Utilisation de modèles par rayons pour les systèmes multi-antennes massifs : analyse statistique et traitements numériques

Matthieu Roy

DOCTORAT BRETAGNE LOIRE / MATHSTIC

INSA RENNES

THESE DE DOCTORAT DE

L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES RENNES

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Télécommunications*

Par

Matthieu ROY

A ray-based Approach for Massive MIMO Systems:

Statistical Analysis and Digital Processing

Thèse présentée et soutenue à Rennes, le 23 Novembre 2020
Unité de recherche : IETR - UMR6164, INSA-Rennes, b<>com
Thèse N° : 20ISAR 19 / D20 - 19

Rapporteurs avant soutenance :

Laura Cottatellucci	Professeure, Université de Erlangen-Nuremberg, Allemagne
Jean-Marie Gorce	Professeur, INSA Lyon

Composition du Jury :

Président :	Alain Sibille	Professeur, Télécom Paris
Examinateurs :	Didier Le Ruyet	Professeur, CNAM, Paris
	Jean-Baptiste Doré	Docteur-Ingénieur de recherche, CEA LETI, Grenoble
Dir. de thèse :	Matthieu Crussière	Maître de conférences, INSA Rennes
Co-encadrant de thèse :	Stéphane Paquelet	Directeur laboratoire IA, B<>COM, Rennes

Remerciements

Tout d'abord, je tiens à remercier chaleureusement mes deux encadrants, Stéphane Paquelet et Matthieu Crussière, pour le temps, la patience et l'énergie qu'ils ont consacré à la supervision de cette thèse. Je voudrais en particulier remercier Stéphane de m'avoir fait partager sa passion, son expertise et son expérience des outils du traitement du signal et des communications numériques. Il a su me conseiller tout au long de la thèse et se rendre disponible lorsque j'avais besoin d'aide.

J'aimerai ensuite remercier les membres du jury d'avoir consacré du temps et de l'attention à la lecture du manuscrit. Leurs remarques et commentaires ont largement contribué à la qualité du document final.

Je voudrais aussi remercier mes collègues à b<>com, à commencer par Malo Manini et Napun, mes voisins de thèse, pour leur bonne humeur et leurs plaisanteries. Tous mes remerciements aussi à Luc Le Magoarou, qui a participé à mon encadrement et a partagé mon bureau. Je remercie aussi l'ensemble des membres du laboratoire Interface Réseau, et plus particulièrement Rodolphe Legouable, Cyril Collineau, Marc Lanoiselée, Patrick Savelli, Jean Dion et Pauline Desnos. Je remercie enfin chaleureusement Guillaume Vercasson qui m'a fait découvrir les subtilités de la modulation OFDM.

Je remercie aussi Jean-François Hélard, d'avoir animé les trois CSI qui ont ponctué la thèse, ainsi que Patrice Pajusco, avec qui j'ai pu échanger sur la problématique de séparation de clusters. Je remercie enfin Renaud Loison, qui a consacré une partie de son temps pour m'expliquer les effets de couplage dans les réseaux d'antennes

Je souhaite remercier tous les habitués des midi jeux, avec qui j'ai passé de très bons moments autour de jeux de société. Je souhaite aussi remercier les pratiquants de la salle d'Aïkido, et tout particulièrement notre professeur Philippe Lemonnier.

Je remercie chaleureusement mes parents, ma soeur, mes cousins ainsi que ma belle famille pour leur soutien pendant la thèse.

Je réserve enfin une place particulière dans ces remerciements à ma femme, Anne-Sophie, qui m'a soutenu et encouragé, tout au long de la thèse.

Abstract

Multi-Input-Multi-Output (MIMO) systems are key enabling technologies that leverage the propagation channel spatial diversity to increase both radio link reliability and data rate. MIMO schemes have already been introduced in modern wireless standards such as WiFi and 4G Long Term Evolution (LTE) with up to 8 antennas at the base station. The envisioned antenna array sizes in 5G New Radio (NR) physical layer are much larger to take advantage of the new massive MIMO paradigm. Dedicated signal processing techniques that scales well with the array sizes have to be designed to ensure optimal performance at constrained complexity. Those algorithms leverage the channel spatial domain properties to perform efficient channel estimation, quantized feedback and precoding.

In this work we have investigated the benefits of using a ray-based channel model to enhance massive MIMO baseband processing. To this end, either deterministic ray-based channel models or Saleh-Valenzuela stochastic models have been used.

The first set of contributions relates to MIMO channel statistical characterization and analysis. We performed an in-dept capacity analysis of single-user MIMO systems under perfect precoding conditions. We have derived a novel formula to quantify the radio link reliability gain of MIMO channels, also called the *channel hardening* phenomenon. We have also studied the spatial multiplexing capacity improvement. In a second time, statistical properties of Saleh-Valenzuela channels have been characterized. Based on those results, we have studied the cluster estimation and separation problem on both theoretical and algorithmic standpoints.

A fundamental problem in MIMO communications is that the information required to perform precoding at the transmit side is only available at the receive side. A workaround consists in deducing the channel state information required for downlink precoding from uplink measurements. We have leveraged the results derived beforehand to evaluate the performance of such approach.

Channel estimation is challenging in massive MIMO systems as the number of parameters to estimate is large. Dedicated techniques take advantage of angular sparsity for better efficiency. We optimized and characterized key steps of one of those algorithms.

Table of Contents

List of Figures

Acronyms

CRLB Cramér-Rao Lower Bound.

CSI Channel State Information.

DOA Direction of Arrival.

DOD Direction of Departure.

EM Expectation-Maximization.

ESPRIT Estimation of Signal Parameter Via Rotational Invariance Technique.

FDD Frequency Division Duplex.

GLRT Generalized likelihood ratio test.

IoT Internet of Things.

LOS Line of Sight.

LTE Long Term Evolution.

MAP Maximum a Posteriori.

MIMO Multi-Input-Multi-Output.

MISO Multi-Input-Single-Output.

MRT Maximum Ratio Transmission.

MSE Mean Square Error.

MUSIC Multiple Signal Classification.

NLOS Non Line Of Sight.

NR New Radio.

OFDM Orthogonal Frequency Division Multiplexing.

OMP Orthogonal Matching Pursuit.

OS Oversampling.

PDP Power Delay Profile.

PMI Prefered Matrix Indicator.

RBG Reduction of Beamforming Gain.

RI Rank Indicator.

RMa Rural Macrocell.

ROC Receiver Operating Characteristic.

SAGE Space Alternating Generalized Expectation-maximization.

SE Spectral Efficiency.

SISO Single-Input-Single-Output.

SNR Signal to Noise Ratio.

TDD Time Division Duplex.

UCA Uniform Circular Array.

UE User Equipments.

ULA Uniform Linear Array.

UMa Urban Macrocell.

UMi Urban Microcell.

UPA Uniform Planar Array.

UWB Ultra Wide Band.

WSS Wide Sense Stationary.

Resumé en Français

Les réseaux de communication actuels doivent répondre à une demande toujours plus importante de débit et de qualité de service. Ce besoin correspond à l'émergence de nouveaux cas d'usage, comme la diffusion de contenus en haute résolution, les applications industrielles et médicales ainsi que l'Internet des Objets.

Le débit maximum théorique atteignable (appelé la capacité du canal C) entre une station de base et un périphérique connecté au réseau est le produit de deux termes, la largeur de bande de fréquence utilisée pour transmettre et l'efficacité spectrale

$$C = B \cdot SE \quad bits/s$$

où la largeur de bande B est la quantité de ressources spectrales utilisées pour la transmission. La capacité du système à utiliser efficacement cette ressource est quantifiée par l'efficacité spectrale SE. A l'heure actuelle, les ressources spectrales restantes s'amenuisent. Pour augmenter la capacité des systèmes communicants, deux solutions se profilent : explorer des régions inutilisées du spectre radiofréquence ou améliorer l'efficacité spectrale des communications.

Les systèmes de transmission mobiles utilisent le plus souvent les fréquences sous $6GHz$, où les conditions de propagation sont plus adaptées à ce cas d'usage. Cependant cette portion du spectre est sur-exploitée et chère, contrairement aux bandes millimétriques. Ces dernières sont de potentielles candidates pour le développement de futurs systèmes de communication.

Les systèmes MIMO sont un composant technologique clé pour l'augmentation de l'efficacité spectrale des réseaux de communication. Ils exploitent la dimension spatiale de la propagation pour transmettre d'avantage de données sur les mêmes ressources temps/fréquence. Contrairement aux systèmes Single-Input-Single-Output (SISO) qui émettent et reçoivent les ondes de façon omnidirectionnelle, ces systèmes peuvent adapter leur diagramme d'antenne au canal de propagation de façon à optimiser la transmission. Ce traitement, appelé *précodage*, est crucial pour le fonctionnement optimal du système de communication. Il est cependant basé sur une connaissance a priori du canal de com-

munication, qui n'est pas triviale à obtenir.

Cette connaissance est naturellement acquise au récepteur. Cependant, l'opération de précodage est tant qu'à elle faite à l'émetteur. Il faut donc ramener tout ou partie de cette information du récepteur à l'émetteur via une voie de retour.

Cette approche est bien adaptée aux systèmes utilisant de petits réseaux d'antennes (jusqu'à 8 antennes). Au delà, le nombre de paramètres à ramener à l'émetteur devient trop important. Des méthodes alternatives de précodage se basent sur les propriétés de réciprocité angulaire du canal de propagation pour réduire le poids de cette voie de retour.

L'étape d'estimation de canal est aussi complexe dans les systèmes MIMO massifs. En effet le nombre de paramètres à estimer est proportionnel à $N_r \times N_t$ et devient particulièrement important dans les systèmes multi-antennes. Il est une fois encore intéressant de considérer les propriétés du canal dans le domaine angulaire de façon à réduire le nombre de paramètres à estimer.

De manière générale, les systèmes MIMO massifs nécessitent de développer de nouvelles techniques de traitement du signal utilisant la dimension angulaire de la propagation. Cela passe par l'introduction de modèles physiques par rayons. Dans ce manuscrit nous étudions ces modèles par rayons et leurs applications au précodage et à l'estimation de canal.

