C. , de ALMEIDA, André L.F., AXNAS, J. MOTA, J.C.M. , "Channel Estimation for Millimeter-Wave Very Large MIMO", . In: European Signal Processing Conference (EUSIPCO), Lisbon, 2014.
  - D. C. Araújo, A. L. F. de Almeida and J. C. M. Mota, "Compressive sensing based channel estimation for massive MIMO systems with planar arrays,"Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on, Cancun, 2015, pp. 413-416. doi: 10.1109/CAMSAP.2015.7383824
  - ARAÚJO, D.C., KARIPIDIS, E., de ALMEIDA, A. L. F., MOTA, J. C. M . "Improving Spectral Efficiency in Large-Array FDD Systems with Hybrid Beamforming", Proc. of IEEE 9th Sensor Array and Multichannel Signal Processing Workshop (SAM 2016), Rio de Janeiro, 2016.
  - ARAÚJO, D.C., KARIPIDIS, E., de ALMEIDA, A. L. F., MOTA, J. C. M . "Hybrid Beamforming Design Using Finite-Resolution Phase-Shifters for Frequency Selective Massive MIMO Channels", submitted on ICASSP 2017.
- Patent
  - SUNDSTROM, L. and Hui, D. and Reial, A. and ARAUJO, D.,"Antenna Beam Control", Reference WO 2015135987 A1.

This work was developed in the context of UFC/ERICSSON Research partnership. The thesis is also product of two research projects over the last 4 years:

- UFC.34 *Advanced MIMO Transceivers and Matrix Decompositions for Wireless Systems*, August/2012 - July/2014;
- UFC.41 *Transceiver Design in MIMO Communication Systems: Distributed Processing and Very-Large-Scale Approaches*, October/2014 - September/2016,

How to Read This Thesis

This thesis has four chapters that cover different aspects of channel estimation. Every chapter is meant to be self-contained so that the reader can read them independently without loosing significant information.

2 COMPRESSIVE SENSING FOR ESTIMATING NARROWBAND CHANNELS

Abstract:

In this chapter, the problem of estimating single-user massive MIMO channels using DL pilots is investigated. In this context, the number of DL pilots tends to increase as the number of antennas scales up. Assuming the channel is flat, the estimator exploits the inherent channel structure which reduces the number of estiamted parameters . This occurs because the channel has a sparse representation in the spatial domain. To derive a sparsity-based estimator, the CS framework is applied. With this approach, we derive three methods: one uses a greedy algorithm, the second uses Lasso criterion, and the third implements the Lasso criterion but in a distributed fashion. The motivation to use a distributed solution comes from the increased computational complexity demanded by the two first methods. The results show that it is possible to reduce the pilot sequence and achieve precise channel estimation by exploiting the channel sparsity. Moreover, the Lasso criterion outperforms the greedy algorithm mainly for low SNR, and its distributed version has a negligible loss Normalized Mean Square Error (NMSE) with respect to the centralized solution. The conclusions and open questions are presented in the end of this chapter.

2.1 Introduction

The promised benefits of massive MIMO systems are strongly dependent on the quality of the CSI available at the receiver and/or transmitter. Conventional channel estimation approaches rely on training sequences [1, 24, 25] and most of them use a least square (LS) approach to estimate the channel. However, this technique requires that the length of the training sequences must be at least equal to the number of transmit antennas which demands precious communication resources and energy.

To circumvent the pilot overhead problem, a channel estimator can be derived based on the channel sparsity. The MIMO channel is a summation of multipath that leads to a channel matrix response that is not sparse. The canonical representation then does not provide a convenient representation for the estimator because the receiver needs to estimate all the entries of the channel matrix. On the other hand, the channel is composed by a set of paths from different scatteres in the environment. This essentially dictates the channel richness. In massive MIMO, it is very unlike that the number of scatters is equal to the number of antennas. Thus, one way to the achieve sparse representation consists of using the inherent channel structure. Once this structure is included in terms of the signal model, the estimator using CS framework is derived. The CS theory provides uniqueness conditions that enable the receiver to design a channel estimator that requires fewer measurements of a vector that has a sparse representation [13, 16]. This topic has

found many different applications such as image processing, radar, sub-Nyquist sampling, just to mention a few [14, 17, 20].

In the channel estimation problem, CS finds many attractive scenarios for its application as for example ultrawideband systems, underwater communication scenarios, mm-wave channels, just to mention a few ones [26–28]. The cited examples have channels that admit a sparse representation on the delay domain, i.e. the power delay profile has very few representative taps. This feature is exploited by the channel estimator to reduce the length of the pilot sequences. In [29], the case of interest is the multi-user scenario where the number of antennas at the BS grows large and the number of UE as well. To achieve channel estimation without vanishing the communication resources, the estimator relies on the channel sparsity to reduce the necessary overhead to estimate such a large channel. This strategy overcomes other techniques that use a LS criterion.

Even though the authors in [29] provide a great contribution in proposing a sparse channel estimator, the solution only has application in crowd environments. The technique relies on the fact that both BS antennas and the number of UEs grow largely, this situation is envisaged as a possible application in 5G [30]. However, there is room for working with single-user scenario. In this case, the BS serves only one user per resource, which means that the multiplexing capabilities of the large array are not enjoyed by the UEs, but rather the array gain. Single-user solutions still find many applications on currently wireless standard systems as for instance, to overcome some coverage limitation issues. Furthermore, the development of receivers with multiples antennas is also considered. They can become very popular with the use of mm-waves because the higher the frequency is, the more the antenna array can be packed [31].

One bottleneck that arises when assuming multiple antennas at the UE is the hardware complexity and the amount of signals to process. Each antenna element at the UE collects a signal which is then sampled and processed in baseband. However, the more antennas are employed, the more processing capabilities to treat the signal are required. In the channel estimation problem, CS helps by reducing the pilot sequence length, and consequently reduces the number of samples to be processed. However, we can further reduce the hardware complexity by using a distributed processing in the baseband part [32]. Specifically, we employ a set of channel estimators that handle small portions of the training sequence. Each estimate is informed in a central node that combines the estimates and report back to each estimator. Such an architecture of an estimator has not been considered yet.

This chapter shows a pilot-assisted technique based on CS theory for channel estimation. To obtain an approximate sparse representation, we consider that angles of departure and arrival are modelled by the Fourier basis. We provide three strategies:

- The first estimator uses a greedy solution called OMP, where the algorithm searches for the most representative directions defined by the Fourier basis. After finding them, the

algorithm calculates the channel projections onto the directions.

- The second solution is derived from the Lasso problem [14, 33]. This formulation tries to find the sparsest solution taking into account the power of the noise.
- The third solution introduces a distributed implementation of the Lasso solution. This becomes an attractive solution for those receivers with limited processing capacity.

All the three methods are compared with the CRLB. From this development, we verified that the set of pilot sequences that leads to a Grassmanian real matrix are necessarily the optimum ones. To the best of author's knowledge, this results have not been addressed yet in the literature.

2.2 Compressive Sensing Background

In recent years, CS has become very popular in many areas such as mathematics, computer science and electrical engineering just to mention a few. The key idea of CS is to recover a sparse signal from very few linear measurements. The theory shows that it is possible to recover N parameters with $M < N$ measurements if the estimator uses a suitable basis that leads to a sparse representation of the data. Thus, the theory offers a framework for simultaneous sensing and compression of finite-dimensional vectors. A lot of results and theorems can be found in the literature, and from this broad literature, we present in this section only the most fundamental concepts and just those that will be used in this thesis [13, 14, 16, 34].

2.2.1 Undetermined Linear Systems

Consider the matrix $\mathbf{\Psi} \in \mathbb{C}^{M \times N}$ with $M < N$ that defines a undetermined linear system $\mathbf{\Psi x} = \mathbf{y}$. This configures a system with less equations than variables, and it has either no solution or infinite solutions. The first case happens when the vector $\mathbf{y}$ is not in the span of the matrix $\mathbf{\Psi}$. The second case happens otherwise. To avoid the situation where we have no solution, let us assume that $\mathbf{\Psi}$ is full rank.

In many areas, it is common to face problems that lead to undetermined linear systems. For instance, we refer to the problem in [14]. The authors investigate the reconstruction of a waveform assuming fewer measurements than the minimum to meet the Nyquist criterion. Even though this classical condition is not achieved, the signal is completely recovered by exploiting the knowledge of signal basis representation. Using this information the authors obtain a compressed representation of the signal and which makes it possible to recover the signal even at sub-Nyquist rates.

2.2.2 General Problem

Let us formally introduce the concept of a sparse vector [13, 16].

Definition 1 *Let $\boldsymbol{\theta} \in \mathbb{C}^N$, if there exist $K < N$ nonzero entries, then it is a K-sparse vector.*

In practice, the vector to be estimated is not always sparse. If this happens, then the estimator has to use a basis that leads the signal of interest into a sparse representation. Thus, consider a given signal $\mathbf{x} \in \mathbb{C}^N$ as the signal of interest and let $\boldsymbol{\Phi} \in \mathbb{C}^{N \times N}$ be an orthonormal basis such that $\boldsymbol{\theta} = \boldsymbol{\Phi}^H \mathbf{x}$ is a K-sparse vector. The matrix $\boldsymbol{\Phi}$ compresses $\mathbf{x}$ if the codebook $\boldsymbol{\Phi}$ sparsifies $\mathbf{x}$. An example where this concept can be applied refers to the communication channel models. In practice, the channels are composed by scatters that define the channel impulse response. If it has only few scatters, then the channel is likely to be sparse. A model that we hardly can rely on sparsity is the estimation of i.i.d channels. In this situation, the problem is unstructured and does not admit a sparse representation in the spatial domain.

The acquisition of $\mathbf{x}$ is a process that is modeled as a linear transformation, where a measurement matrix $\boldsymbol{\Psi}$ is applied over $\mathbf{x}$. The signal observed then is represented as $\mathbf{y} = \boldsymbol{\Psi}\mathbf{x}$ $\mathbb{C}^M$. The number of elements M in $\mathbf{y}$ refers to the number of measurements that are taken from the vector $\mathbf{x}$. The acquired signal can be written as $\mathbf{y} = \boldsymbol{\Psi}\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta}$ [13, 16, 35].

The goal of the CS formulation is to find the sparsest solution $\boldsymbol{\theta}$ such that it attains $\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta}$. Thus, the optimization problem can be stated as

$$\begin{aligned} &\arg\min_{\boldsymbol{\theta}} \ \|\boldsymbol{\theta}\|_0 \\ &\text{s.t. } \mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta}, \end{aligned} \tag{2.1}$$

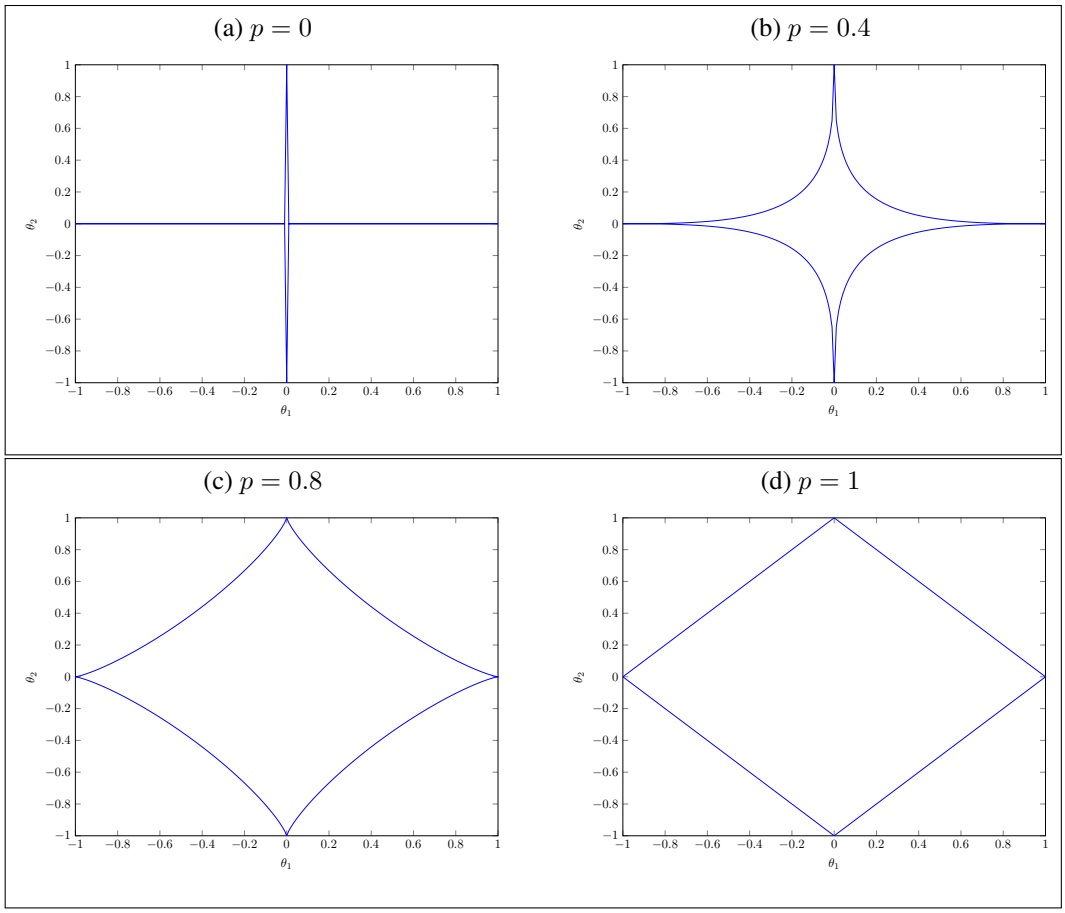
where $\|.\|_0$ stands for the l_0-norm. This norm is of interest to this problem because it returns the number of non-zero entries of $\boldsymbol{\theta}$. However, the problem (2.1) is non-convex and also NP-hard due to the $\|.\|_0$. Therefore, to have a tractable problem, it is modified by changing the l_0-norm to another one that achieves convexity. The $\|.\|_\mathrm{p}$ operator is non-convex for the range $0 < p < 1$, as can be seen in Figure 2.1. Therefore, a reasonable heuristic for relaxing (2.1) consists of approximating $\|.\|_0$ to $\|.\|_1$ representation. Thus, the new problem is

$$\begin{aligned} &\arg\min_{\boldsymbol{\theta}} \ \|\boldsymbol{\theta}\|_1 \\ &\text{s.t. } \mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta}. \end{aligned} \tag{2.2}$$

The l_1-minimization approach is now a convex problem and can be solved by using linear programming if it is real, or second order cone programming if it is complex [13]. The assumptions usually considered when relaxing this problem are:

1. The equality $\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta}$ is too strict. Take the example of communication systems, where the presence of noise introduces a perturbation in the parameter estimation process. To solve this problem, a better approach would be one that allows for some deviation ϵ, as for instance the expression $\|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta}\|_2 < \epsilon$.

Figura 2.1 – Plot of a function $\|\boldsymbol{\theta}\|_{\mathrm{p}}\ \forall p \in [0,1]$, where $\boldsymbol{\theta} \in \mathbb{R}$.



Source: Provided by the author.

2. The sparsity measure is very sensitive to small entries in $\boldsymbol{\theta}$. A good approach consists in defining a threshold to eliminate such values.

2.2.3 Uniqueness Conditions for the Compressive Sensing Problem

A key property to understand the conditions of uniqueness is the $spark$ of the matrix $\boldsymbol{\Upsilon} = \boldsymbol{\Psi}\boldsymbol{\Phi}$. This property provides a way of describing the null-space of $\boldsymbol{\Upsilon}$ using the l_0-norm. This concept is enunciated with the following definition.

Definition 2 *The operator* $spark\{\boldsymbol{\Upsilon}\}$ *is the smallest number of columns from* $\boldsymbol{\Upsilon}$ *that are linearly dependent.*

Note that there is a resemblance between the definition of rank and spark. From linear algebra, the rank is defined as the largest number of columns in $\boldsymbol{\Upsilon}$ that are linearly independent. If we replace the words “largest” with “smallest” and “independent” with “dependent” then we have the spark definition. A natural problem that arises with the calculation of the spark is the computational complexity. To determine the smallest set of linearly dependent columns, we need

to perform an exhaustive search over all the possible subsets formed by the columns of $\boldsymbol{\Upsilon}$. This is more complex than rank calculation [35].

Apart from the complexity issue, the spark is a powerful definition to reveal the uniqueness of $\boldsymbol{\theta}$. The vectors in the null-space of $\boldsymbol{\Upsilon}$, i.e. $\boldsymbol{\Upsilon}\boldsymbol{\theta} = \mathbf{0}$, must combine linearly at least $spark\{\boldsymbol{\Upsilon}\}$ dependent columns of $\boldsymbol{\Upsilon}$. Using the spark definition, the uniqueness condition is ensured by the following theorem:

Theorem 1 *If a set of equations is represented as a linear system in the form $\boldsymbol{\Upsilon}\boldsymbol{\theta} = \mathbf{y}$, its sparsest solution obeys $\|\boldsymbol{\theta}\|_0 < spark\{\boldsymbol{\Upsilon}\}/2$.*

To understand Theorem 1, consider two vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ as null-space elements of $\boldsymbol{\Upsilon}$ and the difference $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$ must also be in the null-space. Using the definition of spark, we get

$$\|\boldsymbol{\theta}_1\|_0 + \|\boldsymbol{\theta}_2\|_0 \geq \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_0 \geq spark\{\boldsymbol{\Upsilon}\}. \tag{2.3}$$

This inequality shows that the number of non-zero entries of $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$ cannot exceed the sum of the non-zero entries of each vector $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, which follows from the triangle inequality. From this observation, and assuming $\|\boldsymbol{\theta}_1\|_0 < spark\{\boldsymbol{\Upsilon}\}/2$, it is possible to conclude that $\|\boldsymbol{\theta}_2\|_0 > spark\{\boldsymbol{\Upsilon}\}/2$.

Because $spark\{\boldsymbol{\Upsilon}\}$ is very difficult to evaluate, its use in uniqueness analysis is very complicated. An alternative is to exploit the mutual coherence concept which is presented in the following definition.

Definition 3 *The mutual-coherence of a given matrix $\boldsymbol{\Upsilon}$ is the largest absolute inner product between all the different columns of $\boldsymbol{\Upsilon}$. Denoting the ith column of $\boldsymbol{\Upsilon}$ as $[\boldsymbol{\Upsilon}]_i$, the mathematical formulation is given by*

$$\mu(\boldsymbol{\Upsilon}) = \max_{1 \leq i,j \leq N, i \neq j} \frac{|\,[\boldsymbol{\Upsilon}]_i^H\,[\boldsymbol{\Upsilon}]_j\,|}{\|[\boldsymbol{\Upsilon}]_i\|_2 \left\|[\boldsymbol{\Upsilon}]_j\right\|_2}. \tag{2.4}$$

Note that for the orthogonal matrix, the mutual coherence is zero. For the case of interest, $M < N$, $\mu(\boldsymbol{\Upsilon})$ is positive, and the smaller is $\mu(\boldsymbol{\Upsilon})$, the closer the estimator performance is to the case where orthogonal matrices are used.

Mutual coherence is far easier to compute than the spark. It is then interesting to establish a relationship between spark and mutual coherence. Following, it can be stated [35]

$$spark\{\boldsymbol{\Upsilon}\} \geq 1 + \frac{1}{\mu(\boldsymbol{\Upsilon})}. \tag{2.5}$$

Using this result and Theorem 1, it is possible to obtain another upper-bound for the uniqueness [35]:

Theorem 2 *The sparsest solution of a linear system* $\boldsymbol{\Upsilon}\boldsymbol{\theta} = \mathbf{y}$ *obeys the relationship*

$$\|\boldsymbol{\theta}\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu\left(\boldsymbol{\Upsilon}\right)}\right). \tag{2.6}$$

Theorems 1 and 2 are similar, but they carry different assumptions. The first one, which relates spark and uniqueness, is more powerful than the second one, which uses the mutual coherence to guarantee uniqueness. The mutual coherence is lower bounded by $\frac{1}{\sqrt{M}}$, thus $\|\boldsymbol{\theta}\|_0$ in Theorem 2 cannot be larger than $\sqrt{M}/2$. However, the spark can be as larger as M. This gives a bound of $M/2$.

2.3 Narrowband Structured Channel Model

Consider a MIMO system with N_{T} antennas at the transmitter and N_{R} antennas at the receiver. Assume that the receiver does not have any knowledge about the communication channel. Thus, the receiver needs to estimate the channel using, e.g. a known training sequence sent by the transmitter. The received training signal is denoted by

$$\mathbf{Y} = \mathbf{S}\mathbf{H} + \mathbf{Z}. \tag{2.7}$$

The matrix $\mathbf{S} \in \mathbb{C}^{T \times N_{\mathrm{T}}}$ contains the training sequences sent by each one of the N_{T} transmit antennas, where T denotes the length of the training sequence, $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{R}}}$ is the channel matrix and $\mathbf{Z}$ is the additive white Gaussian noise matrix whose entries have variance $N_0/2$, and the matrix $\mathbf{Y} \in \mathbb{C}^{T \times N_{\mathrm{R}}}$ defines the received signal.

It is a common practice in the literature to use i.i.d entries for modelling the MIMO channel. However, this assumption leads us to overestimate the spatial degrees of freedom. This is especially true for massive MIMO channels where the number of transmit/receive antennas may be significantly higher than the degrees of freedom provided by the channel [36]. Thus, the i.i.d. case is not considered hereafter but instead a structured channel model.

In [36], a stochastic MIMO channel model, inspired by the Kronecker model and virtual channel model, was proposed. Combining the advantages of both models, the so-called Weichselberger's model shows enhanced capabilities to model the spatial structure of the channel. However, the price for achieving a more accurate channel representation is the knowledge of the transmit and receive one-side correlation matrices. Such matrices are defined as

$$\begin{aligned} \mathbf{R}_{\mathrm{R}} &= \mathrm{E}\left\{\mathbf{H}^{\mathrm{H}}\mathbf{H}\right\} = \mathbf{U}_{\mathrm{r}}\boldsymbol{\Lambda}_{\mathrm{r}}\mathbf{U}_{\mathrm{r}}{}^{\mathrm{H}}, \\ \mathbf{R}_{\mathrm{T}} &= \mathrm{E}\left\{\mathbf{H}\mathbf{H}^{\mathrm{H}}\right\}^{\mathrm{T}} = \mathbf{U}_{\mathrm{t}}\boldsymbol{\Lambda}_{\mathrm{t}}\mathbf{U}_{\mathrm{t}}{}^{\mathrm{H}}, \end{aligned} \tag{2.8}$$

where, the matrices $\mathbf{U}_{\mathrm{r}}$ (resp. $\mathbf{U}_{\mathrm{t}}$) and $\boldsymbol{\Lambda}_{\mathrm{r}}$ (resp. $\boldsymbol{\Lambda}_{\mathrm{t}}$) are the receive (resp. transmit) eigenvector and eingenvalue matrices.

In [36], Weichselberger proposes the following channel decomposition model

$$\mathbf{H} = \mathbf{U}_{\mathrm{t}} \left(\mathbf{\Omega} \odot \mathbf{G} \right) \mathbf{U}_{\mathrm{r}}{}^{\mathrm{H}} \tag{2.9}$$

where $\mathbf{\Omega}$ is a positive and real-valued matrix, being equal to the square-root of the power coupling between transmit and receive eigenmodes, $\mathbf{G}$ has i.i.d entries that accounts for the channel fading, and $\odot$ denotes the Hadamard product. A full-rank matrix $\mathbf{\Omega}$ means a scattering-rich environment with maximum diversity. In this situation, model (2.9) is equivalent to the Kronecker model. In [36, 37], the authors show that if the number of antennas goes to infinity, the discrete Fourier transmform (DFT) matrix is asymptotically equal to the matrix of eigenvectors.

2.4 Compressive Sensing to Massive MIMO Channel Estimation

Channel estimation using a LS criterion applies the pseudoinverse of $\mathbf{S}$ over the received signal. However, such an operation is only possible if $T \geq N_{\mathrm{T}}$. This imposes a practical issue that decreases the spectral efficiency, once N_{T} is very large. The desirable situation is that $T \leq N_{\mathrm{T}}$, but this leads to a system with less equations than variables, and, consequently infinite solutions. On the other hand, using compressed sensing, the receiver can estimate such a large channel with $T \leq N_{\mathrm{T}}$. First, it is necessary to find an appropriate codebook (also known as "dictionary") $\mathbf{\Phi}$ to obtain a sparse representation of $\mathbf{H}$.

Consider Eq. (2.7) and the property $\mathrm{vec}\left\{\mathbf{ABC}\right\} = \left(\mathbf{C}^{T} \otimes \mathbf{A}\right) \mathrm{vec}\left\{\mathbf{B}\right\}$, the received signal $\mathbf{Y}$ as

$$\begin{aligned} \mathrm{vec}\left\{\mathbf{Y}\right\} &= \mathrm{vec}\left\{\mathbf{SH}\right\} + \mathrm{vec}\left\{\mathbf{Z}\right\} \\ &= \left(\mathbf{I}_{N_{\mathrm{R}}}^{T} \otimes \mathbf{S}\right) \mathbf{h} + \mathrm{vec}\left\{\mathbf{Z}\right\}, \end{aligned} \tag{2.10}$$

where $\mathbf{h} \doteq \mathrm{vec}\left\{\mathbf{H}\right\} \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}}}$.Substituting (2.9) into (2.10), we derive the expression

$$\mathrm{vec}\left\{\mathbf{Y}\right\} = \left(\mathbf{I}_{N_{\mathrm{R}}}^{T} \otimes \mathbf{S}\right) \left(\mathbf{U}_{\mathrm{r}}^{*} \otimes \mathbf{U}_{\mathrm{t}}\right) \mathrm{vec}\left\{\left(\mathbf{\Omega} \odot \mathbf{G}\right)\right\} + \mathrm{vec}\left\{\mathbf{Z}\right\}. \tag{2.11}$$

Then, we define

$$\mathbf{y} \doteq \mathrm{vec}\left\{\mathbf{Y}\right\} \in \mathbb{C}^{T N_{\mathrm{R}}}, \tag{2.12}$$

$$\mathbf{z} \doteq \mathrm{vec}\left\{\mathbf{Z}\right\} \in \mathbb{C}^{T N_{\mathrm{R}}}, \tag{2.13}$$

$$\mathbf{\Psi} \doteq \left(\mathbf{I}_{N_{\mathrm{R}}}^{T} \otimes \mathbf{S}\right) \in \mathbb{C}^{T N_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{R}}}, \tag{2.14}$$

$$\mathbf{\Phi} \doteq \left(\mathbf{U}_{\mathrm{r}}^{*} \otimes \mathbf{U}_{\mathrm{t}}\right) \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{R}}}, \tag{2.15}$$

$$\boldsymbol{\theta} \doteq \mathrm{vec}\left\{\left(\mathbf{\Omega} \odot \mathbf{G}\right)\right\} \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{R}}}, \tag{2.16}$$

which allows one to compactly rewrite (2.10) as

$$\mathbf{y} = \mathbf{\Psi \Phi} \boldsymbol{\theta} + \mathbf{z}. \tag{2.17}$$

Note that in this model, the measurement matrix $\mathbf{\Psi}$ depends on the set of pilot sequences contained in $\mathbf{S}$. Their length and the number of antennas at the receiver correspond to the total number of measurements that the receiver has access. This number can be reduced if the channel vector $\mathbf{h}$ admits a sparse representation. This is because we can use CS theory to reduce the length T of the pilot sequences.

The channel sparsity is achieved using the dictionary $\mathbf{\Phi}$ to model $\mathbf{h}$. This representation is obtained from the transmit-receive spatial bases $\mathbf{U}_{\mathrm{t}}$ and $\mathbf{U}_{\mathrm{r}}$ respectively. These matrices model statically the paths that compose the channel. The coupling matrix $\mathbf{\Omega}$ defines the power associated with the scatters and how the paths of the transmit and receive sides connect among themselves. If we consider a full matrix $\mathbf{\Omega}$, it means that all the scatters communicate with each other and the model reduces to Kronecker channel model [36]. However, it is known from the literature [36] that the consideration creates clusters that do not exist, i.e. it overestimates the richness of the channel [38]. Therefore, it is correct to assume that $\mathbf{\Omega}$ is sparse.

The sparsity of the model depends on the basis adopted to represent the channel. For the same channel vec $\{\mathbf{H}\}$, there are more than one sparse representation and they may lead to different sparse representations. The compression provided by the eigenvectors in the model (2.9) can only be achieved if the receiver knows such bases. These orthonormal matrices are obtained from the covariance matrices of the transmit and receive side which means that the receiver needs to acquire such an information. The acquisition demands an extra overhead before the estimation of the instantaneous channel, but $\mathbf{U}_{\mathrm{t}}$ and $\mathbf{U}_{\mathrm{r}}$ are long-term information and remain constant over many time instants. In the case, if we have large arrays at both, transmitter and receiver, these matrices tend to be the DFT. The advantage of this result is that the covariance matrix acquisition can be avoided. Thus, it is possible to derive a sparsity-based channel estimation technique assuming the DFT approximation to reduce the pilot sequence.

2.5 Cramer-Rao Bound Discussion For Estimating Sparse Channels

The CRLB provides a benchmark to analyse how good is the accuracy of a given channel estimator. If it attains the CRLB in linear estimation problems, the solution is equivalent to the minimum variance and unbiased (MVU) estimator [15]. This type of estimator applied to the problem in (2.17) must be capable of extracting the sparse channel $\boldsymbol{\theta}$. The accuracy of such an estimator is compared with the benchmark provided by the CRLB, which makes it possible to obtain insights with respect of the estimator accuracy.

In (2.17), the vectorized representation of the received signal contains a white noise $\mathbf{z}$. Each sample of noise follows the complex multivariate Gaussian distribution and the samples are idealistically uncorrelated. In practice, correlated noise is present because of interference sources. However, this case is not taken into account in this thesis.

Assuming that the received signal is modeled linearly as in (2.17), we wish to determine

an estimator $\hat{\boldsymbol{\theta}}$, so that it applies a linear transformation onto the received samples, i.e. $\hat{\boldsymbol{\theta}} = g(\mathbf{y})$, where $g(.)$ is some function. To determine good estimators, it is important to firstly establish a statistical model for the measurements, so that it describes the randomness due to the presence of noise. The characterization of $\mathbf{y}$ by the probability density function (PDF) $p(\mathbf{y}; \boldsymbol{\theta})$ and parametrizing the PDF with the unknown parameter $\boldsymbol{\theta}$ become a fundamental prerequisite to obtain the CRLB. With this model, a class of PDFs is generated, where each of them differs according to the values of $\boldsymbol{\theta}$.

Assuming the PDF of the noise, Eq. (2.17) are considered, and that the sparsity $\|\boldsymbol{\theta}\|_0 = K$ is known a priori, the PDF of the received signal can be expressed as

$$p(\mathbf{y}; \boldsymbol{\theta}_K, \boldsymbol{\theta}_K^*) = \frac{1}{\left(2\pi\sigma^2\right)^{TN_{\mathrm{R}}/2}} \exp \frac{\left(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\Phi}_K\boldsymbol{\theta}_K\right)^H \left(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\Phi}_K\boldsymbol{\theta}_K\right)}{2\sigma^2}, \tag{2.18}$$

where σ^2 is the noise variance, $\boldsymbol{\theta}_K$ is a vector that contains only the K non-zero entries of $\boldsymbol{\theta}$, and $\boldsymbol{\Phi}_K$ is the range space of $\boldsymbol{\Phi}$. For convenience, we decide to express the PDF as a function of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ as if they were different variables. This type of assumption was already used in other works as for instance in [39].

The multivariate CRLB gives a bound for any unbiased estimator $\hat{\boldsymbol{\theta}}$ in function of the Fisher information matrix

$$\mathbf{C}_{\mathrm{CRLB}} - \mathbf{I}(\boldsymbol{\theta}_K, \boldsymbol{\theta}_K^*)^{-1} \geq \mathbf{0}_{K\times K}, \tag{2.19}$$

where the Fisher information matrix is given by

$$\mathbf{I}(\boldsymbol{\theta}_K, \boldsymbol{\theta}_K^*) = -\mathrm{E}\left\{\frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta}_K, \boldsymbol{\theta}_K^*)}{\partial \boldsymbol{\theta}_K^H \partial \boldsymbol{\theta}_K^T}\right\}. \tag{2.20}$$

To calculate this matrix, the first derivative with respect to $\boldsymbol{\theta}^H$ of the function $\ln p(\mathbf{y}; \boldsymbol{\theta}_K, \boldsymbol{\theta}_K^*)$ is obtained

$$\frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta}_k, \boldsymbol{\theta}_k^*)}{\partial \boldsymbol{\theta}_k^H} = \frac{1}{2\sigma^2} \sum_{m=1}^{M} \left(y[m] - [\boldsymbol{\Psi}]_m^H [\boldsymbol{\Phi}]_k\right) \boldsymbol{\Phi}^H [\boldsymbol{\Psi}]_m .$$

Then, the second derivative is obtained

$$\frac{\partial^2 \ln p(\mathbf{y}; \boldsymbol{\theta}_k, \boldsymbol{\theta}_k^*)}{\partial \boldsymbol{\theta}_k^H \partial \boldsymbol{\theta}_k^T} = -\frac{1}{2\sigma^2} \boldsymbol{\Phi}^H \boldsymbol{\Psi} \boldsymbol{\Psi}^H \boldsymbol{\Phi}.$$

From the second derivative, the Fisher information matrix is given by

$$\mathbf{I}(\boldsymbol{\theta}_K, \boldsymbol{\theta}_K^*) = \frac{1}{2\sigma^2} \boldsymbol{\Phi}^H \boldsymbol{\Psi} \boldsymbol{\Psi}^H \boldsymbol{\Phi}. \tag{2.21}$$

Note that the fisher information matrix defined in (2.21) is composed by the inner product of $\boldsymbol{\Upsilon} = \boldsymbol{\Psi}\boldsymbol{\Phi}$. This means that the off-diagonal elements can be described as

$$[\boldsymbol{\Upsilon}]_i^H [\boldsymbol{\Upsilon}]_j = [\boldsymbol{\Phi}]_i^H \boldsymbol{\Psi} \boldsymbol{\Psi}^H [\boldsymbol{\Phi}]_j . \tag{2.22}$$

From Definition 3, it is possible to conclude that the matrix $\mathbf{I}(\boldsymbol{\theta}_K, \boldsymbol{\theta}_K^*)$ gives also information about the mutual coherence, which is the highest value among all the off-diagonal entries. Besides this, the pilot design also plays an important role in the estimation process. Note that, if the system was able to employ orthogonal pilots, the Fisher information matrix would be diagonal. However, because of overhead issues, the length of the training sequences is shorter than the total number of antennas, which makes it impossible to achieve orthogonality in this case. One way to reduce the loss of performance compared with the orthogonal case consists of designing pilot sequences that have good properties of cross-correlation. It can be measured from the off-diagonal entries of the matrix $\boldsymbol{\Psi}\boldsymbol{\Psi}^H$, where the maximum value among all the entries must be minimized. To achieve such a condition, the mutual-coherence of the pilot sequences must achieve their minimum value.

From CS, if $\boldsymbol{\Psi}$ is a full-rank matrix, its mutual coherence is bounded as shown in [35]. Applying this result to the problem, the minimum value is given by

$$\begin{aligned} \mu\left(\boldsymbol{\Psi}\right) &\geq \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}} - TN_{\mathrm{R}}}{TN_{\mathrm{R}}\left(N_{\mathrm{T}}N_{\mathrm{R}} - 1\right)}} \\ &\geq \sqrt{\frac{N_{\mathrm{T}} - T}{T\left(N_{\mathrm{T}} - 1\right)}}. \end{aligned} \tag{2.23}$$

The equality is only achieved for a family of matrices called Grassmannian [35]. For these matrices, the angles between each pair of columns is the same, and is also the smallest possible. This type of matrix is extremely attractive to produce pilot sequences. However, the problem is to obtain such a Grassmannian matrices. They are usually very difficult to be obtained, but heuristics can be used to achieve approximations as shown in [35].

2.6 Channel Estimators and Receiver Structures

The use of downlink pilots in massive MIMO tends to limit the system spectral efficiency. Classical estimators as LS estimators achieve the minimum error upon the use of orthogonal sequences. However, to achieve this orthogonality the length of such sequences is at least equal to the number of antennas. This indeed demands a lot of communication resources in massive MIMO and consequently reduces the promised spectral efficiency gains. There is a need for techniques that achieve good accuracy, but demanding fewer pilots.

In this section, it is proposed three estimators that lead to three different architectures of receivers. One solution is developed based on the greedy algorithm, called OMP; the second is obtained using the Lasso formulation and ADMM method to derive a primal-dual solution; and the third is developed also from the Lasso formulation, but it is implemented in a distributed fashion.

2.6.1 Greedy Channel Estimation Method

Let us consider in a first moment a scenario without noise so that we can study the problem. Suppose that the matrix $\boldsymbol{\Upsilon}$ has $spark\{\boldsymbol{\Upsilon}\} > 2$, and the solution for the problem

$$\boldsymbol{\theta} = \arg\min \|\boldsymbol{\theta}\|_0 \text{ s.t. } \mathbf{y} = \boldsymbol{\Upsilon}\boldsymbol{\theta}, \tag{2.24}$$

is $\|\boldsymbol{\theta}\|_0 = 1$. This means that the vector of measurements $\mathbf{y}$ has one scalar that multiplies one of the columns of the matrix $\boldsymbol{\Upsilon}$, and this solution is known to be unique [13]. It is possible to identify this column by applying $N_{\mathrm{T}}N_{\mathrm{R}}$ tests, one for each column of $\boldsymbol{\Upsilon}$. The nth test can be performed by minimizing the error

$$e(n) = \| [\boldsymbol{\Upsilon}]_n \omega_n - \mathbf{y}\|_2^2. \tag{2.25}$$

This leads to the solution

$$\omega_n' = \frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}. \tag{2.26}$$

Substituting the equation (2.26) into (2.25) the error expression becomes

$$\begin{aligned} e(n) &= \| [\boldsymbol{\Upsilon}]_n \frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2} - \mathbf{y}\|_2^2 \\ &= \|\mathbf{y}\|_2^2 + \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right)^2 \| [\boldsymbol{\Upsilon}]_n \|_2^2 - \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right)^* [\boldsymbol{\Upsilon}]_n^H \mathbf{y} - \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right) \mathbf{y}^H [\boldsymbol{\Upsilon}]_n \\ &= \|\mathbf{y}\|_2^2 + \frac{\left([\boldsymbol{\Upsilon}]_n^H \mathbf{y}\right)^2}{\| [\boldsymbol{\Upsilon}]_n \|_2^2} - \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right)^* [\boldsymbol{\Upsilon}]_n^H \mathbf{y} - \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right) \mathbf{y}^H [\boldsymbol{\Upsilon}]_n \\ &= \|\mathbf{y}\|_2^2 - \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right)^* [\boldsymbol{\Upsilon}]_n^H \mathbf{y} \end{aligned} \tag{2.27}$$

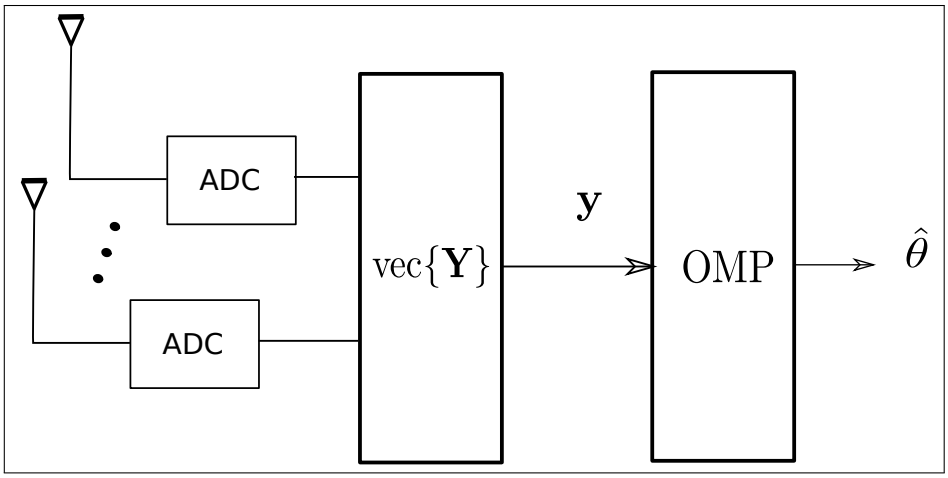
If $e(n) = 0$ in (2.27), the non-zero entry of $\boldsymbol{\theta}$ was found. Thus, the test to identify the entry of interest is

$$\begin{aligned} \|\mathbf{y}\|_2^2 - \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right)^* [\boldsymbol{\Upsilon}]_n^H \mathbf{y} &= 0 \\ \left(\frac{[\boldsymbol{\Upsilon}]_n^H \mathbf{y}}{\| [\boldsymbol{\Upsilon}]_n \|_2^2}\right)^* [\boldsymbol{\Upsilon}]_n^H \mathbf{y} &= \|\mathbf{y}\|_2^2 \\ | [\boldsymbol{\Upsilon}]_n^H \mathbf{y}|^2 &= \|\mathbf{y}\|_2^2 \| [\boldsymbol{\Upsilon}]_n \|_2^2. \end{aligned} \tag{2.28}$$

The expression (2.28) shows that if $\mathbf{y}$ and $[\boldsymbol{\Upsilon}]_n$ are parallel, n is the non-zero entry of $\boldsymbol{\theta}$. The search procedure in this case takes a number of operations on the order of $\mathcal{O}\left(TN_{\mathrm{T}}N_{\mathrm{R}}\right)$ flops.