Modélisation du canal MIMO

Ce premier chapitre permet d'introduire les notations, les concepts, la modulation Orthogonal Frequency Division Multiplexing (OFDM) ainsi que le modèle de canal utilisé tout au long du manuscrit. Il pose les bases sur lesquelles vont s'articuler l'ensemble des contributions suivantes. A partir d'une modélisation géométrique par rayons de la propagation, définie par des directions de départ, des directions d'arrivée, des retards et des gains de rayons, nous construisons un modèle mathématique qui peut se décliner sous différentes formes (matricielle, tensorielle ou vectorisée) suivant le cas d'utilisation.

Enfin nous introduisons le modèle stochastique Saleh-Valenzuela. D'après ce modèle, les rayons du canal sont rassemblés en un faible nombre de clusters définis par des directions et retards proches.

Capacité des systèmes MIMO

Dans ce second chapitre, nous étudions la capacité (débit maximal théorique) des canaux de communications multi-antennes. Ce chapitre est l'occasion d'introduire les concepts de capacité, d'évanouissement dans les canaux multi-trajet et de rigidification du canal dans les systèmes MIMO. Dans ce contexte, nous proposons une nouvelle formule pour quantifier la rigidification du canal dans les canaux par rayons, donnée par l'équation (2.11). Les différents termes de cette formule sont interprétés et illustrés par des exemples. Cette équation peut enfin se voir comme une généralisation de la formule (2.7), valide uniquement pour les modèles Gaussiens, qui modélisent un canal multi-trajet asymptotiquement riche.

Représentation efficace des clusters

Ce troisième chapitre est consacré à la caractérisation des canaux générés avec le modèle Saleh-Valenzuela. Ceux-ci sont constitués de clusters de trajets, dont on va caractériser au premier et second ordre les moments statistiques. En particulier, nous montrons que la matrice de covariance du cluster peut se décomposer en matrices fréquentielles et angulaires. La formule décrivant la matrice de covariance fréquentielle du cluster Saleh-Valenzuela est alors redémontrée suivant une méthode plus simple que celle de l'état de l'art. Nous analysons ensuite le nombre de coefficients nécessaires pour représenter un canal Saleh-Valenzuela, projeté sur une base parcimonieuse. Nous avons comparé l'efficacité de la représentation dans le domaine Fourier à la représentation optimale utilisant les vecteurs propres de la matrice de covariance, aussi bien dans les domaines fréquentiel que angulaire. Enfin, nous comparons le nombre de coefficients nécessaire pour représenter un canal de propagation complet, en utilisant une modélisation classique de type matrice de canal ou une représentation par rayons. Les abaques résultantes permettent de savoir à partir de quel nombre d'antennes la représentation par rayons devient intéressante.

Estimation et séparation des clusters

Les trajets qui constituent les canaux de propagation sont rassemblés en un faible nombre de clusters correspondant à des entités réfléchissant les ondes dans l'environnement. Le

problème de séparation et d'identification des clusters qui constituent un canal de propagation est fondamentalement mal posé sans l'introduction d'une définition du cluster. Dans cette partie, nous mettons à profit notre connaissance du moment d'ordre 2 d'un cluster Saleh-Valenzuela pour lui donner une définition mathématique claire faisant appel à un modèle paramétrique. En se basant sur ce modèle, nous évaluons et interprétons la borne de Cramér-Rao de façon à vérifier le bon conditionnement du problème ainsi que les performances des estimateurs.

La séparation de clusters est un problème à part entière qui pourrait faire l'objet d'un sujet de thèse. Pour des raisons de temps, nous nous sommes uniquement attardés sur la dimension fréquentielle de la propagation.

Précodage utilisant l'information de la voie montante en FDD

Comme précédemment évoqué, dans les systèmes multi-antennes, la détermination du précodeur optimal nécessite la connaissance des conditions de propagation entre émetteur et récepteur. Cette information est le plus souvent ramenée par le biais d'une voie de retour. Quand le nombre d'antennes à l'émetteur augmente, conformément au nouveau paradigme du MIMO massif, la quantité d'information à ramener à l'émetteur par cette voie de retour augmente de façon considérable.

On s'intéresse à des systèmes de communications en duplex complet, composés d'une voie montante (transmission de données du mobile à la station de base) et d'une voie descendante (transmission de données de la station de base au mobile). On s'intéresse au mode FDD (division fréquentielle) caractérisé par des voies montante et descendante placées sur des bandes de fréquence adjacentes. De façon à réduire le poids de cette voie de retour, des chercheurs ont proposé d'extrapoler le canal de propagation de la voie montante (mesuré par la station de base) sur la voie descendante, en utilisant la réciprocité angulaire. Ce processus fonctionne bien lorsque le mobile est en ligne de vue de la station de base. Cependant les performances se dégradent fortement lorsque les phénomènes de multi-trajets sont prédominants.

Dans cette partie, nous nous sommes intéressés à l'erreur d'extrapolation du canal de la voie montante sur la voie descendante. Les rayons qui constituent les canaux multi-trajets sont rassemblés en clusters. Cette proximité les rend difficilement séparables et limite la portée fréquentielle de l'extrapolation. Dans ce chapitre, nous avons évalué cette

portée pour des clusters Saleh-Valenzuela. Encore une fois, nous avons mis à contribution notre connaissance du moment d'ordre 2 d'un cluster Saleh-Valenzuela. Cette étude est complétée par une analyse de la perte de capacité à travers la métrique de la perte de gain de formation de faisceau.

Estimation de canaux MIMO en ondes millimétriques

Les bandes millimétriques sont aujourd'hui peu exploitées. La physique de la propagation des ondes y est moins favorable car le canal souffre de pertes en espace libre élevées, accentuées sur les longues distances par l'absorption des ondes par l'atmosphère (en particulier à 60 GHz). Les tailles d'antennes, proportionnelles à la longueur d'onde, sont plus petites. Les phénomènes de multi-trajets sont moins importants et l'ouverture angulaire des clusters est plus faible. Le canal est ainsi souvent considéré comme *creux*, car il ne contient qu'un faible nombre de rayons spéculaires bien définis dans les domaines des retards et angulaire.

La combinaison des ondes millimétriques, associées à la technologie MIMO massif est particulièrement intéressante. La compacité des antennes permet de constituer des réseaux comportant plus d'éléments rayonnants dans des volumes restreints. Le gain d'antenne considérable gagné par l'utilisation de ces réseaux vient compenser les fortes pertes en espace libre des ondes millimétriques.

L'estimation de canal est une étape importante et nécessaire pour la démodulation des échantillons. Elle consiste à caractériser le canal de propagation entre émetteur et récepteur, pour le retrancher ensuite aux échantillons de données reçus avant de les décoder.

Le nombre de coefficients à estimer augmente avec $N_r \times N_t$. L'estimation de canal devient donc particulièrement complexe lorsque les dimensions des réseaux d'antennes augmentent. Une solution consiste à prendre en compte la structure creuse du canal de propagation de façon à réduire le nombre de paramètres à estimer. Dans ce contexte, des chercheurs ont utilisé l'algorithme OMP de façon à extraire séquentiellement les rayons spéculaires qui constituent le canal. Dans le cadre de cette thèse, nous avons apporté des améliorations à cet algorithme générique de façon à l'adapter aux caractéristiques du problème d'estimation de canal. Nous avons en particulier travaillé sur le critère d'arrêt et la phase d'estimation de rayon de l'algorithme.

INTRODUCTION

Context and motivations

Currently existing wireless network infrastructure has to keep up with the ever increasing demand for data rate and quality of service, fueled by new use cases such as high resolution video streaming, medical and industrial applications and Internet of Things (IoT). The maximum achievable data rate (also called the capacity C) between a base station and a user is the product of two terms, the available bandwidth and the spectral efficiency

$$C = B \cdot SE \quad bits/s$$

where the bandwidth B is the amount of spectral resource used to transmit information and the spectral efficiency SE relates to how efficient we are at exploiting this resource. As the remaining amount of spectral resources shrink, increasing the capacity requires to either explore unused portions of the radio spectrum or to increase the spectral efficiency using advanced wireless techniques.

Mobile broadband communication systems where first deployed at frequencies below 6 GHz, characterized by highly favorable propagation conditions for this application. However this section of the spectrum is overused and expensive. In contrast, a large amount of available spectrum lies above 6 GHz, in particular in millimeter wave bands. First steps towards millimeter wave mobile broadband have already been carried out. In standardization, the numerology (frame structure) used in 4G has been extended in 5G to adapt to the extended bandwidth and propagation characteristics of millimeter waves. At the same time, telecommunication regulators started allocating spectrum in the 26 GHz (EU) and 28 GHz (USA) bands. The 60 GHz unlicensed band is also considered for small cells deployments. The demonstrator designed within the M5HESTIA project targets this particular band.

MIMO systems are a key technological component to increase the spectral efficiency of wireless networks. They exploit the spatial dimension of the propagation to transmit more data over the same time/frequency resources, thus increasing the maximum achiev-

able data rates beyond the theoretical capacity of SISO systems. MIMO systems can shape their antenna pattern based on the propagation characteristics of the channel between transmitter and receiver to optimize the capacity (the *precoding* step) while SISO systems lack this adaptability. Precoding, the technique's most crucial step, has to be designed using prior knowledge on the propagation channel. Acquiring this Channel State Information (CSI) is a key scientific problem.

Operating the system in TDD is the simplest way to solve this problem. The same frequency band is used at transmit and receive sides. The channel estimated from received samples can be reused to design the precoding. However in FDD operations, the transmit and receive bands are different. Only the user equipment receiver can estimate the propagation channel. This CSI has to be sent back to the base station through a feedback loop to design the precoder. However as the dimension of the antenna array scales up, the amount of information to feed back to the transmitter gets overwhelming. Meanwhile scaled up antenna arrays increase the transmission system angular resolution and enables separation of the multipath components that constitute the propagation channel. Alternative precoding techniques leverage the channel angular properties (which does not depend on the frequency band considered) to reduce the feedback requirements.

Channel estimation is also challenging in MIMO systems. As the antenna array sizes increase, the number of parameters that should be estimated gets overwhelming. In order to reduce the estimation process complexity, it is beneficial to leverage the sparsity of the channel in the angular domain and only estimate the parameters of the few rays that govern the propagation instead of the resulting mixture of these.

In general, when moving from small scale MIMO systems to the massive MIMO regime, dedicated signal processing has to leverage the angular sparsity of the propagation channel to reach optimal performance with constrained complexity. In this thesis we studied novel precoding and channel estimation techniques that leverage the angular sparsity of the massive MIMO propagation channel.

Outline of the Thesis

The thesis is made of two main parts; the first one relates to MIMO channel modeling and characterization (Chapters 1, 2, 3, 4) while the second exploits those results to propose solutions for precoding in FDD (Chapter 5) and for MIMO channel estimation (Chapter 6).

Chapter 1 relates to channel modeling. In all the sections of this thesis we adopt a model based approach to derive formulas and to conceive algorithms. We use a realistic ray-based model inspired from ray-tracing techniques to provide the best level of generality. In particular we exploit extensively the Saleh-Valenzuela stochastic channel model which is built upon the ray-based model. This model is widely used in the literature and has been extensively characterized in measurement campaigns.

Within this context we have performed an in-depth capacity analysis of multi-antenna systems, presented in Chapter 2. First we analyze the *channel hardening* phenomenon and propose a new expression for ray-based models, presented at the international conference WiMob 2018 and detailed in Section 2.2. Secondly we shift our attention to spatial multiplexing. We present a way to reshape the capacity of MIMO channels under perfect precoding conditions that depends on a fixed number of parameters for any transmission rank in Section 2.3.

In a later stage, we analyze the properties of channels generated using the Saleh-Valenzuela model in Chapter 3. First we prove the expressions of the statistical covariance functions for both a single Saleh-Valenzuela cluster and the whole channel in Section 3.1. Those expressions are then reused in the next chapters. Knowing the second order statistics, we perform a basis decomposition of the channel and compare the usually preferred Fourier decomposition and the optimal eigendecomposition in Section 3.2. Finally we compare the number of coefficients required to describe the channel using either a cluster-based representation or a matrix-based representation in Section 3.3. This contribution is summarized in a conference paper presented at GLOBECOM 2019.

We study the cluster estimation and separation problem in Chapter 4. To this end we propose a novel approach that exploits the second order statistics of Saleh-Valenzuela channels previously calculated in Chapter 3. The performance of this procedure is measured experimentally then compared with the theoretical lower bounds derived from the model in Section 4.3.

Then we investigate the potential of using information from the Uplink channel to precode on the Downlink channel. This is the expected approach for MIMO systems operating in TDD as the transmission band is the same for Uplink and Downlink and the estimated channel can be directly used for precoding. However, this is not sufficient when Uplink and Downlink operate on different bands. We propose in Chapter 5 an analysis of a precoding scheme using a channel extrapolation procedure from Uplink to Downlink for FDD. The results have been published at the international conference PIMRC 2020.

Then we shift our attention towards channel estimation for massive MIMO systems in Chapter 6. Millimeter wave propagation channels are sparse. In order to exploit this sparsity, new algorithms using tools from compressed sensing are proposed in the literature. The key step of those new channel estimation techniques is the specular ray detection and estimation procedure. We propose in Section 6.4 a scalable specular ray detection and estimation technique. It can be integrated seamlessly in novel channel estimation algorithms, as evidenced in Section 6.5.

The first part of this work has been completed within the European Project One5G. The goal of this project was to propose and investigate new techniques for the lower layers (MAC and PHY) for 5G wireless transmission systems.

Publications

Conference articles

— Matthieu Roy et al., « MIMO Channel Hardening for Ray-based Models », *in*: *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Limassol: IEEE, Oct. 2018 (best paper award)

— Matthieu Roy, Stephane Paquelet, and Matthieu Crussiere, « Degrees of Freedom of Ray-Based Models for mm-Wave Wideband MIMO-OFDM », *in*: *GLOBECOM 2019 - 2019 IEEE Global Communications Conference*, Waikiloa Village, 2019

— Matthieu Roy, Stephane Paquelet, and Matthieu Crussiere, « Cluster Extrapolation for FDD Downlink MIMO Precoding », *in*: *PIMRC 2020*, London: IEEE, Aug. 2020

Others

— Matthieu Roy, Stephane Paquelet, and Luc Le Magoarou, *Massive MIMO: Analytic Capacity, Topology of the antenna arrays*, One5G F2F Meeting, Malaga, Sept. 2018

— Matthieu Roy et al., *Channel Hardening for Ray-Based Models*, Journée des Doctorants, Vannes, June 2018

— Matthieu Roy and Stephane Paquelet, « Procédés et dispositifs de détection d'un signal et d'estimation de canal, et programme d'ordinateur associé » (to be filed)

Part I

Statistical Analysis of MIMO Channels

Chapter 1

Channel Modeling for MIMO Systems

The propagation channel alters the signal traveling between the transmitter and the receiver. Accessing its properties is necessary to determine how the transmitted data should be processed to be decoded efficiently at the receiver. In this chapter, we first introduce the mathematical notations which will be used throughout the manuscript to represent any propagation channel, namely regarding MIMO architectures.

Stochastic channel models aim at reproducing the properties of a particular environment in a statistical sense. The modeled environment is defined by a set of random variables whose realizations span all the possible propagation channels of this environment. This approach is of particular interest for theoretical analyses to predict analytically or using simulations the behavior of a transmission system in a characterized environment. In this manuscript we will cross paths with the two main kinds of MIMO channel statistical models : the Gaussian model and the Geometry-based stochastic models.

We first introduce the Gaussian model, that is a direct matrix generalization of its classical scalar SISO channel model counterpart. It works well when modeling channels with asymptotically high scattering. The mathematical simplicity of this model make it particularly attractive for theoretical papers.

The Geometry-based stochastic models are built upon a ray-based deterministic channel model. Electromagnetic waves are modeled by rays that start from the transmitter, bounces on the environment to reach the receiver. This model is parametrized by complex gains, delays and angles and provides a framework which is valid for any realization. In order to describe accurately a particular kind of environment (indoor, outdoor, rural, urban, ...), the ray parameters are given statistical laws that encompass all propagation medium properties.

Finally, we introduce the generalized tensor and vector channel notations that are useful when considering channels that span over more that two dimensions (for instance

several transmit, receive antennas and multiple subcarriers).

1.1 Mathematical representation

The main propagation mechanism is multipath fading. The received signal consists of a superposition of the main signal with delayed echoes coming from reflexions and diffractions on features of the environment. Figure 1.1 displays a propagation channel between a base station and a user equipment consisting of a Line of Sight (LOS) path and a reflection path from a building. In the general case, the channel is represented as a filter with impulse response

$$h(t) = \sum_{p=1}^{P} \beta_p \delta(t - \tau_p)$$

where β_p denotes the complex amplitude of the p^{th} echo, that depend on the scatterer

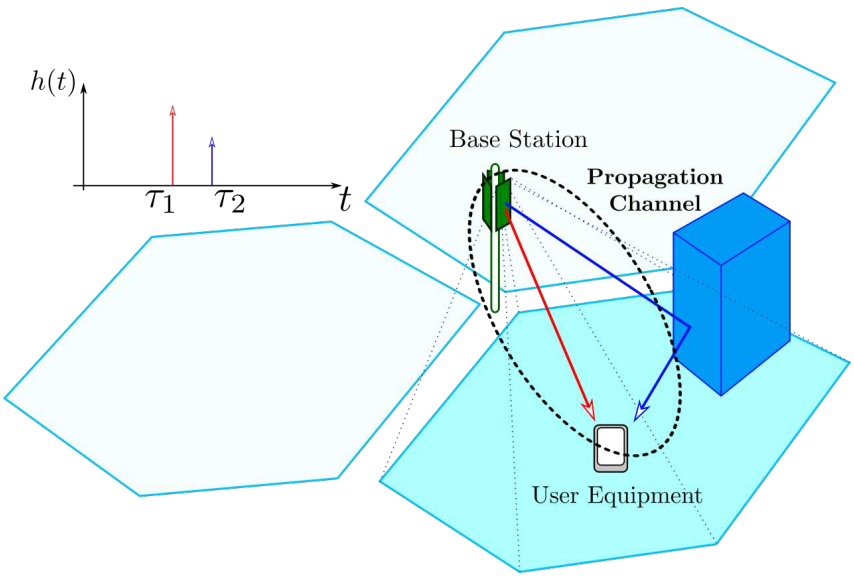

Figure 1.1 – The multipath propagation channel.

material and the angle of incidence of the ray. τ_p denotes the delay of the p^{th} echo that depends on the time of flight between transmitter and receiver along this path. In the frequency domain, the channel is characterized by the complex function

$$H(f) = \mathcal{TF}(h)(f) = \sum_{p=1}^{P} \beta_p \, \mathrm{e}^{-2\pi j f \tau_p} \, . \tag{1.1}$$

When using a transmission bandwidth small enough in comparison with the channel coherence bandwidth (*narrow-band channel*), in the vicinity of a specific carrier frequency f, we can consider that $H(f)$ is constant, *ie* the channel is frequency flat. The received symbols y are simply the emitted symbols x multiplied by the channel gain $H(f)$ plus some thermal noise n

$$y = H(f)x + n \tag{1.2}$$

In Equation (1.1), the echoes can sometimes add up constructively (the received signal power is amplified) or destructively (the received signal power is diminished). This phenomenon is called *multipath fading* or *fast fading*. The average distance between two destructive sums is half a wavelength, that is several centimeters only in commonly used sub 6 GHz bands, 11 mm at 26 GHz and only 5 mm at 60 GHz. Considering a narrowband channel as modeled in Equation (1.2), the received power can vary significantly over short intervals of time because of multipath fading.

Modern communication standards (5G NR, 4G LTE, WiFi) are based on the OFDM modulation technique. The data are transmitted over a set of N_f subcarriers equally spaced of Δf and centered on the carrier frequency. This particular modulation technique can deal with multipath channels efficiently. Indeed, each subcarrier is equivalent to a narrow-band channel and is orthogonal to the other subcarriers. In this case the OFDM symbol duration is $T_s = 1/\Delta f$. The channel applies multiplicatively on each subcarrier k as in Equation (1.2). This enables low complexity channel estimation by only dividing the received samples y by the known emitted signal x (called *pilot signal*) to estimate the channel. The channel on subcarrier k placed at frequency f_k is given by

$$y_k = H(f_k)x_k + n_k \tag{1.3}$$

where y_k denotes the received symbol on subcarrier k obtained by multiplying the transmitted symbol x_k by the channel gain at frequency f_k, $H(f_k)$. Those samples are also mixed with thermal noise n_k typically modeled as a complex centered Gaussian distributed variable.