Consider now $spark\{\boldsymbol{\Upsilon}\} > 2K_0$ and $\|\boldsymbol{\theta}\|_0 = K_0$. The vector $\mathbf{y}$ is a linear combination of at most K_0 columns of $\boldsymbol{\Upsilon}$. Using the same solution, the optimum approach is to test all the $\binom{N_{\mathrm{T}}N_{\mathrm{R}}}{K_0}$ possible combination of columns from $\boldsymbol{\Upsilon}$. The number of operations in this situation is $\mathcal{O}\left({N_{\mathrm{T}}}^{K_0}{N_{\mathrm{R}}}^{K_0}TK_0^2\right)$ flops, which seems prohibitive in any telecommunication setting.

Figura 2.2 – Centralized channel estimation solution using greedy algorithm.



Source: Provided by the author.

A greedy algorithm avoids such combinatorial solution and evaluates a series of locally optimal single-term updates. From these local problems, the algorithm constructs iteratively a sparse vector $\boldsymbol{\theta}_K$ by using the rule in (2.28). The columns corresponding to the non-zero entries of $\boldsymbol{\theta}$ reduces maximally the residual error in approximating $\mathbf{y}$. Using the column of $\boldsymbol{\Upsilon}$ obtained from each iteration, the residual error is evaluated and if it falls below a predefined threshold, the algorithm stops. The pseudocode is shown the Table 1.

The implementation of the algorithm consists of collecting the training signal after the analog-digital conversion, followed by an operation that performs the vectorization of the samples. This vector is the input to the algorithm described in Table 1.

Although the OMP algorithm saves computational processing by avoiding exhaustive search through all $\binom{N_{\mathrm{T}}N_{\mathrm{R}}}{K_0}$ possible combination of columns from $\boldsymbol{\Upsilon}$, the search procedure of finding the strongest components can also lead to a lot operations. The number of multiplications is $\mathcal{O}\left({N_{\mathrm{T}}}^{K_0}{N_{\mathrm{R}}}^{K_0}TK_0^2\right)$ and because we have a massive array the implementation of such an algorithm can be prohibitive on devices with limited processing capabilities. The block diagram of the receiver is shown in Fig. **??**

2.6.2 Distributed Linear Regression for Sparse Channels Using ADMM

Let us consider the problem of l_1 regression known as Lasso [35] which is stated as

$$\min_{\boldsymbol{\theta}} \ (1/2)\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \beta \left\|\boldsymbol{\theta}\right\|_1, \tag{2.29}$$

where $\beta > 0$ is a parameter that controls the importance of sparse solutions. The Lasso formula-

Algorithm 1 Pseudocode of the OMP algorithm.

Require: The dictionary and the measurement matrices that are combined as $\boldsymbol{\Upsilon} = \boldsymbol{\Psi}\boldsymbol{\Phi}$ and the received signal $\mathbf{y}$ containing the measurements

Ensure: Find the sparsest solution of $\boldsymbol{\theta}$ that solves the problem $\boldsymbol{\theta} = \arg\min \|\boldsymbol{\theta}\|_0$ s.t. $\mathbf{y} = \boldsymbol{\Upsilon}\boldsymbol{\theta}$
Start the algorithm with

- the initial solution is $\boldsymbol{\theta}^0 = \mathbf{0}$
- the residual vector $\mathbf{r} = \mathbf{y}$;
- the initial support $\mathcal{S}^0 = Support\{\boldsymbol{\theta}^0\} = \mathbf{0}$

while $\|\mathbf{r}\|_2 > \epsilon$ or $k \leq K_0$ **do**
$n^k = \arg\max_n \| [\boldsymbol{\Upsilon}]_n^H \mathbf{y}\|^2$
$\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{n^k\}$
$\boldsymbol{\theta}^k = (\boldsymbol{\Upsilon}_{\mathcal{S}^k})^\dagger \mathbf{y}$
$\mathbf{r}^k = \mathbf{y} - \boldsymbol{\Upsilon}_{\mathcal{S}^k}\boldsymbol{\theta}^k$
end while

tion can be rewritten into the ADMM framework that is given by

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \ (1/2)\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \beta \|\mathbf{q}\|_1 , \\ &\text{s.t. } \boldsymbol{\theta} - \mathbf{q} = \mathbf{0}. \end{aligned} \tag{2.30}$$

It is possible to recast the problem using the augmented Lagrangian function

$$L_\rho(\boldsymbol{\theta}, \mathbf{q}, \lambda) = (1/2)\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \beta \|\boldsymbol{\theta}\|_1 + \boldsymbol{\lambda}^T (\boldsymbol{\theta} - \mathbf{q}) + (\rho/2)\|\boldsymbol{\theta} - \mathbf{q}\|_2, \tag{2.31}$$

where $\rho > 0$ is a penalty parameter. The insertion of the third term is to bring robustness to the solution. The ADMM is a variation of the dual ascent algorithm [40]. In this method, the algorithm implementation exploits the dual decomposition in (2.31) by alternating between the primal and dual formulations. The dual solution is obtained using the gradient ascent, however when the term with a penalty is added, $(\rho/2)\|\boldsymbol{\theta} - \mathbf{q}\|_2$, the convergence of the algorithm extends under far more general conditions than without using the penalty [41].

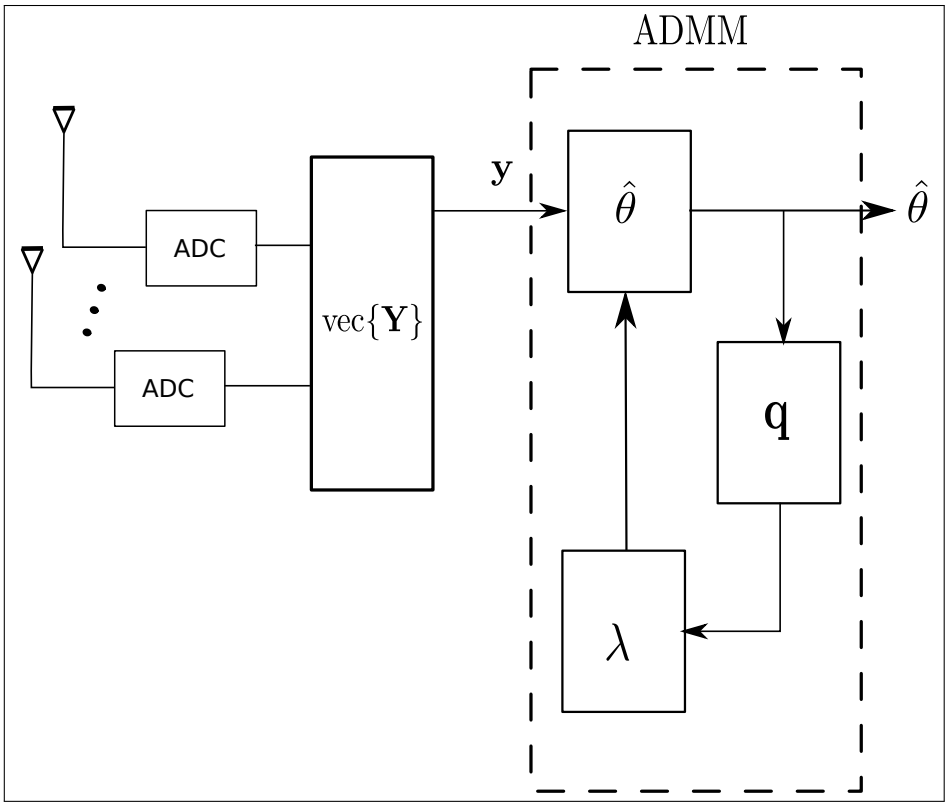
The ADMM expressions per iteration are

$$\begin{aligned} \boldsymbol{\theta}^{k+1} &\doteq \arg\min_{\boldsymbol{\theta}} L_\rho(\boldsymbol{\theta}, \mathbf{q}^k, \boldsymbol{\lambda}^k) \\ \mathbf{q}^{k+1} &\doteq \arg\min_{\mathbf{q}} L_\rho(\boldsymbol{\theta}^{k+1}, \mathbf{q}, \boldsymbol{\lambda}^k) \\ \boldsymbol{\lambda}^{k+1} &\doteq \boldsymbol{\lambda}^k + \boldsymbol{\theta}^{k+1} - \mathbf{q}^{k+1}. \end{aligned}$$

Deriving the expressions, we can say that

$$\begin{aligned} \boldsymbol{\theta}^{k+1} &\doteq \left(\boldsymbol{\Phi}^H\boldsymbol{\Psi}^H\boldsymbol{\Psi}\boldsymbol{\Phi} + \rho\mathbf{I}\right)^{-1}\left(\boldsymbol{\Phi}^H\boldsymbol{\Psi}^H\mathbf{y} + \rho\left(\mathbf{q}^k - \boldsymbol{\lambda}^k\right)\right) \\ \mathbf{q}^{k+1} &\doteq S_{\beta/\rho}\left(\boldsymbol{\lambda}^k + \boldsymbol{\theta}^{k+1}\right) \\ \boldsymbol{\lambda}^{k+1} &\doteq \boldsymbol{\lambda}^k + \boldsymbol{\theta}^{k+1} - \mathbf{q}^{k+1}. \end{aligned}$$

Figura 2.3 – Alternated channel estimation using ADMM solution.



Source: Provided by the author

The Lasso formulation has the advantage of considering the noise using the term $\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2^2$ which provides increased robustness in scenarios with limited SNR. The algorithm is summarized in the Table 2. The receiver architecture is illustrated in Fig. **??**.

The estimation of $\boldsymbol{\theta}$ can be done in a distributed way by splitting the measurement vector $\mathbf{y}$ into smaller ones. In this way, smaller ADMM blocks can perform their own estimate of $\boldsymbol{\theta}$, as shown in Fig. **??**. In this way, the baseband processing architecture can be designed to distribute the computation of all the measurements. In other words, part of the received measurements are allocated to different ADMM. The advantage of this solution is the possibility to use less powerful processors while distributing the computational load across different processors. Each algorithm has its own estimation of $\boldsymbol{\theta}$ and this estimate is reported to a fusion center that combines multiple estimates to update the global $\boldsymbol{\theta}$.

To split the measurements, we consider that

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_I \end{bmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_1 \\ \boldsymbol{\Psi}_2 \\ \vdots \\ \boldsymbol{\Psi}_I \end{bmatrix}.$$

Each subvector $\mathbf{y}_i$ $\forall i \in \{1, \ldots, I\}$ has $\frac{TN_{\mathrm{R}}}{I}$ elements. This means that each ADMM block

Algorithm 2 Primal-dual Lasso solution.

Require: Proper hardware architecture for parallel computation. Moreover, the dictionary and the measurement matrices.

Ensure: Find the sparsest solution of $\boldsymbol{\theta}$ that returns the minimum value $L_\rho(\boldsymbol{\theta}, \mathbf{q}, \lambda) = (1/2)\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \beta \|\boldsymbol{\theta}\|_1 + \boldsymbol{\lambda}^T (\boldsymbol{\theta} - \mathbf{q}) + (\rho/2)\|\boldsymbol{\theta} - \mathbf{q}\|_2$,
Start the algorithm with

- initialize randomly $\mathbf{q}^1 \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}}}$
- initialize randomly $\boldsymbol{\lambda}^1 \in \mathbb{R}^{N_{\mathrm{T}} N_{\mathrm{R}}}$

while $\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2 > \epsilon$ **do**
$\boldsymbol{\theta}^{k+1} \doteq \left(\boldsymbol{\Phi}^H \boldsymbol{\Psi}^H \boldsymbol{\Psi}\boldsymbol{\Phi} + \rho \mathbf{I}\right)^{-1} \left(\boldsymbol{\Phi}^H \boldsymbol{\Psi}^H \mathbf{y} + \rho \left(\mathbf{q}^k - \boldsymbol{\lambda}^k\right)\right)$
$\mathbf{q}^{k+1} = S_{1/\rho}\left(\mathbf{q}^k + \boldsymbol{\theta}^{k+1}\right)$
$\boldsymbol{\lambda}^{k+1} \doteq \boldsymbol{\lambda}^k + \boldsymbol{\theta}^{k+1} - \mathbf{q}^{k+1}$
compute $\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2$
end while

uses matrices $\boldsymbol{\Psi}_i \; \forall i \in \{1, \ldots, I\}$, that are non-overlapping submatrices of $\boldsymbol{\Psi}$. Each one of the subvectors $\mathbf{y}_i$ are allocated to the corresponding ADMM block i. Assuming this distributed architecture, we exploit the l_2 norm property

$$\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2 = \|\boldsymbol{\Psi}_1\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}_1\|_2 + \|\boldsymbol{\Psi}_2\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}_2\|_2 + \ldots + \|\boldsymbol{\Psi}_I\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}_I\|_2, \tag{2.32}$$

to rewrite the Lasso problem as

$$\begin{aligned} \min_{\boldsymbol{\theta}} \; & (1/2) \sum_{i=1}^{I} \|\boldsymbol{\Psi}_i\boldsymbol{\Phi}\boldsymbol{\theta}_i - \mathbf{y}_i\|_2^2 + \beta \|\mathbf{q}\|_1 , \\ \text{s.t. } & \boldsymbol{\theta}_i - \mathbf{q} = \mathbf{0}, \; \forall i \; \in \; \{1, \ldots, I\}. \end{aligned} \tag{2.33}$$

Using ADMM, we can define the expression as

$$\begin{aligned} \boldsymbol{\theta}_i^{k+1} &\doteq \arg\min_{\boldsymbol{\theta}_i} \left((1/2)\|\boldsymbol{\Psi}_i\boldsymbol{\Phi}\boldsymbol{\theta}_i - \mathbf{y}_i\|_2 + \rho/2\|\boldsymbol{\theta}_i - \mathbf{q}^k + \boldsymbol{\lambda}_i^k\|_2\right) \\ \mathbf{q}^{k+1} &\doteq S_{\beta/\rho I}\left(\frac{1}{I}\sum_{i=1}^{I} \boldsymbol{\theta}_i^{k+1} - \frac{1}{I}\sum_{i=1}^{I} \lambda_i\right) \\ \boldsymbol{\lambda}_i^{k+1} &\doteq \boldsymbol{\lambda}_i^k + \boldsymbol{\theta}_i^{k+1} - \mathbf{q}^{k+1}. \end{aligned} \tag{2.34}$$

Each $\boldsymbol{\theta}_i$ update can be solved with the analytic expression

$$\boldsymbol{\theta}_i^{k+1} = \left(\boldsymbol{\Phi}^H \boldsymbol{\Psi}_i^H \boldsymbol{\Psi}_i \boldsymbol{\Phi} + \rho \mathbf{I}\right)^{-1} \left(\boldsymbol{\Phi}^H \boldsymbol{\Psi}_i^H \mathbf{y}_i + \rho \left(\mathbf{q}^k - \boldsymbol{\lambda}_i^k\right)\right). \tag{2.35}$$

or, alternatively,

$$\boldsymbol{\theta}_i^{k+1} = \left(\boldsymbol{\Upsilon}_i^H \boldsymbol{\Upsilon}_i + \rho \mathbf{I}\right)^{-1} \left(\boldsymbol{\Upsilon}_i^H \mathbf{y}_i + \rho \left(\mathbf{q}^k - \boldsymbol{\lambda}_i^k\right)\right). \tag{2.36}$$

Table. 3 summarizes the method presenting the pseudocode for the distributed channel estimation solution.

Algorithm 3 Pseudocode of the distributed Lasso problem using ADMM formulation.

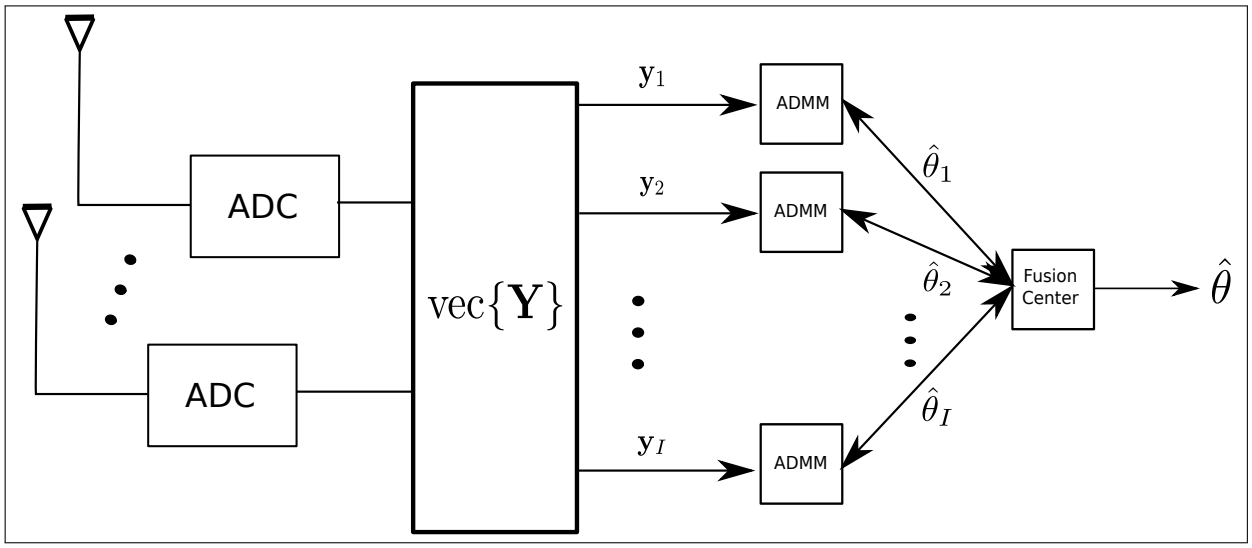
Require: The receiver has a properly architecture in terms of hardware implementation. Updates are processed in a distributed fashion. Moreover, the dictionary and the measurement matrices are assumed to be known penalty ρ

Ensure: Find the sparsest solution of $\boldsymbol{\theta}$ that returns the minimum value $L_\rho(\boldsymbol{\theta}, \mathbf{q}, \lambda) = (1/2)\|\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \beta\|\boldsymbol{\theta}\|_1 + \boldsymbol{\lambda}^T(\boldsymbol{\theta} - \mathbf{q}) + (\rho/2)\|\boldsymbol{\theta} - \mathbf{q}\|_2$,
Start the algorithm with

- initializing $\mathbf{q}^1 \in \mathbb{C}^{N_{\mathrm{T}}N_{\mathrm{R}}}$
- initializing $\boldsymbol{\lambda}_i^1 \in \mathbb{R}^{N_{\mathrm{T}}N_{\mathrm{R}}}$

while $\|\frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\theta}_i^{k+1} - \frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\theta}_i^{k}\|_2 > \epsilon$ **do**
$\boldsymbol{\theta}_i^{k+1} \doteq \left(\boldsymbol{\Phi}^H\boldsymbol{\Psi}_i^H\boldsymbol{\Psi}_i\boldsymbol{\Phi} + \rho\mathbf{I}\right)^{-1}\left(\boldsymbol{\Phi}^H\boldsymbol{\Psi}_i^H\mathbf{y}_i + \rho\left(\mathbf{q}^k - \boldsymbol{\lambda}_i^k\right)\right)$
$\mathbf{q}^{k+1} \doteq S_{\beta/\rho I}\left(\frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\theta}_i^{k+1} + \frac{1}{I}\sum_{i=1}^{I}\lambda_i\right)$
$\boldsymbol{\lambda}_i^{k+1} \doteq \boldsymbol{\lambda}_i^k + \boldsymbol{\theta}_i^{k+1} - \mathbf{q}^{k+1}$
compute $\|\frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\theta}_i^{k+1} - \frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\theta}_i^{k}\|_2$
end while

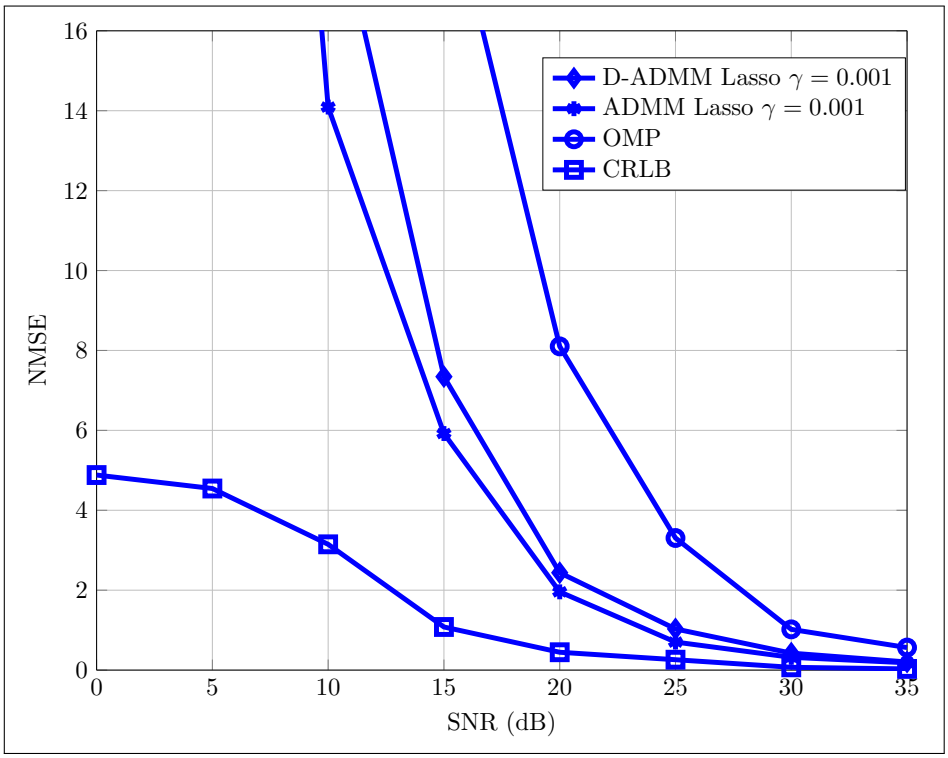
Figura 2.4 – Distributed channel estimation solution using ADMM in each module.



Provided by the author.

The distributed processing can be implemented using different architectures. One possibility consists of splitting the measurements and allocating them into processors so that each one performs individually and report the estimation to the fusion center that combines all the estimates. The method also can be applied to distributed hardware architecture. More specifically, a set of modules with one or multiple antennas could be used to operate separately. Each module exchanges information with the fusion center until each ADMM achieves the convergence.

Figura 2.5 – Channel estimation performance of distributed ADMM, centralized ADMM and OMP for a 128×2 MIMO using system $T = 40$.



Source: Provided by the author.

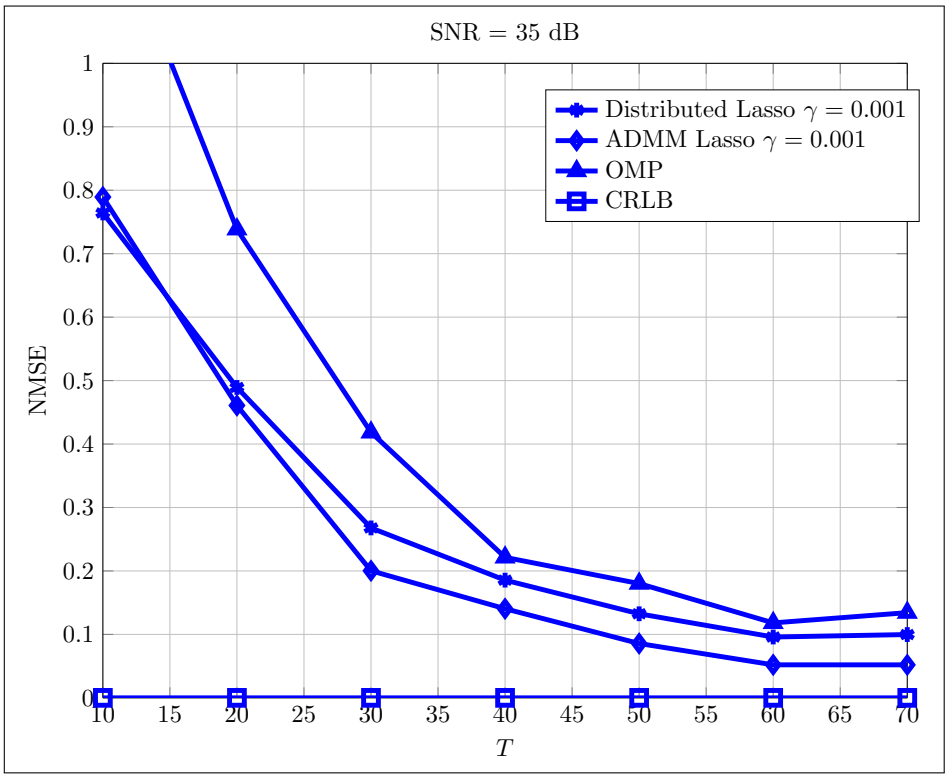
2.7 Numerical Results

This section provides some simulation results to compare the performance of three sparsity-based channel estimators. The number of antennas at the BS is $N_{\mathrm{T}} = 128$ and the number of antennas at the UE is $N_{\mathrm{R}} = 2$. The channel is generated using a Fourier basis to compose the model (2.9) and the coupling matrix $\mathbf{\Omega}$ is generated randomly with 0s and 1s while $\mathbf{G}$ is obtained from a complex Gaussian distribution.

Fig. **??** shows a comparison between the three methods and the CRLB. In this plot, the BS transmits random pilot sequences, obtained from a Bernoulli distribution, with length $T = 40$ and the channel sparsity is $K = 10$. The ADMM methods, which are obtained from Lasso formulation achieve a better performance because they take into account the noise power which provides enhanced performance over the OMP algorithm. The plots also show that the channel estimation error becomes closer to the CRLB as the SNR increases. This means that the methods are successful in finding sparse channel coefficients, corresponding the nonzero entries of $\mathbf{G}$, because the CRLB assumes perfect knowledge of the sparse matrix indexes corresponding to the non-zero entries.

Fig. **??** compares the channel estimation performance of the three methods with respect to the length of the training sequences considering SNR = 35dB. As the length of the sequences increases, the methods get closer to CRLB up to a gap between them and the CRLB. It is impor-

Figura 2.6 – Channel estimation performance for a 128×2 MIMO system considering SNR = 35dB.



Source: Provided by the author.

tant to note that the channel estimation error does not improve even increasing the overhead. If such a saturation happens, the BS must increase the power to improve the estimator performance.

Figures **??** and **??** show that the BS can adjust the power and the pilot length to control the estimation accuracy. With this feature, the system can either prioritize reducing the overhead or saving power transmission.

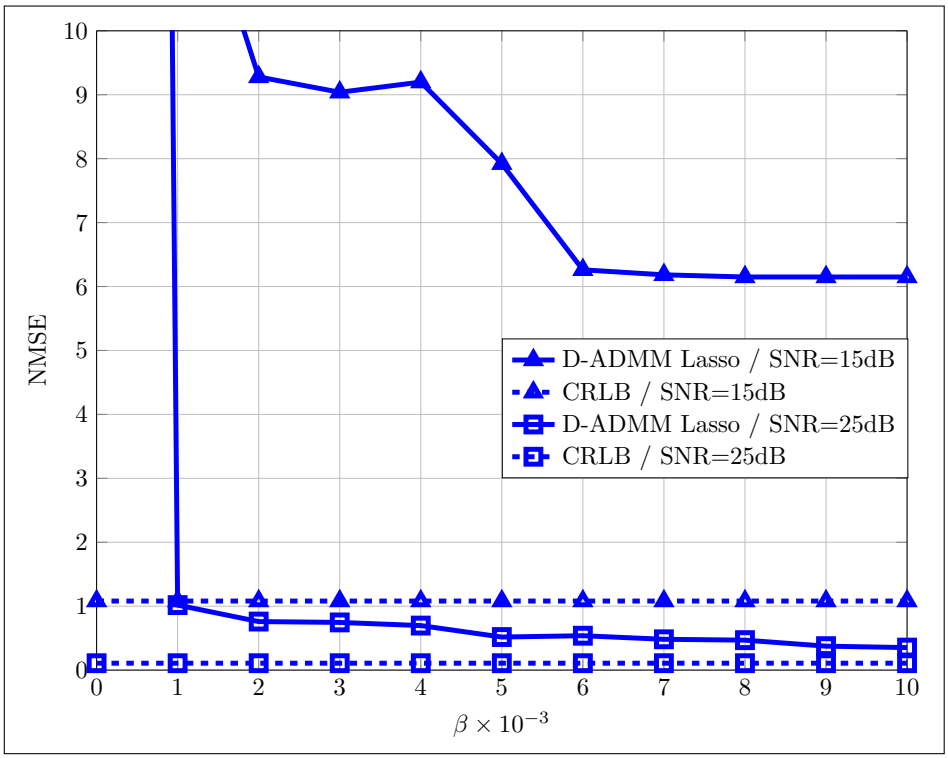
One drawback of Lasso-based techniques is the choice of β. Figure **??** shows that the parameter depends on the SNR. The increase of β means that the estimator is prioritizing sparser solutions.

2.8 Conclusion

Sparsity-based channel estimators provide accurate estimates demanding fewer pilots than the classical LS solution. The methods proposed in this chapter collect DL training sequences and extract the channel sparse representation. This is the key factor that enables the estimator to reduce the pilot overhead.

Comparing the performance among the three methods, the greedy solution implemented through the OMP algorithm achieves CRLB only for high SNR. This is because the optimization problem in such a case does not take into account the additive noise in the received signal. The

Figura 2.7 – Channel estimation performance of D-ADMM in a 128×2 MIMO system. In this plot, we vary the parameter β.



Source: Provided by the author.

solutions derived from the lasso formulation consider the presence of noise which results in a better performance.

The distributed solution shows negligible loss of performance compared with its centralized version. This provides the possibility of implementing distributed architectures at the receiver and still achieving the same centralized channel estimation performance.

Another interesting result regards to the optimum pilot sequence. The discussions with respect to the CRLB and the results indicate that the Grassmanian achieves the CRLB. Thus, Grasmanian pilots are promising for achieving high channel estimation quality.

Although the techniques proposed in this chapter show good performance in certain conditions, most of the current systems have frequency-selective channels. The solutions developed in this chapter do not cope with such scenarios. In the next chapter, we continue to use the CS estimator, but we keep our interest in the use of CS-based techniques for massive MIMO channel estimation, but shift the focus to frequency-selective channels. This opens the possibility of exploiting the multidimensionality of the channel model by means of a tensor CS approach.

3 TENSOR COMPRESSIVE SENSING FOR ESTIMATING FREQUENCY SELECTIVE CHANNELS

Abstract

In this chapter, the study on sparse channel estimation techniques is extended to frequency-selective channels. The dependency of the channel on the frequency is generally modeled by means of a set of channel matrices, one for each frequency bin. Assuming, for instance, multicarrier modulation, pilots symbols are spread across different subcarriers. This procedure can consume a lot of communication resources without properly exploiting the channel structure. The solution proposed in this chapter combines tensor analysis and the CS framework to exploit a joint sparsity resulting from a 3D sparse representation of the frequency-selective massive MIMO channel. A new tensor-OMP algorithm to exploit the tensor sparse representation is proposed, which exploits the tensor factorization to solve a greedy problem in an alternate fashion. This drastically reduces the complexity compared with the traditional OMP solution. Our results corroborate the reduction on the number of pilots in time and frequency. The proposed algorithm not only estimates the channel coefficients associated with the pilot subcarriers, but also interpolates across frequency to estimate the channel coefficients associated with the remaining subcarriers.

3.1 Introduction

As discussed in Chapter **??**, massive MIMO has the potential of increasing the spectral efficiency of current wireless communication systems by employing hundreds of antennas at the BS. However, the promised gains in spectral efficiency are achieved up to CSI knowledge and its acquisition demands a lot of pilots if the BS uses the traditional techniques of channel estimation. To circumvent it, we developed in the last chapter, sparsity-based techniques that use a basis to achieve a sparse channel representation.

The choice of a basis is fundamental for the good performance of sparsity-based channel estimators. Imposing a basis means that the estimator represents the channel response by a structured model. In general, most papers in the massive MIMO literature consider an independent identically distributed (i.i.d) channel. However, the spatial structure of the channel which is related to the scatterers distribution have a direct impact on the channel capacity and should be taken into account in more realistic propagation environments [42].

The use of spatially structured models is very important in scenarios where the environment has a poor scattering of multipath propagation, which motivates the use of low-rank channel models. These models can be exploited by sparsity-based estimators as shown in [43–46]. The authors of [43] present a CS-based study for sparse channel estimation and its impact on the de-

sign of efficient pilot assisted techniques. For instance, the works [45,46] capitalize on structured channel models to obtain the angles of arrival and departure to reconstruct the CSI. Examples of structured MIMO channels are the Kronecker model [47], the Weichselberger's model [36] or the virtual channel representation [48]. They spatially describe the paths without considering their delays. Thus, those models are only applied to narrowband channels. Most of modern wireless communication systems perform their transmissions over frequency-selective channels. In such a case, the channel frequency response is not flat over the bandwidth which means that the models used by sparse channel estimators need to consider the frequency selectiveness. This means to use sparse models that describe angle and delay domains. For the MIMO case, the narrowband channel is modelled as a matrix because only the spatial domains are taken into account. For the frequency-selective channel, a three-dimensional representation can be used.

A useful tool to model multidimensional channels is the tensor analysis. This is a multilinear algebra tool that can be applied to problems involving multidimensional modelling. For instance, [49] adopted this tool to extend the Kronecker model to the wideband case. Following that work, in [50], a tensor-based wideband channel model built up from the Weichselberger's model [36] was proposed and validated experimentally. Recently, [51] introduced a CSI tensor expression and a full correlation model for wideband MIMO channels by means of a general coupling-based framework built from the relationship between spatial and frequency channel dimensions. Beyond the importance of channel modelling, as stated previously, tensor models such as the Parallel Factors (PARAFAC) model and the Tucker model [52] have interesting properties that make possible to exploit sparse representations. Among these multilinear models, the Tucker model has a special interest in this work. It provides us with a trilinear sparse channel representation that can be exploited by a CS estimator to reduce the pilot overhead.

In this chapter, we propose a sparsity-based channel estimation for wideband channels by exploiting the Tucker3 model to obtain the tensor sparse representation of the channel. Using this representation, it is possible to map the entries of a tensor into a vector. However, such a mapping uses the Kronecker basis representation to obtain the sparse vector. This basis has very large columns and rows which create an issue with limited processing receivers. A new greedy algorithm exploiting the Tucker3 model is proposed, where the solution uses the unfolding representations of the model to drastically simplify the search of the sparse tensor strongest entries.

Besides the computational complexity aspects, the solution uses the pilot placement of a time-frequency system. Including a space-frequency channel model in the estimator, it is possible to reduce the length of the pilot sequences and the number of subcarriers used for transmitting them. The method uses the measurements from the subcarriers and obtains the the sparse tensor. Using such a representation and the basis, the estimator can jointly estimate and interpolate the channel frequency response.

3.2 Tensor Prerequisites

Before proceeding with the tensor characterization of the frequency selective MIMO channel, it is convenient to provide some basic concepts and definitions of tensor algebra, which will be used in later sections.

Tensors are natural extensions of vectors and matrices to a multidimensional space (i.e. with more than two dimensions). A possible definition of a tensor $\mathcal{A}$ is given as follows.

Definition 4 *A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is a N-th order array whose typical element $a_{i_1,i_2,\ldots,i_N}$ is accessed via N indices, $i_n \in \{I_1, I_2, \ldots I_N\}$.*

According to Definition 4, a zero-order tensor is a scalar, a first-order tensor is a vector, a second-order tensor is a matrix and a third-order tensor is a "cube". A tensor with more than three dimensions is generally referred to as a higher-order tensor [53]. A tensor of order N can also be interpreted as the result of a multilinear projection by means of a tensor (Kronecker) product of N vector spaces.

When manipulating tensors, conventional operations such as sums and products can be efficiently applied by rearranging the tensor into equivalent matrices. This process is usually referred to as "matricization" or "unfolding" [52, 53].

3.2.1 n-Mode product

To define a multiplication between a tensor and a matrix in the tensor space, it is necessary to specify which mode (dimension) of the tensor is affected by that operation.

Definition 5 *The n-mode product of a tensor $\mathcal{X} \in \mathbb{C}^{J_1 \times J_2 \ldots \times J_N}$ and a matrix $\mathbf{A} \in \mathbb{C}^{I_n \times J_n}$ is symbolized by $\mathcal{Y} = \mathcal{X} \times_n \mathbf{A}$, where $\mathcal{Y} \in \mathbb{C}^{J_1 \times J_2 \ldots \times J_{n-1} \times I_n \times J_{n+1} \ldots \times J_N}$, each element being defined as [53]:*

$$y_{j_1,j_2,\ldots,j_{n-1},i_n,j_{n+1},\ldots,j_N} = \sum_{j_n=1}^{J_n} x_{j_1,j_2,\ldots,j_{n-1},j_n,j_{n+1},\ldots,j_N} a_{i_n,j_n}. \tag{3.1}$$

The n-mode product can be applied successively along several modes. This operation enjoys the commutative property, i.e.

$$(\mathcal{X} \times_n \mathbf{A}) \times_m \mathbf{B} = (\mathcal{X} \times_m \mathbf{B}) \times_n \mathbf{A} \ (m \neq n). \tag{3.2}$$

In the case $m = n$, we have

$$(\mathcal{X} \times_n \mathbf{A}) \times_n \mathbf{B} = \mathcal{X} \times_n \mathbf{BA}. \tag{3.3}$$

It is possible to perform multiplication in all modes of the tensor using a set of matrices $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times J_n}$, as follows:

$$\mathcal{X} \times \left\{ \bigcup_{i=1}^{N} \mathbf{A}^{(i)} \right\} = \mathcal{X} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \ldots \times_N \mathbf{A}^{(N)}. \tag{3.4}$$

The tensor transformation in Eq. (3.4) can alternatively be represented as a function of its n-mode matrix unfolding, i.e.:

$$\left[\mathcal{X} \times \left\{ \bigcup_{i=1}^{N} \mathbf{A}^{(i)} \right\} \right]_n = \mathbf{A}^{(n)} \mathbf{X}_{(n)} [\mathbf{A}^{(N)} \otimes \ldots \mathbf{A}^{(n+1)} \otimes \ldots \otimes \mathbf{A}^{(1)}]^T, \tag{3.5}$$

where $\mathbf{X}_{(n)} \in \mathbb{C}^{J_n \times J_1 \cdots J_{n-1} J_{n+1} \cdots J_N}$ is the n-mode matrix unfolding of $\mathcal{X} \in \mathbb{C}^{J_1 \times J_2 \ldots \times J_N}$.

To illustrate the above concepts, consider the case $N = 3$ (third-order tensor), which will be relevant to the context of this work. In this case, (3.4) can be rewritten in scalar form as follows

$$y_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} x_{pqr} a_{i,p} b_{j,q} c_{k,r}, \tag{3.6}$$

where y_{ijk} is an element of the fourth-order tensor $\mathcal{Y} \in \mathbb{C}^{I \times J \times K}$, g_{pqr} is an element of the third-order tensor $\mathcal{X} \in \mathbb{C}^{P \times Q \times R}$, and $a_{i,p}$, $b_{j,q}$, and $c_{k,r}$ are typical elements of the factor matrices $\mathbf{A} \in \mathbb{C}^{I \times P}$, $\mathbf{B} \in \mathbb{C}^{J \times Q}$, and $\mathbf{C} \in \mathbb{C}^{K \times R}$, respectively. Using the n-mode product (tensor) representation, Eq. (3.4) translates directly to

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}. \tag{3.7}$$

The matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, perform, respectively, 1-mode, 2-mode, and 3-mode products across the input tensor $\mathcal{X}$, yielding the output tensor $\mathcal{Y}$. Such an operation can also be interpreted as a third-dimensional filtering (e.g. compression or decompression) that maps $\mathcal{X}$ into $\mathcal{Y}$ and is represented Fig. 3.1.