Hence considering the propagation channel on a specific subcarrier at frequency f_k of an OFDM communication system, the radio link between the transmit antenna and the receive antenna is simply modeled in the frequency domain by a complex gain $H(f_k)$. Now increasing the number of antennas (N_t transmit antennas, N_r receive antennas) at both the transmit and receive sides, the channel is then modeled as a complex-valued matrix,

called *channel matrix*, with coefficients

$$\mathbf{H}(f_k) = \begin{bmatrix} H_{1,1}(f_k) & H_{1,2}(f_k) \\ H_{2,1}(f_k) & H_{2,2}(f_k) \end{bmatrix}$$

where each coefficient $H_{i,j}(f_k)$ matches the frequency domain channel gain of the radio link from the jth transmit antenna to the ith receive antenna (see Figure 1.2).

As we focus our analysis on the OFDM waveform, we will always (otherwise explicitly mentioned) work in the frequency domain. We adopt the common convention that represent vectors as lower case bold symbols and matrices as upper case bold symbols. We denote vector and matrix elements as lower case and upper case standard symbols.

Numerology

The OFDM waveform is parametrized by the number of subcarriers N_f, the subcarrier spacing Δf and the total occupied bandwidth B. This set of parameters is called the *numerology*. In LTE only one numerology, optimized for centimeter waves, has been defined. To cope with millimeter bands, additional numerologies have been defined in the new standard 5G NR.

Standard	N_f	Δf	B	Reference
LTE	1200	$15kHz$	$20MHz$	[1]
5G NR Num #0	3000	$15kHz$	$50MHz$	[2]
5G NR Num #1	3000	$30kHz$	$100MHz$	[2]
5G NR Num #2	3000	$60kHz$	$200MHz$	[2]
5G NR Num #3	3000	$120kHz$	$400MHz$	[2]

The received symbols $\mathbf{y_k}$ on the receive antennas (column vector of size N_r) of an OFDM subcarrier at frequency f_k are then given by the equation

$$\mathbf{y_k} = \mathbf{H}(f_k)\mathbf{x_k} + \mathbf{n_k} \tag{1.4}$$

where $\mathbf{H}$ is the channel matrix (size $N_r \times N_t$). The complex valued vector $\mathbf{x_k}$ contains the transmitted symbols on each of the antennas (vector of size N_t). $\mathbf{n_k}$ is the noise vector at the receiver (size N_r).

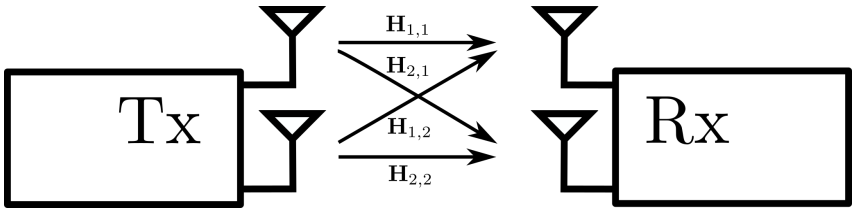

Figure 1.2 – The MIMO channel matrix.

1.2 The Gaussian model

The most commonly used multipath propagation model is the *Gaussian channel model*. It models well the distribution of the channel gain in rich scattering environments (multiple echoes with similar power). It is also called *Rayleigh fading channel* because the amplitude of a complex Gaussian distributed variable is Rayleigh distributed.

1.2.1 The SISO Gaussian model

As stated previously, a multipath channel from a transmitting antenna to a receiving antenna can be written at a specific frequency f as

$$H(f) = \sum_{p=1}^{P} |\beta_p| \, \mathrm{e}^{j\phi_p}$$

where, for the sake of simplicity, we merged the phase of β_p and the phasor $\mathrm{e}^{-2\pi j f \tau_p}$ within the independent $[0, 2\pi]$-uniformly distributed phases ϕ_p. Assuming that all the gains $|\beta_p|$ have identical amplitude, the Central Limit Theorem states that the distribution of the channel gain converges to a Gaussian distribution

$$H(f) = \sum_{p=1}^{P} |\beta_p| \, \mathrm{e}^{j\phi_p} \xrightarrow[P\to\infty]{} \mathcal{CN}(0, \sigma^2)$$

For SISO communications, the asymptotic distribution is easily reached in channel measurements [3]. This model is particularly useful because it offers analytically tractable expressions and it fits well to actual measurements in high scattering situations.

1.2.2 The MIMO Gaussian model

The MIMO Gaussian model is a natural extension of the SISO Gaussian channel model presented previously. It is widely used in theoretical studies as well as numerical evaluations. In this model, each radio link $H_{i,j}$ follows a complex Gaussian distribution. Thus the vector vect($\mathbf{H}$) is Gaussian distributed as

$$\text{vect}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$$

where $\mathbf{R}$ denotes the covariance matrix of the channel matrix coefficients. This model is also called the correlated Rayleigh fading channel model in the literature. Several variations around this main model have been proposed such as :

- **the Gaussian i.i.d. model :** A simplified model where channel matrix coefficients are independent identically distributed ($\mathbf{R} = \sigma^2\mathbf{I}$).
- **the Kronecker Model :** Covariance matrices at transmitter and receiver sides are assumed to be separable ($\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_{tx}} \otimes \mathbf{R_{rx}})$).

This model offers a simple framework that can provide analytically tractable expressions for algorithms and theoretical capacity analysis. However it has several limitations. First the influences of propagation conditions and antenna array topologies are not evidenced. As well as the SISO Gaussian model, the MIMO generalization requires asymptotically high scattering environments to be valid. While the convergence to the asymptotic distribution happens quickly for SISO channels, the scattering must be richer as the number of antennas increase to verify the model. This is evidenced in the research paper [3] based on channel measurements performed at the campus of Brigham Young University.

1.3 Ray-Based Channels

The received signal consists of a constructive or destructive sum of echoes of the main signal that bounces on features of the surrounding propagation environment. This is modeled as a sum of phasors in Equation (1.1). Here we extend this notation for MIMO channels by adding an angular dimension to the propagation, according to directions of departure and arrival of rays, onto and from the antenna set. In that perspective, it is useful to introduce a proper 3D spatial basis and define on it the geometrical representation of antenna arrays, and rays.

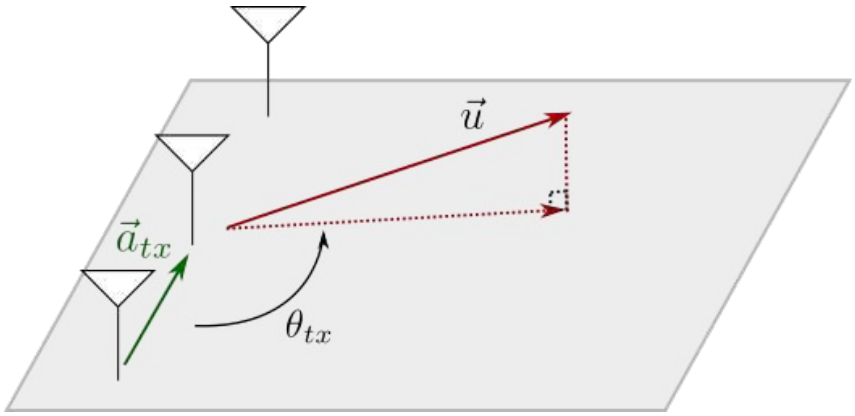

Figure 1.3 – The Uniform Linear Array.

1.3.1 Antenna arrays

The antennas at transmit and receive sides are arranged in geometrically organized structures called antenna arrays. Throughout this manuscript, N_t antennas are used in the transmit array and N_r antennas are used at the receive side. The positions of antennas within the transmit array are identified by the 3D vectors $\vec{a}_{tx,j}$ ($\vec{a}_{rx,i}$ at the receive side, respectively). Using those notations, we can model any arbitrary antenna array shape.

In particular in this thesis we will consider particular structures of antenna arrays, namely ULA where antennas are uniformly spaced along a line (Figure 1.3), Uniform Planar Array (UPA) where antennas are uniformly placed on a rectangle (Figure 1.4) and more rarely Uniform Circular Array (UCA) where antennas are uniformly spaced on a circle. We usually use a half-wavelength antenna spacing, unless explicitly mentioned otherwise.

In a ULA, the antenna positions are defined by

$$\vec{a}_{tx,j} = (j-1)\vec{a}_{tx} \tag{1.5}$$

where $\vec{a}_{tx}$ is the array primitive vector (it defines the smallest structure that repeats itself). The distance between two antennas is given by $\|\vec{a}_{tx,j}\|$. Note that the antennas coordinates defined by the vectors $\vec{a}_{tx,j}$ are always taken relatively to an origin point. In Equation (1.5), the origin is antenna $j = 1$ ($\vec{a}_{tx,1} = \vec{0}$). We can actually set the array origin arbitrarily, for instance in the middle of the array leading to $\vec{a}_{tx,j} = (j-1-(N_t-1)/2)\vec{a}_{tx}$. We checked that the results provided in this thesis are always independent with regards to the origin point.

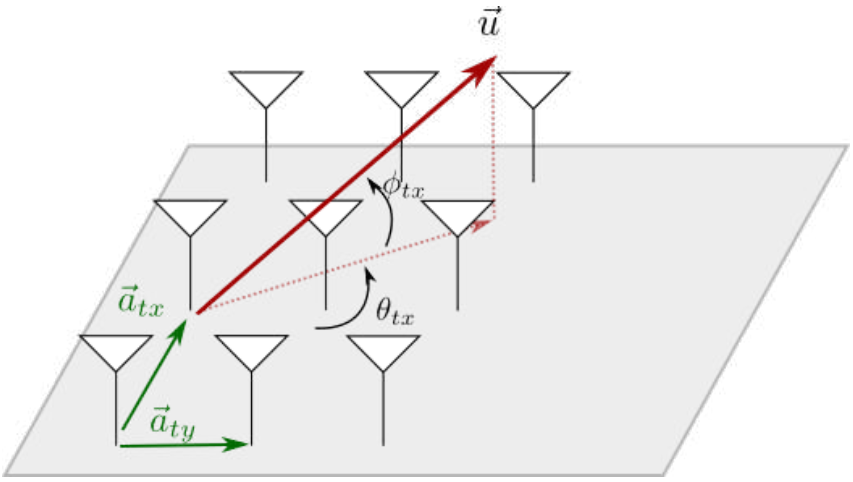

Figure 1.4 – The Uniform Planar Array.

In a UPA, the antennas are arranged in a rectangular fashion, defined by the antenna count on the main axis N_{tx} and on the side axis N_{ty} ($N_t = N_{tx}N_{ty}$). Each antenna can be identified either by a single index j or by two indexes j_x, j_y denoting the antenna coordinates in the rectangular grid where $j = j_x + N_{tx}j_y$. The antenna positions for the UPA are then given by

$$\vec{a}_{tx,j} = (j_x - 1)\vec{a}_{tx} + (j_y - 1)\vec{a}_{ty} \tag{1.6}$$

1.3.2 Propagation ray definition

The direction of departure (direction of arrival, respectively) of the pth ray of the channel is represented by a unit norm 3D vector $\vec{u}_{tx,p}$ ($\vec{u}_{rx,p}$, respectively). Each ray p is also characterized by a complex gain c_p.

A direction in a 3D space is naturally defined by two parameters (azimuth and elevation). The ULA only permits the measurement of one parameter, the angle between the direction $\vec{u}$ and the array main axis $\vec{a}_{tx}$ (azimuth if the array main axis is horizontal, elevation if the array main axis is vertical). Both parameters can be measured simultaneously when using an UPA.

Those notations are summed up in Figure 1.5 that features a propagation environment with 3 rays. The signal traveling on each ray impinge on the antennas with a path difference that generates a phase difference in baseband processing. The path difference induced by a ray of direction $\vec{u}$ between antenna i and the origin point of the array (see Section 1.3.1) is $\vec{a}_{tx,j} \cdot \vec{u}$ (showcased in the zoomed circle of Figure 1.5).

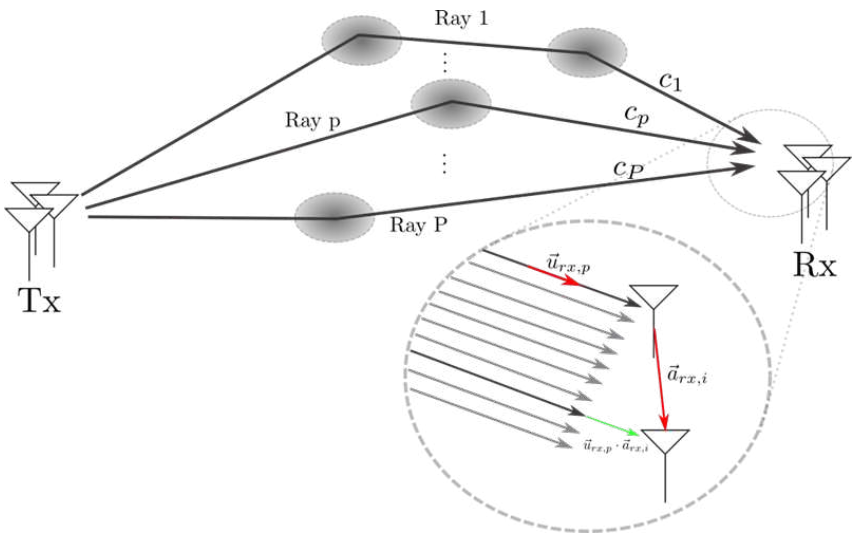

Figure 1.5 – The multipath MIMO channel.

This effect generates phase differences between the signals transmitted and received over the antenna arrays. The contribution of the pth ray to the element of the channel matrix $H_{i,j}$ (propagation channel between the emitting antenna j and the receiving antenna i) is given as

$$H_{i,j} = \sum_{p=1}^{P} \beta_p \, \mathrm{e}^{-2\pi j f \tau_p} \, \mathrm{e}^{-2\pi \frac{\vec{a}_{rx,i} \cdot \vec{u}_{rx,p}}{\lambda}} \, \mathrm{e}^{2\pi \frac{\vec{a}_{tx,j} \cdot \vec{u}_{tx,p}}{\lambda}}$$

Note that this model assumes a planar wavefront of each ray at both transmitter and receiver sides. This means that all scatterers are far enough, with respect to the array sizes, from both transmit and receive antenna arrays. The plane wave assumption breaks when the array sizes are no longer negligible with respect to the distance to the nearest scatterer. The border between both regimes is given by the Rayleigh distance. More information on the plane wave hypothesis and on the spherical wave model (generalized model without the plane wave hypothesis) can be found in the papers [4, 5, 6].

We can use a more generic and compact way to describe the contribution of a ray using structures called *steering vectors*. Those vectors store the phases differences between the antennas for a given direction of departure or arrival. The transmit steering vector is given

by

$$\mathbf{e_t}(\vec{u}) = \frac{1}{\sqrt{N_t}} \begin{bmatrix} \mathrm{e}^{-2\pi \frac{\vec{a}_{tx,1} \cdot \vec{u}}{\lambda}} \\ \mathrm{e}^{-2\pi \frac{\vec{a}_{tx,2} \cdot \vec{u}}{\lambda}} \\ \vdots \\ \mathrm{e}^{-2\pi \frac{\vec{a}_{tx,N_t} \cdot \vec{u}}{\lambda}} \end{bmatrix}$$

and the receive steering vector by

$$\mathbf{e_r}(\vec{u}) = \frac{1}{\sqrt{N_r}} \begin{bmatrix} \mathrm{e}^{-2\pi \frac{\vec{a}_{rx,1} \cdot \vec{u}}{\lambda}} \\ \mathrm{e}^{-2\pi \frac{\vec{a}_{rx,2} \cdot \vec{u}}{\lambda}} \\ \vdots \\ \mathrm{e}^{-2\pi \frac{\vec{a}_{rx,N_r} \cdot \vec{u}}{\lambda}} \end{bmatrix}$$

The narrowband multipath channel matrix at frequency f can then be written as

$$\mathbf{H}(f) = \sqrt{N_r N_t} \sum_{p=1}^{P} \beta_p \, \mathrm{e}^{-2\pi j f \tau_p} \, \mathbf{e_r}(\vec{u}_{rx,p}) \mathbf{e_t}^H(\vec{u}_{tx,p}) \tag{1.7}$$

Steering vectors for ULAs and UPAs

The steering vector of a ULA is given by

$$\mathbf{e_t}(\theta_{tx}) = \frac{1}{\sqrt{N_t}} \begin{bmatrix} 1 \\ \mathrm{e}^{-2\pi \frac{\|\vec{a}_{tx}\| \cos(\theta_{tx})}{\lambda}} \\ \vdots \\ \mathrm{e}^{-2\pi (N_t - 1) \frac{\|\vec{a}_{tx}\| \cos(\theta_{tx})}{\lambda}} \end{bmatrix} \tag{1.8}$$

where θ_{tx} is the angle between the antenna array main axis and the direction of departure $\vec{u}$ (Figure 1.3). Note that $t = \cos(\theta_{tx})$ is then an appealing variable substitution. We can generalize this expression to UPAs. The steering vector $\mathbf{e_t}(\vec{u}_{tx})$ is then the Kronecker product of the two equivalent ULAs in the two dimensions of the array

$$\mathbf{e_t}(\vec{u}_{tx}) = \mathbf{e_{tx}}(\vec{u}_{tx}) \otimes \mathbf{e_{ty}}(\vec{u}_{tx}).$$

1.3.3 Clustering

The paths that constitute the propagation are due to diffractions and reflections on a limited set of objects (called *scatterers*) of the environment. As a consequence the directions of departure and arrival of the rays are grouped around a small set of main directions. Each group of rays corresponding to a particular scatterer is called a *cluster* of rays or a *multipath component*. The clustering phenomenon has been observed in numerous channel sounding experiments [7, 8, 9, 10, 11]. The most commonly used method for clustered channel modeling is the Saleh-Valenzuela model.

1.4 The Saleh-Valenzuela channel model

The original model proposed by Saleh and Valenzuela in 1997 [12] aimed at characterizing the indoor frequency selective wideband channel. As only SISO transmission was considered at the time, directions of departure and arrival were omitted. This overlooked feature was added later on in the papers [9, 7] in what is called the *angular Saleh-Valenzuela channel model*.

1.4.1 The original Saleh-Valenzuela model

The original Saleh-Valenzuela model was designed to characterize SISO channels in multipath environments. The impulse response of a generic P-paths multipath SISO channel is

$$h(\tau) = \sum_{p=1}^{P} \beta_p \, \mathrm{e}^{j\phi_p} \, \delta(\tau - \tau_p)$$

The Saleh-Valenzuela model assumes that the paths are grouped into Q clusters. Each cluster consists of P sub-paths. This can be written

$$h(\tau) = \sum_{q=1}^{Q} \sum_{p=1}^{P} \beta_{q,p} \, \mathrm{e}^{j\phi_{q,p}} \, \delta(\tau - \tau_{q,p} - \tau_q) \tag{1.9}$$

The model states that

— the path gains $\beta_{q,p}$ are centered complex Gaussian distributed

— the path gain variances are given by $\mathbb{E}\left\{|\beta_{q,p}|^2\right\} = \mathbb{E}\left\{|\beta_{1,1}|^2\right\} \mathrm{e}^{-\tau_q/\Gamma} \, \mathrm{e}^{-\tau_{q,p}/\gamma}$ where Γ and γ are the cluster decay rate and intra-cluster path decay rate

— the cluster delays are given by a Poisson process with rate of arrival Λ

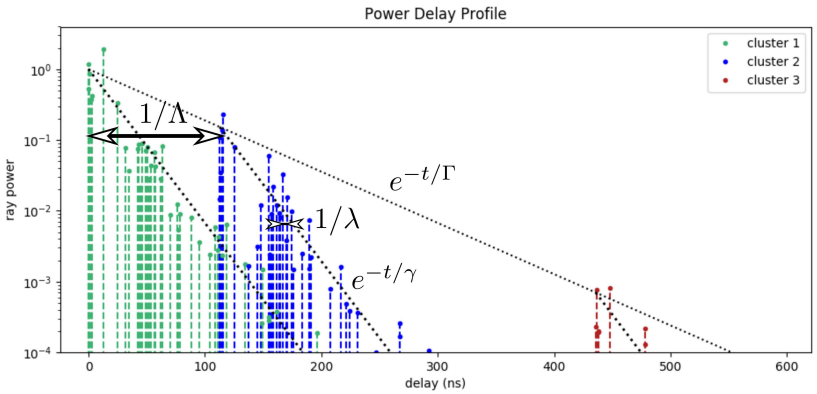

Figure 1.6 – A realization of the Saleh-Valenzuela channel model using the original Saleh-Valenzuela parameters.