Another useful way of representing the tensor transformation (3.7) consists of rewriting as three different (but equivalent) unfolded forms. According to Eq. (3.5), we have

$$\mathbf{Y}_{(1)} = \mathbf{A}\mathbf{X}_{(1)} [\mathbf{B} \otimes \mathbf{C}]^T \tag{3.8}$$

$$\mathbf{Y}_{(2)} = \mathbf{B}\mathbf{X}_{(2)} [\mathbf{C} \otimes \mathbf{A}]^T \tag{3.9}$$

$$\mathbf{Y}_{(3)} = \mathbf{C}\mathbf{X}_{(3)} [\mathbf{A} \otimes \mathbf{B}]^T. \tag{3.10}$$

Another way of representing the above equations is to use their vector form

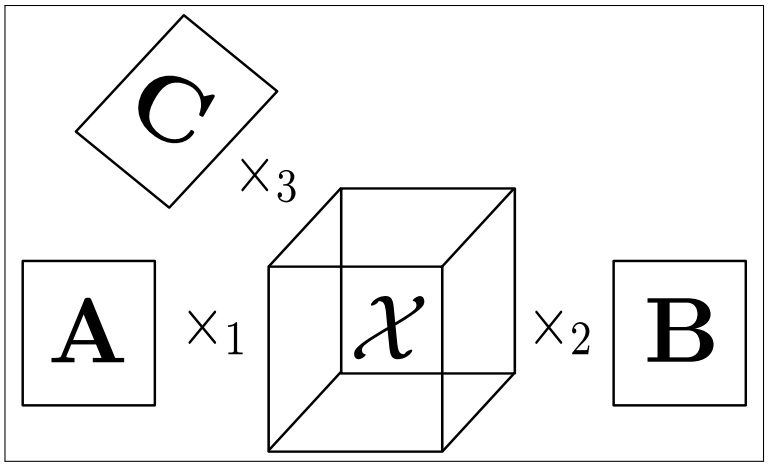
$$\mathbf{y}_{(1)} = (\mathbf{B} \otimes \mathbf{C} \otimes \mathbf{A})\, \mathbf{x}_{(1)} \tag{3.11}$$

$$\mathbf{y}_{(2)} = (\mathbf{C} \otimes \mathbf{A} \otimes \mathbf{B})\, \mathbf{x}_{(2)}, \tag{3.12}$$

$$\mathbf{y}_{(3)} = (\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C})\, \mathbf{x}_{(3)}, \tag{3.13}$$

where $\mathbf{x}_{(i)} \doteq \text{vec}\{\mathbf{X}_{(i)}\}$, and $\mathbf{y}_{(i)} \doteq \text{vec}\{\mathbf{Y}_{(i)}\}$ $\forall i \in \{1, 2, 3\}$.

Figura 3.1 – n-Mode product representation for a third order tensor model.



Source: Provided by the author.

Equation (3.7) is known in the tensor literature as the third-order Tucker decomposition [53]. In this chapter, the tensor formalism is used to model a frequency-selective MIMO channel by means of a sparse multidimensional representation. More specifically, the factor matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, in (3.7) will play the role of the fixed bases while the core tensor $\mathcal{X}$ will reveal the sparse structure of the channel in the joint angle-delay domain. Later, this structure will be exploited for channel estimation to derive a tensor-based sparse channel estimator.

3.2.2 Kronecker compressed sensing

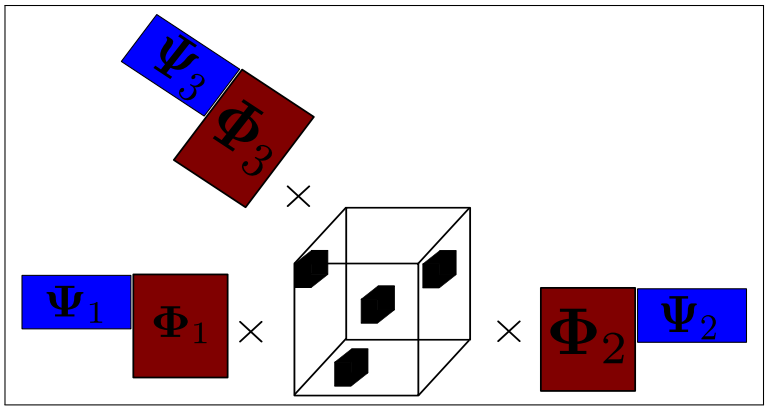
Kronecker CS refers to problems where we have to exploit sparsity in multidimensional signals [54]. In this type of problem, the measurements are concatenated into a big vector which produces a particular Kronecker structure to the measurement and the basis matrices. More specifically, the basis and the measurement matrices are defined as $\mathbf{\Phi} = \mathbf{\Phi}_1 \otimes \mathbf{\Phi}_2 \otimes \ldots \otimes \mathbf{\Phi}_D$ and $\mathbf{\Psi} = \mathbf{\Psi}_1 \otimes \mathbf{\Psi}_2 \otimes \ldots \otimes \mathbf{\Psi}_D$, respectively. A given matrix $\mathbf{\Psi}_i$ and $\mathbf{\Phi}_i$ corresponds to the measurement and the basis with respect the dimension i.

The Kronecker structure leads to a tensor representation of the signal. Equation (3.11), for instance, reveals that if a vector $\mathbf{y}_1$ holds a relationship with $\mathbf{x}_1$ by a linear transformation defined as a Kronecker matrix, then $\mathbf{y}_1$ has a tensor representation defined by (3.7). This is true because $\mathbf{y}_1$ is the result of a vectorization over $\mathbf{Y}_{(1)}$. This matrix is, in turn, a representation of a third-order tensor that has matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ performing 1-mode, 2-mode, and 3-mode products, respectively, across a core tensor. This relationship is clear by comparing (3.11) and (3.7). Thus, $\mathbf{y}_1 = \mathbf{\Psi}\mathbf{\Phi}\mathbf{x}_1$ can be represented using the tensor formalism, where the vector $\mathbf{y}_1$ is rearranged as a tensor $\mathcal{Y}$ and, in turn, it admits a sparse representation $\mathcal{X}$ as shown in

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{\Psi}_1\mathbf{\Phi}_1 \ldots \times_D \mathbf{\Psi}_D\mathbf{\Phi}_D, \tag{3.14}$$

where the $M_i \times N_i$ measurement matrix $\mathbf{\Psi}_i$, $M_i < N_i$, compresses the tensor $\mathcal{X}$ along its i-th

Figura 3.2 – Third-order tucker decomposition of a sparse tensor.



Soruce: Provided by the author.

dimension, $\forall\, i = \{1, 2, \ldots, D\}$. The tensor representation for (3.14) is a D-order Tucker model. Figure 3.2 illustrates a third-order Tucker decomposition with a sparse core tensor.

As an extension to the traditional (vector) CS problem, in the case of multidimensional signals, the mutual coherence is considered per dimension. For a given dimension i, we can define $\boldsymbol{\Upsilon}_i = \boldsymbol{\Psi}_i \boldsymbol{\Phi}_i$ and the corresponding mutual coherence $\mu(\boldsymbol{\Upsilon}_i)$. The equivalent mutual coherence can be obtained from the relationship [54]

$$\mu\left(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_D\right) = \prod_{i=1}^{D} \mu(\boldsymbol{\Upsilon}_i). \tag{3.15}$$

From (3.15), we can note that if multiple measurements are employed over different modes, the mutual coherence of the equivalent vector problem tends to decrease. This is because the individual terms of the product in (3.15) are defined between 0 and 1. The multiplication of many terms within such an interval decreases the overall mutual coherence $\mu\left(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_D\right)$. Consequently, less measurements are necessary to recover the sparse tensor.

The tensor formalism provides a powerful tool to solve Kronecker CS problems. Using the Tucker decomposition, it is possible to develop algorithms that exploit the tensor structure of the underlying sparse problem. Since all dimensions of the data are jointly exploited, a better compression rate can be achieved in comparison with traditional sparse reconstruction algorithms.

In the next section, a frequency-selective MIMO channel model is developed by using the tensor formalism. Then, a new sparse channel estimator exploiting the Tucker decomposition is proposed.

3.3 Tensor-Based Channel Model

Consider a MIMO system with I and J antennas at the transmitter and receiver, respectively. The channel between them is frequency selective and modeled as a low-rank channel. The rank is defined by the number of clusters as shown in [48]. The channel matrix associated with the f-th frequency is given by:

$$\mathbf{G}(f) = \sum_{n=1}^{N_p} \beta_n \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^H(\theta_{T,n}) e^{-\jmath 2\pi \tau_n f}, \tag{3.16}$$

where, $\mathbf{G}(f) \in \mathbb{C}^{I \times J}$, N_p represents the number of paths, β_n denotes a random variable with complex Gaussian distribution, $\theta_{R,n}$ and $\theta_{T,n}$ denote the angle of arrival (AoA) and angle of departure (AoD), respectively, $\tau_n \in [0, \tau_{max}]$ is the delay related to the n-th path. The $J \times 1$ vector $\mathbf{a}_T(\theta_{T,n})$ and the $I \times 1$ vector $\mathbf{a}_R(\theta_{R,n})$ are the array steering vectors for the transmitter and receiver, respectively.

The model in (3.16) imposes a nonlinear dependence on the number of parameters. However, it can be approximated using the virtual channel representation proposed in [48]. Adopting a uniform sampling in each dimension, the angle-delay grid has a resolution $(\Delta\theta_R, \Delta\theta_T, \Delta\tau) = (1/I, 1/J, 1/W)$, yields [43]

$$\begin{aligned} \mathbf{G}(f) \;\approx\; & \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{m=1}^{M} \mathcal{H}^{\mathrm{v}}(i,j,m) \\ & \mathbf{a}_R\left(\frac{i-1}{I}\right) \mathbf{a}_T^H\left(\frac{j-1}{J}\right) \times \\ & e^{-\jmath 2\pi \frac{m-1}{W} f}, \end{aligned} \tag{3.17}$$

with

$$\begin{aligned} \mathcal{H}^{\mathrm{v}}(i,j,m) \;\approx\; & \sum_{n \,\in\, S_{R,i} \cap S_{T,j} \cap S_{\tau,m} \cap S_{v,d}} \beta_n \\ & \times f_I(i/I - \theta_{R,n}) f_J^*(j/J - \theta_{T,n}) \\ & \times \mathrm{sinc}(m - W\tau_n), \end{aligned} \tag{3.18}$$

where $\mathcal{H}^{\mathrm{v}}(i,j,m)$ denotes the (i,j,m)-th element of the tensor $\mathcal{H}^{\mathrm{v}} \in \mathbb{C}^{I \times J \times M}$. Note that the region $\mathbb{H} = \{(\theta_{T,n}, \theta_{R,n}, \tau_n) \in \mathbb{R}^3 \; \forall \; n \in \{1, \ldots, N_p\}\}$ is sampled into discrete values (i,j,m). Therefore, each tensor element describes an angle-delay bin $\Delta\theta_R \times \Delta\theta_T \times \Delta\tau$. The variable W defines the one-sided bandwidth of the transmitted signal, T is the symbol duration, $f_\alpha(\gamma) = \frac{1}{\alpha} \sum_{i=0}^{\alpha-1} e^{-j2\pi i\gamma}$ is the Dirichlet kernel and $\mathrm{sinc}(x) = e^{-j\pi x} \sin(\pi x)/(\pi^2 x)$ is the sinc kernel [43]. The sets over which the summation in (3.18) is defined are given by

$$\begin{aligned} S_{R,i} &= \{n \;:\; \theta_{R,n} \;\in\; (i/I - \Delta\theta_R/2, i/I + \Delta\theta_R/2]\} \\ S_{T,j} &= \{n \;:\; \theta_{T,n} \;\in\; (j/J - \Delta\theta_T/2, j/J + \Delta\theta_T/2]\} \\ S_{\tau,m} &= \{n \;:\; \tau_n \;\in\; (m/W - \Delta\tau/2, m/W + \tau/2]\}. \end{aligned}$$

Comparing (3.17) and (3.6) and using the n-mode formulation presented in Section 3.2 (see Eq. (3.4)), the transfer function $\mathbf{G}(f)$ can be recast as a tensor, which is decomposable as

$$\mathcal{G} \approx \mathcal{H}^{\text{v}} \ \times_1 \ \mathbf{A}_R \ \times_2 \ \mathbf{A}_{\text{T}} \ \times_3 \ \mathbf{W}_{\text{delay}}, \tag{3.19}$$

where $g_{ijl} = [\mathbf{G}(lW)]_{ij}$ is its element, and

$$\begin{aligned} \mathbf{A}_R &= \left[\mathbf{a}_R(0) \ \ \mathbf{a}_R\left(\frac{1}{I}\right) \ \ \ldots \ \ \mathbf{a}_R\left(\frac{I-1}{I}\right)\right], \\ \mathbf{A}_{\text{T}} &= \left[\mathbf{a}_T(0) \ \ \mathbf{a}_T\left(\frac{1}{J}\right) \ \ \ldots \mathbf{a}_T\left(\frac{J-1}{J}\right)\right]. \end{aligned}$$

From (3.18) and (3.19), we can note that the first expression shows the sampling of the multipath environment in the angle-delay domain, while the latter one provides a low-rank approximation of the channel tensor $\mathcal{G}$. Note that due to the fixed sampling of the angle-delay-Doppler grid, the virtual channel representation is completely characterized by the virtual channel index unlike the Weichselberger's model. Therefore, the number of parameters to be estimated in this model is much less than that of the virtual channel model [43, 48].

3.3.1 Channel Compression Discussion

The upper-bound on the degree of freedom (DoF) of the fourth-order virtual channel tensor $\mathcal{H}^{\text{v}}$ is given by

$$P = IJM, \tag{3.20}$$

which corresponds the maximum number of resolvable bins of the virtual channel tensor. Most of the actual propagation channels are not characterized by a "full" tensor $\mathcal{H}^{\text{v}}$ and the virtual channel is likely to be sparse [44–46]. In this case, the channel capacity is not dictated by P, but instead by the number of coefficients $p = |(i,j,l) : |h^{\text{v}}_{i,j,m}| > \epsilon|$, where $|.|$ means the cardinality.

Some works have already analyzed the structure of $\mathcal{H}^{\text{v}}$, such as [43]. Therein, the authors propose the use of a cluster-based sparse representation of physical wireless channels. Along the same lines, we can say that the smaller $\Delta\theta_R$, $\Delta\theta_I$, and $\Delta\tau$ are, the more resolution the virtual channel tensor $\mathcal{H}^{v}$ has. Consequently, multipath channels tend to have far fewer than P dominant virtual channel coefficients, which means that $\mathcal{H}^{\text{v}}$ is sparse. For convenience, in the following we provide the definition of a sparse channel given in [43].

Definition 6 *Let $\mathbb{H}_p = \{(i,j,l) : |h^{v}_{i,j,m}| > \epsilon\}$, $\forall\ \epsilon > 0$, be the set of dominant virtual channel coefficients of channel $\mathcal{H}$. A channel is said to be effectively sparse if its DoF p is smaller than P (defined in Eq.* (3.20)*), where $p = |\mathbb{H}_p|$ is the cardinality of $\mathbb{H}_p$. Furthermore, we say that the channel is exactly sparse if the same holds for $\epsilon = 0$.*

Let us now recast the P coefficients of $\mathcal{H}^{\text{v}}$ in a decreasing magnitude order $|h_{o(1)}|$, $|h_{o(2)}|$, ..., $|h_{o(P)}|$, where the p-th term is given by

$$|h_{o(p)}| < Cp^{-1/s}, \tag{3.21}$$

with $C > 0$ and $s \leq 1$. Channels that follow (3.21) are called s-compressible.

3.4 Proposed Channel Estimator

Let us consider a BS employing J transmit antennas and a UE using I receive antennas, with $I << J$. The BS antennas transmit independent pilot sequences, which are repeated across L subcarriers using multicarrier modulation, where

$$\mathbf{x}_n = [x_{0,n} \;\; x_{2,n} \;\; \ldots \;\; x_{J-1,n}]^T \;\; \in \;\; \mathbb{C}^{J\times 1}, \tag{3.22}$$

where $x_{j,n}$ is the pilot symbol transmitted by the jth antenna during the nth discrete-time instant. After multicarrier demodulation at the UE, each data block is converted back to the frequency domain yielding

$$\mathbf{y}_{l,n} = [y_{0,l,n} \;\; y_{2,l,n} \;\; \ldots \;\; y_{I-1,l,n}]^T \;\; \in \;\; \mathbb{C}^{I\times 1}, \tag{3.23}$$

where $y_{i,l,n}$ is the signal associated with the ith receive antenna, lth subcarrier, and nth discrete-time instant. The input-output relation in the frequency domain is given as

$$\mathbf{y}_{l,n} = \mathbf{G}_{l,n}\mathbf{x}_n + \mathbf{z}_{l,n} \tag{3.24}$$

where $\mathbf{G}_{l,n}$ is the $I \times J$ time-frequency "slice" of the channel tensor whose definition is given by (3.19), and $\mathbf{z}_{l,n}$ is the zero mean Gaussian noise vector with variance $N_0/2$, associated with the (l, n) time-frequency resource. In the following developments, consider only the useful signal part of the model (3.24)(i.e. without the noise term), for the sake of simplicity. Stacking $\{\mathbf{y}_{l,n}\}$, $l \in \{0, \;\; 1 \;\; \ldots \;\; L-1\}$, yields

$$\mathbf{y}_n = \begin{bmatrix} \tilde{\mathbf{y}}_{0,n} \\ \tilde{\mathbf{y}}_{1,n} \\ \ldots \\ \tilde{\mathbf{y}}_{L-1,n} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{0,n} \\ \mathbf{G}_{1,n} \\ \ldots \\ \mathbf{G}_{L-1,n} \end{bmatrix} \mathbf{x}_n \;\; \in \;\; \mathbb{C}^{IL\times 1}, \tag{3.25}$$

where $\tilde{\mathbf{y}}_n$ is the equivalent received signal vector without noise. After collecting the samples of the received signal vector along the frequency domain, the data is concatenated along the time domain, yielding the following data matrix:

$$\begin{aligned} \tilde{\mathbf{Y}} \;\; &= \;\; [\tilde{\mathbf{y}}_0, \;\; \ldots \;\; \tilde{\mathbf{y}}_{N-1}] \\ &= \;\; [(\mathbf{G}_0\mathbf{x}_0), \;\; \ldots \;\; (\mathbf{G}_{N-1}\mathbf{x}_{N-1})] \;\; \in \;\; \mathbb{C}^{IL\times N}, \end{aligned} \tag{3.26}$$

where

$$\mathbf{G}_n = \left[(\mathbf{G}_{0,n})^T \;\; (\mathbf{G}_{1,n})^T \;\; \ldots \;\; (\mathbf{G}_{L-1,n})^T\right]^T \;\; \in \;\; \mathbb{C}^{IL\times J},$$

and $n = 0, \;\; 1, \;\; \ldots, \;\; N-1$. The matrix $\tilde{\mathbf{Y}}$ represents a set of training symbols collected after N discrete time instants in L subcarriers. The total number of communication resources used is NL.

Assume that during the whole observation window of N symbol periods, the channel remains constant. This implies that $\mathbf{G}_0 = \mathbf{G}_1 = \mathbf{G}_2 = \ldots = \mathbf{G}_N = \mathbf{G}$ and means that the matrix $\tilde{\mathbf{Y}}$ can be expressed as

$$\tilde{\mathbf{Y}} = \mathbf{G}\mathbf{X} = \begin{bmatrix} \mathbf{G}_{0,0} \\ \mathbf{G}_{1,0} \\ \cdots \\ \mathbf{G}_{L-1,0} \end{bmatrix} \mathbf{X} \in \mathbb{C}^{IL \times N}, \tag{3.27}$$

which can be interpreted as the 2-mode product between the received signal tensor and the pilot symbol matrix, as follows

$$\tilde{\mathcal{Y}} \approx \mathcal{G} \times_2 \mathbf{X}. \tag{3.28}$$

Note that the tensor $\mathcal{G}$ is a concatenation of L spatial-frequency channel responses. They are allocated for pilot transmission whereas another set of subcarriers carry data. Because the receiver has no knowledge of the channel yet, it needs to estimate the CSI collecting the training signal from the subcarriers allocated for pilot transmission. After estimating the channel, the receiver performs an interpolation to obtain the complete channel response. Thus, the optimal strategy in the frequency selective channel consists of exploiting the pilot placement to obtain the whole channel frequency response.

Using the tensor formalism, the pilot placement is implemented as a 3-mode product. The matrix in the third dimension is a $L \times D$ matrix $\mathbf{F}$ over the complete tensor channel $\tilde{\mathcal{G}} \in \mathbb{C}^{I \times J \times D}$. This tensor represents the spatial frequency channel response with respect to all the D subcarriers . Formally, this is represented as

$$\mathcal{G} = \tilde{\mathcal{G}} \times_3 \mathbf{F}.$$

Thus, the training signal is

$$\tilde{\mathcal{Y}} = \tilde{\mathcal{G}} \times_2 \mathbf{X} \times_3 \mathbf{F}. \tag{3.29}$$

Note that the tensor $\tilde{\mathcal{G}}$ is compressed across two dimensions: the spatial and the frequency. This is because of the lack of measurements imposed in both dimensions. In the spatial domain, the length of the training sequence is $N << J$; and in the frequency domain, the number of subcarriers is $L << D$.

It is possible to perform a compression in the first dimension as well. Recently, we have seen a type of beamforming called HB [12, 46]. Different from the full digital beamforming, the number of RF chains is less than the number of antennas. In this context, the number of measurements in the spatial domain reduces. In this context, the $M \times I$ matrix $\mathbf{Q}$ represents the HB matrix that is applied over the first (spatial) dimension of the tensor $\tilde{\mathcal{G}}$. We express the received pilot signal after receive beamforming as

$$\tilde{\mathcal{Y}}_{\mathrm{c}} \approx \mathcal{G} \times_1 \mathbf{Q} \times_2 \mathbf{X} \times_3 \mathbf{F}.$$

By adding the noise term, we get:

$$\mathcal{Y}_{\mathrm{c}} \approx \mathcal{G} \times_1 \mathbf{Q} \times_2 \mathbf{X}_0^T \times_3 \mathbf{F} + \tilde{\mathcal{Z}}, \tag{3.30}$$

where $\mathcal{Z} = \tilde{\mathcal{Z}} \times_1 \mathbf{Q}$ and $\tilde{\mathcal{Z}} \in \mathbb{C}^{I \times N \times L}$ is the noise tensor. Substituting (3.19) into (3.30), we get:

$$\mathcal{Y}_{\mathrm{c}} \approx \mathcal{H}^{\mathrm{v}} \times_1 (\mathbf{Q}\mathbf{A}_{\mathrm{R}}) \times_2 (\mathbf{X}_0\mathbf{A}_{\mathrm{T}}) \times_3 (\mathbf{F}\mathbf{W}_{\text{delay}}) + \mathcal{Z}. \tag{3.31}$$

This decomposition expresses the received signal tensor as a third-order tensor decomposed in terms of the virtual channel tensor $\mathcal{H}^{\mathrm{v}}$ jointly capturing the sparse angle-delay (or the space-frequency) profile of the channel. By analogy with (3.7), the received signal is a third-order (Tucker3) tensor, and the correspondence is:

$$\begin{aligned} &\Big[\mathcal{X}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \Big] \longleftrightarrow \\ &\Big[\mathcal{H}^{\mathrm{v}}, (\mathbf{Q}\mathbf{A}_{\mathrm{R}}), (\mathbf{X}_0^T\mathbf{A}_{\mathrm{T}}), (\mathbf{F}\mathbf{W}_{\text{delay}})\Big]. \end{aligned} \tag{3.32}$$

3.4.1 Channel estimation via sparse vector representation

Based on the formulation presented in this section, one solution for estimating $\mathcal{H}^{\mathrm{v}}$ consist of vectorizing the tensor $\mathcal{Y}$ and then applying a CS estimator. For instance, it is possible to apply the well-known OMP alogrithm to extract the vectorized version of the $\mathcal{H}^{\mathrm{v}}$.

From tensor $\mathcal{Y}$, it is possible to achieve three different vector representations. This is because there are three different dimensions that can be used as a reference to stack the entries of $\mathcal{Y}$. We unfold the tensor (3.31) along the 2-mode, which can be done according to Eq. (3.9), with the analogies (3.32) in mind. It follows that

$$\mathbf{Y}_{(2)} = [(\mathbf{F}\mathbf{W}_{\text{delay}}) \otimes (\mathbf{Q}\mathbf{A}_{\mathrm{R}})]\mathbf{H}^{\mathrm{v}}_{(2)}(\mathbf{A}_{\mathrm{T}}^T\mathbf{X}_0) + \mathbf{Z}_{(2)}.$$

Now, by applying the vec$(\cdot)$ operator on both sides of (3.33) yields:

$$\mathbf{y}_{(2)} = \Big[(\mathbf{X}_0^T\mathbf{A}_{\mathrm{T}}) \otimes (\mathbf{F}\mathbf{W}_{\text{delay}}) \otimes (\mathbf{Q}\mathbf{A}_{\mathrm{R}})]\mathbf{h}^{\mathrm{v}}_{(2)} + \mathbf{z}_{(2)}, \tag{3.33}$$

where $\mathbf{h}^{\mathrm{v}}_{(2)} = \mathrm{vec}\{\mathbf{H}^{\mathrm{v}}_{(2)}\}$, $\mathbf{y}_{\mathrm{c};(2)} = \mathrm{vec}\{\mathbf{Y}_{\mathrm{c};(2)}\}$, $\mathbf{z}_{(2)} = \mathrm{vec}\{\mathcal{Z}\}$. Due to the Kronecker product structure of the vectorized channel model (3.33), the sparse channel vector $\mathbf{h}^{\mathrm{v}}_{(2)}$ can be estimated by applying the Kronecker compressed sensing framework of [54]. In this case, (3.33) can be rewritten as

$$\mathbf{y}_{(2)} = \mathbf{\Psi}\mathbf{\Phi}\mathbf{h}^{\mathrm{v}}_{(2)} + \mathbf{z}_{(2)}, \tag{3.34}$$

where

$$\mathbf{\Psi} \doteq \mathbf{X}_0^T \otimes \mathbf{F} \otimes \mathbf{Q}, \in \mathbb{C}^{NLM \times JDI}, \tag{3.35}$$

$$\mathbf{\Phi} \doteq \mathbf{A}_T \otimes \mathbf{W}_{\text{delay}} \otimes \mathbf{A}_{\mathrm{R}} \in \mathbb{C}^{JDI \times JLI}. \tag{3.36}$$

The above formulation corresponds to a CS reconstruction model introduced in Section 3.2.2. Note that, in this model, the measurement matrix $\mathbf{\Psi}$ is constructed by the known pilot sequence

matrix $\mathbf{X}_0$ using the knowledge of the frequency bins allocated to these pilots. The required number of measurements depends on two parameters: the first is the pilot sequence length, which is assumed to be much shorter than the number of transmit antennas, i.e. $N < J$. The second is the number of frequency bins allocated to the pilots. The dictionary matrix is given by the compound (Kronecker) basis, herein assumed to be generated by DFT matrices.

To evaluate uniqueness of the sparse representation (3.34), we apply (3.15) to the corresponding basis and measurement matrices defined in (3.35) and (3.36), respectively. Uniqueness is achieved if

$$\begin{aligned} ||\mathbf{g}_{(2)}||_0 &\leq 1 + \frac{1}{\mu(\overline{\mathbf{\Phi}\mathbf{\Psi}})} \\ &\leq 1 + \frac{1}{\mu(\mathbf{X}_0^T\mathbf{A}_T)\mu(\mathbf{F}\mathbf{W}_{\text{delay}})\mu(\mathbf{Q}\mathbf{A}_{\text{R}})}. \end{aligned}$$

Since (3.37) is a product of three mutual coherence measures, the global coherence can achieve lower values whenever the channel admits a sparse multidimensional structure, as it is the case in this work. For instance, when $\mathbf{X}_0^T$, $\mathbf{F}$, and $\mathbf{Q}$ are designed in such way that their columns are taken randomly from orthonormal matrices of compatible dimensions, the fundamental result of [35] implies $1/\sqrt{N} \leq \mu(\mathbf{X}_0^T\mathbf{A}_T) \leq 1$, $1/\sqrt{L} \leq \mu(\mathbf{F}\mathbf{W}_{\text{delay}}) \leq 1$, and $1/\sqrt{A} \leq \mu(\mathbf{Q}\mathbf{A}_{\text{R}}) \leq 1$.

The problem (3.34) can be solved by means of linear programming [55] or using second order cone programs [56]. Note, however, that working on the vectorized model (3.34) may result in a higher computational complexity due to the higher dimensionality of the equivalent channel vector $\mathbf{h}_{(2)}^{\text{v}}$. If the channel tensor is frequency-selective, its sparse vector representation is of length IJD. In the massive MIMO case, where the dimension J is large the complexity of the equivalent (vectorized) channel estimation algorithm may be an issue, since the size of global dictionary scales exponentially with the number of channel dimensions. In the next section, we present an alternative lower-complexity solution that capitalizes on the tensor-based model developed in the previous sections by solving smaller (per dimension) sparse recovery problems.

3.4.2 Kronecker-Basis Channel Estimation Using Tensor Unfolding

Although the conventional OMP algorithm (or a variant of it) has been widely used in the literature, it was designed for recovering sparse vectors (i.e. 1D signals). For recovering a higher-order tensor, which is the case of interest, the use of OMP algorithm directly to a vectorized form of the tensor can potentially yield serious problems in terms of memory usage and processing. The high number and complexity of the operations involved, including inversion of very big matrices calls for a more efficient way to circumvent these shortcomings. One way is to avoid the channel vectorization step and explicitly operate on each dimension of the channel tensor by exploiting its multidimesional structure.

Kronecker-OMP algorithm uses the tensor modes $\mathbf{\Upsilon}_1 = \mathbf{Q}\mathbf{A}_{\text{R}}$, $\mathbf{\Upsilon}_2 = \mathbf{X}\mathbf{A}_{\text{T}}$, and $\mathbf{\Upsilon}_3 = \mathbf{F}\mathbf{W}_{\text{delay}}$ by exploiting the three unfolding forms of tensor $\mathcal{Y}$ defined in (3.31). The relationship

between the channel $\mathcal{H}^{\text{v}}$ and $\mathcal{Y}$ is established by the algorithm using the unfolding definitions in (3.8), (3.9) and (3.9). From that, we obtain

$$\mathbf{Y}_{(1)} = \boldsymbol{\Upsilon}_1 \mathbf{H}^{\text{v}}_{(1)} \left[\boldsymbol{\Upsilon}_2 \otimes \boldsymbol{\Upsilon}_3\right]^T, \tag{3.37}$$

$$\mathbf{Y}_{(2)} = \boldsymbol{\Upsilon}_2 \mathbf{H}^{\text{v}}_{(2)} \left[\boldsymbol{\Upsilon}_3 \otimes \boldsymbol{\Upsilon}_1\right]^T, \tag{3.38}$$

$$\mathbf{Y}_{(3)} = \boldsymbol{\Upsilon}_3 \mathbf{H}^{\text{v}}_{(3)} \left[\boldsymbol{\Upsilon}_1 \otimes \boldsymbol{\Upsilon}_2\right]^T. \tag{3.39}$$

Without losing the generality, let us take the mode-3 representation and vectorize the matrix $\mathbf{Y}_{(3)}$ as shown in

$$\mathbf{y}_{(3)} = \left(\boldsymbol{\Upsilon}_1 \otimes \boldsymbol{\Upsilon}_2 \otimes \boldsymbol{\Upsilon}_3\right) \mathbf{h}^{\text{v}}_{(3)}, \tag{3.40}$$

where $\mathbf{h}^{\text{v}}_{(3)}$ is the vectorized version obtained from the matrix $\mathbf{H}^{\text{v}}_{(3)}$. The problem (3.40) is an undetermined linear system which means that the number of equations is less than the number of variables ($NML < IJD$). In such a case, we cannot use the LS solution to obtain $\mathbf{h}^{\text{v}}_{(3)}$, but we can formulate it as a new problem taking into account the fact that $\mathbf{H}^{\text{v}}_{(3)}$ is sparse. This helps us because some of the variables are zero and, consequently, some columns of $\boldsymbol{\Upsilon} = \left(\boldsymbol{\Upsilon}_1 \otimes \boldsymbol{\Upsilon}_2 \otimes \boldsymbol{\Upsilon}_3\right)$ can be removed from the formulation.

To find the strongest entries of $\mathcal{H}^{\text{v}}$ [22],the greedy algorithm computes the optimization problem defined at the kth iteration

$$\begin{aligned}\left[i_1^{(k)}, i_2^{(k)}, i_3^{(k)}\right] = \arg \max_{i_1^{(k)}, i_2^{(k)}, i_3^{(k)}} & \\ ||\mathcal{R}^{(k-1)} \times_1 \boldsymbol{\Upsilon}_1^H(i_1) \times_2 \boldsymbol{\Upsilon}_2^H(i_2) & \\ \times_3 \boldsymbol{\Upsilon}_3^H(i_3)||_F^2, & \end{aligned} \tag{3.41}$$

where $\mathcal{R}^{(k-1)}$ is a residual tensor at the $(k-1)$th iteration, $\boldsymbol{\Upsilon}_1(i_1)$, $\boldsymbol{\Upsilon}_2(i_2)$, and $\boldsymbol{\Upsilon}_3(i_3)$ are the i_1, i_2 and i_3 columns of $\boldsymbol{\Upsilon}_1$, $\boldsymbol{\Upsilon}_2$ and $\boldsymbol{\Upsilon}_3$, respectively. The residual tensor is updated by removing the components obtained in (3.41) from the original tensor $\mathcal{Y}$. The algorithm performs this task by means of one of its unfolding representations. To keep the consistence in our analysis let us choose the mode-3 as we did before. The $(i_1^{(k)}, i_2^{(k)}, i_3^{(k)})$th sparse entry is calculated as

$$\hat{\mathbf{h}}^{\text{v}}_{i_1^{(k)}, i_2^{(k)}, i_3^{(k)}} = \left(\boldsymbol{\Upsilon}_1^H(i_1) \diamond \boldsymbol{\Upsilon}_2^H(i_2) \diamond \boldsymbol{\Upsilon}_3^H(i_3)\right) \mathbf{r}^{(k)}_{(3)}, \tag{3.42}$$

where $\diamond$ is the Katri-Rao product [22] and $\hat{\mathbf{h}}^{\text{v}}_{i_1^{(k)}, i_2^{(k)}, i_3^{(k)}}$ is a scalar. Using (3.42) the algorithm tests all the possible combinations of $(i_1^{(k)}, i_2^{(k)}, i_3^{(k)})$ and chooses the strongest one at kth iteration. An estimate of the sparse tensor using the components obtained until iteration k is expressed as

$$\hat{\mathbf{h}}^{(k)}_{(3)} = \left(\boldsymbol{\Upsilon}_1^{(k)} \diamond \boldsymbol{\Upsilon}_2^{(k)} \diamond \boldsymbol{\Upsilon}_3^{(k)}\right)^{\dagger} \mathbf{y}_{(3)}, \tag{3.43}$$

where $\boldsymbol{\Upsilon}_1^{(k)}$, $\boldsymbol{\Upsilon}_2^{(k)}$, and $\boldsymbol{\Upsilon}_3^{(k)}$ are the subset of k columns from their original $\boldsymbol{\Upsilon}_1$, $\boldsymbol{\Upsilon}_2$, and $\boldsymbol{\Upsilon}_3$; and $\hat{\mathbf{h}}^{(k)}$ is the estimated sparse channel after k iterations. Note that the pseudoinverse is valid if $k \leq MNL$. Otherwise, the system becomes undetermined as in (3.40).

Tabela 1 – Summary of the Kronecker-OMP algorithm.

Require:	Received signal tensor $\mathcal{Y}$ and $\{\boldsymbol{\Upsilon}_i\}$, $i = 1, 2, 3, 4$ $\mathbb{I} = \{\}$
Step 1	Set the residual tensor $\mathcal{R} = \mathcal{Z}$
Step 2	At the k-th iteration, solve Eq. (3.41) by looking for the possible combinations of $[i_1^{(k)}, i_2^{(k)}, i_3^{(k)}]$ using (3.42)
Step 3	Calculate $\hat{\mathcal{Y}}$ by solving Eq. (3.43)
Step 4	Update the residual tensor using Eq. (3.44)
Step 5	If $\|\|\mathcal{R}\|\|_F^2 < \epsilon$, then stop. Otherwise, repeat step 2 for $k = k + 1$.

Using $\hat{\mathbf{h}}_{(3)}^{(k)}$, the k components are removed from $\mathbf{y}_{(3)}$ as shown in

$$\mathbf{r}_{(3)}^{(k)} = \mathbf{y} - \left(\boldsymbol{\Upsilon}_1^{(k)} \diamond \boldsymbol{\Upsilon}_2^{(k)} \diamond \boldsymbol{\Upsilon}_3^{(k)}\right) \hat{\mathbf{h}}_{(3)}^{(k)}, \tag{3.44}$$

This process is repeated until $||\mathcal{R}||_F^2 < \epsilon$ is satisfied, where ϵ defines a given threshold. Table **??** shows a summary of the algorithm.

3.5 Proposed Tensor-OMP

The algorithm described in **??** improves the memory usage of the classical OMP, but the method still faces problems when the sparse tensor grows large. Even using the Kronecker-OMP, the solution of (3.41) still involves an exhaustive search. For instance, in massive MIMO, the dimension connected with the number of antennas tends to be very large which creates many possible entries in $\mathcal{H}^{\text{v}}$ and, consequently, more processing capacity is needed in the receiver. However, the mobile unit has limited processing capacity which poses difficulties to exploit the tensor modelling in the downlink channel scenario. In what follows, we propose an algorithm to enjoy the Tucker3 decomposition and meet limited-processing receiver requirements.

To understand the proposed method, let us consider again only the portion of the signal that represents the pilots and recall the 1-mode representation of the tensor $\mathcal{Y}$ which is

$$\begin{aligned} \mathbf{Y}_{(1)} &= \boldsymbol{\Upsilon}_1 \mathbf{H}_1^{\text{v}} \left[\boldsymbol{\Upsilon}_2 \otimes \boldsymbol{\Upsilon}_3\right]^T \\ &\doteq \boldsymbol{\Upsilon}_1 \tilde{\mathbf{H}}_1^{\text{v}}, \end{aligned} \tag{3.45}$$

In (3.45), the matrix $\tilde{\mathbf{H}}_1^{\text{v}}$ can be viewed as a collection of sparse vectors that are represented by the same $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Psi}_1$, and they define $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Psi}_1 \boldsymbol{\Phi}_1$. Thus, $\mathbf{Y}_{(1)}$ is a collection of measurements obtained in the 1-mode and the algorithm is interested to know which component is the strongest one of that dimension. To obtain it, we need to solve the problem

$$i_1' = \arg\max_{i_1} \sum_{j=1}^{JL} |\left[\boldsymbol{\Upsilon}_1(i_1)\right]^H \mathbf{Y}_{(1)}(j)|^2, \tag{3.46}$$

where $\boldsymbol{\Upsilon}_1(i_1)$ and $\mathbf{Y}_{(1)}(j)$ are the i_1th and jth columns of $\boldsymbol{\Upsilon}_1$ and $\mathbf{Y}_{(1)}$, respectively.

The solution of (3.46) is used to discard other components that are not described by i_1' in the first mode. To do such an operation, we apply the vector $\boldsymbol{\Upsilon}_1^*(i_1')$ over the 1-mode dimension. This gives a 2-mode unfolding that is expressed as

$$\begin{aligned}\tilde{\mathbf{Y}}_{(3)} &= \boldsymbol{\Upsilon}_2\mathbf{H}_2^{\mathrm{v}}\left(\boldsymbol{\Upsilon}_3 \otimes \boldsymbol{\Upsilon}_1\right)^T\left(\mathbf{I} \otimes \boldsymbol{\Upsilon}_1^*(i_1')\right)\\ &\doteq \boldsymbol{\Upsilon}_2\mathbf{H}_2^{\mathrm{v}}\left(\boldsymbol{\Upsilon}_3 \otimes \left[\boldsymbol{\Upsilon}_1(i_1')\right]^H\boldsymbol{\Upsilon}_1\right)^T\\ &\doteq \boldsymbol{\Upsilon}_2\tilde{\mathbf{H}}_2^{\mathrm{v}}. \end{aligned} \tag{3.47}$$

The resulting vector $\left[\boldsymbol{\Upsilon}_1^*(i_1')\right]^H\boldsymbol{\Upsilon}_1$ fixes the 1-mode in i_1 and cuts off any other 3D component for which $i_1 \neq i_1'$ resulting in $(I-1)L$ less components. After this, the algorithm uses the same metric as in (3.46), and the problem can be expressed in 2-mode as

$$i_2' = \arg\max_{i_2}\sum_{j=1}^{L}|\left[\boldsymbol{\Upsilon}_2(i_2)\right]^H\tilde{\mathbf{Y}}_{(2)}(j)|^2. \tag{3.48}$$

Using i_1' and i_2', we can remove any 3D components for which $i_1 \neq i_1'$ and $i_2 \neq i_2'$. Thus, the third unfolding is expressed as

$$\begin{aligned}\tilde{\mathbf{y}}_{(3)} &= \boldsymbol{\Upsilon}_3\mathbf{H}_3^{\mathrm{v}}\left(\boldsymbol{\Upsilon}_1 \otimes \boldsymbol{\Upsilon}_2\right)^T\left(\boldsymbol{\Upsilon}_1^*(i_1') \otimes \boldsymbol{\Upsilon}_2^*(i_2')\right)\\ &\doteq \boldsymbol{\Upsilon}_3\mathbf{H}_3^{\mathrm{v}}\left(\left[\boldsymbol{\Upsilon}_1(i_1')\right]^H\boldsymbol{\Upsilon}_1 \otimes \left[\boldsymbol{\Upsilon}_2(i_2')\right]^H\boldsymbol{\Upsilon}_2\right)^T\\ &\doteq \boldsymbol{\Upsilon}_3\tilde{\mathbf{h}}_3^{\mathrm{v}}. \end{aligned} \tag{3.49}$$

Finally, the 3-mode reduces to finding the strongest component for the vector case. Note that, doing local decisions after estimating separately i_1', i_2', and i_3', we avoid the exhaustive search performed by OMP and Kronecker-OMP. This dramatically reduces the computational complexity of the algorithm. The summary of the algorithm is stated in Table (**??**).

3.6 Numerical Results

This section shows the performance of the proposed tensor-OMP algorithm. The number of antennas at the BS is $N_{\mathrm{T}} = 128$, the number of antennas at the UE is $N_{\mathrm{R}} = 2$, the system has a total number of subcarriers is $D = 68$ and the spacing among them is 15kHz. The channel is implemented by assuming $N_p = 9$ multipaths and the power delay profile is the one shown in Fig. 3.3. The channel is implemented using Eq. (3.16).

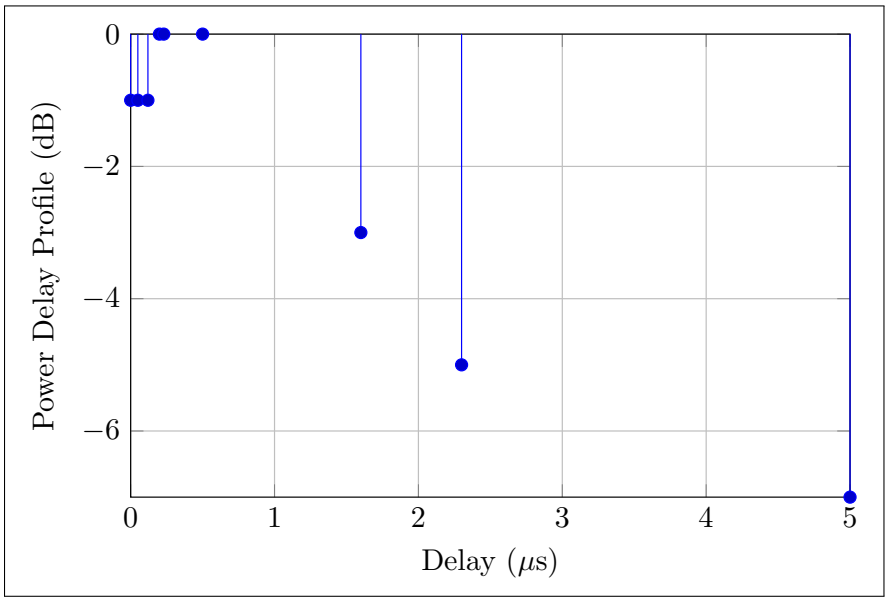
The trilinear basis $\boldsymbol{\Phi}$ defined in (3.36) follows a 3D-Fourier one. The measurement matrix $\mathbf{X}_0$ is randomly drawn from a Bernoulli distribution; the matrix $\mathbf{Q}$ is assumed to be an identity; and the matrix $\mathbf{F}$ defines a pilot placement with subcarriers regularly spaced, this means that the measurement matrix is obtained from subset of an identity one $\mathbf{I}_{D\times D}$.

In Fig. 3.4, the pilots are placed at every two subcarriers, thus the total number of pilot subcarriers is $L = 34$. It can observed, the method accurately estimates the frequency response

Tabela 2 – Summary of the modified Kronecker-OMP algorithm.

Require:	Received signal tensor $\mathcal{Y}$ and $\{\boldsymbol{\Upsilon}_i\}$, $i = 1, 2, 3, 4$ $\mathbb{I} = \{\}$
Step 1	Set the residual tensor $\mathcal{R} = \mathcal{Z}$
Step 2	Solve $i'_1 = \arg\max_{i_1} \sum_{j=1}^{JL} \lvert [\boldsymbol{\Upsilon}_1(i_1)]^H \mathbf{Y}_{(1)}(j)\rvert^2$, $i'_2 = \arg\max_{i_2} \sum_{j=1}^{L} \lvert [\boldsymbol{\Upsilon}_2(i_2)]^H \tilde{\mathbf{Y}}_{(2)}(j)\rvert^2$, $i'_3 = \arg\max_{i_3} \lvert [\boldsymbol{\Upsilon}_2(i_3)]^H \tilde{\mathbf{y}}_{(3)}\rvert^2$, where $\mathbf{Y}_{(1)} = \boldsymbol{\Upsilon}_1 \mathbf{H}_1^{\mathrm{v}} (\boldsymbol{\Upsilon}_2 \otimes \boldsymbol{\Upsilon}_3)^T$ $\tilde{\mathbf{Y}}_{(2)} = \boldsymbol{\Upsilon}_2 \mathbf{H}_2^{\mathrm{v}} \left(\boldsymbol{\Upsilon}_3 \otimes [\boldsymbol{\Upsilon}_1(i'_1)]^H \boldsymbol{\Upsilon}_1\right)^T$ $\tilde{\mathbf{y}}_{(3)} = \boldsymbol{\Upsilon}_3 \mathbf{H}_3^{\mathrm{v}} \left([\boldsymbol{\Upsilon}_1(i'_1)]^H \boldsymbol{\Upsilon}_1 \otimes [\boldsymbol{\Upsilon}_2(i'_2)]^H \boldsymbol{\Upsilon}_2\right)^T$
Step 3	Calculate $\hat{\mathcal{Y}}$ by solving Eq. (3.43)
Step 4	Update the residual tensor using Eq. (3.44)
Step 5	If $\lvert\lvert\mathcal{R}\rvert\rvert_F^2 < \epsilon$, then stop. Otherwise, repeat step 2 for $k = k + 1$.

Figura 3.3 – Power delay profile considered to create a wideband channel.

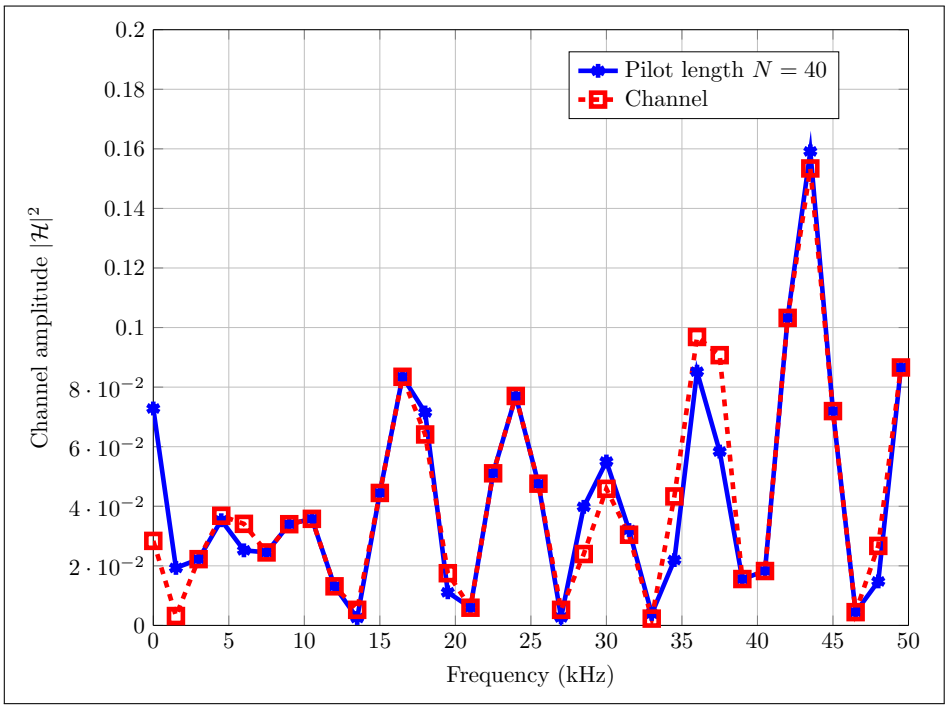


Source: Provided by the author.

of the channel. However, as the number of pilot subcarriers is reduced, the estimator tends to lose performance as shown in Fig. 3.5. In such a simulation, the pilots are placed at every 6 subcarriers, which correspond to $L = 11$ pilot subcarriers. The estimator has less measurements which lead to reducing the accuracy of the estimates.

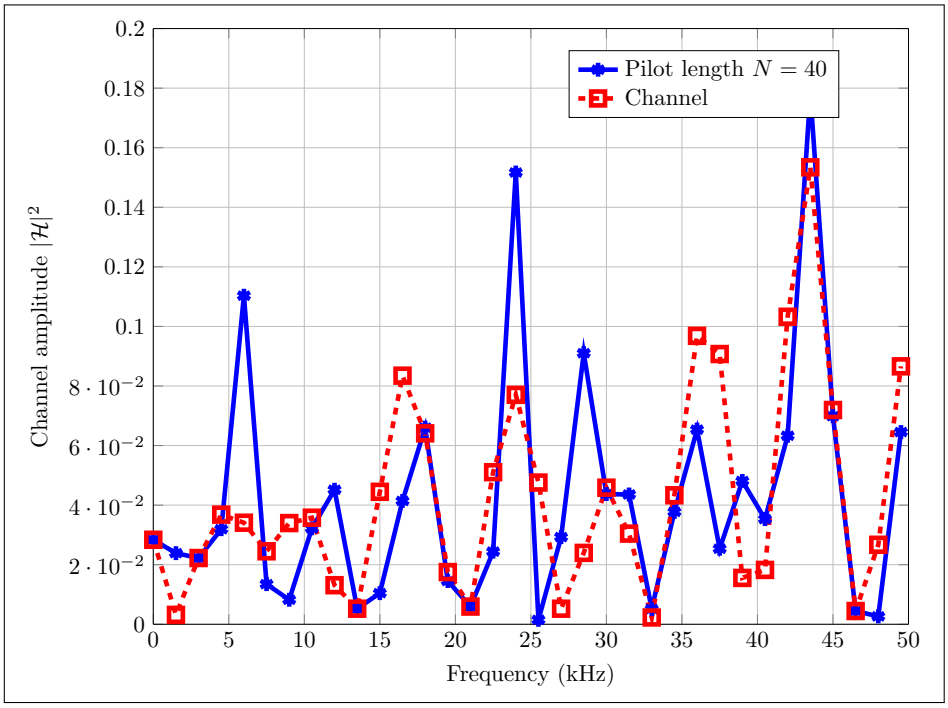
The advantage of the proposed method is the flexibility of exploiting the time-frequency frame by compensating the lack of pilot subcarrier with larger sequences and vice-versa. This is possible thanks to the estimator that takes into account the joint sparsity of the channel. The

Figura 3.4 – Channel estimation result with a pilot placement configuration that uses 34 regularly spaced pilot subcarriers which carry sequences with $N = 40$.



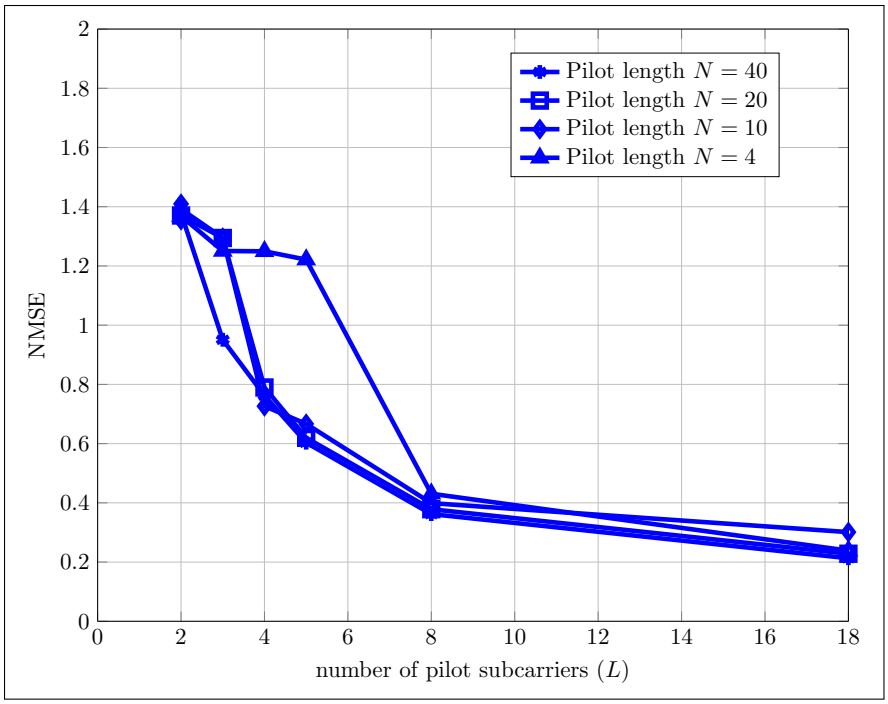
Source: Provided by the author.

Figura 3.5 – Channel estimation result with a pilot placement configuration that uses 10 regularly spaced pilot subcarriers which carry sequences with $N = 40$.



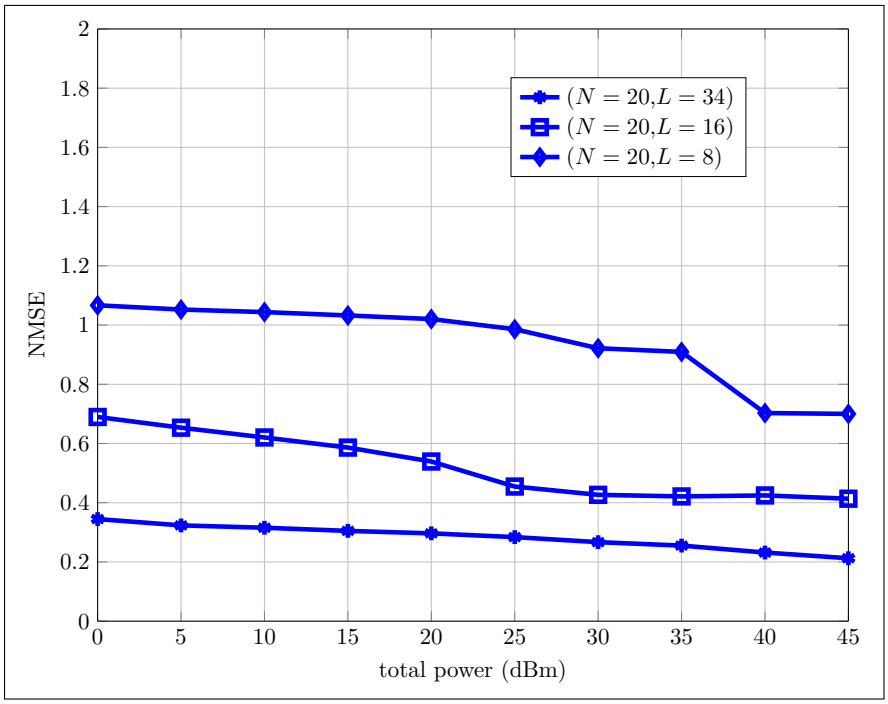
Source: Provided by the author.

Figura 3.6 – Channel estimator performance varying the number of subcarriers.



Source:Provided by the author.

Figura 3.7 – Channel estimator performance by varying the transmit power for some pilot placement configurations.



Source: Provided by the author.

tensor model enables the receiver to exploit multiple dimensions and then inferring the sparse tensor entries. The relationship between the multidimensional massive MIMO channel and its

sparse representation is achieved due to the 3D Fourier basis. Using mode-1, mode-2, mode-3 unfolding forms, the algorithm obtains an estimate of the channel frequency response. Figure 3.6 shows the performance for various configurations of N and L.

Fig. 3.7 shows the estimator performance for some pilot placement configurations which are defined by (N, L). The configuration $(N = 20, L = 34)$ is the densest among the three curves, thus the receiver has more measurements to processing which provides better accuracy in the estimation. The more power is used, the less is the error. However, the variation of estimation error with respect the total power show that there are little benefits in increasing power when the system has denser pilot schemes. On the hand, using fewer pilot measurements, the estimation error with respect the power tends to increase. In this case, there are benefits of inserting more power for the pilots. Thus, there is a trade-off between overhead and power that the BS can adjust according to the scenario. The BS can use denser pilot placements yet without vanishing its communication resources with overhead.

It is important to highlight that the computational complexity becomes cumbersome using the Kronecker-OMP described in Table **??**. The simulations crash because of the memory available on the machine. The steps 2 and 3 complexity limit the implementation of such a method in devices with low computational complexity requirements.

Even though the estimator uses the 3D Fourier basis, other representations can also be used. Once many channel estimations are performed, the UE can use them to adapt the trilinear basis $\mathbf{\Phi}$. This potentially increases the channel sparsity and consequently enables to reduce the number of pilots.

3.7 Conclusion

The tensor-OMP solution capitalizes on sparse multidimensional decomposition model to estimate the channel. Using this formalism, the tensor-OMP algorithm can exploit the joint sparsity between angle and delay domains. The plots show that the method provides reliable channel estimation and that the power spending for pilot transmission can be reduced and compensated with a correct choice of the pilot placement.

An important aspect that is not covered in this chapter is the optimization of the basis. The Fourier basis assumes regularly spaced discrete directions and delays. This causes a power leakage that reduces the sparsity in the tensor core [57]. One way of improving the sparsity is to develop a method that updates the basis, i.e. solutions that can obtain the angles and delays of the channel. In the next chapter, we deal with the problem of accurately estimating the departure and arrival angles in the case of indoor mm-wave channels. This scenario has become a hot topic over the last years because of its applications in wireless communication of the next generation.

4 SPARSE CHANNEL ESTIMATION FOR MILLIMETER-WAVE MASSIVE MIMO SYSTEMS

Abstract

In this chapter, we investigate the channel estimation problem for applications in mm-wave communications. The pilot overhead is still the main issue to be handled. However, herein there are two aspects taken into account: i) the mismatching between the sparse basis and the multipath structure; ii) the losses due to user mobility. To correct the mismatching between the sparse basis the multipath, the proposed method splits the estimation into two steps: first, the algorithm performs a coarse estimation; then, a refinement process to align the basis with the direction of arrival and departure is proposed. To correct the losses due to the user mobility, the access node insert short sequences within the frame transmission to enable the user to correct the estimated directions. It is shown in the results that the proposed sparse channel estimation method provides quasi-optimum performance in terms of throughput.

4.1 Introduction

The ever growing demand for higher data rates motivates researches in wireless communication to develop new solutions to increase the spectral efficiency. One of the approaches considered recently goes on the direction of using the unlicensed spectrum region of mm-wave channels. Some standards have been proposed to operate around 60 GHz such as ECMA-37, IEEE 802.15.3c and IEEE 802.11.ad [58], especially in future indoor communications. In this band, there is a large chunk of available frequencies that can be used to drastically increase system throughput capacity. However, there are limitations that need to be solved to achieve the promised gain. One issue to be addressed is the limited cover area of the AN. The pathloss in this band drastically limits the cell size. However, the recent literature in mm-wave communications envisages a potential combination between massive MIMO and mm-wave transmission to increase the cell size [10, 59].

Massive MIMO is considered as a key enabler of mm-wave future systems. With hundreds of antennas at the AN, it is possible to create very narrow beams that can deal with the severe signal attenuation. Moreover, at this frequency, it is possible to shrink the array size into very small areas. However, the channel characterization differs from that used in cellular communication systems [31].

In regard to the channel model, the electromagnetic wave in millimeter range has similar behavior to the light, making ray-tracing model a very attractive solution to channel modeling. Besides this, other effects can be observed such as the reduced diffraction and the substantial losses presented at each reflection. This means that just a few rays compose the channel between

the AN and UE. The poor scattering environment leads to a sparse representation in the angle domain. This can be used to reduce the number of cell specific pilots [60] in the communication system and to increase the spectral efficiency. The inherent channel sparsity is exploited in this chapter by means of CS technique to achieve a reduction in the pilot overhead.

Different from Chapter **??** and 3, where the bases are fixed to DFT matrices, herein we are interested in optimizing them. Otherwise stated, in previous chapters, the basis mismatch is neglected, although this is a real issue because it causes a power leakage that affects the limits the sparsity degree of the channel to be estimated. Another aspect that calls for a constant update of the direction of arrival (DoA) and the direction of departure (DoD), it is the user mobility. Because the channel between the AN and UE varies the set of beams must be corrected so that the AN can continually serve the UE.

The problem of correcting the basis representation and channel tracking is addressed in [57]. Therein, the authors consider a DL channel estimation, the system is operating on the 60 GHz band, and the AN has a large number of antennas. The solution proposed in this chapter extends that work to multiple-antenna UEs and also for OFDM transmission.

The proposed technique consists of placing two types of pilot sequences. The first one is longer and used by the UE to perform a two phase sparse channel estimation solution while the second sequence is very short and comes in the middle of the transmission frame to be used in the tracking procedure. The estimation process uses the longer sequences and works to acquire the channel from "scratch". The process is split into two steps: the first one coarsely estimates the angles and the second one refines the estimated angles. The shorter sequences are used to estimate the angle variations based on latest information of the channel. After acquiring the angle variations, the UE reports them to the AN, which can perform the antenna beam control [61].

The evaluation of the channel estimation and compressive beamtracking are based on the downlink system throughput considering a single-user scenario. Our evaluations consider different precoding schemes and receiver algorithms. Basically, three precoders are implemented: a wideband precoder (WP), a phase-constrained wideband precoder (PCWP) and a frequency dependent precoder (FP). The first one is a vector of weights which is equal for all subcarriers. The second one also assumes the same beamforming weights for all subcarriers, but consider finite-resolution phase-shifters, where the phase adjustment of the beamformer is discrete and chosen from a codebook of four elements. The third one is designed from the singular value decomposition (SVD) of the channel matrix for each subcarrier, resulting in as many SVDs as the number of used subcarriers.

4.2 Milimeter Wave Channel Characterization and Motivation

The mm-wave band holds to the frequencies between 30-300 GHz, which means a wavelength in the range of 1-10 mm. The use of such frequencies is expected to increase in the

next generation of wireless technologies. This is because the number of applications and users using the licensed frequencies are growing fast, and the spectrum shortage is becoming an issue. A promising way to avoid such an issue consists of utilizing the mm-wave band that typically has not been used by consumer applications.

To propose a new technology that meets that ever growing data rate demands, we need to understand the channel characterization in the mm-wave band. An important difference between mm-wave and other systems that operate at lower frequencies is that the radio waves experience greater attenuation in higher frequencies. The coverage is thus affected by strong pathloss, a higher penetration loss and strong diffraction loss. These effects confine the applications that use this band of frequencies to very short communication ranges, such as a single room. The advantage of this feature is the cell isolation that the channel itself naturally provides, this allows systems with higher frequency re-use factor [62, 63].

Another interesting aspect is the possibility of housing multiple antennas into a very small area. Since the antenna size is related to the wavelength of the frequency carrier, the separation among the antenna radiators and their size become smaller as the wavelength becomes short. With this possibility, many applications envisaged in the next generation of wireless communication assume massive MIMO transmitters to circumvent the coverage limitations. The large number of antennas indeed provides enhanced array directive which makes possible the transmitted signal reaches farther distances.

Every channel model in wireless communication has a trade-off between accuracy and simplicity. The model must capture the aspects that are important without being complex, so that the model is easy to implement. Aspects such as frequency, bandwidth and propagation environments play an important role the on channel characterization. There are two types of channel model: stochastic and deterministic. Stochastic models aim to reproduce the statistical behavior of the channel. Deterministic models are more site-specific, which means that they are better suited for those cases to a specific environment.

There are many types of deterministic models, but we only focus on the use of a ray-tracing model. In this model, each multipath is represented as a geometric-ray. This model is a very interesting tool to represent the quasi-optical nature of the eletromagnectic-wave [31]. The time-domain massive MIMO channel matrix is

$$\mathbf{H}(\tau) = \mathbf{H}_{ray-tracing}(\tau) + \rho\mathbf{H}_{fading}(\tau), \tag{4.1}$$

where, $\mathbf{H}_{ray-tracing}(\tau)$ is the non-fading component generated by the ray-tracing model including the line-of-sight (LOS) ray and reflected rays, $\mathbf{H}_{fading}(\tau)$ is the normalized fading component generated by the sum of multiple random rays coming from a number of predefined point scatters in random directions while ρ is a pre-defined parameter which assigns the power ratio between $\mathbf{H}_{ray_tracing}(\tau)$ and $\mathbf{H}_{fading}(\tau)$. A more detail description about the channel model can be found in [64].

It is important to highlight that, although the channel model from [64] has been used in simulations in the 60 GHz band, there are two specific points not implemented yet on the simulator. First, according to the experimental measurements discussed in [62, 63] for each reflected path there are several other closely spaced rays in the time-angular domain due to the fine irregular structure. Second, the power loss due to the polarization mismatch between the antennas and the channel can be as high as 10dB-20dB [62, 63]. These two considerations are envisaged for future investigations.

4.3 System Model

Consider an indoor environment with an AN deployed with N_t antennas and a UE with N_r antennas. At each link end a rectangular array of multiple antennas is considered. We assume that the AN and the UE are perfectly synchronized and operate according to an FDD protocol. This means that a cell search procedure was performed resulting in a frequency and symbol synchronization to a cell [60]. Given that the communication between the AN and the UE is synchronized and assuming an OFDM modulation, the AN needs to acquire the channel to design a beamforming and then serving the UE. The way to acquire this information is to place some pilots in the OFDM symbols so that the UE can use these specific symbols to estimate the channel. The received signal is the input of channel estimator that estimates the channel.

To propose the channel estimator, we first describe the received signal model associated with the k-th discrete time instant and l-th subcarrier at the UE as

$$\begin{aligned}\mathbf{y}_r(k,l) &= \sum_{n=1}^{N_p} \beta_n \mathbf{v}_R(\theta_{R,n},\phi_{R,n})\mathbf{v}_T^H(\theta_{T,n},\phi_{T,n}) \\ &\quad \mathbf{s}(k,l)e^{-\jmath 2\pi\tau_n l\Delta f} + \mathbf{z}(k,l),\end{aligned} \tag{4.2}$$

where $k \in \mathcal{S}_{time}$ and $l \in \mathcal{S}_{subcarrier}$, and their cardinality are K and L respectively. These sets define the frequency and time indices chosen for the transmission of the pilot symbols. The variable N_p denotes the number of paths, β_n is a complex factor with Gaussian distribution along the n-th ray, $\phi_{R,n}$ and $\theta_{R,n}$ are respectively the azimuth and the inclination of the incoming rays relative to $x-y$ plane. $\phi_{T,n}$ and $\theta_{T,n}$ denote, respectively, the azimuth and inclination of the rays at the transmitter. The parameter $\tau_n \in [0, \tau_{max}]$ defines the delay related to the n-th path, T_s is the OFDM symbol period and Δf is frequency distance between two adjacent subcarriers. The $N_t \times 1$ vector $\mathbf{v}_T(\theta_{T,n})$ and the $N_r \times 1$ vector $\mathbf{v}_R(\theta_{R,n})$ are the array steering vectors for the transmitter and receiver, respectively. The vector $\mathbf{s}(k,l)$ contains the pilot symbols used at the k-th time instant and l-th subcarrier. Finally, the term $\mathbf{z}(k,l)$ represents the additive noise whose entries are independent and follow a zero mean complex-valued Gaussian distribution with variance σ^2 [27].

A given element of the steering vectors describe the phase response to a planar wave in

the azimuth θ and elevation ϕ. The ith element of the steering vector $\mathbf{v}$ is given by

$$[\mathbf{v}]_i = e^{\jmath(\omega_x x_i + \omega_y y_i)}, \tag{4.3}$$

where $\omega_x = \frac{2\pi d}{\lambda}\cos(\theta)\cos(\phi)$, and $\omega_y = \frac{2\pi d}{\lambda}\cos(\theta)\sin(\phi)$, with x_i and y_i defining the spatial position of the i-th antenna element on the plane $x-y$. In this chapter, we call ω_x and ω_y as spatial frequencies and their units are radians/mm.

4.4 Channel Estimation

The channel estimation process starts with the receiver probing the channel in a set of pre-defined pairs spatial frequencies, $(\omega^p_{R,x}, \omega^p_{R,y}) \in \mathcal{D}$, $\forall\, p \in \{1, 2, \ldots, P\}$. The set $\mathcal{D}$ has cardinality $|\mathcal{D}| = P$ and the index p identifies the p-th pair of the set $\mathcal{D}$. Each pair defines one direction, where the receiver measures the associated energy. For the pth spatial frequency pair, there are KL measurements taken on the time-frequency grid given by:

$$r_p(k,l) = \mathbf{w}_p^H \mathbf{y}_r(k,l), \quad p = 1, \ldots, P, \tag{4.4}$$

where $r_p(k,l)$ is the received signal measured on the p-th direction, k-th time instant and l-th subcarrier and $\mathbf{w}_p \in \mathcal{W}$ is the combining vector associated with the p-th direction, $|\mathcal{W}| = P$. By inserting (4.2) into (4.4), we can rewrite $r_p(k,l)$ as:

$$r_p(k,l) = \mathbf{s}^T(k,l)\mathbf{V}_T^* \mathbf{F}_p \mathbf{b}(l) + \tilde{z}_p(k,l), \tag{4.5}$$

where $\mathbf{V}_T = [\mathbf{v}_T(\theta_{T,1}, \phi_{T,1}),\ \ldots,\ \mathbf{v}_T(\theta_{T,N_p}, \phi_{T,N_p})]$, $\mathbf{F}_p$ is a diagonal matrix whose n-th diagonal element is given by $[\mathbf{F}_p]_{n,n} = \mathbf{w}_p^H \mathbf{v}_R(\theta_{R,n}, \phi_{R,n})$, i.e. they account for the correlation between spatial filter and the incoming rays, $\mathbf{b}(l) = \left[\beta_1 e^{-\jmath 2\pi\tau_1 l\Delta f}, \ldots, \beta_{N_p} e^{-\jmath 2\pi\tau_{N_p} l\Delta f}\right]^T$, and $\tilde{z}_p(k,l) = \mathbf{w}_p^H \mathbf{z}(k,l)$. Defining the $K \times 1$ vector $\mathbf{x}_p(l) = [r_p(0,l), \ldots, r_p(K-1,l)]^T$ concatenating K time-domain measurements leads to

$$\mathbf{x}_p(l) = \mathbf{S}_l^T \mathbf{V}_T^* \mathbf{F}_p \mathbf{b}(l) + \tilde{\mathbf{z}}_p(l), \tag{4.6}$$

where $\mathbf{S}_l = [\mathbf{s}(0,l),\ \mathbf{s}(1,l),\ \ldots,\ \mathbf{s}(K-1,l)]$ is the $N_t \times K$ pilot matrix, and $\tilde{\mathbf{z}}_p(l) = [\tilde{z}_p(0,l),\ \ldots, \tilde{z}_p(K-1,l)]^T$.

The estimation accuracy of $\mathbf{b}(l)$ depends on the number of measurements acquired by the receiver. Such a number is very important to be analyzed because in massive MIMO the number of downlink pilots tends be prohibitive. To avoid the loss in spectral efficiency due to the overhead, it is important to exploit additional information on the channel structure that enables us to develop channel estimators capable of estimating with less overhead.

The authors in [31] show in their measurements that mm-wave channels are sparse in the angular domain. This is because the propagation mechanisms of reflection and diffraction produce severe attenuation on the signal. This reduces not only the coverage area, but also affects

the channel richness. Therefore, it is reasonable to consider that the number of multipaths N_p is much smaller than the number of transmit antennas. Under this assumption, the receiver aims to estimate only those N_p entries of $\mathbf{b}(l)$ instead of the $N_r N_T >> N_p$ entries of the full massive MIMO channel matrix. This approach reduces the number of parameters to be estimated which means that the transmitter can use less pilot symbols (i.e. shorter pilot sequences). This is a very attractive solution for massive MIMO because it potentially reduces the system overhead.

To have a better intuition on how to exploit the sparsity, consider a hypothetical case where the receiver knows the arrival angles of the incoming paths and designs a receive beamforming vector $\mathbf{w}_p$ pointing towards the direction of the strongest path. Beside this, consider also that the spatial frequencies seen by the transmitter are integer multiples of $2\pi/N_t$, i.e. $\omega_{T,x}$, $\omega_{T,y} \in \{0, \quad \frac{2\pi}{N_t}, \quad \ldots, \quad \frac{(N_t-1)2\pi}{N_t}\}$. This assumption implies that $\mathbf{V}_T$ is a 2-D DFT and a sparse basis for the channel. Using this information, the BS can exploit such a sparse representation and recover $\mathbf{F}_p\mathbf{b}(l)$ from (4.6) with high probability [27].

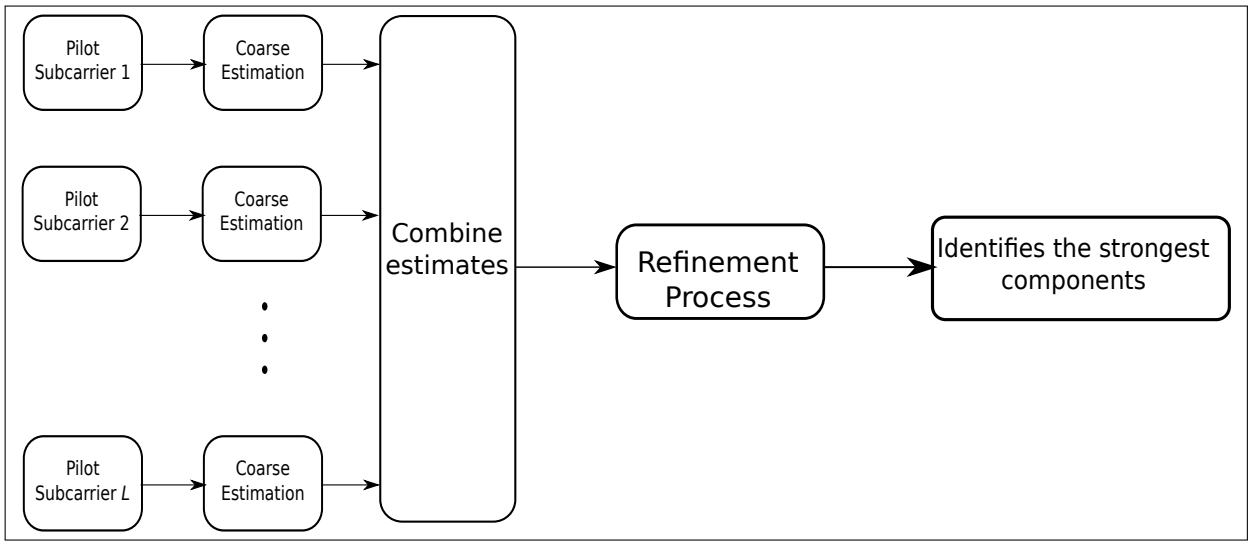
Even though the example gives us an insight on how to exploit the channel structure, the above assumptions are not realistic. Some of them do not hold in reality. For instance the use of the DFT matrix to obtain the sparsity is true only if the paths were regularly spaced. In reality, the use such a basis may cause power leakage because the spatial frequencies come from the continuum space representation, and it is unlike they are integer multiples of $\frac{2\pi}{N_t}$. Despite these observations, we could ignore them and apply compressive sensing algorithms by assuming that the spatial frequencies are integer multiples of $\frac{2\pi}{N_t}$. However, the power leakage among the spatial frequencies limits the CS performance, because it reduces the sparsity level.

4.4.1 Compressive Estimation

In this section, the algorithm to estimate spatial frequencies is stated. The proposed solution is obtained by exploiting the sparse structure of the channel. The approach used in this solution has mainly two stages: first, there is a uniform quantization of the spatial frequencies on each side of the transmission link, creating a 2-D spatial frequency grid for the transmitter and other for the receiver, where $\omega_{R,x}, \omega_{R,y}, \omega_{T,x}, \omega_{T,y} \in [0, 2\pi]$. The algorithm searches for the best combination $(\omega_{R,x}, \omega_{R,y}, \omega_{T,x}, \omega_{T,y})$ which characterizes the first procedure that is called as coarse estimation. After this, there is a refinement step that adjusts the coarse estimates of the spatial frequencies.

The coarse estimation gives a preliminary information of where $(\omega_{R,x}, \omega_{R,y}, \omega_{T,x}, \omega_{T,y})$ sits in the 4D space representation whereas the refinement procedure seeks to find the optimal point $(\omega_{R,x}, \omega_{R,y}, \omega_{T,x}, \omega_{T,y})$ that precisely describes the rays. The procedure is performed sequentially: i) coarse estimation and ii) the refinement. The method is applied per subcarrier, and the outcomes of each estimation process are combined so that the receiver can identify the dominant directions. The Fig. **??** shows the block diagram of the proposed scheme.

Figura 4.1 – Proposed channel estimation to identify strongest multipaths using multiple subcarriers.



Source: Provided by the author.

4.4.2 Coarse Estimation

In reality, the receiver does not know neither the directions of the paths nor the number of paths. Additionally, the real spatial frequencies that synthesize the beamforming vector $\mathbf{w}_p$ and the array response matrix $\mathbf{V}_T$ in Eqs. (4.4) and (4.6) respectively, are also not known at the receiver. Therefore, as a starting point, the estimator creates a set of discrete spatial frequencies for both sides of the link and for each x, y axis, i.e. $\omega_{R,x},\ \omega_{R,y},\ \omega_{T,x},\ \omega_{T,y} \in \{0,\ \frac{2\pi}{P},\ \ldots,\ \frac{(P-1)2\pi}{P}\}$, where $P > \max(N_T, N_R)$ [65]. The receiver collects the received signal and process the signal to find the most relevant directions. The search for the strongest rays is performed according to the expression

$$\mathbf{w}_o = \arg\max_p \sum_l \sum_k |\mathbf{w}_p^H \mathbf{y}_r(k,l)|^2, \quad \mathbf{w}_p \in \mathcal{W}. \tag{4.7}$$

The vector $\mathbf{w}_p$ is generated according to the discrete spatial frequencies $\omega_{R,x}^{(p)}$ and $\omega_{R,y}^{(p)}$ as shown below

$$[\mathbf{w}_p]_i = \frac{1}{\sqrt{N_r}} e^{\jmath(\omega_{R,x}^{(p)} x_i + \omega_{R,y}^{(p)} y_i)}, \quad i = \{1,\ \ldots,\ N_R\}, \tag{4.8}$$

where x_i and y_i are the coordinates of the ith antenna element at the receive array.

Based on $\mathbf{w}_o$ from (4.7), the incoming signal along the strongest path during K consecutive time instants is given by:

$$\begin{aligned} \mathbf{y}_{beam}(l) &= [\mathbf{w}_o^H \mathbf{y}_r(0,l),\ \ldots,\ \mathbf{w}_o^H \mathbf{y}_r(K-1,l)]^T \\ &= \mathbf{S}_l^T \mathbf{V}_{D,T}^* \tilde{\mathbf{F}}_o \tilde{\mathbf{b}}(l) + \tilde{\mathbf{z}}_o(l), \end{aligned} \tag{4.9}$$

where $\mathbf{V}_{D,T}$ is a $N_t \times P$ matrix with each column describing the array response for a possible combination of $\omega_{T,x}^{(p)}$ and $\omega_{T,y}^{(p)}$, $\tilde{\mathbf{b}}(l) = [\tilde{\beta}_1 e^{-\jmath 2\pi\tau_1 l\Delta f},\ \ldots,\ \tilde{\beta}_P e^{-\jmath 2\pi\tau_P l\Delta f}]^T$, and $\tilde{\beta}_1, \ldots, \tilde{\beta}_P$ are

independent random variables following a zero-mean unitary-variance Gaussian distribution. The matrix $\tilde{\mathbf{F}}_o$ is a $P \times P$ diagonal matrix, with the mth diagonal element $[\tilde{\mathbf{F}}_o]_{m,m} = \mathbf{w}_o^H \mathbf{w}_m$, $m = 1, \ldots, P$. We call attention to the difference between Eqs. (4.6) and (4.9). The first one is based on the exact number of multipaths, therefore the dimensions of vector $\mathbf{b}(l)$ and matrix $\mathbf{F}_p$ are functions of N_p. The second one is based on the number P of predefined directions taken from the 2-D continuum of spatial frequencies $\omega_{R,x}$ and $\omega_{R,y}$.

The traditional LS estimation method imposes a minimum number of pilots $K = \mathcal{O}(N_t)$ to be used. However, such a condition in a massive MIMO system results in huge overhead to the system. It is known that a channel estimator based on CS circumvents this problem by exploiting the sparse channel structure. The estimator enables more reduction of the pilot sequence length as sparser as the channel is. From (4.9), more specifically in the matrix $\tilde{\mathbf{F}}_o$, the sparsity of the channel is controlled by the inner product between $\mathbf{w}_o$ and the other incoming rays. In mm-wave, there is the possibility of packing more antennas at the receiver, which creates the possibility of improving the spatial resolution. This induces sparsity in the vector $\tilde{\mathbf{F}}_o \tilde{\mathbf{b}}(l)$. If the receiver had a single omnidirectional antenna, as in [65], $\tilde{\mathbf{F}}_o$ would be the identity matrix and $\mathbf{b}_{\text{filt}}(l) = \tilde{\mathbf{b}}(l)$. This means that CS estimation performance would depend on the "natural" channel sparsity. In the solution proposed in this work, the receiver uses its multiple antennas to spatially filter other paths. This increases the degree of channel sparsity which makes CS estimation more attractive.