— the intra-cluster delays are given by a Poisson process with rate of arrival λ

A realization of the Saleh-Valenzuela channel is presented on Figure 1.6. The role played by each parameter Γ, γ, Λ, λ is displayed on the figure. Note that $1/\Lambda$ represents the average delay between two cluster and $1/\lambda$ represents the average delay between two intra-cluster rays. Saleh-Valenzuela channels have been extensively studied and characterized in the literature. We present in Table 1.1 some parameters available in either published articles or industrial reports that will serve as a baseline for further evaluations.

Name	f_0 (GHz)	λ (ns^{-1})	γ (ns)	Reference
Original Saleh-Valenzuela	1.5	0.2	20	[12]
Crabtree building	$7GHz$	0.14	82	[7]
Clyde building	$7GHz$	0.2	29	[7]
IEEE 802.15.3c (Residential)	$60GHz$	1.22	6.25	[13]
IEEE 802.15.3c (Office)	$60GHz$	0.760	59	[13]
IEEE 802.15.3c (Library)	$60GHz$	4	7	[13]
Lund University	$60GHz$	4.7	1.1	[14]

Table 1.1 – Saleh-Valenzuela parameters for various environments.

The first line corresponds to the characterization results presented in the original paper from Saleh and Valenzuela. The next two sets of parameters correspond to measurement campaigns performed by Brigham Young University in two buildings of their campus (the Crabtree and Clyde buildings) [7]. The following measurement sets cover millimeter wave propagation scenarios in the $60GHz$ band. They consist of model calibration results proposed in the 802.15.3c TG3c Channel Modeling Sub-committee Final Report document

[13] and a measurement campaign performed at Lund University [14].

1.4.2 The angular Saleh-Valenzuela model

The initial Saleh-Valenzuela model has been extended later on by adding statistical distributions to the Direction of Departure (DOD) and Direction of Arrival (DOA) of multipath components in the papers [7, 9, 10]. Indeed, measurements from channel sounding showed that the rays constituting the multipath channel are clustered in both the angular domain and the delay domain.

According to the angular Saleh-Valenzuela model, the propagation channel for a single carrier MIMO setup with ULA arrays at both ends is given as

$$\mathbf{H}(f) = \sqrt{N_r N_t} \sum_{q=1}^{Q} \sum_{p=1}^{P} \beta_{p,q}\, \mathrm{e}^{-2\pi j f(\tau_q + \tau_{p,q})}\, \mathbf{e_r}(\theta_{rx,q} + \Delta\theta_{rx,p,q}) \otimes \mathbf{e_t}^*(\theta_{tx,q} + \Delta\theta_{tx,p,q}) \quad (1.10)$$

where $\theta_{rx,q}$ and $\theta_{tx,q}$ denote the main DOA and DOD of cluster q, respectively. $\Delta\theta_{rx,p,q}$ and $\Delta\theta_{tx,p,q}$ denote the intra-cluster rays offset directions relatively to the corresponding main cluster DOA and DOD, respectively.

The cluster main angles of departure and arrival $\theta_{tx,q}$ and $\theta_{rx,q}$ are found to be uniformly distributed in $[0, 2\pi]$ (no preferred direction) while intra-cluster offsets ($\Delta\theta_{rx,p,q}$ and $\Delta\theta_{tx,p,q}$) follow Laplacian distributions

$$p(\Delta\theta) = \frac{1}{\sqrt{2}\sigma_\theta}\, \mathrm{e}^{-\sqrt{2}\frac{|\Delta\theta|}{\sigma_\theta}}$$

characterized by the angular spread (standard deviation) σ_θ. The angular spread varies from 5° to more than 20° depending on propagation conditions [15, 7] (centimeter waves or millimeter waves, indoor or outdoor).

1.5 Tensor channel model

We saw in Section 1.1 that the received samples on subcarrier k of an OFDM symbol is given by Equation (1.3) as

$$y_k = H(f_k)x_k + n_k.$$

Using the same matrix structure formerly introduced in section 1.3, we can write this equation as

$$\mathbf{y} = \mathbf{h} \odot \mathbf{x} + \mathbf{n}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the column vectors (size N_f) containing the transmitted and received symbols on all subcarriers while $\mathbf{n}$ contains the noise samples. The channel $\mathbf{h}$ can be rewritten by means of characteristic vectors $\mathbf{e_f}$ as

$$\mathbf{h} = \sqrt{N_f} \sum_{p=1}^{P} \beta_p \mathbf{e_f}(\tau_p) \tag{1.11}$$

where $\mathbf{e_f}$ is defined as $[\mathbf{e_f}(\tau)]_k = \frac{1}{\sqrt{N_f}} \mathrm{e}^{-2\pi j f_k \tau}$ in a similar fashion as the steering vectors presented in Section 1.3. We would need a 3-dimensional channel matrix to extend the narrowband channel representation in Equation (1.7) to wideband OFDM. This can be achieved by using the tensor-based notation

$$\mathbf{h} = \sqrt{N_f N_r N_t} \sum_{p=1}^{P} \beta_p \mathbf{e_f}(\tau_p) \times \mathbf{e_r}(\vec{u}_{rx,p}) \times \mathbf{e_t}^*(\vec{u}_{tx,p}) \tag{1.12}$$

where $\times$ denotes the tensor product. The wideband MIMO-OFDM propagation channel can be rewritten using tensors and Einstein notations as

$$y_{i,k,m} = h^j_{i,k} x_{j,k,m} + n_{i,k,m} \tag{1.13}$$

where $y_{i,k,m}$ denotes the received samples on antenna i, subcarrier k of the mth OFDM symbol. Similarly, $x_{j,k,m}$ denotes the transmitted symbol on antenna j on the same time-frequency resource.

The model presented in Equation (1.12) can be expanded when using UPA antenna arrays. Indeed we saw that the steering vector of an UPA can be written as the product of two ULA steering vectors $\mathbf{e_t}(\vec{u}_{tx}) = \mathbf{e_{tx}}(\vec{u}_{tx}) \times \mathbf{e_{ty}}(\vec{u}_{tx})$. From those notations the channel tensor (1.12) can be rewritten as

$$\mathbf{h} = \sqrt{N_f N_{rx} N_{ry} N_{tx} N_{ty}} \sum_{p=1}^{P} \beta_p \mathbf{e_f}(\tau_p) \times \mathbf{e_{rx}}(\vec{u}_{rx,p}) \times \mathbf{e_{ry}}(\vec{u}_{rx,p}) \times \mathbf{e_{tx}}^*(\vec{u}_{tx,p}) \times \mathbf{e_{ty}}^*(\vec{u}_{tx,p}) \tag{1.14}$$

for UPA arrays on both sides. To simplify the notations we will usually only study the ULA case. The results can always be easily extended to the UPA case. Analytical ten-

sor manipulations are not trivial so we reshape the multidimensional tensor into a one dimensional vector, operation which is achieved by replacing the tensor products $\times$ by Kronecker products $\otimes$ as

$$\mathbf{h} = \sqrt{N_f N_r N_t} \sum_{p=1}^{P} \beta_p \mathbf{e_f}(\tau_p) \otimes \mathbf{e_r}(\vec{u}_{rx,p}) \otimes \mathbf{e_t}^*(\vec{u}_{tx,p}). \tag{1.15}$$

This simple yet comprehensive expression of the channel model is the starting point of Chapters 3, 4, 5 and 6.

1.6 Other aspects

In the ray-based model depicted above we assumed that all antennas have isotropic radiation pattern. This is not true in reality. The non-isotropic radiation pattern changes the received signal according to the angle of arrival. Assuming that all antennas have identical patterns and that coupling is negligible, the ray-based model (1.7) can be rewritten as

$$\mathbf{H} = \sqrt{N_r N_t} \sum_{p=1}^{P} G_{tx}(\vec{u}_{tx,p}) G_{rx}(\vec{u}_{rx,p}) \beta_p \, \mathrm{e}^{-2\pi j f \tau_p} \, \mathbf{e_r}(\vec{u}_{rx,p}) \mathbf{e_t}^H(\vec{u}_{tx,p}) \tag{1.16}$$

where $G_{tx}(\vec{u}_{tx,p})$ and $G_{rx}(\vec{u}_{rx,p})$ denote the transmit and receive antenna radiation patterns. Note that we can define $\tilde{\beta}_p = G_{tx}(\vec{u}_{tx,p}) G_{rx}(\vec{u}_{rx,p}) \beta_p$ which reduces Equation (1.16) to the isotropic model in Equation (1.7). The algorithms developed with isotropic antennas are also valid for arrays of identical non-isotropic antennas.

Another feature that we didn't studied in this thesis is polarization. Indeed we focused on the directional properties of MIMO propagation channels. Using the polarization would increase the achievable channel capacity by a factor 2 [16].

More details on those features can be found in the paper [17].

1.7 The NYUSIM channel simulator

Outdoor millimeter wave channels are sparse [18, 19, 20], as they consists in only a few specular paths. In order to generate accurate channel realizations, the NYU Research team created the NYUSIM channel simulator software. It has been developed based on extensive measurement campaigns at millimeter wave frequencies ranging from 28 GHz

to 73 GHz in various outdoor environments in Urban Microcell (UMi), Urban Macrocell (UMa) and Rural Macrocell (RMa) scenarios [21].

This simulator is reviewed in more details in Chapter 6 where is is used to evaluate the performance of a channel estimation algorithm.

1.8 Conclusion

In this chapter we have introduced the channel model which is used throughout this manuscript. It constitutes the starting point of this document. All the other sections are constructed on the bases that we just set up.