4.4.3 Combining Estimates from Multiples Subcarriers

The coarse estimation is performed per subcarrier as shown in Fig. **??**. More specifically, the receiver combines the estimates $\hat{\mathbf{b}}_{\text{filt}}(l)\ \forall l \in \mathcal{S}_{subcarrier}$ to assign the best direction. The estimator then concatenates these estimates $\hat{\mathbf{b}}_{\text{filt}}(l)$ from every subcarrier into a matrix $\mathbf{B}_{\text{filt}}$. This matrix is the input of a function $f : \mathbb{C}^{P \times L} \rightarrow \mathbb{C}^{P \times L}$ that generates the vector

$$\mathbf{b}_f = f(\mathbf{B}_{\text{filt}}). \tag{4.10}$$

In this thesis, three functions are considered. The first one is

$$f(\mathbf{B}_{\text{filt}}) = 1/L \sum_{l=1}^{L} \hat{\mathbf{b}}_{\text{filt}}(l), \tag{4.11}$$

where, the estimation of $\mathbf{b}_{\text{filt}}(l)$ is obtained from average estimates of each pilot subcarrier. The second function is:

$$f(\mathbf{b}_{\text{filt}}) = 1/L \sum_{l=1}^{L} |\hat{\mathbf{b}}_{\text{filt}}(l)|^2 \tag{4.12}$$

which takes into account the pointwise absolute value operation and add up all the estimates. The third function is

$$f(\mathbf{B}_{\text{filt}}) = \max_{\tau} 1/L \sum_{l=1}^{L} \hat{\mathbf{b}}_{\text{filt}}(l) e^{j2\pi l \Delta f \tau}. \tag{4.13}$$

Eq. (4.11) is the simplest way to combine the estimates. However, it does not consider the phase shift caused by the delay. Thus, summing up complex values may is more suited to

cause destructive interference due to this fixed phase. This means that (4.11) is mainly used for low delay spread scenarios, where the phase rotation due to channel path delays is negligible. The second function presented in Eq. (4.12) overcomes the phase incoherence problem by taking into account the the average power of $\hat{\mathbf{b}}_{\text{filt}}(l)$ in L subcarriers. This method exploits the fact that the Fourier transform does not change the power of the paths. This makes possible identify the most relevant paths. A cumbersome situation for this approach would be in a low SNR scenario. To handle this scenario, the channel delay must be taken into account. This situation is covered in (4.13).

After combining the estimates from different subcarriers, we choose largest value $\mathbf{b}_f$ which is defined as

$$i^* = \max_i [\mathbf{b}_f]_i. \tag{4.14}$$

The value i^* is associated with the i^*th direction at the transmitter (i.e. the i^*-th column of $\mathbf{V}_T$), corresponding to the strongest path. Table 3 summarizes the coarse estimation step of the proposed algorithm.

Tabela 3 – Coarse estimation

Step 1	Quantize the spatial frequencies components: $\omega_{R,x},\ \omega_{R,y},\ \omega_{T,x},\ \omega_{T,y} \in \{0,\ \frac{2\pi}{R_{samp}},\ \ldots,\ \frac{(R_{samp}-1)2\pi}{R_{samp}}\}$, where, $R_{samp} > N_t$.
Step 2	Find $\hat{\omega}_{R,x}$ and $\hat{\omega}_{R,y}$ which give the strongest direction. $[\mathbf{w}]_i = \frac{1}{\sqrt{N_t}} e^{\jmath(\omega_{R,x} x_i + \omega_{R,y} y_i)}$, $J = \max_{\mathbf{w}} \{\sum_{l=1}^{L} \sum_{k=1}^{K} \lvert\mathbf{w}^H \mathbf{y}(k,l)\rvert^2\}$
Step 3	Calculate the beam signal coming from the direction $(\hat{\omega}_{R,x}, \hat{\omega}_{R,y})$ $[\mathbf{w}]_i = \frac{1}{\sqrt{N_r}} e^{\jmath(\hat{\omega}_{R,x} x_i + \hat{\omega}_{R,y} y_i)}$, $\mathbf{y}_{beam}(l) = [\mathbf{w}^H \mathbf{y}(0,l),\ \mathbf{w}^H \mathbf{y}(1,l),\ \ldots,\ \mathbf{w}^H \mathbf{y}(K-1,l)]$ $\mathbf{y}_{beam}(l) = \mathbf{S}_l^T \mathbf{V}_T \tilde{\mathbf{F}}_o \tilde{\mathbf{b}}(l) + \mathbf{z}(l)$
Step 4	Use OMP, described in Table 1, to estimate $\mathbf{b}_{\text{filt}}(l) = \tilde{\mathbf{F}}_o \tilde{\mathbf{b}}(l)$ $\hat{\mathbf{b}}_{\text{filt}}(l) = \lvert\mathbf{b}_{\text{filt}}(l)\rvert_1 + \lvert\mathbf{y}_{beam}(l) - \mathbf{S}_l^T \mathbf{V}_T \mathbf{b}_{\text{filt}}(l)\rvert^2 < \sigma^2$
Step 5	Repeat steps 3 and 4 for $l \in \mathcal{P}_{subcarrier}$ $\mathbf{B}_{\text{filt}} = [\hat{\mathbf{b}}_{\text{filt}}(1),\ \ldots,\ \hat{\mathbf{b}}_{\text{filt}}(L)]$
Step 6	Use the function $f : \mathbb{C}^{P \times L} \rightarrow \mathbb{C}^{P \times L}$ to combine the $\hat{\mathbf{b}}_{\text{filt}}(l)$ $\forall\ l \in \mathcal{P}_{subcarrier}$. $\mathbf{b}_f = f(\hat{\mathbf{b}}_{\text{filt}}(l))$
Step 7	Choose the index of $\mathbf{b}_f$ whose absolute value is the largest. $i^* = \arg\max_i [\mathbf{b}_f]_i$ i^*th column of $\mathbf{V}_T$ is the strongest direction of the transmitter.

4.4.4 Refinement of the Estimates

After the coarse estimation step, an optimization is performed for adjusting the coarse estimate of the directions $\omega_{R,x}^{coarse}$ and $\omega_{R,y}^{coarse}$. The achieved adjustment is based on the

maximization of the received signal energy. The refinement problem can be stated as:

$$[\omega_{R,x}^{\text{ref}}(l), \omega_{R,y}^{\text{ref}}(l)] = \arg \max_{(\omega_{R,x}, \omega_{R,y}) \in \mathbb{R}^2} J(\omega_{R,x}, \omega_{R,y}, l)$$

where

$$J(\omega_{R,x}, \omega_{R,y}, l) \doteq \sum_{k=1}^{K} |\mathbf{w}^H(\omega_{R,x}, \omega_{R,y})\mathbf{y}(k,l)|^2, \tag{4.15}$$

and $\mathbf{w}(\omega_{R,x}, \omega_{R,y})$ is the steering vector associated with the pair $(\omega_{R,x}, \omega_{R,y})$, and each element $[\mathbf{w}(\omega_{R,x}, \omega_{R,y})]_i$ follows the same formulation as in (4.8). The final estimates are given by averaging over the L subcarriers, i.e. $\overline{\omega}_{R,x}^{\text{ref}} = (1/L) \sum_{l=1}^{L} \omega_{R,x}^{\text{ref}}(l)$, and $\overline{\omega}_{R,y}^{\text{ref}} = (1/L) \sum_{l=1}^{L} \omega_{R,y}^{\text{ref}}(l)$.

In practice, the adjustment of the spatial frequencies $\omega_{R,x}$ and $\omega_{R,y}$ is done through small steps Δ. This can be done using the so-called "search region" algorithm, which basically consists of evaluating, in a given iteration, the cost function (4.15) for a set of Δ-spaced points around the one selected in the previous iteration. The method seeks for the largest value of this cost function among the candidate points and selects the one leading to the maximum value. The algorithm stops according to the desired granularity. Differently from the coarse estimation step, the refinement step does not take into account the whole set of discrete directions initially defined by the Fourier matrix. It takes one predefined direction, $(\omega_{R,x}^{\text{coarse}}, \omega_{R,y}^{\text{coarse}})$, and searches for the best pair $(\omega_{R,x}, \omega_{R,y}) \in \mathbb{R}^2$ that optimizes (4.15).

Since $\overline{\omega}_{R,x}^{\text{ref}}$ and $\overline{\omega}_{R,y}^{\text{ref}}$ have already been obtained, this information is then used to refine the transmit spatial frequencies, which is based on the maximization of the following function:

$$[\omega_{T,x}^{\text{ref}}(l), \omega_{T,y}^{\text{ref}}(l)] = \arg \max_{\omega_{T,x}, \omega_{T,y}} |\mathbf{v}^H(\omega_{T,x}, \omega_{T,y})\mathbf{S}_l^* \mathbf{y}_{beam}(l)|^2, \tag{4.16}$$

where $(\omega_{T,x}, \omega_{T,y}) \in \mathbb{R}^2$, $\mathbf{v}(\omega_{T,x}, \omega_{T,y})$ follows the same formulation as in (4.8), except by the index p and the number of antennas, denoted by N_T instead of N_R. As for the receive spatial frequencies, the final estimates of $\overline{\omega}_{T,x}^{\text{ref}}$ and $\overline{\omega}_{T,y}^{\text{ref}}$ are obtained also by averaging over the L subcarriers.

In principle, the refinement of receive and transmit spatial frequencies could be performed jointly by updating the spatial frequencies at the transmitter and receiver in an iterative (ping-pong) process. However, such an approach would imply a higher computational complexity. In this work, we consider a low complexity approach, where this refinement is done separately, i.e. in a decoupled fashion, at each link end. According to our results, good performance can be obtained with this approach for LOS scenarios. Table 4 summarizes the procedure used in the refinement of the transmit spatial frequencies. A similar procedure is used in the refinement of the receive spatial frequencies, except that step 2 is not used and step 4 performs the maximization defined in (4.7) as shown in Table 5.

Tabela 4 – Refinement of the transmitter spatial frequencies

Step 1	From the coarse estimation, use the spatial frequencies at the transmitter, $\omega_{T,x}^{coarse}$ and $\omega_{T,y}^{coarse}$, as initial values: $\psi_{0,x} = \omega_{T,x}^{coarse}$ and $\psi_{0,y} = \omega_{T,y}^{coarse}$ step size: Δ granularity adjustment constant : ϵ $[\mathbf{w}]_i(\omega_{R,x}^{refined}, \omega_{R,y}) = \frac{1}{\sqrt{N_r}} e^{\jmath(\omega_{R,x} x_i + \omega_{R,y} y_i)}$, $i = \{1, \ \ldots, \ N_t\}$
Step 2	Calculate the beam signal coming from the direction $(\omega_{R,x}^{refined}, \omega_{R,y}^{refined})$ $\mathbf{y}_{beam}(l) = [\mathbf{w}^H(\omega_{R,x}^{refined}, \omega_{R,y})\mathbf{y}(0,l), \ \ldots, \ \mathbf{w}^H(\omega_{R,x}^{refined}, \omega_{R,y})\mathbf{y}(K-1,l)]^T$
Step 3	Update $\psi_{t,x}$ and $\psi_{t,y}$ $\psi_{+\Delta,x} = \psi_{t,x} + \Delta$ $\psi_{-\Delta,x} = \psi_{t,x} - \Delta$ $\psi_{+\Delta,y} = \psi_{t,y} + \Delta$ $\psi_{-\Delta,y} = \psi_{t,y} - \Delta$
Step 4	Assign the new spatial frequency $[\psi_x, \psi_y] = \arg\max_{(\psi_x,\psi_y)} \lvert\mathbf{v}^H(\psi_x,\psi_y) S_l^* \mathbf{y}_{beam}(l)\rvert^2$, where, $[\mathbf{v}]_i(\psi_x,\psi_y) = \frac{1}{\sqrt{N_t}} e^{\jmath(\psi_x x_i + \psi_y y_i)}$, $i = \{1, \ \ldots, \ N_t\}$ $\psi_x \in \{\psi_{+\Delta,x}, \psi_{-\Delta,x}\}$ $\psi_y \in \{\psi_{+\Delta,y}, \psi_{-\Delta,y}\}$ $[\psi_{t,x}, \psi_{t,y}] = \arg\max_{(\psi_x,\psi_y)} J(\psi_x,\psi_y)$
Step 5	If $[\psi_{t,x}, \psi_{t,y}] = [\psi_{t-1,x}, \psi_{t-1,y}]$ do $\Delta = \Delta - \epsilon$ and go back to step 2. Otherwise, go forward.
Step 6	If the number of iterations $t < T_o$, set $t = t + 1$ and go back to Step 3, otherwise $\omega_{R,x}^{refined} = \psi_{t,x}$ $\omega_{R,y}^{refined} = \psi_{t,y}$

4.4.5 Channel Tracking

The beams produced by the large array are quite narrow which increases the directivity and, consequently, the transmitted signal can reach farther distances. However, because of its narrowness, if the user slightly moves to any direction the power delivered decreases fast. Therefore, the AN must have a solution to correct the beam to keep serving the UE. To estimate the new direction, one possible solution is to repeat the process we developed based on CS. However, this is not interesting because the algorithm estimates the channel without exploiting any initial information of the channel. In the channel tracking problem, the AN already knows the last estimate of the beam and can use this information to devise a solution that demands less pilots.

The advantage of channel tracking is that the AN can maintain the block of pilots to estimate the channel with a large period, and between them send very small blocks containing very few pilots responsible for tracking the channel variation between the two blocks of estimation. This enables to increase the period between two large blocks which reduces the overhead.

The key to support the reduction of pilots for channel tracking is related to the partial

Tabela 5 – Refinement of the receiver spatial frequencies

Step 1	From the coarse estimation, use the spatial frequencies at the receiver, $\omega_{R,x}^{coarse}$ and $\omega_{R,y}^{coarse}$, as initial values: $\psi_{0,x} = \omega_{R,x}^{coarse}$ and $\psi_{0,y} = \omega_{R,y}^{coarse}$ step size: Δ granularity adjustment constant : ϵ
Step 2	Update $\psi_{t,x}$ and $\psi_{t,y}$ $\psi_{+\Delta,x} = \psi_{t,x} + \Delta$ $\psi_{-\Delta,x} = \psi_{t,x} - \Delta$ $\psi_{+\Delta,y} = \psi_{t,y} + \Delta$ $\psi_{-\Delta,y} = \psi_{t,y} - \Delta$
Step 3	Assign the new spatial frequencies $[\mathbf{w}]_i(\psi_x, \psi_y) = \frac{1}{\sqrt{N_t}} e^{\jmath(\psi_x x_i + \psi_y y_i)}$, $i = \{1, \ \ldots, \ N_t\}$ $J(\psi_x, \psi_y, l) = \{\sum_{k=1}^{K} \lvert\mathbf{w}(\psi_x, \psi_y)^H \mathbf{y}(k,l)\rvert^2\}$ where, $\psi_x \in \{\psi_{+\Delta,x}, \psi_{-\Delta,x}\}$ $\psi_y \in \{\psi_{+\Delta,y}, \psi_{-\Delta,y}\}$ $[\psi_{t,x}, \psi_{t,y}] = \arg\max_{(\psi_x,\psi_y)} J(\psi_x, \psi_y)$
Step 4	If $[\psi_{t,x}, \psi_{t,y}] = [\psi_{t-1,x}, \psi_{t-1,y}]$ do $\Delta = \Delta - \epsilon$ and go back to step 2. Otherwise, go forward.
Step 5	If the number of iterations $t < T_o$, set $t = t + 1$ and go back to Step 2, otherwise $\omega_{R,x}^{refined} = \psi_{t,x}$ $\omega_{R,y}^{refined} = \psi_{t,y}$

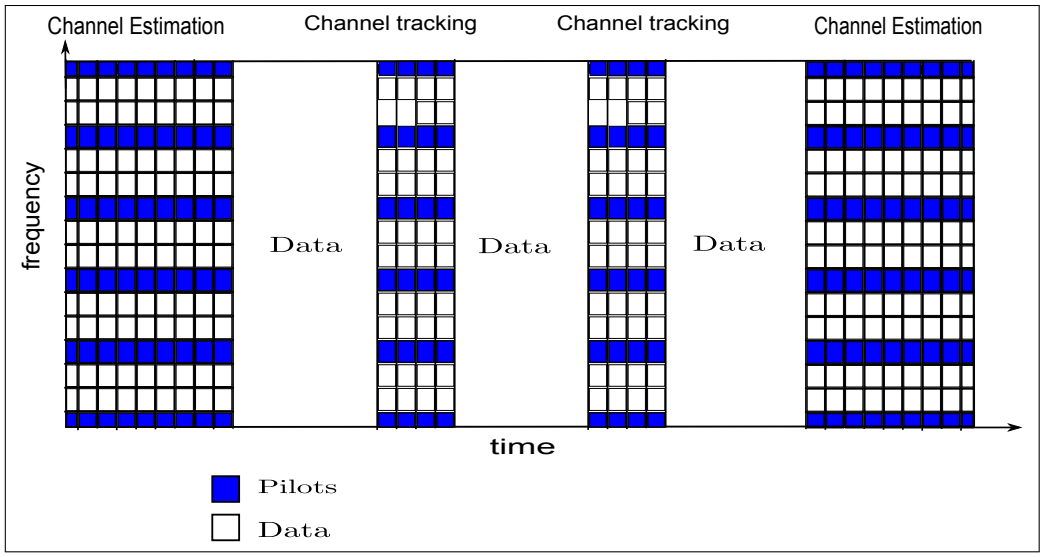
knowledge of channel. The receiver uses the last beam direction used to communicate with the AN as an input of the channel tracking algorithm. The estimator collects the measurements obtained from small blocks of pilots an the UE corrects beam directions on both sides of the link and reports them to the AN.

The solution applied to channel tracking is similar to the one considered on the refinement process. The tracking algorithm basically corrects the spatial frequencies of the receiver and the transmitter based on the criteria stated on (4.15) and (4.16), respectively. This means that the same algorithms employed for the refinement of the transmit and receive directions can be used for the tracking. The only difference is the input of each algorithm, instead of being the coarsely estimated spatial frequencies, the tracking is done by initializing the refinement algorithm with the spatial frequencies estimated at the last channel estimation. Fig. 4.2 shows alternation between channel estimation and tracking on the time-frequency grid.

4.5 Performance evaluation methodology

After the channel beam direction estimation or beam direction adjustment (tracking), the UE reports the estimates through the feedback channel to the AN. The latter assigns the

Figura 4.2 – Pilot distribution on the time-frequency grid.



Source: Provided by the author.

beamforming weights and transmits data to the UE. At the receiver side, a maximum ratio combiner (MRC) is used [66].

The proposed algorithms are evaluated using Shannon's formula

$$C_n = \log_2(1 + \text{SNR}(n)). \tag{4.17}$$

where $\text{SNR}(n)$ is the signal noise ratio The system throughput is determined as follows

$$C = \sum_{n=1}^{N_{subcarrier}} \frac{C_n}{T_{symbol}}. \tag{4.18}$$

where, T_{symbol} is the OFDM period symbol.

In this work, three beamforming algorithms are used, namely:

- Eigenbeamforming: derived from the right singular vector of the frequency-dependent channel matrix (i.e. each subcarrier has different beamforming weights). Perfect knowledge of the full instantaneous channel matrix is assumed;
- Steering vector: designed from the only knowledge of the estimated spatial frequencies, i.e.

$$[\mathbf{v}(\omega_{T,x}, \omega_{T,y})]_i = e^{j(\omega_{T,x} x_i + \omega_{T,y} y_i)}; \tag{4.19}$$

- Round-phase: The design of the steering vector is constrained to four different predefined phases only. Specifically, the phases associated with each entry of the steering vector are rounded to the closest phase among the four ones.

Tabela 6 – Simualtion Default Parameter

Carrier Frequency	60 GHz
Multiplexing Scheme	OFDM
Subcarrier Bandwidth	360 kHz
Number of Subcarriers	512
System Bandwidth	0.18 GHz
Number of pilots Subcarriers	32
FFT size	1024
Payload period	1.389 μs
Cyclic prefix	347.22 ns
OFDM symbol period	3.1252 μs
Fading Power Ratio	$\rho = 11.2$ dB
Function f	It is defined in Eq. 4.12
Maximum Tx Power per AN	2 mW
Thermal Noise Level	-174 dBm/Hz
Noise Figure	6 dB
Number of Tx Antennas	64
Number of Rx Antennas	16
Distance Between the Antennas	$\lambda/2$
UE speed	1 m/s

4.6 Results

This section presents the results of the channel estimation and channel tracking algorithms described in this chapter. The simulations are performed according to the parameters of the Table 6 and channel coefficients are generated according to the ray-tracing model used in [64]. The performance is evaluated in terms of the system throughput considering the transmission of a single data stream.

4.6.1 Channel Estimation

Figure 4.3 shows a LOS scenario which has three ANs and one UE in a corridor and, initially, the AN need to determine the beamforming to serve the UE. As a benchmark, we use the classical eigenbeamforming solution which is defined as the eigenvector of the channel matrix associated to the strongest eigenvalue. We show that the steering-vector and round-phase solutions can provide very similar performance to the benchmark one in indoor mm-wave communication systems. The beams are constructed based on the channel angular information which is acquired using quite low overhead by applying the proposed algorithms.

Figure 4.4 shows the system throughput for the three beamforming schemes. Note that the difference between the eigenbeamforming and steering vector is almost negligible. This is because the LOS component retains great part of the received power and other rays have low contribution. This condition is very attractive because the channel has a very sparse representation

which enables us to exploit this sparsity to reduce the system overhead by using the proposed technique.

Another interesting aspect lays on the beamforming implementation. Although the eigenbeamforming achieves the optimum throughput, the BS needs to calculate one per subcarrier which leads to a more complex transceiver architecture. This is an important hardware drawback that added with other implementation issues associated to mm-wave communication systems [46, 67] make eigenbeamforming an unattractive solution from the implementation perspective. Figure 4.4 shows yet that we can further simplify the steering vector and use one that uses finite discrete alphabet. The simulations shows that the round-phase solutions achieve similar performance compared to eigenbeamforming one which provides a promising solution because the AN has the possibility of employing phase-shifter with finite resolutions. Although this is a very attractive solution, more investigations to cover other aspects that are out of the scope of this thesis are still needed.

The algorithm performance is evaluated in terms of pilot overhead by comparing Figures 4.4 and 4.5. The first plot shows the system throughput (bps) resulting from an estimation that uses 20 OFDM symbols with a time interval of 0.1s between two large blocks to estimate the LOS component. The total number of resources used to carry pilots in such a case is 0.0039% of the total communication resources. In the second figure, the AN uses only 5 OFDM symbols to estimate the same channel with the same time interval between two blocks. This leads to a number of resources equal to $9.74 \times 10^{-4}\%$ of the total communication resource available.

Although Fig. 4.5 shows that the channel estimation algorithm provides quite low overhead, there is a trade-off between overhead, distance to cover, and transmitted power as can be noted by comparing Figures 4.6 and 4.7. They show the system performance when the AN reduces its transmitted power. In the first figure, the system acquires reliable information of the channel estimates until the UE reaches 10 meters away from the AN, whereas in the second figure, the UE loses performance around 5 meters away from the AN. Therefore, the AN has the possibility of adjusting the power and the overhead to meet coverage requirements.

4.6.2 Channel Estimation Using Multiple ANs

Although coverage limitations can be circumvented by adjusting power and overhead as discussed lately, this solution has its limitations. Communication systems must follow power policy settings which means that increasing power above values determined by the policies is not an option. Beside this, increasing the number of pilots to cover large areas can results in large overhead. In such a situation, the UE needs to use multiple ANs.

Figure 4.8 shows the channel estimation exploiting the handover between AN 1 and AN 2. In this solution, the UE constantly receives pilot sequences associated to each AN and uses them to evaluate which link has a better SNR. After this, the UE reports the angles to the

intended AN and that designs the steering vector to serve the user. The curves also show that the power must be above a given threshold, otherwise the angle estimation is not accurate and, consequently, the UE cannot decide correctly which AN has the strongest ray. As shown in chapter 2, CS techniques tend to have poor accuracy in low SNR environments.

Because handover gain can be exploited, the ANs can provide high throughput to the UE without overloading the system with large pilot sequences. On the other hand, we can use more pilots to increase the distances between the ANs. In this case, consider that the system wants to perform a handover between AN 1 and AN 3. The number of pilots increases because the distance between those is larger than between AN 1 and AN 2. This strategy can be particularly useful if the AN 2 has a malfunction and the handover can only be performed by ANs 1 and 3.

4.6.3 Combining Pilot Frequency Subcarriers

The channel estimation performances presented on the last figures consider only the combiner defined in Eq. (4.12). Figure 4.10 shows the channel estimation performance using all the three possible choices for f that are defined in Eq.(4.11) (fixed delay), Eq. (4.12) (average power) and in Eq. (4.13) (optimum). The function defined in (4.11) has the worst performance, mainly after 10 meters, whereas in this points the multipath delay becomes more important. The function defined in (4.12) and (4.13) keep the channel estimation close to the ideal performance. This is a very interesting result because the function defined in (4.12) does not need any knowledge of the delay.

4.6.4 Channel Estimation in NLOS Scenario

Figure 4.11 shows the UE moving from the corridor to the room which means that there is a transition between a LOS channel to a NLOS one. The evaluation of this scenario is shown in Fig. 4.12 by means of the system throughput. Note that, there are two regions on the picture: the first one is before the time instant 1.3s and the other one is after this time instant. The former has a better throughput than the last, but it shows many throughput variation due the constructive and destructive interference between the rays caused by the environment geometry. Note that there is a natural loss of performance from region 1 to 2. This is explained by the number of reflections that each ray had before arriving at the receiver when it goes into room. With a number of reflection higher in the NLOS scenario than in a LOS one, the system throughput decreases. A possible solution to improve it is the deployment of a new AN inside the room. This would make possible the handover practice.

Although there is a lot of SNR variation, the channel estimation has the same performance for three different values of overhead as shown on Fig. 4.12. This means that the we can still rely on the sparsity assumption even though the channel is not LOS anymore. However, the use of very short pilot sequences to estimate the channel from “scratch” does not give robustness against

in NLOS. The number of multipaths that are summed up constructively and destructively creates a very high angle variation in receiver side which reduces the channel sparsity representation. Because of it, we need to user more pilots. To avoid it the UE has track the angle variations as shown in the next section.

4.6.5 Channel Tracking

Figure 4.13 shows the performance only of the beam tracking algorithm, i.e there are only small blocks. In the beginning of the transmission $(t = 0)$, we consider perfect knowledge of channel. The beam tracking algorithm is ran at the NLOS scenario shown on Fig. 4.11. At the first part (before 1.3 s), the proposed algorithm provides a system throughput similar to the benchmark one. The performance is the same as in channel estimation case, but the main advantage of tracking is the low algorithm complexity and the low overhead demanded. This last is because the method estimates the channel using information estimated in a moment before. On the other hand, this method does not deal with the death and born of rays [57] which may lead to a situation where the UE is served.

Although the proposed beam tracking approach can not handle with the death of rays, the green curve shows an increasing of the throughput. A possible explanation is that the local optimum point has become weak and the algorithm starts to look for a maximum and by chance the algorithm found one. However, it is not possible to guarantee that.

Since the beam correction is not capable to handle with the death of rays, the combination between channel estimation and beam tracking becomes suitable. From the point of view of beam correction, channel estimation algorithm can re-estimate the channel making possible the tracking algorithm to adjust the beam direction. From the point of view of channel estimation, beam tracking makes possible to set a higher time interval between blocks of estimation. This is shown on Fig. 4.14 where two configurations are tested. One (with red circles) has large blocks with 50 OFDM symbols at every 0.1s and smaller blocks with 20 symbols for tracking at every 0.02s. The other configuration (with green circles) has large blocks with 50 OFDM symbols at every 0.2s and smaller blocks with 20 symbols for tracking at 0.01s. Because the high channel variation in the transition from the LOS to NLOS scenario, the configuration with smaller blocks at every 0.02s lost the tracking, but the system could recover it after a large block.

4.7 Conclusion

Using the proposed method, the AN is possible to accurately acquire the angular information of the channel and use it to adjust a beam to serve the UE. We evaluate this technique in different situations and analysed the trade-off among overhead, power and coverage limitations. This solution enables the AN and UE to save power by increasing the overhead, or to reduce the overhead by increasing the power, or to adjust the AN coverage whenever is possible. The

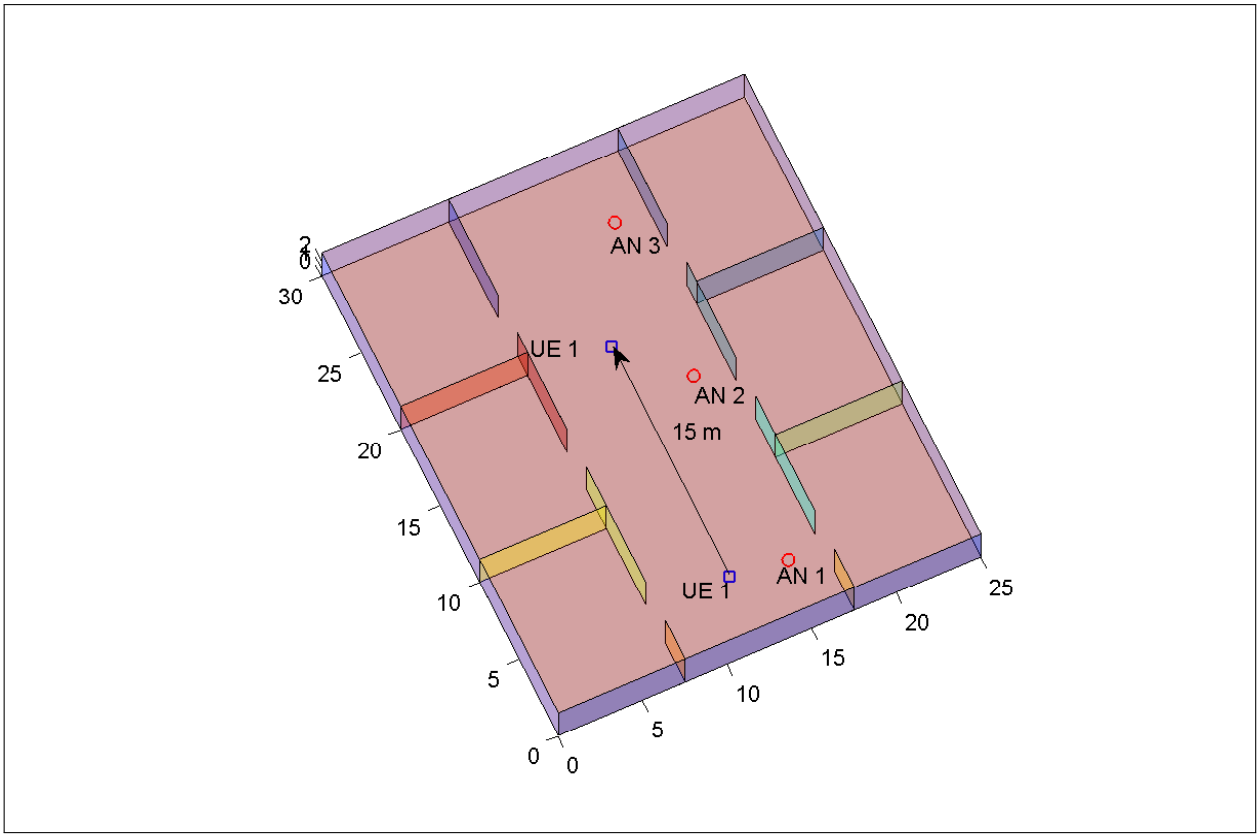
method can be used in practical situations of handover between ANs, since it provides accurate estimation of the arrival and departure directions of multiple LOS components from different ANs.

An important aspect investigated in this chapter is the hardware limitations. The simulations show that eigenbeamforming (frequency dependent) and the steering vector (Non-frequency dependent) has similar performances in this scenario. This is because the mm-wave channel has low angular spread and consequently it remains almost flat even though the channel has a 2GHz bandwidth. Another aspect considered is the phase-shifter resolution. The round-phase beamforming has lower resolution in its phase-shifters, the loss in terms of throughput is small compared with the continuous steering vector phase.

Regarding to the channel tracking solution, we can conclude that it is complementary solution of the channel estimation. It gives the benefit of farther reduce the overhead because it avoids to send larger blocks to re-estimate the channel.

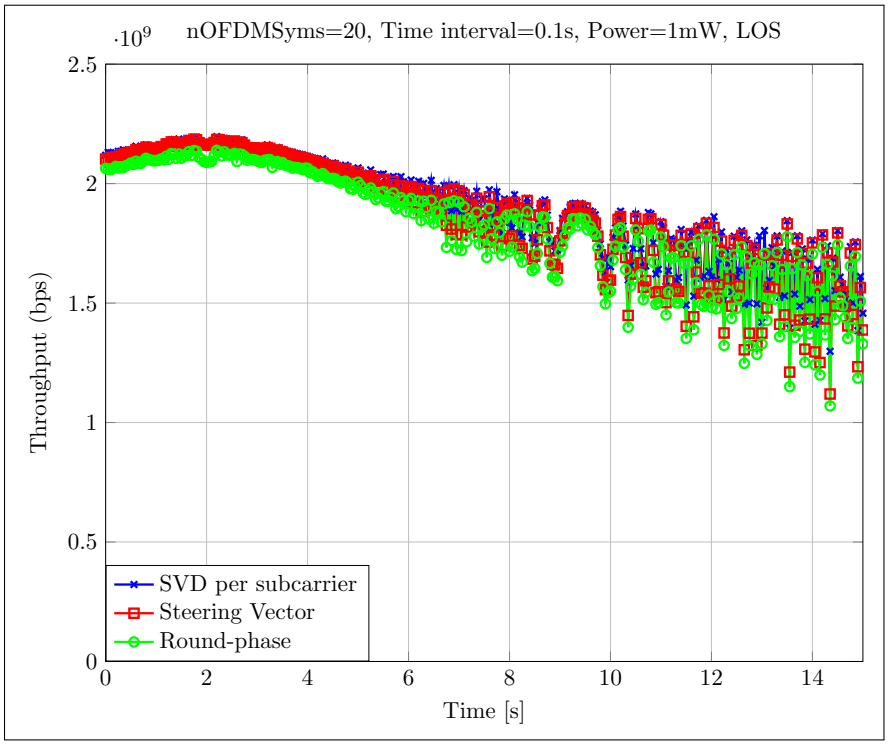
4.8 Figures

Figura 4.3 – UE is going trough the corridor.



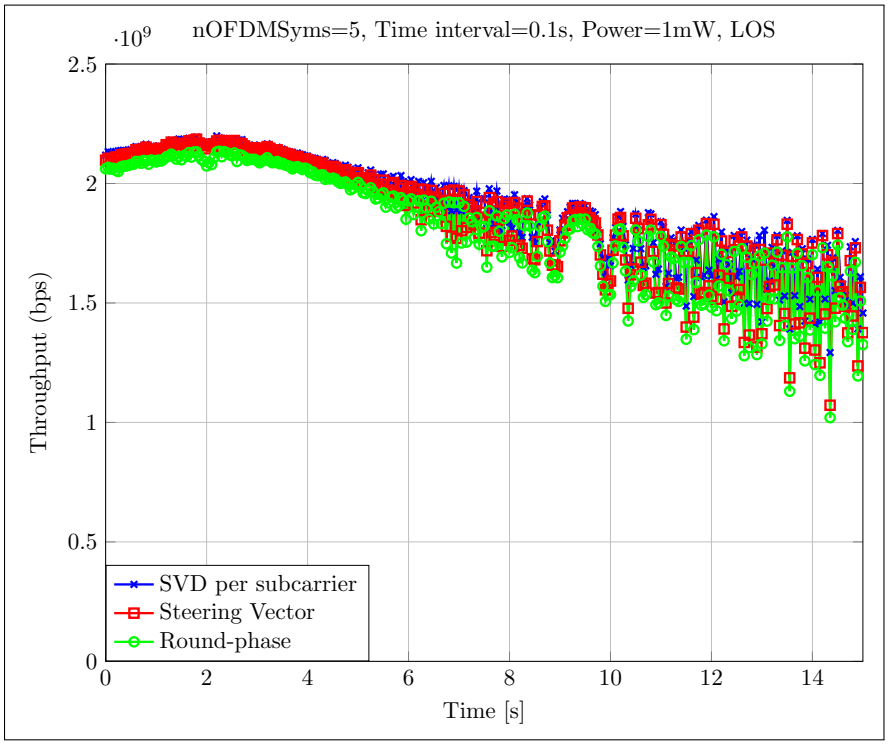
Source: Provided by the author.

Figura 4.4 – System throughput considering three different type of beamforming: SVD per subcarrier, steering vector and round-phase vector. The time interval between two consecutive large blocks is 0.1 s amd the number of OFDM symbols is 20.



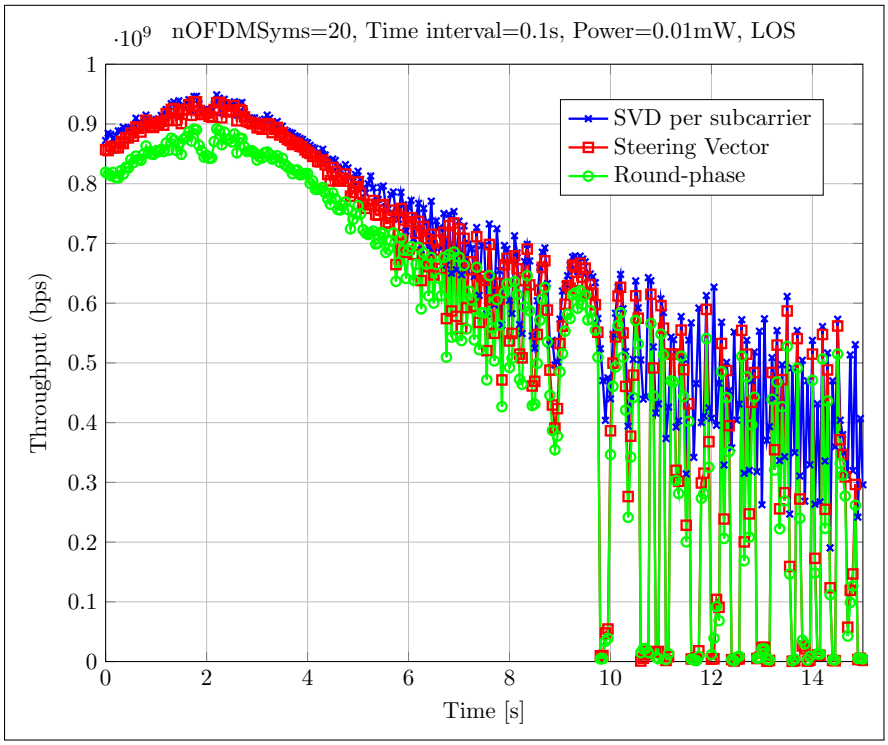
Source: Provided by the author.

Figura 4.5 – System throughput considering three different type of beamforming: SVD per subcarrier, steering vector and round-phase vector.



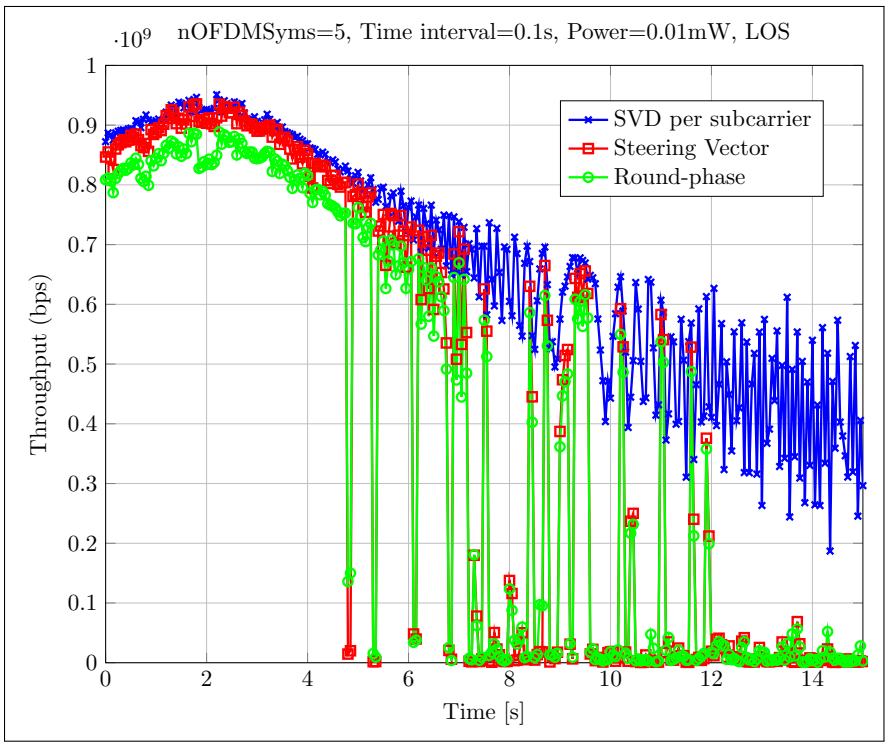
Source: Provided by the author.

Figura 4.6 – System throughput considering three different type of beamforming: SVD per subcarrier, steering vector and round-phase vector.



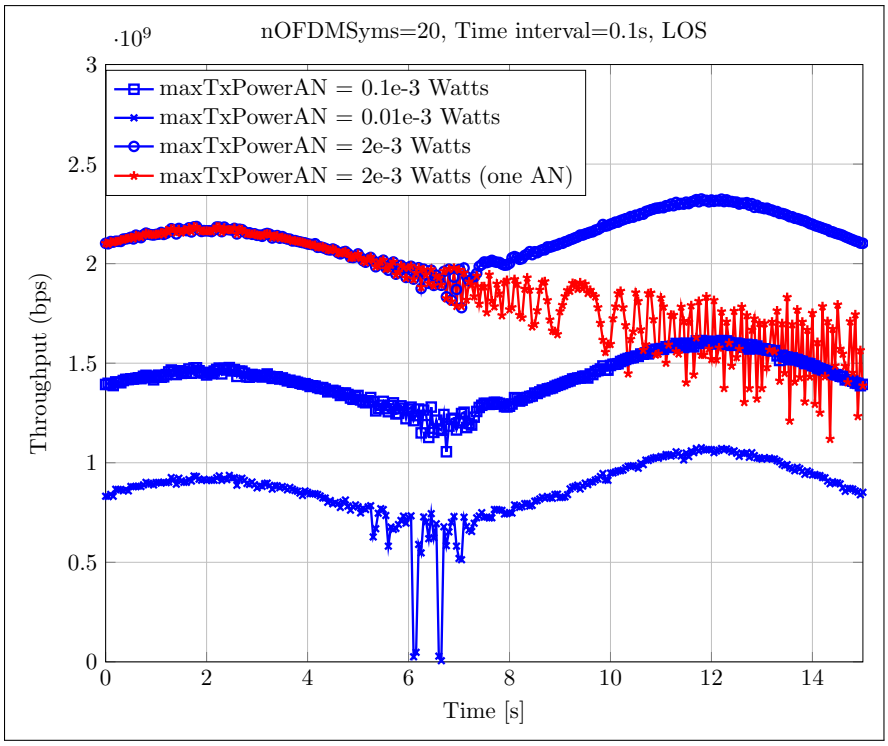
Source: Provided by the author.

Figura 4.7 – System throughput considering three different type of beamforming: SVD per subcarrier, steering vector and round-phase vector.



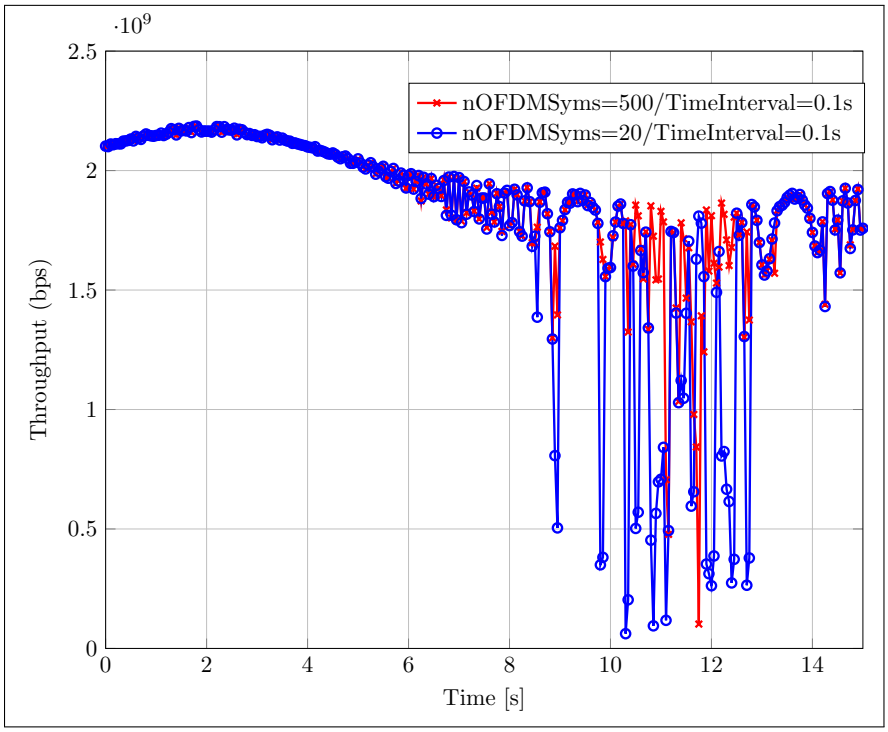
Source: Provided by the author.

Figura 4.8 – Channel estimation considering handover between AN 1 and AN 2 deployed in the corridor.



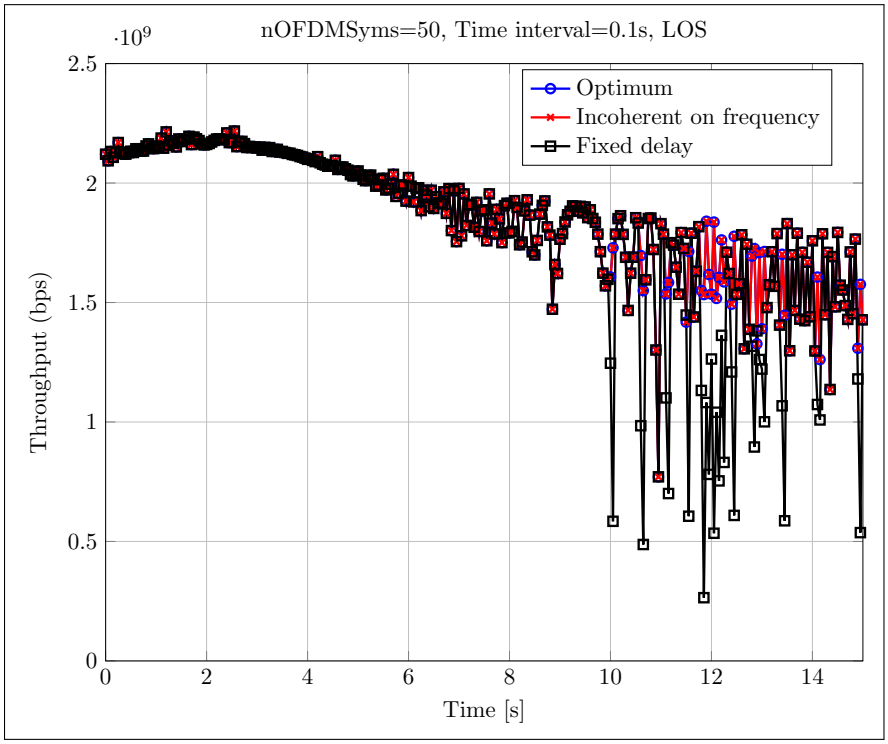
Source: Provided by the author.

Figura 4.9 – Channel estimation considering handover between AN 1 and AN 3 deployed in the corridor.



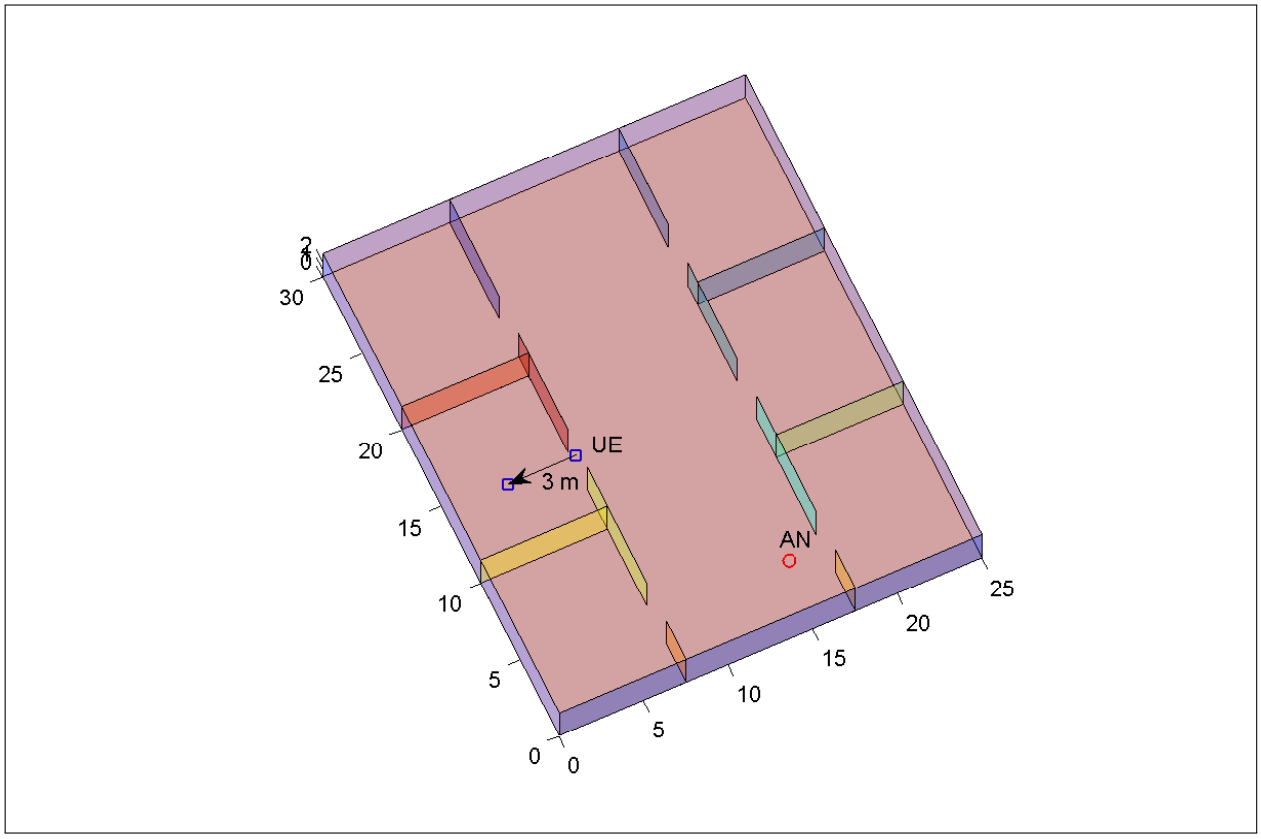
Source: Provided by the author.

Figura 4.10 – System throughput for power fading ratio $\rho = 0$ and for three functions f's. Fixed delay is defined in Eq.(4.11), incoherent in frequency is defined in Eq. (4.12) and optimum is defined in Eq. (4.13).



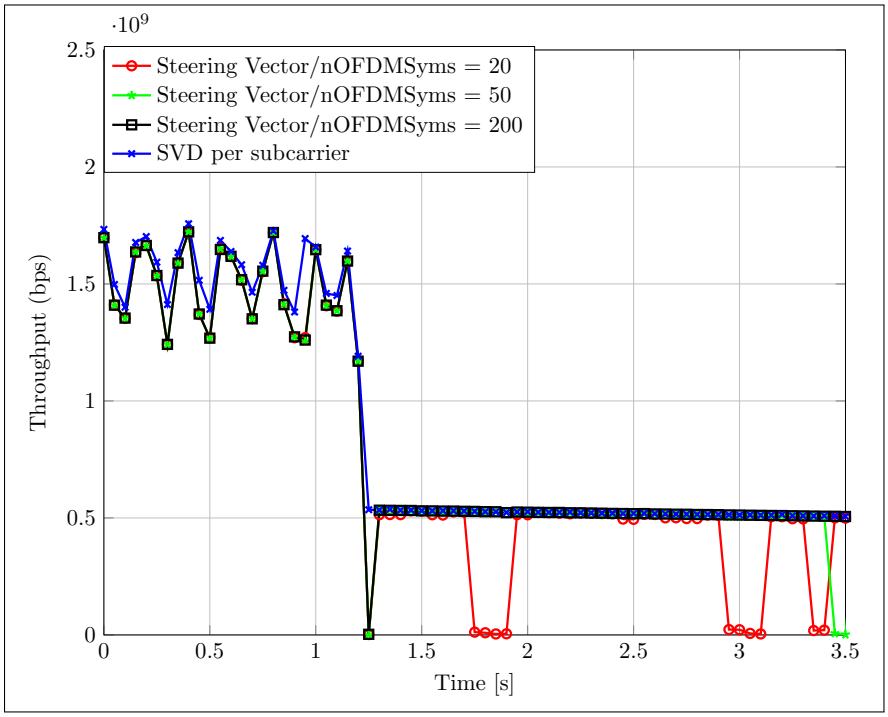
Source: Provided by the author.

Figura 4.11 – UE is going into the room.



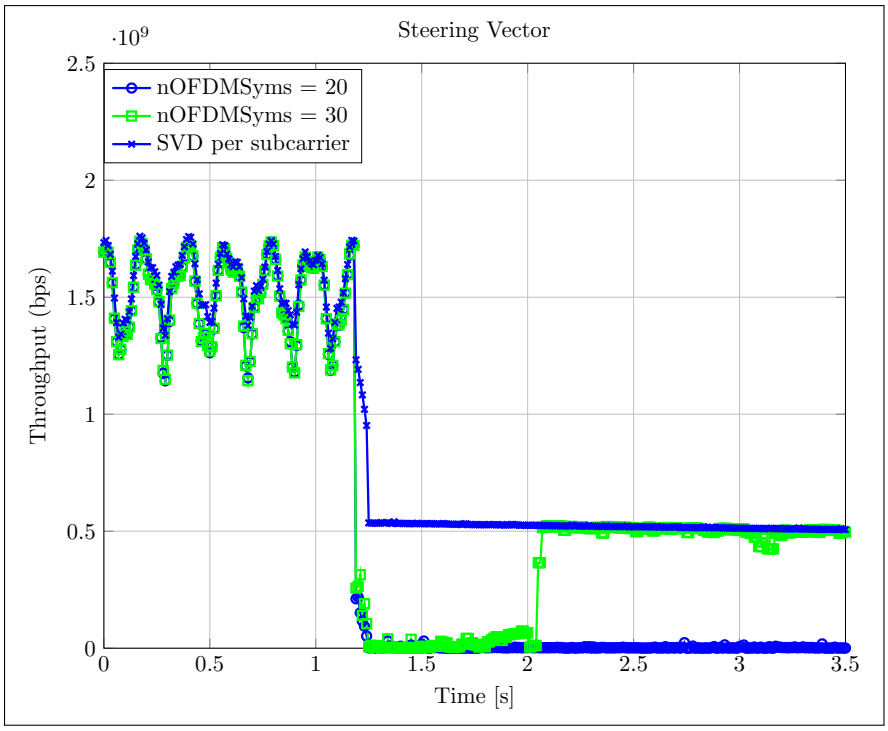
Source: Provided by the author.

Figura 4.12 – System throughput considering three different amount of pilots: 20, 50 and 200 OFDM symbols.



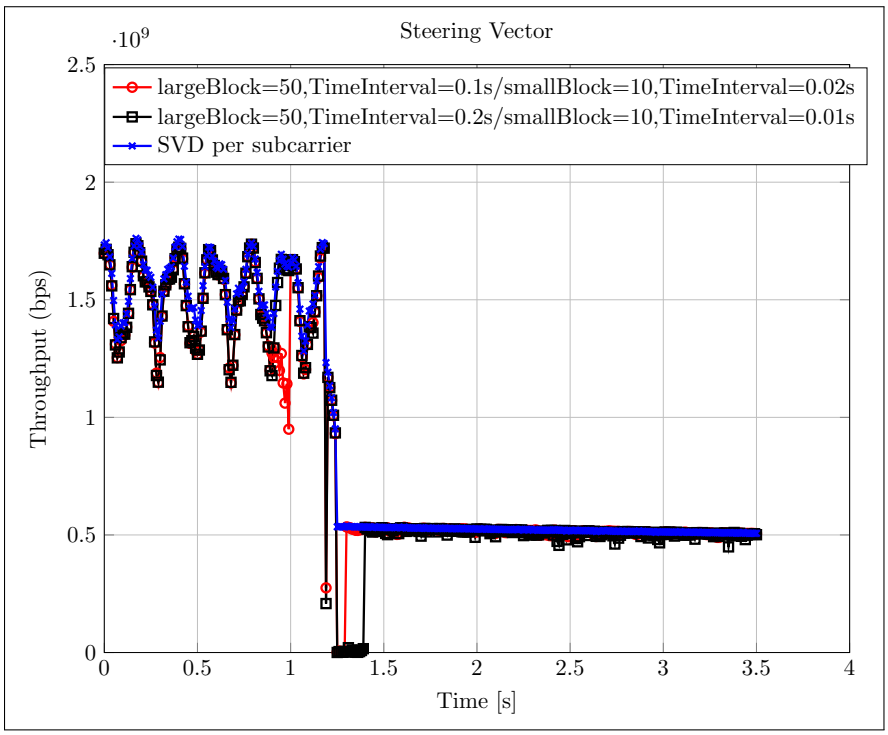
Source: Provided by the author.

Figura 4.13 – System throughput at NLOS considering two different amount of pilots: 20, 30 OFDM symbols.



Source: Provided by the author.

Figura 4.14 – Channel estimation and channel tracking evaluation.



Source: Provided by the author.

5 STATISTICAL HYBRID BEAMFORMING FOR MASSIVE MIMO CHANNEL ESTIMATION

Abstract

In this chapter, we address the channel estimation problem in massive MIMO systems by adopting a hybrid (analog-digital) beamforming architecture that exploits second-order statistics. Differently from the strategies presented on previous chapters, the channel is represented in terms of its low-dimensional subspace extracted from the spatial covariance matrix. The proposed technique consists of spatially multiplexing the pilots using a beamforming scheme. In the first moment, we consider a downlink transmission in time-frequency coherent channel blocks, containing both the pilots and data. Considering FDD and taking the pilot overhead into account, we show that hybrid beamforming provides better spectral efficiency performance than full digital beamforming in cases with many antenna elements. We also optimize the number of used radio chains for hybrid beamforming in FDD and show it is possible to achieve comparable spectral efficiency as with full digital beamforming in TDD. In the second moment, we focus on designing the optimum HB and the channel estimator according to the MMSE criterion. In this problem, we derive a new HB beamforming architecture and provide conditions to determine the optimum number of RF chains and pilot power allocation according to that criterion.

5.1 Introduction

In mobile communication systems, the channel acquisition depends on the duplexing mode, which can be either FDD or TDD. In FDD, the BS sends downlink pilots and each UE estimates the CSI. This information is then fed back to the BS so that one can design a beamforming solution. However, both channel acquisition and report demand more overhead as the number of BS antenna elements becomes large. In TDD, the UEs send uplink pilots and the BS estimates the uplink channel. Assuming channel reciprocity, the downlink CSI is obtained from a hermitian version of the uplink channel. In this case, the overhead increases proportionally to the number of total UE antennas [68]. This is the main reason most massive MIMO related works assume operation under TDD. However, cellular networks in large parts of the world operate in paired spectrum and this naturally creates a demand for solutions under FDD that enjoy the enhanced spectral efficiency provided by massive MIMO.

Beamforming techniques are effective solutions to improve SNR and consequently the system spectral efficiency. Full digital beamforming is the natural solution adopted in massive MIMO systems because it can achieve the maximum degree of freedom for signal processing. In [69], the authors propose a two-stage digital beamforming by concatenating two digital beamformers, where the first depends on the channel statistics and the second is designed from the instantaneous knowledge of the channel. Although such a structure reduces the overhead,

scaling up a complete digital solution as a function of the number of antennas becomes prohibitive. Its implementation requires separate RF chains consisting of mixers, power amplifiers and a digital to analog converter (DAC) per antenna element. This architecture results in high power consumption, cooling demands, and cost, mainly due to the high number of DACs [70–72].

Another solution is to use the so-called hybrid analog-digital beamforming. In this hybrid structure, the array has a power amplifier per antenna element, but the signals are steered by a large-scale phase shifter network and a much smaller set of baseband processing chains [73]. This structure brings the possibility to upgrade the BS by keeping the current baseband hardware and exchanging the existing fixed-pattern antenna array with a large-scale phased array. Most of the HB literature targets mm-wave frequencies. In [74], HB is designed based on perfect CSI for supporting point-to-point MIMO. The analog part is composed by the steering vectors of the most relevant multipaths and the digital part is obtained from the SVD of the equivalent baseband channel. In [46], the analog part is designed based on the angle of arrival (AoA) obtained perfectly by uplink measurements and the digital part is obtained from the maximization of the system rate. However, mm-wave channels have very poor scattering characteristics, whereas in current (< 6 GHz) cellular systems, the channels have typically many relevant paths due to the low reflection attenuation. In [72], cellular frequencies are considered and the analog beamforming is designed as the phase-only version of a frequency-flat channel. However, in cellular communication systems the channels are frequency-selective, the analog and digital parts have to be designed accordingly.

In this chapter, we propose a HB design that can operate in flat and frequency-selective channels. We use the spatial covariance matrix to design the wideband analog beamforming and the instantaneous channel to design the narrowband digital beamforming. Assuming a single-user scenario, we first study the optimization of the number of RF chains based on an SNR criterion. Then, we extend our analysis for a multilayer transmission and investigate the optimum channel estimator from a MMSE criterion. We also solve a convex problem to obtain the power allocation across the RF chains, which is then used to beamform the downlink pilots in FDD. Taking into account the overall overhead, we show that by optimizing the power allocation over the RF chains, the spectral efficiency of HB in FDD is similar to that achieved with digital beamforming in TDD.

5.2 FDD System Model

Consider a BS with N antenna elements transmitting an OFDM signal to a single-antenna UE using a FDD transmission. The channel is block-fading and defined by a coherence time of K OFDM symbols and coherence bandwidth of L subcarriers. This means that the channel is approximately constant in each block and changes only block-to-block[1] as shown in Fig. 5.1.

[1] For simplicity, only blocks in time dimension are shown.

The communication resources inside every time-frequency block are allocated either for pilot or data transmission. The training sequences occupy K_τ OFDM symbols and the remaining $K - K_\tau$ symbols are allocated to data transmission. The pilots are used by the the UE to estimate the channel. The estimates are reported from the UE to the BS using K_f symbols in the UL channel.

Given this framework, it is possible to define the received signal associated with the jth channel in the ith coherence block as

$$y_i[j] = \mathbf{h}_i^T \mathbf{x}_i[j] + z_i[j], \;\; K_\tau L + 1 \leq j \leq KL, \tag{5.1}$$

where $y_i[j]$ is the received signal, $\mathbf{h}_i \in \mathbb{C}^N$ is the channel vector, $\mathbf{x}_i[j]$ is the transmitted signal, and $z_i[j] \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise.

In the first $K_\tau + K_\mathrm{f}$ OFDM symbols of the ith block, the BS does not have knowledge of $\mathbf{h}_i$ since the channel estimate has not been acquired yet. Thus, in the data transmission phase, we assume that the BS continues using the beamforming vector $\mathbf{w}_{i-1}$ designed in the previous block. The beamforming vector $\mathbf{w}_i$ is designed upon reception of $\mathbf{h}_i$, the channel estimate vector from the feedback. Therefore, the beamformed signal during the data transmission phase is

$$\mathbf{x}_i[j] = \begin{cases} \mathbf{w}_{i-1} s_i[j] & , K_\tau L + 1 \leq j \leq L(K_\mathrm{f} + K_\tau) \\ \mathbf{w}_i s_i[j] & , L(K_\mathrm{f} + K_\tau) + 1 \leq j \leq LK, \end{cases} \tag{5.2}$$

where $\|\mathbf{w}_i\|^2 = p$ and $s_i[j]$ is the unit-power data symbol.

Based on (5.2), the spectral efficiency for the ith block is

$$\begin{aligned} C_i^{\mathrm{FDD}} = & \frac{(T_s - T_{cp})}{T_s K} [K_\mathrm{f} \log_2(1 + \Gamma_i(\mathbf{w}_{i-1})) \\ & + (K - K_\mathrm{f} - K_\tau) \log_2(1 + \Gamma_i(\mathbf{w}_i))], \end{aligned} \tag{5.3}$$

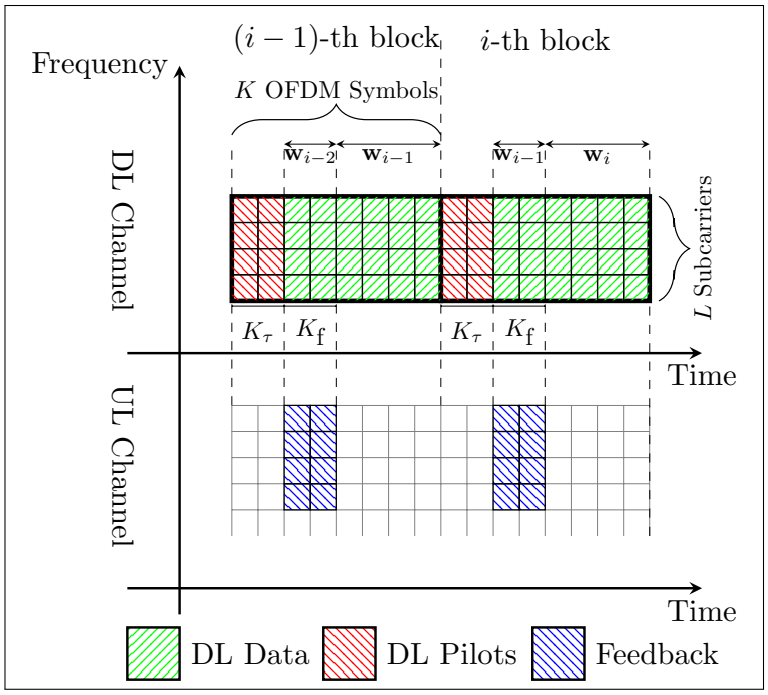
where T_s is the OFDM symbol duration, T_{cp} is the cyclic prefix length, and $\Gamma_i(\mathbf{w}_i) = |\mathbf{h}_i^T \mathbf{w}_i|^2 / \sigma^2$ is the received SNR of the signal beamformed by $\mathbf{w}_i$. Note that the length K_τ of the training phase reduces linearly the spectral efficiency. Moreover, we expect that $\Gamma_i(\mathbf{w}_{i-1}) < \Gamma_i(\mathbf{w}_i)$, due to the channel aging. Hence, the spectral efficiency decreases logarithmically with the length K_f of the feedback.

5.3 Hybrid Beamforming Design

In the first part of this chapter, the focus is to design a HB using the knowledge of the spatial channel covariance matrix. The HB is defined by the architecture implemented in the BS. It is considered that there is a phase-shifter network that connects N antennas with M RF chains. The concatenation of the analog and digital beamformers defines the overall beamforming which can be expressed as

$$\mathbf{w}_i = \mathbf{W}_{\mathrm{RF}} \mathbf{f}_i, \tag{5.4}$$

Figura 5.1 – Simplified representation of time-frequency blocks in FDD.



Source: Provided by the author.

where $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N \times M}$ is the wideband analog part with $|[\mathbf{W}_{\text{RF}}]_{n,m}| = 1$ and $\mathbf{f}_i \in \mathbb{C}^M$ is the block-specific digital part. Note that each element of $\mathbf{f}_i$ rotates and amplifies the beam defined by the corresponding column of $\mathbf{W}_{\text{RF}}$.

5.3.1 Analog Beamforming

Let $\mathbf{R} = \mathrm{E}\left\{\mathbf{h}_i \mathbf{h}_i^H\right\}$ be the spatial channel covariance matrix, where the expectation operator $\mathrm{E}\{.\}$ is over both time and frequency. Considering the SVD of $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, the first $M^\star \leq M$ eigenvectors $\mathbf{U}_M$ corresponding to the strongest eigenvalues of $\mathbf{\Lambda}$ are used to design the analog beamforming matrix. The choice of $M^\star$ will be discussed later, where it will be shown that the spectral efficiency is improved by properly adjusting this parameter. Because the phase-shifter network only controls the signal amplitude, the analog beamforming is represented as

$$\mathbf{W}_{\text{RF}} = \tilde{\mathbf{U}}_{M^\star}. \tag{5.5}$$

where $\tilde{\mathbf{U}}_{M^\star}$ is defined by the phase-only version of $\mathbf{U}_{M^\star}$. Note that the $\mathbf{W}_{\text{RF}}$, representing the (analog) phase-shifter network, does not depend on the channel frequency response. In other words, $\mathbf{W}_{\text{RF}}$ represents the wideband beamforming part, shared by the whole range of frequencies.

In FDD, the estimation of $\mathbf{R}$ can be performed by the UE using a beam sweeping inside each time-frequency coherence block. The covariance matrix of the received signal is computed by averaging out samples over such blocks. The spatial covariance matrix is then obtained

using a LS estimator. Because the second order statistics do not vary fast, this estimation can be performed with low time periodicity, thus not demanding substantial overhead. The practical effect is that analog beamforming matrix does not need to be updated at every time block.

5.3.2 Number of RF Chains

Considering the channel decomposition $\mathbf{h}_i = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{v}_i$ for $\mathbf{v}_i \in \mathbb{C}^N$ and using the HB definition in (5.4), the SNR can be expressed as

$$\begin{aligned}\Gamma_i(\mathbf{w}_i) &= |\mathbf{w}_i^H \mathbf{h}_i|/\sigma^2 \\ &= |\mathbf{f}_i^H \mathbf{W}_{\text{RF}}^H \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{v}_i|^2/\sigma^2 \\ &\approx |\sum_{m=1}^{M^\star} \lambda_m^{1/2} f_{m,i} v_{m,i}|^2/\sigma^2, \end{aligned} \tag{5.6}$$

where $f_{m,i}$ and $v_{m,i}$ are the mth entries of $\mathbf{f}_i$ and $\mathbf{v}_i$, respectively, and the approximation is because $\mathbf{W}_{\text{RF}} \approx \mathbf{U}_M$. The eigenvectors tend asymptotically to have constant modulus because they tend to a DFT matrix [36] as the number of antennas increases. Using such a result and defining $\|\mathbf{W}_{\text{RF}}\mathbf{f}_i\|_2^2 = p$, it is possible to write $\|\mathbf{f}_i\|_2^2 \approx p/N$. The terms of the summation in (5.6) are weighted by the eigenvalues, and the BS chooses the $M^\star$ strongest eigenvalues. If $M^\star < M$, the BS sets $f_{m,i} = 0 \; \forall m \; M^\star < m \leq M$. Note that in (5.6), $v_{m,i}$ defines the dependency on the small-scale fading. However, this information is not known *a priori*, it is acquired by the BS after the UE estimates and reports $\mathbf{v}_i$ back through feedback channel.

The dimension of $\mathbf{v}_i$ is defined by the BS. Using the eigevalues of $\mathbf{R}$, the BS establishes the relationship between the number of RF chains and the number of relevant eigenmodes using the proposed heuristic convex optimization

$$\begin{aligned}\max_{\mathbf{d}} \; & \sum_{m=1}^{M} \lambda_m^{1/2} d_m \\ \text{s.t } & \|\mathbf{d}\|_2^2 \leq p/N, \end{aligned} \tag{5.7}$$

where $\mathbf{d} \in \mathbb{R}^N$ selects the eigenmodes to be used in the transmission. The parameter $M^\star$ is defined from the solution $\mathbf{d}^\star$ as $M^\star = \|\mathbf{d}^\star\|_0$. Although the BS must spend extra overhead to acquire $\mathbf{R}$, the estimation happens after long periods. Clearly, $M^\star$ and $\mathbf{W}_{\text{RF}}$ depend on $\mathbf{R}$ meaning that the periodicity to solve (5.7) is similar to estimating $\mathbf{R}$. Note that the eigenvalues can be used to define the environment richness and $M^\star$ determines the equivalent channel dimension. Since the system spectral efficiency is affected by the overhead that is proportional to the equivalent channel dimension, the proposed method provides a way to adapt the system spectral efficiency by exploiting the environment richness obtained from the second order statistics.

5.3.3 Digital Beamforming

Due to the analog beamforming, the effective channel in the ith block, as evidenced in the baseband is

$$\mathbf{g}_i = \mathbf{W}_{\text{RF}}^T \mathbf{h}_i. \tag{5.8}$$

Since $\mathbf{g}_i \in \mathbb{C}^{M^\star}$ and $M^\star \ll N$, far fewer pilots are required for estimating $\mathbf{g}_i$ instead of $\mathbf{h}_i$. Specifically, only $M^\star$ orthogonal pilot sequences are used, as defined by the matrix $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_{K_\tau L}] \in \mathbb{C}^{M^\star \times K_\tau L}$, and they are beamformed by $\mathbf{W}_{\text{RF}}^T$. The received training sequence $\mathbf{y}_i = [y_i[1], \ldots, y_i[K_\tau L]]^T$ at the UE is then given by

$$\mathbf{y}_i = \mathbf{S}^T \mathbf{D} \mathbf{W}_{\text{RF}}^T \mathbf{h}_i + \mathbf{z}_i, \tag{5.9}$$

where $\|\mathbf{s}_j^T \mathbf{D} \mathbf{W}_{\text{RF}}^T\|^2 = p$ and $\mathbf{z}_i = [z_i[1], \ldots, z_i[K_\tau L]]^T$. Note that $\mathbf{D} = \text{diag}[d_1, \ldots, d_{M^\star}]$ is weighting the beams according to the solution found in (5.7). The MMSE estimator $\hat{\mathbf{g}}_i = \text{E}\{\mathbf{g}_i | \mathbf{y}_i\}$ is calculated as [75]

$$\hat{\mathbf{g}}_i = \tilde{\mathbf{R}} \tilde{\mathbf{S}}^* \left(\sigma^2 \mathbf{I}_\tau + \tilde{\mathbf{S}}^T \tilde{\mathbf{R}} \tilde{\mathbf{S}}^* \right)^{-1} \mathbf{y}_i, \tag{5.10}$$

where $\tilde{\mathbf{R}} = \mathbf{W}_{\text{RF}}^T \mathbf{R} \mathbf{W}_{\text{RF}}$ and $\tilde{\mathbf{S}} = \mathbf{D}\mathbf{S}$. The estimated effective channel is then quantized and fed back to the BS, which can calculates the digital part of the beamforming, defined in (5.4), according to the maximum ratio transmission (MRT) criterion as $\mathbf{f}_i = \hat{\mathbf{g}}_i^*$.

5.3.4 CSI Acquisition Overhead

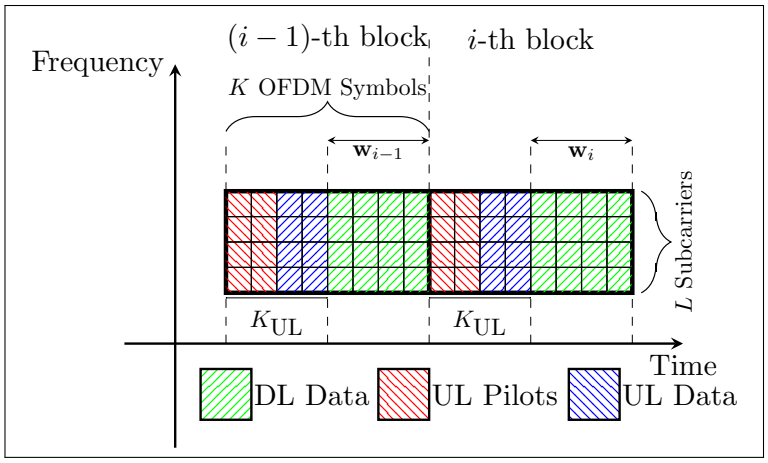
The system overhead comprises the total number of OFDM symbols allocated to pilot transmission and feedback, which are K_τ and K_{f} respectively, as shown in Fig. 5.1. The feedback load increases proportionally to the dimension of $\hat{\mathbf{g}}_i$, and is given by

$$K_{\text{f}} = \left\lceil \frac{b M^\star}{\log_2 (N_{\text{order}}) L_{\text{f}}} \right\rceil, \tag{5.11}$$

where b refers to the number of bits employed to quantize one channel vector entry, N_{order} is the modulation order used to transmit data on the feedback channel, L_{f} is the number of subcarriers that composes the feedback channel. As established in (5.2), the BS spends K_{f} time instants using $\mathbf{w}_{i-1}$. Also, as discussed after (5.3), it is of interest to keep K_{f} as low as possible to minimize the downlink spectral efficiency loss due to channel aging.

With HB, the BS has the flexibility to control the system overhead by adapting the use of the RF chains according to the UE's requirements and channel conditions. This is important for cell-edge UEs with poor SNR for channel estimation. Note that for multi-antenna UEs, supporting multi-layer transmission, would require more RF chains.

Figura 5.2 – Simplified representation of the time-frequency frame in TDD.



Source: Provided by the author.

5.4 Spectral Efficiency in TDD

In TDD, unlike FDD, the coherence block is split into uplink and downlink parts that share the same bandwidth and occupy complementary time intervals, as shown in Fig. 5.2. Channel estimation is performed by the BS from the uplink training sequences sent by the UE. The overall overhead scales linearly as the number of antennas at the UE [76].

Similar to (5.3), the downlink spectral efficiency is defined in a TDD system as

$$C_i^{\text{TDD}} = \frac{(T_s - T_{cp})(K - K_{\text{UL}})}{T_s K} \log_2(1 + \Gamma_i(\mathbf{w}_i)) \tag{5.12}$$

where K_{UL} is the number of OFDM symbols in the uplink transmission, containing both pilot and potentially uplink data [68]. In TDD, the downlink spectral efficiency is essentially determined by the uplink-downlink ratio. Assuming full digital beamforming according to the MRT criterion, i.e., $\mathbf{w}_i = \hat{\mathbf{h}}_i^*$, the entire propagation channel needs to be estimated, e.g., by considering the MMSE estimator [75]

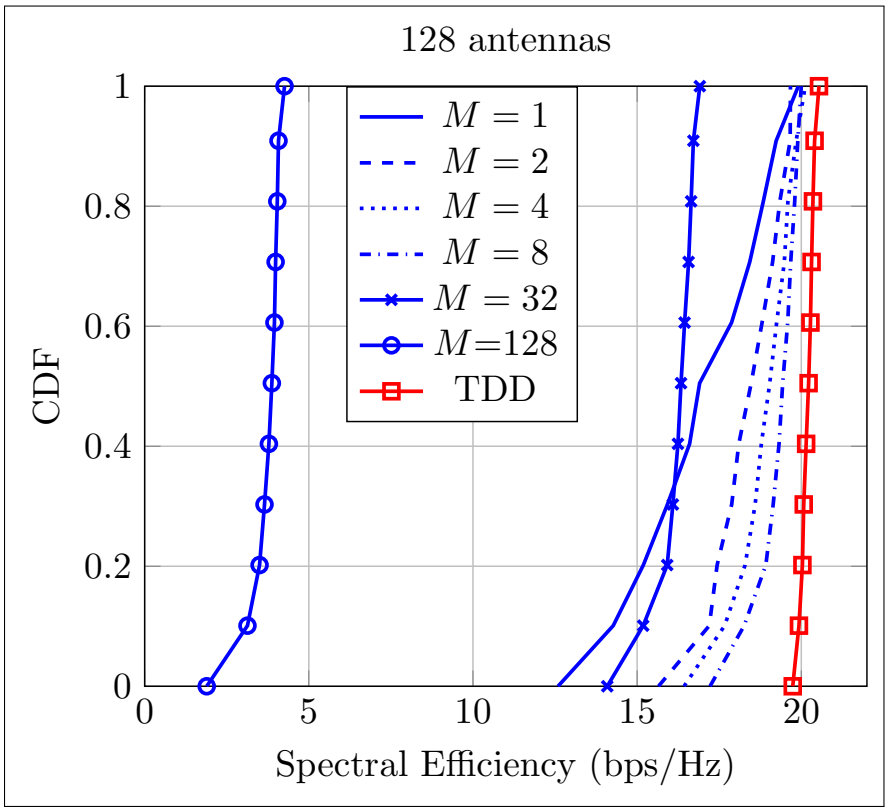
$$\hat{\mathbf{h}}_i = \mathbf{h}_i + \mathbf{C}^{1/2}\mathbf{n}, \tag{5.13}$$

where $\mathbf{C} = \mathbf{R}\left(\mathbf{I}_N + \frac{pK_\tau L}{\sigma^2}\mathbf{R}\right)^{-1}$ is the error covariance matrix and $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}_N)$ is the estimation error.