We have seen that all dimensions (direction of departure, direction of arrival, delay) can be studied using the same tools and methodology. Sometimes, we will only consider one or two dimensions to simplify the notations and spare some place. The reader is invited to keep in mind that in general, those results and algorithms can be expanded to the other dimensions.

BIBLIOGRAPHY

[1] *3GPP TS 36.21: "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation"*, 2020.

[2] *3GPP TS 38.211: "NR; Physical channels and modulation"*, 2020.

[3] T. Svantesson and J.W. Wallace, « Tests for assessing multivariate normality and the covariance structure of MIMO data », *in*: IEEE, 2003.

[4] David Asztély, Bjorn Ottersten, and A. Lee Swindlehurst, « A generalized array manifold model for local scattering in wireless communications », *in*: *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, IEEE, 1997.

[5] Antoine Le Calvez, Luc Le Magoarou, and Stéphane Paquelet, « Massive MIMO channel estimation taking into account spherical waves », *in*: *arXiv:1811.05669 [cs]* (Nov. 2018).

[6] Sohail Payami and Fredrik Tufvesson, « Channel measurements and analysis for very large array systems at 2.6 GHz », *in*: IEEE, Mar. 2012.

[7] Q.H. Spencer et al., « Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel », *in*: *IEEE Journal on Selected Areas in Communications* (Mar. 2000).

[8] Hao Xu, Vikas Kukshya, and Theodore S. Rappaport, « Spatial and temporal characteristics of 60-GHz indoor channels », *in*: *IEEE Journal on selected areas in communications* (2002).

[9] T. Zwick, C. Fischer, and W. Wiesbeck, « A stochastic multipath channel model including path directions for indoor environments », *in*: *IEEE Journal on Selected Areas in Communications* (Aug. 2002).

[10] N. Czink et al., « Cluster Angular Spreads in a MIMO Indoor Propagation Environment », *in*: *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, Berlin, Germany: IEEE, 2005.

[11] Ruisi He et al., « On the Clustering of Radio Channel Impulse Responses Using Sparsity-Based Methods », *in*: *IEEE Transactions on Antennas and Propagation* (June 2016).

[12] Adel AM Saleh and Reinaldo Valenzuela, « A statistical model for indoor multipath propagation », *in*: *IEEE Journal on selected areas in communications* (1987).

[13] Su-Khiong Yong et al., *G3c channel modeling sub-commitee final report*, Mar. 2007.

[14] Carl Gustafson et al., « On mm-Wave Multipath Clustering and Channel Modeling », *in*: *IEEE Transactions on Antennas and Propagation* (Mar. 2014).

[15] C. Liu, E. Skafidas, and R.J. Evans, « Angle of arrival extended S-V model for the 60 GHz wireless indoor channel », *in*: *2007 Australasian Telecommunication Networks and Applications Conference*, Christchurch, New Zealand: IEEE, Dec. 2007.

[16] V. Eiceg, H. Sampath, and S. Catreux-Erceg, « Dual-polarization versus single-polarization MIMO channel measurement results and modeling », *in*: *IEEE Trans. Wireless Commun.* (Jan. 2006).

[17] P. Almers et al., « Survey of Channel and Radio Propagation Models for Wireless MIMO Systems », *in*: *J Wireless Com Network* (Dec. 2007).

[18] Mustafa Riza Akdeniz et al., « Millimeter Wave Channel Modeling and Cellular Capacity Evaluation », *in*: *IEEE J. Select. Areas Commun.* (June 2014).

[19] M. Shehata et al., « Angular Based Beamforming and Power Allocation Framework in a Multi-User Millimeter-Wave Massive MIMO System », *in*: *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, Porto: IEEE, June 2018.

[20] Z. Muhi-Eldeen, L.P. Ivrissimtzis, and M. Al-Nuaimi, « Modelling and measurements of millimetre wavelength propagation in urban environments », *in*: *IET Microw. Antennas Propag.* (2010).

[21] Shu Sun, George R. MacCartney Jr., and Theodore S. Rappaport, « A Novel Millimeter-Wave Channel Simulator and Applications for 5G Wireless Communications », *in*: *arXiv:1703.08232 [cs, math]* (Mar. 2017).

Chapter 2

MIMO Capacity

Being able to put an upper bound on the achievable data rate of a transmission system is of particular interest. It enables system dimensioning without assuming a particular implementation and it provides a baseline to evaluate the performance of a device. This upper bound on the data rate is called the capacity.

This chapter introduces the expressions of SISO and MIMO capacity, which depends largely on a Signal to Noise Ratio (SNR) term. As previously mentioned in Chapter 1, the SNR can change rapidly due to multipath channel propagation mechanics. This *small-scale fading* effect is counterbalanced in MIMO systems by the *channel hardening* phenomenon. The first contribution of this thesis is a derivation of a formula to characterize the *channel hardening* phenomenon for ray-based channels, featured in Section 2.2 and presented at WiMob 2018 [1]. The second contribution is the characterization of the MIMO spatial multiplexing gain in Section 2.3.

2.1 Generalities

Channel capacity is the theoretical upper bound on the achievable error-free data rate. It is a property of the channel that does not depend on the transmission system implementation. The concept of channel capacity was first defined by Shannon in his seminal paper [2].

Definition. *The capacity of a noisy channel can be stated as*

$$C = B \log_2 \left(1 + \frac{P_r}{N}\right). \tag{2.1}$$

where C is the capacity of the channel in bits/s, B is the channel bandwidth, P_r is the received power and N is the noise power. The ratio P_r/N is called the SNR, and represent the radio link quality.

Throughout this manuscript we will use the notation $\rho = P_r/N$. Applied to the multipath frequency selective channel $H(f)$ introduced in Equation (1.1), the channel capacity can be rewritten as

$$C_{SISO} = \int_B \log_2 \left(1 + \frac{P_e|H(f)|^2}{N_0 B}\right) \mathrm{d}f \ (bits/s) \tag{2.2}$$

where

— P_e denotes the emitted power
— $H(f)$ denotes the frequency selective channel gain
— B denotes the channel bandwidth
— N_0 denotes the power spectral density of thermal noise (W/Hz).

Definition. *The achievable channel capacity per Hertz is called the Spectral Efficiency. The SE of a frequency selective channel at frequency f is the capacity formula integrand in Equation (2.2), stated as*

$$SE_{SISO}(f) = \log_2 \left(1 + \frac{P_e|H(f)|^2}{N_0 B}\right) (bits/s/Hz) \tag{2.3}$$

The capacity of the whole channel is obtained by integration of the spectral efficiency over the transmission bandwidth. Frequency selective channels exhibit deep fading phenomenons where the gain $|H(f)|$ vanishes, leading locally to very poor capacity. In other words, OFDM (Section 1.1) subcarriers whose channel gain fades will suffer from poor SNR.

The reliability of the radio link depends on its SNR which fluctuates because of multipath fading phenomenons. Thus quantifying the variation range of the channel gain $|H(f)|$ is particularly meaningful to predict the radio link reliability.

Definition. *A simple yet useful measure of the channel gain variations is the normalized coefficient of variation defined by the channel gain variance normalized by its mean, as*

$$CV^2_{SISO} = \frac{\mathbb{V}ar\left\{|H(f)|^2\right\}}{\mathbb{E}\left\{|H(f)|^2\right\}^2}$$

Stability of Gaussian channels

In the high scattering regime, the channel gain can be modeled as a complex Gaussian distribution (see Section 1).

$$H(f) = \sum_{p=1}^{P} \alpha_p \, \mathrm{e}^{j\phi_p} \, \mathrm{e}^{-2\pi j f_0 \tau_p} \xrightarrow[P\to\infty]{} \mathcal{CN}(0, \sigma^2)$$

Thus the channel frequency domain power is exponentially distributed

$$|H(f)|^2 \sim \exp\left(\frac{1}{\sqrt{2}\sigma}\right).$$

As the coefficient of variation of an exponentially distributed random variable is always 1, we have

$$CV_{SISO}^2 = 1$$

2.1.1 Capacity of MIMO channels

A MIMO communication system uses several antennas at both ends of the transmission. The radio link from transmit antenna j to receive antenna i is modeled in the frequency domain by the complex gain $H_{i,j}(f)$. The channel is then modeled as a complex-valued matrix, called the *channel matrix*, with coefficients (in a 2×2 setup)

$$\mathbf{H}(f) = \begin{bmatrix} H_{1,1}(f) & H_{1,2}(f) \\ H_{2,1}(f) & H_{2,2}(f) \end{bmatrix}.$$

See Section 1.1 for more details on this notation. The received symbols $\mathbf{y}$ are then given by

$$\mathbf{y} = \mathbf{H}(f)\mathbf{x} + \mathbf{n}$$

where $\mathbf{x}$ are the emitted symbols and $\mathbf{n}$ denote the thermal noise vector. The emitted power is given by $P_e = \mathrm{Tr}(\mathbf{x}\mathbf{x}^H)$. The noise vector is complex Gaussian distributed $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$.

In 1999, Emre Telatar proposed an expression for the spectral efficiency (capacity of

a $1Hz$ channel) of such multi-antenna systems in [3]. This expression is

$$SE_{MIMO} = \log_2 \left(\det \left(\mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{H}^H(f) \mathbf{Q} \mathbf{H}(f) \right) \right) \tag{2.4}$$

where $\mathbf{Q} = \mathbf{x}\mathbf{x}^{\mathbf{H}}$ is the transmit covariance matrix. This covariance matrix is controlled by the transmitter and should be chosen in order to maximize the channel capacity. This critical operation is called *precoding*. The spectral efficiency can be rewritten as

$$SE_{MIMO} = \log_2 \left(\det \left(\mathbf{I} + \frac{P_e \|\mathbf{H}\|_F^2}{\sigma_n^2} \bar{\mathbf{H}}^H \bar{\mathbf{Q}} \bar{\mathbf{H}} \right) \right) \tag{2.5}$$

where the normalized channel matrix and covariance are given by $\bar{\mathbf{H}} = \mathbf{H}/\|\mathbf{H}\|_F$ and $\bar{\mathbf{Q}} = \mathbf{Q}/\|\mathbf{Q}\|_F$ The former term $\rho = \frac{P_e \|\mathbf{H}\|_F^2}{\sigma_n^2}$ is the generalization of the SNR for MIMO channels. The latter term is interpreted as the spectral efficiency gain from spatial multiplexing.