5.5 Comparison Between FDD and TDD Schemes

In the simulations, TDD and FDD spectral efficiencies are compared using the COST 2100 channel model extension for massive MIMO proposed by the EU project MAMMOET in [77]. A 10 MHz bandwidth channel operating at 2.6 GHz is considered in this comparison. Consider also $K = 14$ OFDM symbols and $L = 12$ subcarriers with subcarrier spacing $B_c = 78.12$ KHz. The feedback overhead K_{f} is calculated assuming $b = 4$, $N_{\text{order}} = 4$, and $L_{\text{f}} = 12$. For this setup, we generated 500 consecutive coherent blocks, 10 in the frequency-domain and 50

Figura 5.3 – Spectral efficiency distribution comparison between full digital in TDD and HB with EP in FDD for various numbers of RF chains



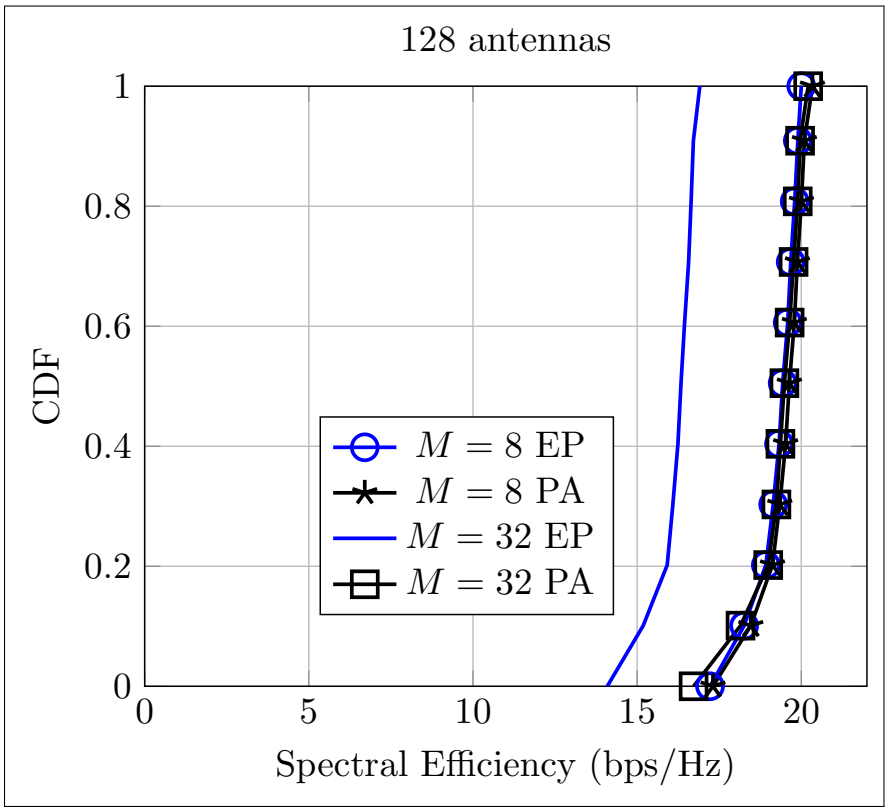
Source: Provided by the author.

in the time-domain. In our comparisons between FDD and TDD, the total transmit power is set 20 dBm for both BS and UE. The plots refer to EP for equal power and PA to power allocation according to (5.7).

In Fig. 5.3, FDD and TDD spectral efficiencies are obtained according to equations (5.3) and (5.12), respectively, for $N = 128$ antennas. The comparison shows a HB scheme with $M = \{1, 2, 4, 8, 32, 128\}$ RF chains transmitting in FDD and a digital beamforming transmitting in TDD. We assume the most favorable uplink-downlink ratio in TDD; namely, that $K_{\text{UL}} = 1$, essentially providing an upper bound on the achievable spectral efficiency. In FDD, as the number of RF chains increases up to $M = 8$, the spectral efficiency increases since the BS is choosing more eigenmodes and the overhead is not so significant. However, the gains of the large-array are "vanished" by using more RF chains as can be seen for $M = 32$. Comparing both duplexing schemes, it can be concluded that the "gap" between them can be reduced, when the number of RF chains is chosen so that the BS can efficiently exploit the channel modes without increasing substantially the overhead.

In Fig. 5.4, we evaluate the gains due to the optimization in (5.7) for $M = \{8, 32\}$. With $M = 32$, only the eight first eigenvectors are activated, meaning that the BS can approximately maximize the SNR using much less RF chains than the value used as input. For $M = 8$, the

Figura 5.4 – This plot compares $M = \{8, 32\}$ RF chains with equal power (EP) and power allocation (PA) accordingly with (5.7).



Source: Provided by the author.

result is essentially the same as for equal power. In such a case, the large-array provides high SNR gains enforcing equal power allocation among the RF chains. Applying the optimization in (5.7), the number of RF chains converges to the same value in both cases, $M = \{8, 32\}$. They have quite similar curves showing that $M^\star = 8$ is a good choice for this scenario.

The plots give an insight that the BS can use the statistical representation to reduce the dimension of the channel and so the pilot overhead. However, the analysis presented until this moment does not show the relationship between the channel error and the RF chains. As we show in the next section, where a multiple-antenna UE is considered, several RF chains are connected to the analog beamforming filter. Moreover, the angular spread also plays an important role in the determination of the number of RF chains to be used in the transmission.

5.6 System Model with Multiple-Antenna UE

Consider a BS and a single UE employing N and $M \ll N$ antennas, respectively. Let us assume an OFDM transmission. The frequency response is represented by a set of block-fading channels $\mathbf{H}(k) \in \mathbb{R}^{M \times N}\ \forall k\ k = \{1, 2, \ldots, K\}$. Each one has a coherence time and bandwidth of T OFDM symbols and F subcarriers, respectively. Within a block, there are TF resource elements and they are split either into pilot or data transmission. We define I and D as

the number of resources for pilot and data, respectively. It is important to highlight that, each resource element of a given block k is represented by the same channel matrix $\mathbf{H}(k) \in \mathbb{R}^{M \times N}$ with a spatial correlation matrix given by $\mathrm{E}\left\{\mathbf{H}(k)^H \mathbf{H}(k)\right\} = \mathbf{R}$, $\forall k$.

We assume a MIMO point-to-point system where the BS transmits multiple streams to a single UE using a HB solution. As discussed before, there are two stages to calculate the HB. The first one is the phase-shifter network defined by the matrix $\mathbf{A} \in \mathbb{C}^{N \times L}$ whose entries have constant modulus; the second stage is the digital part defined by the matrix $\mathbf{D}(k) \in \mathbb{C}^{L \times J}$. The phase-shifters do not depend on the frequency index k, and $\mathbf{D}(k)$ precodes the data streams into J layers associated with the frequency k.

The mathematical definition of HB is commonly represented as $\mathbf{W}(k) = \mathbf{A}\mathbf{D}(k)$ [78]. However, the constant modulus limitation in $\mathbf{A}$ causes a performance loss compared with that achieved by digital beamforming (DB). When the number of streams increases, the gap between the HB and DB performances tends to grow as higher as the angular spread. One way of compensating for this loss is increasing the number of RF chains [79]. However, such a solution employs more DACs. This component consumes considerable amount of power as shown in [80] and the system can reach prohibited power consumption levels, if a high number of RF chains are used. Therefore, it is important to devise HB solutions that can operate using the same number of RF chains as the number of layers, i.e. when $L = J$.

Here, we propose to exploit the factorization $\mathbf{D}(k) = \mathbf{B}\mathbf{C}(k)$, where the rows of $\mathbf{B}$ combine the columns of $\mathbf{A}$ to enable adjustments on the amplitude signal. The matrix $\mathbf{C}(k)$ is used to precode data or pilot in every coherence block. Figure 5.5 shows the commonly used HB and the proposed one architecture.

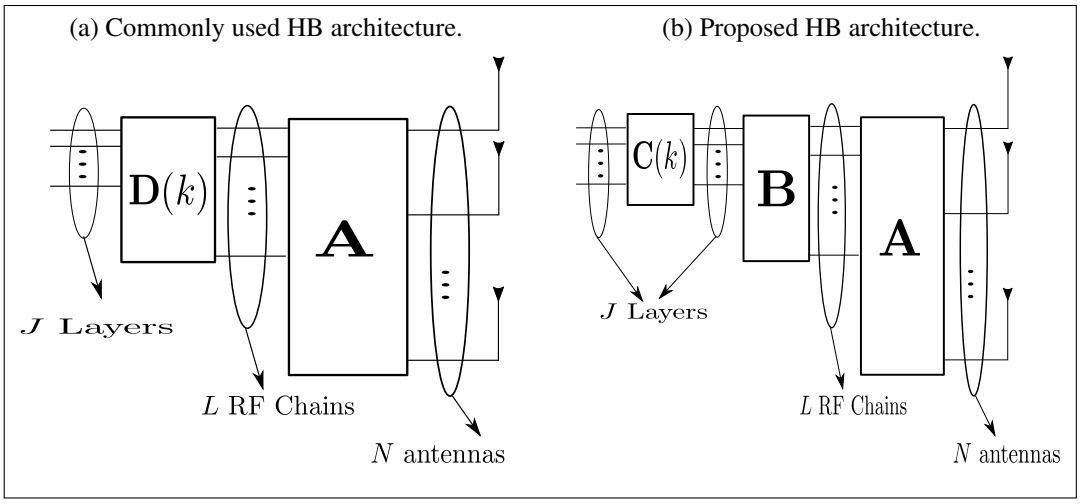
The received signal model writes

$$\begin{aligned} \mathbf{Y}(k) &= \mathbf{H}(k)\mathbf{W}(k)\mathbf{S}(k) + \mathbf{Z}(k), \\ &= \mathbf{H}(k)\mathbf{A}\mathbf{B}\mathbf{C}(k)\mathbf{S}(k) + \mathbf{Z}(k), \end{aligned} \tag{5.14}$$

where $\mathbf{S}(k) \in \mathbb{C}^{J \times D}$ collects J independent signals in D time-frequency elements, $\mathbf{Y}(k) = [\mathbf{y}_1(k), \ldots, \mathbf{y}_D(k)] \in \mathbb{C}^{M \times D}$, and $\mathbf{Z}(k) = [\mathbf{z}_1(k), \ldots, \mathbf{z}_D(k)] \in \mathbb{C}^{M \times D}$ whose covariance matrix across the time-frequency elements is defined as $\mathbf{R}_{\mathbf{z}} = \mathrm{E}\left\{\mathbf{Z}^H(k)\mathbf{Z}(k)\right\} \in \mathbb{C}^{D \times D}$.

The curves in Section 5.5 show that HB can be used to compress the full massive MIMO channel into a baseband one. This reduces the overall overhead and, consequently, increases the system spectral efficiency. In the next section, we derive the optimum design for the proposed HB, the MMSE estimator, the optimum number of RF chains, and the pilot power allocation that lead to MMSE bound.

Figura 5.5 – Comparison between the standard HB and proposed HB representations.



Source: Provided by the author.

5.7 Joint Design of MMSE estimator and Beamforming

Consider a set of J_p pilot sequences where their length is defined by the variable I. We concatenate them into a matrix $\mathbf{P}(k) \in \mathbb{C}^{J_p \times I}$, meaning that each row of the matrix corresponds to one sequence with length I. The BS then multiplexes the sequences using a HB filter defined by $\mathbf{W}(k)$. This means that the transmitter separates the pilots into $J = J_p$ number of layers.

Let us assume that the BS has no knowledge of the CSI, but it knows the second order statistics of the channel. As we did in the single-antenna UE case, the BS can use this information to smartly design a HB to transmit the pilots. The difference is that the solution proposed hereafter deals with the constant modulus limitation of the analog beamforming. This limitation negatively affects the sum-rate as we will show in our results.

The knowledge of the spatial covariance matrix brings the possibility of designing a statistical HB to transmit the pilots. This is because the matrix $\mathbf{R}$ describes statistically a group of time-frequency coherent blocks, i.e. the statistics remain almost constant over many frequency channel blocks. Using this fact, we can then simplify the baseband matrix $\mathbf{C}(k)$ to be independent of index k, i.e. $\mathbf{C}(k) = \mathbf{C}$. With this assumption, we can now derive the set of beams using $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, and assuming that $\mathbf{R}$ is constant over the transmit bandwidth.

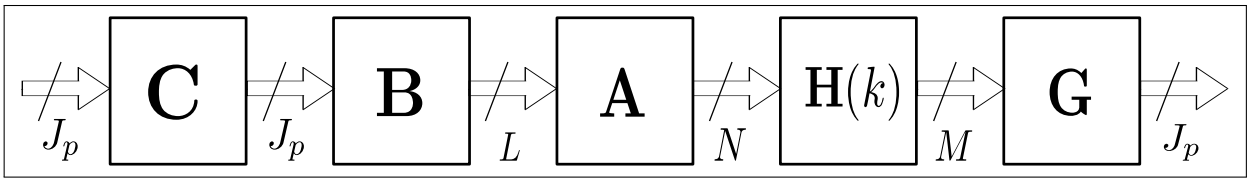
The BS creates the set of beams by multiplying $\mathbf{A}$ and $\mathbf{B}$. Both matrices are parts of the effective channel defined as

$$\mathbf{H}_{\text{e}}(k) \triangleq \mathbf{H}(k)\mathbf{A}\mathbf{B}, \tag{5.15}$$

where $\mathbf{H}_{\text{e}}(k) \in \mathbb{C}^{M \times J_p}$. Assuming that the number of sequences $J_p < N$, the UE does not estimate the full dimension channel $\mathbf{H}(k)$, but instead a reduced version of it.

It is important to highlight that using multiplexed pilots offers spectral efficiency gains over the well beam sweeping pilot scheme. The latter uses a sequence of OFDM symbols to

Figura 5.6 – System model block diagram.



Source: Provided by the author.

transmit one pilot sequence towards to a given direction. Thus, the total number of OFDM symbols to transmit all the J_p sequence is J_pI [60, 81], while the number of OFDM symbols using the multiplexed pilots corresponds to I. Such a gain comes to the cost of knowing the spatial covariance matrix in advance, which implies the implementation of a protocol to estimate $\mathbf{R}$. Nevertheless, beam sweeping only needs one RF chain, meaning that this solution needs a much simpler transmitter architecture.

The set of beams is obtained by designing $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. To optimize the number of pilot sequences and, consequently, reduce massive MIMO channel from $\mathbf{H}(k)$ to $\mathbf{H}_e(k)$, it is very import to design an estimator that uses fewer pilots and still obtains a very accurate CSI estimation. This is possible by designing the estimator $\mathbf{G} \in \mathbb{C}^{I \times J_p}$ according to the MMSE criterion. The estimated effective channel can be expressed as

$$\hat{\mathbf{H}}_{\mathrm{e}}(k) = \mathbf{Y}(k)\mathbf{G}. \tag{5.16}$$

Note that the received signal depends on the beamforming design, i.e. it depends on $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and J_p. In the next section, we use the MMSE criterion to assign the beamforming matrices and the number of sequences to minimize the channel estimator error. Fig. 5.6 represents the block diagram of the system model showing all the variables we must determine.

5.7.1 Channel Estimation

Let us define the error matrix associated with the estimator $\mathbf{G}$ as

$$\boldsymbol{\Delta}(k) \triangleq \mathbf{Y}(k)\mathbf{G} - \mathbf{H}_{\mathrm{e}}(k), \tag{5.17}$$

where $\boldsymbol{\Delta}(k) \in \mathbb{C}^{M \times J_p}$. The channel estimation error covariance matrix is then given by

$$\begin{aligned} \mathrm{E}\left\{\boldsymbol{\Delta}(k)^H \boldsymbol{\Delta}(k)\right\} = \; & (\mathbf{G}^H\mathbf{P}^H\mathbf{C}^H - \mathbf{I})\mathbf{R}_e(\mathbf{CPG} - \mathbf{I}) \\ & + \mathbf{G}^H\mathbf{R}_z\mathbf{G}, \end{aligned} \tag{5.18}$$

where $\mathbf{R}_{\mathrm{e}} = \mathbf{B}^H\mathbf{A}^H\mathbf{RAB}$. Using Eq. (5.18) and assuming the knowledge of the spatial covariance matrix, the solution for the MMSE estimator is given by

$$\mathbf{G}_{\mathrm{MMSE}} = (\mathbf{P}^H\mathbf{C}^H\mathbf{R}_{\mathrm{e}}\mathbf{CP} + \mathbf{R}_z)^{-1}\mathbf{P}^H\mathbf{C}^H\mathbf{R}_{\mathrm{e}}. \tag{5.19}$$

The procedure to obtain the estimator is shown in the Appendix **??**. Notice that the pilot matrix $\mathbf{P}$ now appears without the frequency index k. This is a choice in our system model, which implies using the same set of pilots over multiple the time-frequency blocks. With this assumption, we can conveniently remove the dependence of our channel estimator on the index k. In practice, this means that our estimator $\mathbf{G}_{\text{MMSE}}$ is a wideband filter that remains constant until covariance matrix be updated.

Substituting (5.19) into (5.18), the error covariance matrix becomes

$$\mathrm{E}\left\{\boldsymbol{\Delta}^H(k)\boldsymbol{\Delta}(k)\right\} = \mathbf{R}_{\mathrm{e}} - \mathbf{R}_{\mathrm{e}}\mathbf{C}\mathbf{P}\mathbf{Q}\mathbf{P}^H\mathbf{C}^H\mathbf{R}_{\mathrm{e}}, \tag{5.20}$$

where

$$\mathbf{Q} \triangleq (\mathbf{P}^H\mathbf{C}\mathbf{B}^H\mathbf{A}^H\mathbf{R}\mathbf{A}\mathbf{B}\mathbf{C}\mathbf{P} + \mathbf{R}_z)^{-1}. \tag{5.21}$$

The MMSE problem to calculate $\mathbf{A}$ and $\mathbf{B}$ is formulated as

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B}} \ & \mathrm{trace}\{\mathrm{E}\left\{\boldsymbol{\Delta}^H(k)\boldsymbol{\Delta}(k)\right\}\} \\ \text{s.t} \ & \mathrm{trace}\{\mathbf{A}\mathbf{B}\mathbf{C}\mathbf{C}^H\mathbf{B}^H\mathbf{A}^H\} \leq P, \end{aligned} \tag{5.22}$$

where P is the power allocated to a given subcarrier. It is important to highlight that the solution of (5.22) is conditioned to a given pilot sequence $\mathbf{P}$. Before discussing this solution, we enunciate our assumption about the design of the pilot sequences. The matrix $\mathbf{P}$ depends on the statistics of the noise as we enunciate in the next lemma.

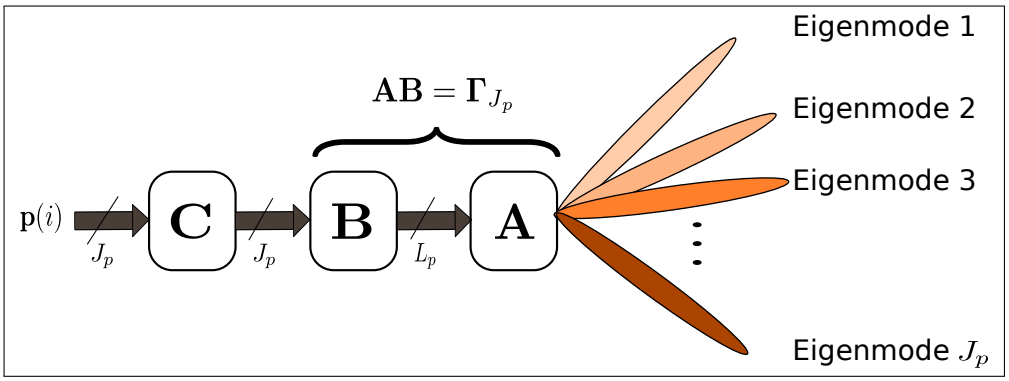
Lemma 1 *If the noise covariance matrix $\mathbf{R}_z$ is the identity, any orthogonal sequence leads to the MMSE bound. However, if the noise is correlated over the time, the optimum $J_p \times I$ pilot sequence matrix is given by $\mathbf{P} = \mathbf{U}_{\mathbf{z}}^H$, where $\mathbf{R}_z = \mathbf{U}_{\mathbf{z}}\boldsymbol{\Lambda}_{\mathbf{z}}\mathbf{U}_{\mathbf{z}}^H$.*

In some scenarios, the UE is strongly affected by interference sources coming from neighbour cells which generates a correlated noise, i.e. $\mathbf{R}_z$ cannot be described by an identity matrix. When this happens, Lemma 1 shows that the BS must exploit the eigen-decomposition of $\mathbf{R}_z$ to determine the set of pilots. This shows that the matrix $\mathbf{P}$ works as a filter to whiten the incoming interference the UE.

In practice, the UE needs to measure the second order statistics of the noise and report it back to the BS. This means that a closed-loop system must be employed so that the feedback channel provides such an information to the BS. In a open-loop system, the BS does not have $\mathbf{R}_z$ and the pilot sequences do not deal optimally with the interference. From Lemma 1, we can still conclude that the open-loop and closed-loop schemes have the same performance if and only if $\mathbf{R}_z$ is the identity. In such a case, any orthonormal $\mathbf{P}$ is optimal because it does not change the second order statistics of a white process. Hereafter, we assume a scenario without interference, but a closed-loop channel estimation scheme will be an object of a future investigation.

Assuming the pilots are designed based on Lemma (1), the wideband beamforming matrix $\mathbf{AB}$ can be determined by

Figura 5.7 – Implementation of the optimum beams to minimize the MSE of the channel estimator.



Source: Provided by the author.

Theorem 3 *The optimum wideband hybrid beamforming is achieved if* $\mathbf{AB} = \mathbf{\Gamma}_{J_p}$*, where* $\mathbf{\Gamma}_{J_p}$ *corresponds to the first* J_p *eigenvectors of* $\mathbf{R}$*.*

Fig. 5.7 illustrates Theorem 3 by showing a set of beams created from the HB design. Note that the Theorem 3 only states the optimum beams, but it does not impose any specific structure $\mathbf{A}$ and $\mathbf{B}$. They actually admit more than one solution as long as they satisfy Theorem 3. Another important aspect is that, the solution employed to implement $\mathbf{A}$ and $\mathbf{B}$ are transparent to UE, i.e. from the perspective of estimator, the only requirement that must be met is $\mathbf{AB} = \mathbf{\Gamma}_{J_p}$.

Essentially, when the BS implements $\mathbf{AB} = \mathbf{\Gamma}_{J_p}$, this means that the beams are aligned with the eigenmodes of channel covariance matrix. Therefore, the effective covariance matrix can be represented as $\mathbf{R}_{\text{e}} = \mathbf{\Lambda}_{J_p}$, where $\mathbf{\Lambda}_{J_p}$ is a $J_p \times J_p$ diagonal matrix containing the J_p first eigenvalues. The diagonalization of $\mathbf{R}_{\text{e}}$ means that the BS creates J_p parallel channels to transmit J_p pilot sequences. Using this result, we can obtain an expression for the Using the result of Theorem 3, MSE in corollary.

Corollary 1 *Assuming that the BS implements* $\mathbf{AB} = \mathbf{\Gamma}_{J_p}$*, the MSE expression of the proposed estimator is given by*

$$MSE = N - \sum_{i=0}^{J_p-1} \frac{\lambda_i^2 c_i^2}{\lambda_i c_i^2 + \sigma^2}. \tag{5.23}$$

The proof can be found in Appendix 7.1. Clearly, the MSE depends on the number of pilots employed in the channel estimation process. In the next section, we discuss how to obtain the optimum J_p.

5.7.2 Pilot Power Loading

The minimization of Eq. (5.23) depends on the BS choice for $c_i \; \forall i \in \{1, \ldots, J_p\}$, which are the diagonal elements of the matrix $\mathbf{C}$. The variable c_i^2 can be interpreted as the power

allocated to the pilot beamormed by the eigenmode i. Therefore, the BS needs to choose the optimum pilot power allocation that minimizes the MSE.

As a second theorem, we enunciate the power allocation condition to minimize the MSE.

Theorem 4 *Given that the BS uses* $\mathbf{AB} = \mathbf{\Gamma}_{J_p}$*, the optimal pilot power allocation for the minimization of the MSE is determined by the water-filling solution.*

The derivation is found in Appendix 7.2. Given the choice of beams based on Theorem 3, the BS uses the traditional water-filling algorithm to assign the pilot power allocation. Other than this, the Theorem 4 also provides us a method to decide how many sequences we can employ on the estimation phase. For instance, if we have low SNR the algorithm tends to set zero on those eigen directions with low eigenvalues. This decision is taken based on the ratio between σ^2/λ_j and prevents the BS of sending pilots into components where the estimation will be likely poor. Beside this, it becomes a mechanism to assign the number of pilot sequences J_p based on a given ratio σ^2/λ_j. Using this mechanism, the number of pilot sequences are adapted according the SNR and the eigenvalues.

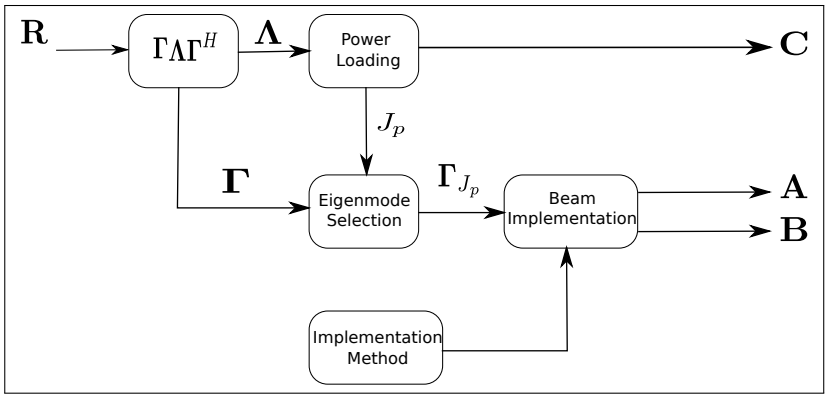
The parameter J_p has impact mainly in two aspects: (i) the overhead and (ii) HB implementation. About the overhead, the J_p dictates the dimension of the effective channel $\mathbf{H}_{\mathrm{e}}(k)$ which determines the overhead demanded to report this information. About the implementation aspect, J_p determines the number of beams that must be implemented by the product $\mathbf{AB}$. With respect the latter, there are different solutions to implement the beams and each of them imposes different architectures to the HB. Each of the solutions holds are relationship between J_p and number of RF chains L as will be shown ahead.

Fig. 5.8 summarizes the HB design. Initially, the BS uses $\mathbf{R}$ to calculate $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$. Using the power loading solution, the BS defines J_p and, automatically, chooses which eigenvectors to be employed for beamforming the pilot schemes. The block, called "implementation method", has a set of methods to calculate $\mathbf{A}$ and $\mathbf{B}$ and each of them is chosen based on a criterion defined by the BS. The calculations of those matrices will be discussed in the next section, where we show the limitations and advantage of each method.

5.8 Hybrid Beamforming Design

Theorem 3 stats the optimum beams to be used by the BS for transmitting the pilot sequences. However, the design of $\mathbf{A}$ and $\mathbf{B}$ is an open problem. In this section, we describe two solutions: (i) one is a baseline method proposed and discussed in the literature [79] and (ii) the second one is proposed in this paper. Although [79] proposed a good HB solution, therein the authors consider data transmission. In our method, pilot transmisison is assumed. Moreover, the authors in [79] worked with a flat channel, but thanks to our statistical description we could also extend this method to a frequency selective one.

Figura 5.8 – The block diagram describes the design of the HB for pilot transmission.



Source: Provided by the author.

5.8.1 Paired Phase-shifter

The first solution uses the fact that any vector is obtained from a combination of two other vectors with constant modulus entries. Using this result, it is possible to combine two columns of $\mathbf{A}$ using two rows of $\mathbf{B}$ to implement one eigenvector. The equation to be solved is given by [79]:

$$\mathbf{A}(n,l)\mathbf{B}(l,j) + \mathbf{A}(n,l+1)\mathbf{B}(l+1,j) = \mathbf{\Gamma}(n,j). \tag{5.24}$$

Using the procedure described in [79], the solution for $\mathbf{B}$ is as follows

$$\mathbf{B}(l,j) = \frac{\gamma_{\max}(j) + \gamma_{\min}(j)}{2}, \tag{5.25}$$

$$\mathbf{B}(l+1,j) = \frac{\gamma_{\max}(j) - \gamma_{\min}(j)}{2}, \tag{5.26}$$

where $\gamma_{\max}(j) = \max_n\{|\mathbf{\Gamma}(n,j)|\}$ and $\gamma_{\min}(j) = \min_n\{|\mathbf{\Gamma}(n,j)|\}$. The solution for the phase shifters are then given by [79]

$$\phi_{n,l} = \angle\mathbf{\Gamma}(n,j) + \cos^{-1}\left(\frac{|\mathbf{\Gamma}(n,j)| + \gamma_{\max}(j)\gamma_{\min}(j)}{|\mathbf{\Gamma}(n,j)|(\gamma_{\max}(j) + \gamma_{\min}(j))}\right), \tag{5.27}$$

$$\phi_{n,l+1} = \angle\mathbf{\Gamma}(n,j) + \cos^{-1}\left(\frac{|\mathbf{\Gamma}(n,j)| - \gamma_{\max}(j)\gamma_{\min}(j)}{|\mathbf{\Gamma}(n,j)|(\gamma_{\max}(j) - \gamma_{\min}(j))}\right), \tag{5.28}$$

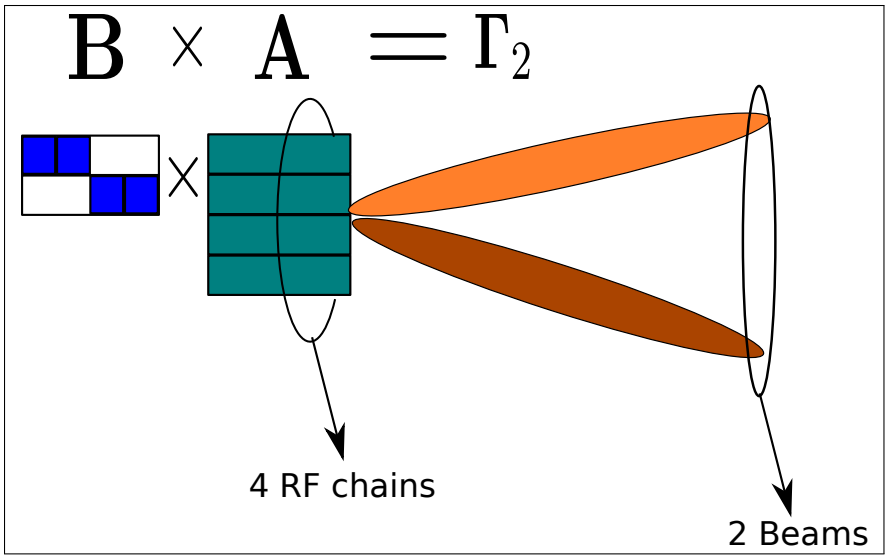
where $\phi_{n,l} = \angle\mathbf{A}(n,l)$ and $\phi_{n,l+1} = \angle\mathbf{A}(n,l+1)$. In this technique, each eigenvector is implemented by combining two RF chains. Therefore, the total number of RF chains to transmit the L layers is scaled by two, i.e. $L = 2J_p$. This means the BS can eliminate the hardware limitations of the analog beamforming at the cost of twice more RF chains.

5.8.2 LS criterion

Another way of determining $\mathbf{A}$ and $\mathbf{B}$ consists of minimizing the euclidean norm between $\mathbf{AB}$ and $\mathbf{\Gamma}_L$. To obtain a solution for $\mathbf{A}$ and $\mathbf{B}$, we propose to formulate the problem as shown:

$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{\Gamma}_L - \mathbf{AB}\|_{\mathcal{F}} \text{ s.t. } \mathbf{A} \in \mathcal{A}. \tag{5.29}$$

Figura 5.9 – The figure shows an example using 4 RF chains to create 2 beams.



Source: Provided by the author.

An attractive solution to this problem is given by the alternating least-square with projection (ALSP) algorithm proposed earlier in [82] as a in the context of blind separation of co-channel signals. Herein, we choose this algorithm as a solution to the HB problem at hand. More specifically, the solution of problem (5.29) yields the analong and digital wideband beamforming matrices, under the constraint that the analog part satisfies a constant modulus restriction.

The algorithm starts by assuming an initial guess $\hat{\mathbf{A}}$ to obtain $\hat{\mathbf{B}}$ in the LS sense, yielding

$$\hat{\mathbf{B}} = \hat{\mathbf{A}}^{\dagger}\boldsymbol{\Gamma}_L. \tag{5.30}$$

Given $\hat{\mathbf{B}}$, we then conditionally update $\hat{\mathbf{A}}$ in the LS sense, which gives

$$\tilde{\mathbf{A}} = \boldsymbol{\Gamma}_L\hat{\mathbf{B}}^{\dagger}. \tag{5.31}$$

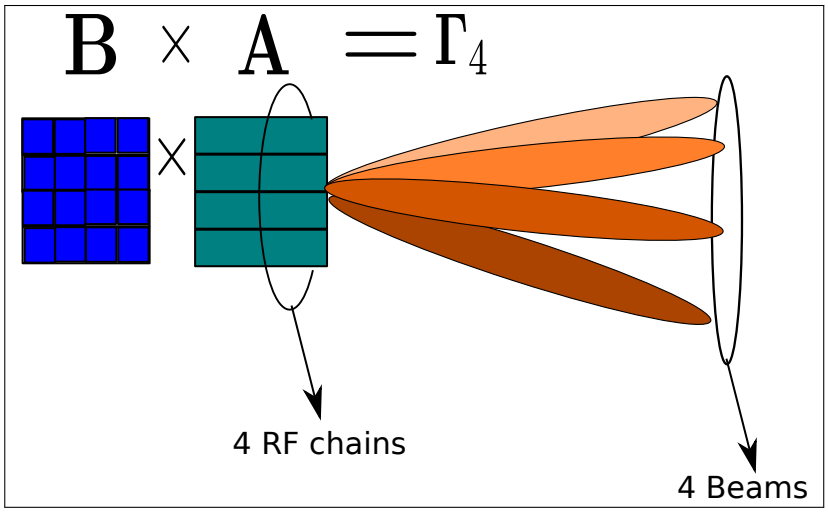
Since the solution found for $\hat{\mathbf{A}}$ in the equation above is unstructured and does not meet the finite resolution requirement for the phase shifters, we project the elements of $\hat{\mathbf{A}}$ onto the admissible set $\mathcal{A}$, i.e.

$$\hat{\mathbf{A}} = \text{proj}_{\mathcal{A}}\left[\tilde{\mathbf{A}}\right] \tag{5.32}$$

The overall algorithm is summarized in Table 4.

Note that, in the proposed solution, the number of RF chains scales with the rank of the covariance matrix, i.e. $L = J_p$. Thus, this technique requires half of the number of RF chains needed in the paired phase shifter solution. As an example, Fig. **??** shows a BS using 4 RF chains to create 4 beams by applying the ALSP algorithm. The results show that this solution offers a better performance than the solution discussed previously, since the proposed one enables us to spatially multiplex more pilots per unit of time.

Figura 5.10 – The figure shows an example using 4 RF chains to create 4 beams.



Source: Provided by the author.

The next subsection discusses more about implementation aspects and how they can be considered int the framework of Eq. (5.29).

5.8.3 Dealing with Implementation Issues Using The ALSP Framework

Hybrid beamforming performance is limited by the phase-shifter network limitations. In the framework stated by the problem (5.29), the set $\mathcal{A}$ contain matrices with constant amplitude entries and continuous phases. However, such an assumption leads to very cost phase-shifters in practice, since the need for an accurate phase control demands expensive components. The use of phase-shifters with infinite precision is not always practical in massive MIMO because they proportionally increase to the number of antennas. Given this limitation, low-resolution phase-shifters are very attractive in terms of implementation aspects. They offer simplicity and cost reduction compared with the HB using high resolution phase-shifters or/and digital beamforming.

Another important aspect is the number of phase-shifters which varies according to the HB architecture. This is determined by the number of connections that every RF chain has with the antenna array. The most common approach is the full connected phase-shifter network, where all the antenna elements have connections to all the RF chains. However, this leads to a large RF circuit that is susceptible to losses. To solve this issue, a sub-connected network can be implemented instead. In this case, the RF chains are connected only to a subset of antenna elements. This indeed reduces the number of phase-shifters, but also reduces also the degree of processing which leads also to loss performance comparing to other beamforming schemes such as digital beamforming or HB with full connected phase-shifter network.

The above issues can be incorporated into the ALSP framework by defining a set $\mathcal{A}$ that contains Q points. Each point stands for a possible analog beamforming solution $\mathbf{A}$ obtained

Algorithm 4 ALSP

Require: $\mathbf{A} \in \mathcal{A}$
Ensure: $\min_{\mathbf{A},\mathbf{B}} \|\mathbf{\Gamma} - \mathbf{AB}\|_F^2$
  Start the algorithm with a random choice of $\hat{\mathbf{A}}$
  while $\|\hat{\mathbf{A}}_i - \hat{\mathbf{A}}_{i-1}\|_F^2 \geq \epsilon$ **do**
    $\hat{\mathbf{B}}_{i+1} = \mathbf{A}_i^{\dagger}\mathbf{\Gamma}$
    $\hat{\mathbf{A}}_{i+1} = \mathbf{\Gamma}\mathbf{B}^{\dagger}$
    $\hat{\mathbf{A}}_{i+1} = \text{proj}\left[\hat{\mathbf{A}}_{i+1}\right]$
  end while

using a set of phase-shifter with X values of resolution. The phase-shifter network also obeys a given configuration of connection between the RF chain and the set of antennas which defines the architecture of the analog part.

5.8.4 Convergence Aspects For Wideband Beamforming

Before discussing convergence results, we review some important definitions that are part of the framework applied on our convergence analysis. As we know, an algorithm is an iterative process that generates a sequence of points $\{\mathbf{V}_k\}_{k=1}^{\infty}$, where $\mathbf{V}_k \in \mathcal{V}$. The points are generated using a mapping function $\mathcal{T} : \mathcal{V} \to \mathcal{V}$ that maps a point from a space $\mathcal{V}$ to the same space $\mathcal{V}$, i.e. $\mathbf{V}_k = \mathcal{T}(\mathbf{V}_{k-1})$. For the mapping function $\mathcal{T}$, there is a fixed point $\mathbf{V}^{\star}$ such that $\mathbf{V}^{\star} = \mathcal{T}(\mathbf{V}^{\star})$. The set of all possible fixed points is defined as $\mathcal{C}$.

Given the definition of the mapping function, we now state the definition of the descent function. A function f is descent for $\mathcal{T}$ if it satisfies the conditions

1. $f : \mathcal{V} \to \mathbb{R}$ is non-negative and continuous,
2. $f(\mathbf{V}_{k+1}) < f(\mathbf{V}_k)$ for $\mathbf{V}_{k+1} = \mathcal{T}(\mathbf{V}_k)$ and $\mathbf{V}_{k+1} \notin \mathcal{C}$,
3. $f(\mathbf{V}_{k+1}) \leq f(\mathbf{V}_k)$ for $\mathbf{V}_{k+1} = \mathcal{T}(\mathbf{V}_k)$ and $\mathbf{V}_{k+1} \in \mathcal{C}$.

Using this framework in the context of the problem under investigation, we define the set $\mathcal{V}$, the mapping function $\mathcal{T}$, and the descent function f with respect to the algorithm stated in the Table 4. Therefore, the ALSP algorithm is the mapping function $\mathcal{T}$, the pair $\mathbf{V}_k = (\mathbf{A}_k, \mathbf{B}_k)$ is an element of the set $\mathcal{V} = \mathcal{A} \times \mathbb{C}^{L \times J_p}$, and the objective function is given by $f(\mathbf{A}, \mathbf{B}; \mathbf{\Gamma})$. For the sake of simplicity, we skip the sub-index J_p in $\mathbf{\Gamma}_{J_p}$.

To prove that $f(\mathbf{A}, \mathbf{B}; \mathbf{\Gamma})$ is a descent function, it has to meet the three conditions. For the first condition, it is natural to say that $f(\mathbf{A}, \mathbf{B}; \mathbf{C})$ is non-negative and continuous in $\mathbf{A}$ and

$\mathbf{B}$. To prove the second and third conditions note that [82]

$$\begin{aligned} f(\mathbf{A}_k, \mathbf{B}_k; \mathbf{\Gamma}) &= \|\mathbf{\Gamma} - \mathbf{A}_k\mathbf{B}_k\|_F^2 \\ &= \min_{\mathbf{B}} \|\mathbf{\Gamma} - \mathbf{A}_k\mathbf{B}\|_F^2 \\ &\leq \|\mathbf{\Gamma} - \mathbf{A}_k\mathbf{B}_{k-1}\|_F^2 \qquad (5.33) \\ &= \min_{\mathbf{A}\in\mathcal{A}} \|\mathbf{\Gamma} - \mathbf{A}\mathbf{B}_{k-1}\|_F^2 \\ &\leq \|\mathbf{\Gamma} - \mathbf{A}_{k-1}\mathbf{B}_{k-1}\|_F^2 \qquad (5.34) \\ &= f(\mathbf{A}_{k-1}, \mathbf{B}_{k-1}; \mathbf{\Gamma}) \qquad (5.35) \end{aligned}$$

The inequality (5.33) is strict unless $\mathbf{B}_k = \mathbf{B}_{k-1}$ and the inequality (5.34) is also strict if $\mathbf{A}_k = \mathbf{A}_{k-1}$.This means that the objective function decreases unless $(\mathbf{A}_{k-1}, \mathbf{B}_{k-1}) = \mathcal{T}(\mathbf{A}_{k-1}, \mathbf{B}_{k-1})$. This shows that the function is decreasing until the algorithm reaches the fixed-point which essentially comprises the conditions two and three.

The algorithm convergence is guaranteed by analyzing the set $\mathcal{A}$ that defines the matrix $\mathbf{A}$. Let us assume that $\mathcal{A}$ is a finite set. This means that the phase-shifters are finite and that they have a set of possible phases. The total number of $\mathbf{A}$s is given by the cardinality of $\mathcal{A}$, which we define as $|\mathcal{A}| = Q$, and each element of $\mathcal{A}$ is associated uniquely to a given $\mathbf{B}_k$. This association is always true when each LS formulation leads to an unique solution which happens if and only if the matrices $\mathbf{A}$ and $\mathbf{B}$ are not rank deficient. The solutions obtained from each LS problem creates a pair $(\mathbf{A}_k, \mathbf{B}_k)$ and there are a total of Q different pairs. Let us consider that after $Q+1$ iterations, the algorithm generated a sequence $\{(\mathbf{A}_k, \mathbf{B}_k)\}_{k=1}^{Q+1}$. We can affirm that there are at least two points in this sequence that must be the same. Taking into account this fact, let us assume that the indices of such points are j and $j+l$, which means that

$$f(\mathbf{A}_j, \mathbf{B}_j) = f(\mathbf{A}_{j+l}, \mathbf{B}_{j+l}). \qquad (5.36)$$

If $l = 1$ then $(\mathbf{A}_j, \mathbf{B}_j)$ is a fixed-point because $\mathcal{T}(\mathbf{A}_j, \mathbf{B}_j) = (\mathbf{A}_{j+1}, \mathbf{B}_{j+1}) = (\mathbf{A}_j, \mathbf{B}_j)$. Now, let us assume $l > 1$, i.e. $(\mathbf{A}_j, \mathbf{B}_j)$ is not a fixed point. Hence,

$$f(\mathbf{A}_j, \mathbf{B}_j) > f(\mathbf{A}_{j+l}, \mathbf{B}_{j+l}). \qquad (5.37)$$

However, (5.36) and (5.37) are contradictory because we have the same pair, which means that $(\mathbf{A}_j, \mathbf{B}_j)$ must be a fixed point. Hence, a fixed point is obtained with finite number of iterations.

5.9 Numerical Results and Discussions

In our simulations, we use the channel model COST 2100 extended for massive MIMO, and proposed by the EU project MAMMOET in [77]. In the simulations, Two types of scenario are considered, which correspond to high and low angular spreads. The channel is generated at 2.6 GHz with 10 MHz bandwidth.

5.9.1 Performance of the MMSE Estimator and Its Implications

Figure (5.11) shows the MSE curve of the estimator defined in Eq. (5.19) using the ALSP algorithm to implement the optimum beams. After finding the matrices $\mathbf{A}$ and $\mathbf{B}$ from this algorithm, the BS uses the eigenvalues of the covariance matrix $\mathbf{R}$ to assign the power associated with each beam. The BS tests equal and waterfilling power allocations and they are shown in Figures 5.11a and 5.11b, respectively.

Figure 5.11b shows that the MSE tends to decrease as the number of beams increases. This behaviour comes from the fact that the estimator collects more components of the channel covariance matrix which results in a more accurate estimation according to Eq. (5.23). However, there is a region with very high values of MSE where the estimation fails. This happens because the individual SNR of each beam becomes very small and no power is allocated.

Notice that there is a point where the channel estimator does not improve. In this case, the number of beams employed in the pilot transmission has reached the rank of the covariance matrix. This means that the estimator collects energy from all possible eigenmodes that represent the channel.

As expected, the equal power solution shown in Fig. 5.11a achieves the same performance as the case with power allocation when the system is operating under high SNR conditions. Different from the optimum solution, the method does not take into account the eigenvalues of the spatial covariance matrix, which results in a suboptimal performance in terms of MSE. To compensate this, the BS needs to increase the transmitted power to guarantee high channel estimation quality. The current Long Term Evolution (LTE) standard uses a power boost solution to guarantee the estimation quality [60]. But the proposed technique, saves transmitted power by exploiting the array gain of the massive MIMO system and also the knowledge of the covariance matrix.

5.9.2 Beam Implementation

The use of beams plays an important role on the proposed channel estimation technique. As discussed before, there is more than one method of designing the HB. The ones considered in the simulations are i) the $\mathbf{W}(k) = \mathbf{ABC}(k)$ and ii) $\mathbf{W}(k) = \mathbf{AD}(k)$. In model (i), two methods are considered to design $\mathbf{A}$ and $\mathbf{B}$, namely the paired phase-shifter and the ALSP algorithm. In model (ii), $\mathbf{A}$ is designed as the phase-only version of $\mathbf{\Gamma}_{J_p}$ and $\mathbf{D}(k)$ is a diagonal matrix that contains the weights of the beams, where they are obtained from the power allocation algorithm. We compare the three solutions in terms of the sum rate (bps/Hz), which is calculated by using the singular vectors of $\hat{\mathbf{H}}_{\mathrm{e}}$ to design $\mathbf{C}(k)$ and $\mathbf{D}(k)$.

Figure 5.12 shows results for two scenarios: low and high angular spread. The implementation using $\mathbf{W}(k) = \mathbf{ABC}(k)$ outperforms $\mathbf{W}(k) = \mathbf{AD}(k)$ because the wideband filter $\mathbf{AB}$ achieves a better approximation of the eigenvectors than using only the phase-only version of

the eigenvector matrix to design $\mathbf{A}$. Moreover, the first model combines multiple phase-shifters to adjust the signal amplitude which reduces the performance loss of the HB.

Comparing the performances of ALSP and paired phase-shifter in Fig. 5.12, the curves show that the first achieves better performance than the latter (shown in Section 5.8.1). This is because ALSP produces more beams with the same number of RF chains used in the paired method, and consequently spatially multiplexing more pilots. Using more beams, the channel estimator uses more eigenmodes in the channel estimation process which naturally reduces the MSE, as can be seen in Eq. (5.23). The estimation of more modes enables to capture more energy from the channel which leads to a better sum-rate performance.

The performance in scenarios with high and low angular spreads are compared in terms of the sum-rate (bps/Hz) in Fig. 5.12. The gain of the proposed solution is clearly in scenarios with a high angular spread, as shown in 5.12a. On the other hand, the three methods achieve similar performances in a low angular spread.

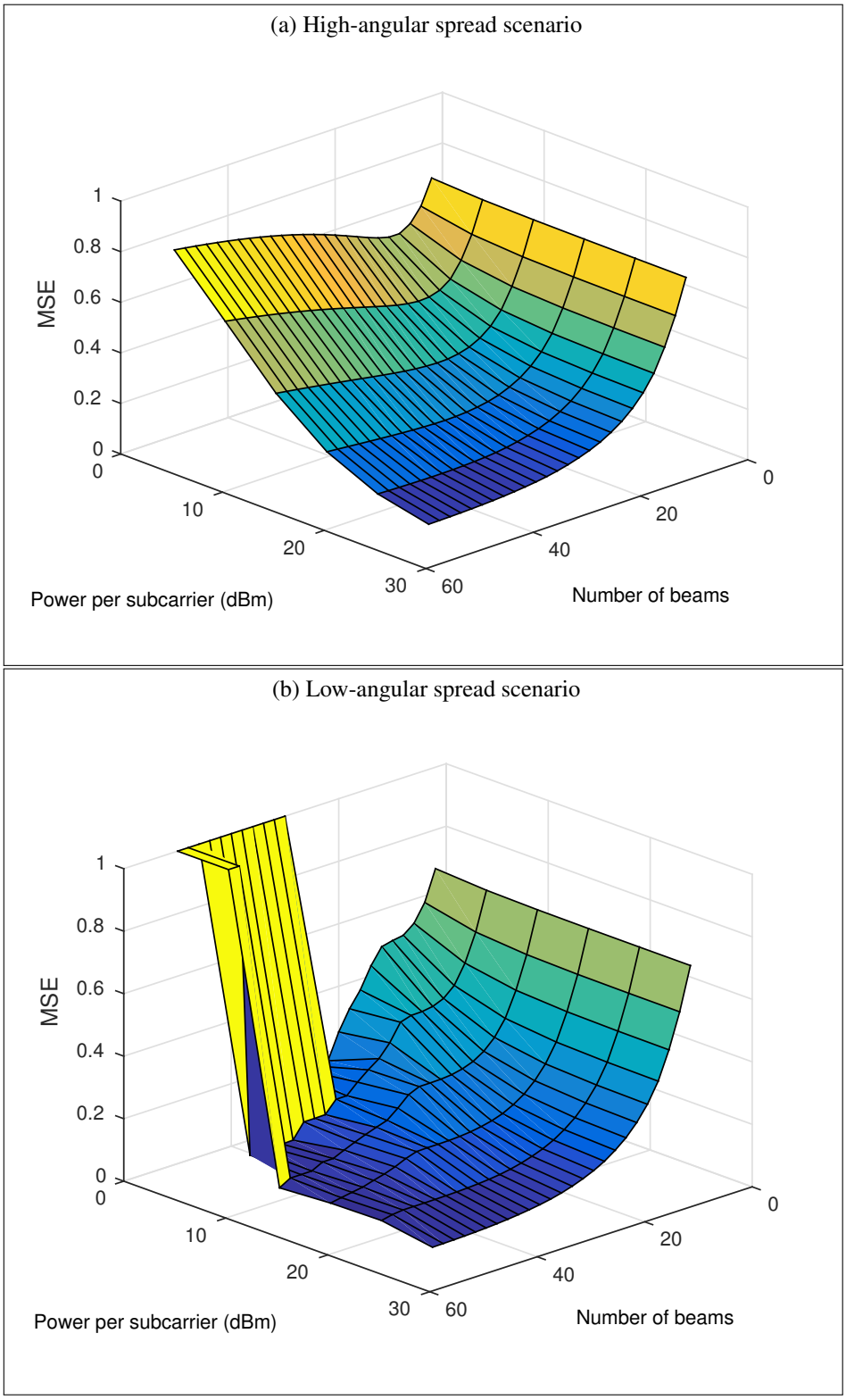
5.10 Conclusion

In this chapter, we developed strategies to desgin the HB exploiting the knowledge of the spatial covariance matrix of the channel. We showed that the spectral efficiency of a large-array HB system in FDD can be significantly increased by adapting the number of RF chains.

In the second part of this chapter, we generalized the channel model by adding multiple antennas at the UE. Exploiting again the second order statistics of the channel, it is possible to use the subspace of the channel covariance matrix to design a set of beams that optimally assists channel estimation. The waterfilling solution gives us the optimum pilot power allocation, which in turn defines the number of pilot sequences to be used for the channel estimation. The HB and the channel estimator are jointly derived from the MMSE criterion which produces improved estimation performance in low SNR scenarios.

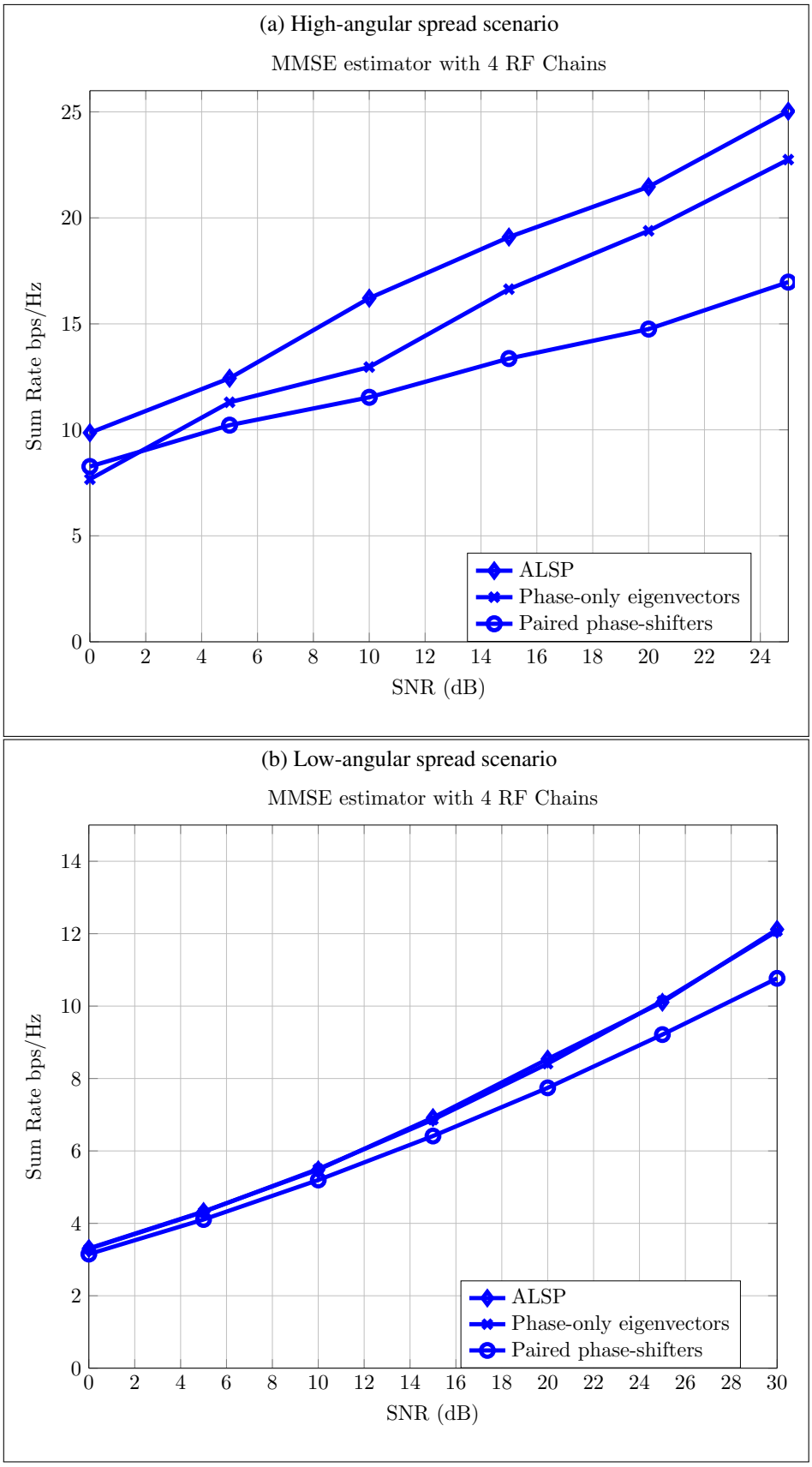
From the analysis developed in this chapter, the proposed method using the ALSP algorithm shows a better performance compared with that using the paired phase-shifter solution and the phase-only version of the eigenvectors. The gains are more pronounced in high-angular spread channels.

Figura 5.11 – Plot of the MSE superficie with respect the number of beams and the power per subcarrier for a scenario with high angular spread.



Source: Provided by the author.

Figura 5.12 – System sum-rate using HB solutions. Fig. (a) and Fig. (b) show the performance in high-angular and low-angular spread scenarios, respectively.



Source: Provided by the author.

6 CONCLUSIONS

In this thesis, we developed channel estimation techniques that do not depend on the channel reciprocity and afforded pilot overhead reduction in massive MIMO systems. This was possible because the proposed estimators exploit a structured environment to obtain a compressed representation of the massive MIMO channel. The summary of our main conclusions are:

- Chapter 2 : CS gave us a powerful framework to derive new sparsity-based estimators for massive MIMO channel. The proposed technique based on lasso-ADMM achieved better performance in terms of estimation error than the well-known OMP algorithm. This is because the optimization problem solved by latter does not take into account any source of noise. On the other hand, lasso formulation has the parameter β that controls channel estimator output between the sparsest solution and the one that minimizes the least-square error. The comparison between the techniques showed that both tend to achieve the CRLB in a high SNR scenario. Moreover, we also obtained that the CRLB depends on the Grassmmanian pilot sequence matrix.

- Chapter 3 : The proposed tensor-based CS formalism provided accurate estimation of frequency-selective channels. The approach capitalizes on a sparse multidimensional decomposition model, for which the channel was formulated by means of a tensor algebra formalism. This modeling approach yields the formulation of a tensor-based CS channel estimator that exploits the three channel dimensions (space(transmitter), space(receiver) and frequency) by solving the problem per dimension of the channel tensor. The proposed tensor-OMP algorithm has a lower complexity than its vectorized version, which makes possible the use of such an estimator in receivers with limited processing capabilities.

- Chapter 4 : Using CS framework, we proposed solution for mm-wave indoor systems that provides high accuracy on estimating the sparse channel in this type of scenario. We showed that the method has enhanced capabilities of dealing with the user mobility in indoor scenarios. More specifically, the UE can track with accuracy the channel variations and report the new angular information to the AN. We also show that simpler beamforming tehcniques can achieve near optimum performance because the environment propagation characteristics provided such a condition.

- Chapter 5: Using the spatial covariance matrix, it is possible to design a set of transmitter beams for transmitting pilots to reduce the system overhead. Our results showed that the proposed scheme drastically enhances the spectral efficiency of FDD systems, which are mainly affected by the overhead issue. After this, we investigated the optimum beam and channel estimator design to achieve the MMSE bound. From that analysis,we concluded

that eigenvectors must be used for transmitting pilots and the eigenvalue must be used to determine the number of sequences to be spatially multiplexed. The results showed that our HB and estimator provide enhanced sum-rate performance in scenarios with high-angular spread.

In terms of future perspectives, we can list:

1. extension of CS methods for estimating the spatial covariance matrix;
2. devise new HB transceivers with multiuser functionality;
3. investigate the hardware implications of HB transceivers in the system spectral efficiency performance;

The first topic refers to the problem of estimating the spatial covariance matrix. To the best of author's knowledge, there is no work in massive MIMO literature that deals with this issue. The challenge that arises here is similar to the one we treated in this thesis, which is the overhead. Either using uplink or downlink, the number of pilots increases because of the large covariance matrix dimensions. To solve it, the idea is to use the CS framework to derive new sparsity-based covariance matrix estimators. This will give us solutions that will require fewer pilots. We believe sparsity is an interesting approach because it gives us a relationship between the channel measurements and the angular information described by the spatial covariance matrix.

The second topic goes on the investigation of HB transceivers for multiuser MIMO. There are works in the literature that propose solutions using HB transmitters in that scenario, but they are very few. There a lot of open questions yet which gives us the opportunity to propose new multiuser transmitters and receivers based on HB architecture. For instance, we can cite the multiuser detection schemes, that are mainly derived based on full digital receivers. To the best of author's knowledge, HB has not been considered yet in multiuser detection algorithms which means that there are opportunities to propose new solutions.

The third topic investigates the implementation aspects of a HB. The hardware implications have not been covered by this thesis, but they must be considered in order to guarantee good performance and also ensure the promised benefits of massive MIMO systems. For instance, we cite the ADC and DAC resolutions that were assumed infinite in this work. In practice, we have instead a set of finite resolution converters which lead to undesirable non-linear effects over the signal. However, by modelling such an effect, we can then propose new schemes that provide robustness against this degradation. Moreover, we cite the high number of phase-shifters that come by using a full connected analog beamforming. To tackle it, one idea consists of using distributed hardware solutions to reduce the number of phase-shifters. That is, the transceiver uses small modules that implement a HB with fewer antennas. When the modules are connected,

we can create a massive array with a distributed architecture. In such a case, the number of phase-shifters is smaller because the analog network does not connect every antenna element to each RF chain. This novel architecture would open the possibility of developing new distributed channel estimators, detection and beamforming solutions that could be derived using the ADMM framework.

This thesis gives its contribution mainly on the channel estimation aspects. However, as the reader may have noticed, there are other problems to be investigated in massive MIMO area. As a future perspective of this work, we intend to deeply investigate the three aforementioned topics, since they are a natural continuation of this work.

REFERENCES

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590 –3600, november 2010.

[2] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?," in *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 545 –550, Sep. 2011.

[3] P. Larsson, "Large-scale cooperative relaying network with optimal coherent combining under aggregate relay power constraints," 2003.

[4] H. D. Nguyen and S. Sun, "Massive MIMO versus small-cell systems: Spectral and energy efficiency comparison," *CoRR*, vol. abs/1509.03998, 2015.

[5] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *ICT Mobile Summit 2010, 19th Future Network & Mobile Summit, June 16-18, 2010, Florence, Italy*, (Florence, ITALY), 06 2010.

[6] C. Shepard, H. Yu, N. An, L. E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations."

[7] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[8] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician MIMO channels with rician disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807–1820, 2010.

[9] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive mimo systems," *IEEE Transactions on Communications*, vol. 61, pp. 5016–5029, Dec. 2013.

[10] D. C. Araujo, A. L. F. de Almeida, and J. C. M. Mota, "Compressive sensing based channel estimation for massive MIMO systems with planar arrays," in *6th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2015, Cancun, Mexico, December 13-16, 2015*, pp. 413–416, 2015.

[11] D. Araujo, A. de Almeida, J. Axnas, and J. Mota, "Channel estimation for millimeter-wave very-large mimo systems," in *2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO),*, pp. 81–85, Sept 2014.

[12] D. Araujo, E. Karipidis, and A. F. d. Almeida, "Improving spectral efficiency in large-array FDD systems with hybrid beamforming," *IEEE Signal Processing Letters*, submitted.

[13] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[14] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Mag*, pp. 118–120, 2007.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[16] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489 – 509, Feb. 2006.

[17] Y. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Transactions on Signal Processing*, vol. 57, pp. 2986 –2997, aug. 2009.

[18] L. Shi, Z. Zhou, L. Tang, H. Yao, and J. Zhang, "Ultra-wideband channel estimation based on bayesian compressive sensing," in *2010 International Symposium on Communications and Information Technologies (ISCIT)*, pp. 779 –782, oct. 2010.

[19] J. Paredes, G. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 383 –395, oct. 2007.

[20] C.-Y. Chen and P. Vaidyanathan, "Compressed sensing in MIMO radar," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, pp. 41 –44, oct. 2008.

[21] J. Choi, D. Love, and P. Bidigare, "Downlink Training Techniques for FDD Massive MIMO Systems: Open-Loop and Closed-Loop Training With Memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 802–814, Oct 2014.

[22] C. Caiafa and A. Cichocki, "Block sparse representations of tensors using kronecker bases," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 2709–2712, 2012.

[23] N. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *Signal Processing Letters, IEEE*, vol. 19, no. 11, pp. 757–760, 2012.

[24] T. Marzetta, "How Much Training is Required for Multiuser MIMO?," in *Fortieth Asilomar Conference on Signals, Systems and Computers* , pp. 359 –363, Nov. 2006.

[25] J. Jose, A. Ashikhmint, P. Whiting, and S. Vishwanath, "Scheduling and pre-conditioning in multi-user MIMO tdd systems," in *IEEE International Conference on Communications, 2008. ICC '08.*, pp. 4100 –4105, may 2008.

[26] J. Joseph, M. R. Devi, and N. S. Kumar, "Compressive sensing based underwater communication system," *Procedia Technology*, vol. 24, pp. 964 – 971, 2016. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015).

[27] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, pp. 1058–1076, June 2010.

[28] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1708–1721, Mar. 2010.

[29] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser mimo systems," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2890–2895, Apr. 2013.

[30] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5g mobile and wireless communications: the vision of the metis project," *IEEE Communications Magazine*, vol. 52, pp. 26–35, May 2014.

[31] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[32] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1227–1231, Jun. 2015.

[33] E. J. Candes and Y. Plan, "A probabilistic and ripless theory of compressed sensing," *IEEE Trans. Inf. Theor.*, vol. 57, pp. 7235–7254, Nov. 2011.

[34] E. J. Candès and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities," *Comm. Pure Appl. Math.*, vol. 57, no. 2, pp. 219–266, 2004.

[35] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer Publishing Company, Incorporated, 1st ed., 2010.

[36] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic mimo channel model with joint correlation of both link ends," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 90 – 100, Jan. 2006.

[37] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2563 – 2579, Oct 2002.

[38] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. New York, NY, USA: Cambridge University Press, 1st ed., 2008.

[39] P. P. Vaidyanathan, S.-M. Phoong, and Y.-P. Lin, *Signal Processing and Optimization for Transceiver Systems*. New York, NY, USA: Cambridge University Press, 1st ed., 2010.

[40] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[41] H. Zhu, A. Cano, and G. B. Giannakis, "Distributed consensus-based demodulation: algorithms and error analysis," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 2044–2054, Jun. 2010.

[42] N. Costa and S. Haykin, *Multiple-Input Multiple-Output Channel Models: Theory and Practice*. Adaptive and Learning Systems for Signal Processing, Communications and Control Series, Wiley, 2010.

[43] W. Bajwa, J. Haupt, A. 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.

[44] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems," in *2014 Globecom Workshops (GC Wkshps)*, pp. 416–421, Dec 2014.

[45] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 273–277, Nov 2014.

[46] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 831–846, Oct 2014.

[47] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 1211 – 1226, aug 2002.

[48] A. M. Sayeed, "Deconstructing multiantenna fading channels," *Signal Processing, IEEE Transactions on*, vol. 50, pp. 2563–2579, Oct. 2002.

[49] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wideband MIMO radio channels based on NLOS indoor measurements," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 3, pp. 655–665, 2004.

[50] N. Costa and S. Haykin, "A novel wideband MIMO channel model and the wideband MIMO software defined radio," in *WiCOM 2006.International Conference on Wireless Communications, Networking and Mobile Computing, 2006.*, pp. 1–4, 2006.

[51] Y. Zhang, O. Edfors, P. Hammarberg, T. Hult, X. Chen, S. Zhou, L. Xiao, and J. Wang, "A general coupling-based model framework for wideband MIMO channels," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 574–586, 2012.

[52] A. Cichocki, R. Zdunek, A. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, 2009.

[53] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, pp. 455–500, September 2009.

[54] M. Duarte and R. Baraniuk, "Kronecker compressive sensing," *IEEE Transactions on Image Processing*, vol. 21, pp. 494 –504, feb. 2012.

[55] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[56] D. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289 –1306, april 2006.

[57] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive tracking with 1000-element arrays: A framework for multi-gbps mm wave cellular downlinks," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pp. 690–697, 2012.

[58] B. Ye and Z. Zhang, "Improved pilot design and channel estimation for 60ghz ofdm based on ieee 802.11.ad," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, pp. 4129–4133, 2013.

[59] M. Jo, D. Araújo, T. Maksymyuk, A. Almeida, and J. M. Tarcisio F Maciel, "Massive MIMO: Survey and future research topics," *IET Communications*, May 2016.

[60] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 1st ed., 2011.

[61] L. Sundstöm, D. Hui, A. Reial, and D. Araújo, "Antenna beam control," Sep. 2015. WO Patent App. PCT/EP2015/055,071.

[62] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Lomayev, and A. Khoryaev, "Statistical channel model for 60 ghz wlan systems in conference room environment," in *Antennas and Propagation (EuCAP), 2010 Proceedings of the Fourth European Conference on*, pp. 1–5, 2010.

[63] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Khoryaev, and A. Lomayev, "Experimental investigations of 60 ghz wlan systems in office environment," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1488–1499, 2009.

[64] D. Hui and J. Axnäs, "Interference-aware routing for self-backhaul in ultra-dense networks - a centralized approach," tech. rep., Ericsson Research, •.

[65] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Information Theory and Applications Workshop (ITA), 2012*, pp. 234–239, 2012.

[66] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[67] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *2013 Information Theory and Applications Workshop (ITA)*, pp. 1–5, Feb 2013.

[68] E. Bjornson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, pp. 114–123, Feb. 2016.

[69] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, pp. 6441–6463, Oct. 2013.

[70] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4131–4143, Aug 2010.

[71] J. Mo and R. W. Heath, "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," pp. 1–5, Feb. 2014.

[72] D. Ying, F. Vook, T. Thomas, and D. Love, "Hybrid structure in massive MIMO: Achieving large sum rate with fewer RF chains," in *IEEE International Conference on Communications (ICC)*, pp. 2344–2349, Jun 2015.

[73] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, pp. 186–194, Jan 2015.

[74] T. Obara, S. Suyama, J. Shen, and Y. Okumura, "Joint fixed beamforming and eigenmode precoding for super high bit rate massive MIMO systems using higher frequency bands," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 607–611, Sep 2014.

[75] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[76] T. Marzetta, "Massive MIMO: An introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[77] X. Gao, F. Tufvesson, and O. Edfors, "Massive MIMO channels - measurements and models," in *Asilomar Conference on Signals, Systems and Computers*, pp. 280–284, Nov 2013.

[78] T. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *2014 IEEE Global Communications Conference (GLOBECOM)*, pp. 4066–4071, Dec 2014.

[79] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale MIMO systems," in *ICASSP,*, pp. 2929–2933, April 2015.

[80] V. Venkateswaran and A. J. van der Veen, "Partial beamforming to reduce adc power consumption in antenna array systems," in *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*, pp. 146–150, Jul. 2008.

[81] C. Vithanage, R. Cepeda, J. Coon, and J. McGeehan, "MIMO-OFDM pilot placement algorithms for wideband indoor communications," *IEEE Transactions on Communications*, vol. 59, pp. 466–476, Feb. 2011.

[82] J. Laurila, K. Kopsa, R. Schurhuber, and E. Bonek, "Semi-blind separation and detection of co-channel signals," in *Communications, 1999. ICC '99. 1999 IEEE International Conference on*, vol. 1, pp. 17–22 vol.1., 1999.

7 APPENDIX - MMSE ESTIMATOR

Consider the minimization problem

$$\min_{\mathbf{G}} \mathrm{E}\left\{\|\mathbf{Y}(k)\mathbf{G} - \mathbf{H}_e(k)\|_F^2\right\}. \tag{7.1}$$

This problem can be rewritten in function of the channel error matrix. Using the definition in Eq. (5.17), we define the cost function as

$$\begin{aligned} f(\mathbf{G}) &\triangleq \mathrm{E}\left\{\|\mathbf{\Delta}(k)\|_F^2\right\} \\ &= \mathrm{E}\left\{\mathrm{trace}\{\Delta(k)^H\Delta(k)\}\right\} \\ &= \mathrm{trace}\{(\mathbf{G}^H\mathbf{P}^H\mathbf{C}^H - \mathbf{I})\mathbf{R}_e(\mathbf{CPG} - \mathbf{I})\} \\ &\quad +\mathrm{trace}\{\mathbf{G}^H\mathbf{R}_z\mathbf{G}\}. \end{aligned}$$

To solve the problem in (7.1), we calculate the derivative of the cost function $f(\mathbf{G})$

$$\frac{\partial f(\mathbf{G})}{\partial \mathbf{G}^H} = \mathbf{P}^H\mathbf{C}^H\mathbf{R}_e(\mathbf{CPG} - \mathbf{I}) + \mathbf{R}_z\mathbf{G}. \tag{7.2}$$

Consider $\frac{\partial f(\mathbf{G})}{\partial \mathbf{G}^H} = \mathbf{0}$, thus

$$\begin{aligned} \mathbf{0} &= \mathbf{P}^H\mathbf{C}^H\mathbf{R}_e(\mathbf{CPG} - \mathbf{I}) + \mathbf{R}_z\mathbf{G} \\ \mathbf{0} &= \left(\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e\mathbf{CP} + \mathbf{R}_z\right)\mathbf{G} - \mathbf{P}^H\mathbf{C}^H\mathbf{R}_e \\ \mathbf{G}_{\mathrm{MMSE}} &= \left(\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e\mathbf{CP} + \mathbf{R}_z\right)^{-1}\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e. \end{aligned} \tag{7.3}$$

7.1 Minimization of the MSE

Using the expression of the MMSE estimator, we can obtain the spatial covariance matrix of the estimated effective channel given by

$$\mathbf{R}_{\hat{\mathbf{H}}_e} = \mathbf{R}_e\mathbf{CP}\left(\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e\mathbf{CP} + \mathbf{R}_z\right)^{-1}\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e. \tag{7.4}$$

Using this result, we express the cost function as

$$\begin{aligned} \mathrm{MSE}(\mathbf{R}_e) &= \mathrm{trace}\{\mathbf{R}_e - \\ &\quad \mathbf{R}_e\mathbf{CP}\left(\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e\mathbf{CP} + \mathbf{R}_z\right)^{-1}\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e\} \end{aligned}$$

Using the Eq. (7.3), we can obtain a closed-form to the minimum MSE expression. Our interest is mainly on the second term of Eq. (7.5). In this way, we define

$$\mathrm{MSE}_1(\mathbf{R}_e) = \mathrm{trace}\{\mathbf{R}_e\} \tag{7.5}$$

$$\mathrm{MSE}_2(\mathbf{R}_e) = \mathrm{trace}\{\mathbf{R}_e\mathbf{CP}\left(\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e\mathbf{CP} + \mathbf{R}_z\right)^{-1}\mathbf{P}^H\mathbf{C}^H\mathbf{R}_e\} \tag{7.6}$$

We can rewrite $\text{MSE}_2(\mathbf{R}_e)$ as

$$\text{MSE}_2(\mathbf{R}_e) = \text{trace}\{\mathbf{R}_e\mathbf{C}\left(\mathbf{C}^H\mathbf{R}_e\mathbf{C} + \mathbf{P}^H\mathbf{R}_z\mathbf{P}\right)^{-1}\mathbf{C}^H\mathbf{R}_e\} \tag{7.7}$$

Consider that we design the pilot sequence so that $\mathbf{P}^H\mathbf{R}_z\mathbf{P} = \Lambda_{\mathbf{z}}$. Thus,

$$\text{MSE}_2(\mathbf{R}_e) = \text{trace}\{\mathbf{R}_e\mathbf{C}\left(\mathbf{C}^H\mathbf{R}_e\mathbf{C} + \mathbf{\Lambda}_z\right)^{-1}\mathbf{C}^H\mathbf{R}_e\}. \tag{7.8}$$

Assuming that the BS knows the channel covariance matrix $\mathbf{R}$, we design $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{\Gamma} = \mathbf{AB}. \tag{7.9}$$

Using this design the equivalent spatial covariance matrix is given by

$$\mathbf{R}_e = \mathbf{\Lambda}_{J_p}, \tag{7.10}$$

where $\mathbf{\Lambda}_{J_p}$ is a $J_p \times J_p$ diagonal matrix created from the J_p first eigenvalues of the channel covariance matrix $\mathbf{R}$. Assuming the design in (7.9), the Eq. (7.8) is rewritten as

$$\begin{aligned}\text{MSE}_2^\star &= \text{trace}\{\mathbf{\Lambda}_{J_p}\mathbf{C}\left(\mathbf{C}^H\mathbf{\Lambda}_{J_p}\mathbf{C} + \mathbf{\Lambda}_z\right)^{-1}\mathbf{C}^H\mathbf{\Lambda}_{J_p}\} \\ &= \text{trace}\{\left(\mathbf{C}^H\mathbf{\Lambda}_{J_p}\mathbf{C} + \mathbf{\Lambda}_z\right)^{-1}\mathbf{C}^H\mathbf{\Lambda}_{J_p}^2\mathbf{C}\}.\end{aligned}$$

Assuming that $\mathbf{C}$ is a $J_p \times J_p$ diagonal matrix, the expression is can be reduced tp

$$\text{MSE}_2^\star = \sum_{i=0}^{J_p-1}\frac{\lambda_i^2 c_i^2}{\lambda_i c_i^2 + \sigma^2}. \tag{7.11}$$

The minimum MSE is

$$\begin{aligned}\text{MSE}^\star &= \text{MSE}_1^\star + \text{MSE}_2^\star \\ &= N - \sum_{i=0}^{J_p-1}\frac{\lambda_i^2 c_i^2}{\lambda_i c_i^2 + \sigma^2}\end{aligned} \tag{7.12}$$

7.2 Power Allocation for Pilot Sequences

Consider the minimization problem

$$\begin{aligned}\min_{\mathbf{c}}\ & N - \sum_{i=0}^{J_p-1}\frac{\lambda_i^2 p_i}{\lambda_i p_i + \sigma^2} \\ \text{s.t.}\ & \sum_{i=0}^{J_p-1} p_i = \beta \\ & p_i \geq 0,\ \forall i\ \in\ \{0,\ldots,J_p-1\}\end{aligned} \tag{7.13}$$

The Lagrangian function is defined as

$$L(\gamma,\mu,\alpha) = N - \sum_{i=0}^{J_p-1}\frac{\lambda_i^2 p_i}{\lambda_i p_i + \sigma^2} + \mu\left(\sum_{i=0}^{J_p-1} p_i - \beta\right) - \sum_{i=0}^{J_p-1}\alpha_i p_i, \tag{7.14}$$

where $\gamma = \left[\gamma_0, \ldots, \gamma_{J_p-1}\right]^T$. The derivative over p_i is given by

$$\begin{aligned}\frac{\partial L(\gamma,\mu)}{\partial \gamma_i} &= \frac{\lambda_i^2\left(\lambda_i p_i + \sigma^2\right) - \lambda_i^2 p_i \lambda_i}{\left(\lambda_i p_i + \sigma^2\right)^2} + \mu - \alpha_i, \\ &= \frac{\lambda_i^2 \sigma^2}{\left(\lambda_i p_i + \sigma^2\right)^2} + \mu - \alpha_i \end{aligned} \tag{7.15}$$

To assign the optimal point, consider $\frac{\partial L(\gamma,\mu)}{\partial \gamma_i} = 0$. Thus

$$\begin{aligned}\frac{\lambda_i^2 \sigma^2}{\left(\lambda_i p_i + \sigma^2\right)^2} + \mu - \alpha_i &= 0 \\ \frac{\lambda_i^2 \sigma^2}{\left(\lambda_i p_i + \sigma^2\right)^2} &= \alpha_i - \mu \\ \lambda_i p_i + \sigma^2 &= \sqrt{\frac{\sigma^2}{\alpha_i - \mu}} \\ p_i &= \sqrt{\frac{\lambda_i^2 \sigma^2}{\lambda_i^2(\alpha_i - \mu)}} - \frac{\sigma^2}{\lambda_i} \\ p_i &= \sqrt{\frac{\sigma^2}{(\alpha_i - \mu)}} - \frac{\sigma^2}{\lambda_i}\end{aligned} \tag{7.16}$$

Because of the orthogonality condition from Karush Kuhn Tucker (KKT) conditions, we know that for any i γ_i and α_i cannot both be nonzero. If $\gamma_i \neq 0$ then $\alpha_i = 0$, and we get the equation Thus,

$$p_i = \begin{cases} K - \frac{\sigma^2}{\lambda_i} & \text{if } 0 < \mu \leq 1 \\ 0, & \text{otherwise,} \end{cases} \tag{7.17}$$

where $K = \sqrt{\frac{\sigma^2}{\mu}}$ is a constant. To calculate this parameter, we use the total power constraint. Thus,

$$\begin{aligned}\sum_{i=0}^{J_p-1} K - \sum_{i=0}^{J_p-1} \frac{\sigma^2}{\lambda_i} &= \beta \\ K &= \frac{1}{J_p}\left(\sum_{i=0}^{J_p-1} \frac{\sigma^2}{\lambda_i} + \beta\right).\end{aligned} \tag{7.18}$$

The Lagrangian multiplier is

$$\mu = -\frac{\sigma^2}{\frac{1}{J_p^2}\left(\sum_{i=0}^{J_p-1} \frac{\sigma^2}{\lambda_i} + \beta\right)^2}. \tag{7.19}$$

7.3 Minimum MSE Employing The Optimum Pilot Power Allocation

To calculate the minimum MSE, we firstly substitute (7.17) into the total power constraint stated in minimization problem (7.13). Thus

$$\begin{aligned}
\text{MSE}^{\star} &= N - \sum_{i=0}^{J_p-1} \frac{\lambda_i^2 p_i^{\star}}{\lambda_i p_i^{\star} + \sigma^2} \\
&= N - \sum_{i=0}^{J_p-1} \frac{\lambda_i^2 \left(K - \frac{\sigma^2}{\lambda_i}\right)}{\lambda_i \left(K - \frac{\sigma^2}{\lambda_i}\right) + \sigma^2} \\
&= N - \sum_{i=0}^{J_p-1} \frac{\lambda_i^2 \left(K - \frac{\sigma^2}{\lambda_i}\right)}{\lambda_i K} \\
&= N - \sum_{i=0}^{J_p-1} \frac{\lambda_i K - \sigma^2}{K} \\
&= N - \sum_{i=0}^{J_p-1} \lambda_i - \sum_{i=0}^{J_p-1} \frac{\sigma^2}{K} \\
&= N - \sum_{i=0}^{J_p-1} \lambda_i - \sum_{i=0}^{J_p-1} \frac{J_p \sigma^2}{\left(\sum_{i=0}^{J_p-1} \frac{\sigma^2}{\lambda_i} + \beta\right)}
\end{aligned}$$