As well as for SISO communications, the MIMO radio link reliability gain depends on the stability of the SNR ρ. In Section 2.2, the stability of MIMO system is studied through the analysis of the so-called *channel hardening* effect using a ray-based model. In Section 2.3, we propose a modified formula for the capacity of MIMO channels in perfect precoding conditions.

2.2 Channel Hardening

In classical SISO systems, the fast fading generated from the multipath propagation channels results in strong capacity fluctuations. MIMO systems average the fading phenomenon over the antennas so that the channel gain varies much more slowly. This effect is called *channel hardening*. The results of this section have been presented at the conference WiMob 2018 [1] where it received the *best paper award*.

Definition. *The channel hardening phenomenon is usually quantified using the relative variations of the channel gain $\|\mathbf{H}\|_F^2$, also called coefficient of variation (CV) [4, 5]*

$$CV_{MIMO}^2 = \frac{\mathbb{V}ar\{\|\mathbf{H}\|_F^2\}}{\mathbb{E}\{\|\mathbf{H}\|_F^2\}^2} = \frac{\mathbb{E}\{\|\mathbf{H}\|_F^4\} - \mathbb{E}\{\|\mathbf{H}\|_F^2\}^2}{\mathbb{E}\{\|\mathbf{H}\|_F^2\}^2} \tag{2.6}$$

In Equation (2.6) *the statistical means are obtained upon the model which governs the entries of $\|\mathbf{H}\|^2$ given random positions of the transmitter and the receiver.*

2.2.1 Gaussian Models

The channel hardening measure CV^2_{MIMO} has been computed for the classical Gaussian model (introduced in Section 1.2.2) in the papers [4], [5, p. 231] . As a reminder this model assumes that channels are Gaussian distributed with covariance matrix $\mathbf{R}$ ($\text{vect}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$).

Theorem. *Using the classical Gaussian model (also called Rayleigh fading model) with correlation matrix* $\mathbf{R}$*, the variations of the MIMO channel gain* (2.6) *can be rewritten as*

$$CV^2_{MIMO} = \frac{\mathbb{E}\left\{|\mathbf{h}^H\mathbf{h}|^2\right\} - \text{Tr}(\mathbf{R})^2}{\text{Tr}(\mathbf{R})^2} = \frac{\text{Tr}(\mathbf{R}^2)}{\text{Tr}(\mathbf{R})^2}, \tag{2.7}$$

where the rightmost equality comes from the properties of Gaussian vectors [5, Lemma B.14].

This result only depends on the covariance matrix $\mathbf{R}$, from which the influences of antenna array topology and propagation conditions are not explicitly identified. Moreover, small-scale and large-scale phenomena are not easily separated either [6, 7, 8].

This expression of the channel hardening phenomenon is upper and lower bounded by

$$\frac{1}{N_r N_t} \leq CV^2_{Rayleigh} = \frac{\text{Tr}(\mathbf{R}^2)}{\text{Tr}(\mathbf{R})^2} \leq 1$$

where the rightmost inequality is attained for $\text{rank}(\mathbf{R}) = 1$ and the leftmost one for $\mathbf{R} = \mathbf{Id}$. With such model, the measure of channel hardening is completely defined by the distribution of the correlation matrix eigenvalues.

The unit rank case can for example be obtained when the antenna spacing is reduced. It is indeed well-known that antenna spacing below the half-wavelength reduces the correlation matrix rank thus reducing the channel hardening effect. In contrast, the full rank situation corresponds to the limit obtained for uncorrelated entries of matrix $\mathbf{H}$, that is when the number of rays becomes sufficiently large. Such situation is obtained with sufficiently spaced antennas and sufficiently rich scattering channels. Channel hardening depends both on antenna array topology and propagation characteristics. Studying channel hardening with a physical channel model is necessary to provide more insights by separating the influences of antenna array topologies and propagation characteristics.

2.2.2 Channel Hardening in ray-based models

We propose a new expression for channel hardening that doesn't rely on Gaussian models assumptions. The ray-based multipath channel model described in Section 1 relies on several parameters (ray gains β_p, ray directions $\vec{u}_{rx,p}$ and $\vec{u}_{tx,p}$) upon which we intend do derive the new expression. We apply statistical laws to those parameters in order to scan all the possible channels of the environment. Our aim is to provide an analytical analysis of CV while relying on the weakest possible set of assumptions on the channel model. Hence, we will consider that

— For each ray, gain, DOD and DOA are independent.
— The phases $\arg(\beta_p) \sim \mathcal{U}[0, 2\pi]$ are i.i.d.
— Both DOD $\vec{u}_{tx,p}$ and DOA $\vec{u}_{rx,p}$ are i.i.d. with distributions $\mathcal{D}_{tx}$ and $\mathcal{D}_{rx}$.

The first hypothesis is widely used and simply says that no formal relation exists between the gain and the DoD/DoA of each ray. The second one reasonably indicates that each propagated path experiences independent phase rotation without any predominant angle. The last one assumes that all the rays come from independent directions, with the same distribution (distributions $\mathcal{D}_{tx}$ at the emitter, $\mathcal{D}_{rx}$ at the receiver). Note that such distribution may be directly governed by the propagation environment. To simplify the notations we also introduce the ray gain vector $\boldsymbol{\beta} = \left[\beta_1, \cdots, \beta_P\right]$. $\|\boldsymbol{\beta}\|^2 = \sum_{p=1}^{P} |\beta_p|^2$ is the aggregated power from all rays, corresponding to large-scale fading due to path-loss and shadowing.

To derive the channel hardening measure in Equation (2.6) we need to compute the channel power expectation and variance.

Expectation of the channel power. From the ray-based channel model expression Equation (1.7), the channel power $\|\mathbf{H}\|_F^2 = \mathrm{Tr}(\mathbf{H}^H\mathbf{H})$ can be written as

$$\|\mathbf{H}\|_F^2 = N_t N_r \sum_{p,p'} \beta_p^* \beta_{p'} \gamma_{p,p'}, \tag{2.8}$$

where the term $\gamma_{p,p'}$ is given by

$$\gamma_{p,p'} = \langle \mathbf{e_r}(\vec{u}_{rx,p}), \mathbf{e_r}(\vec{u}_{rx,p'}) \rangle \langle \mathbf{e_t}(\vec{u}_{tx,p}), \mathbf{e_t}(\vec{u}_{tx,p'}) \rangle^*$$

where $\langle ., . \rangle$ denotes the inner product between two vectors. Using the hypothesis $\arg(\beta_p) \sim \mathcal{U}[0, 2\pi]$ i.i.d. introduced in the channel model and $\gamma_{p,p} = 1$, the channel power expectation

can further be expressed as

$$\mathbb{E}\left\{\|\mathbf{H}\|_F^2\right\} = N_t N_r \mathbb{E}\left\{\|\boldsymbol{\beta}\|^2\right\} \tag{2.9}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_P]$. Thus the average channel gain increases linearly with N_r and N_t, which is consistent with the expected beamforming gain N_t and the fact that the received power depends on N_r linearly.

Coefficient of variation. The coefficient of variation CV^2_{MIMO} is derived using the previous hypotheses and Equation (2.8).

Definition. *We introduce the angular domain ray covariances*

$$\begin{cases} \mathcal{E}^2(\mathcal{A}_{tx}, \mathcal{D}_{tx}) = \mathbb{E}\left\{|\langle \mathbf{e_t}(\vec{u}_{tx,p}), \mathbf{e_t}^*(\vec{u}_{tx,p'})\rangle|^2\right\} \\ \mathcal{E}^2(\mathcal{A}_{rx}, \mathcal{D}_{rx}) = \mathbb{E}\left\{|\langle \mathbf{e_r}(\vec{u}_{rx,p}), \mathbf{e_r}(\vec{u}_{rx,p'})\rangle|^2\right\} \end{cases} \tag{2.10}$$

which depend on the topologies of the antenna arrays defined by the sets of vectors previously defined $\mathcal{A}_{tx}$, $\mathcal{A}_{rx}$ and on the distributions of the DoDs and DoAs $\mathcal{D}_{tx}$, $\mathcal{D}_{rx}$.

These quantities are the second moments of the inner products of the transmit/receive steering vectors associated to two distinct rays. They represent the correlation between two rays as observed by the system. They can also be interpreted as the average inability of the antenna arrays to discriminate between two rays given a specific topology and ray distribution.

Theorem. *From such definitions, CV^2 can be expressed as a sum of two terms,*

$$CV^2_{MIMO} = \mathcal{E}^2(\mathcal{A}_{tx}, \mathcal{D}_{tx})\mathcal{E}^2(\mathcal{A}_{rx}, \mathcal{D}_{rx}) \frac{\mathbb{E}\left\{\|\boldsymbol{\beta}\|^4 - \|\boldsymbol{\beta}\|_4^4\right\}}{\mathbb{E}\left\{\|\boldsymbol{\beta}\|^2\right\}^2} + \frac{\mathbb{V}ar\{\|\boldsymbol{\beta}\|^2\}}{\mathbb{E}\left\{\|\boldsymbol{\beta}\|^2\right\}^2}. \tag{2.11}$$

This formula has been derived from the ray-based model expression in Appendix A. Note that this result only relies on the assumptions introduced at the beginning of the section. The first term represents the contribution of small-scale fading due to ray recombinations. It involves both antenna array topologies and propagation characteristics. The second term can be identified as the contribution of the spatial large-scale phenomena since it simply consists in the coefficient of variation of the previously defined large-scale fading parameter $\|\boldsymbol{\beta}\|^2$ of the channel. This expression is illustrated by a simple example on Figure 2.7, where the contributions of both small-scale and large-scale fading are evidenced.

To allow local channel behavior interpretation, conditioning the statistical model by $\|\boldsymbol{\beta}\|^2$ is required. It results in the cancellation of the large-scale variations contribution of CV^2_{MIMO} which reduces to the first term identified as what is called hereafter *small-scale fading*.

2.2.3 Simulations and interpretations

In this section, small-scale and large-scale components of the channel variations are analyzed. The influence of propagation conditions and antenna array topology on the small-scale contribution are evidenced.

Large Scale fading. The contribution of large-scale fading in CV^2 is basically the coefficient of variation of the total aggregated power $\|\boldsymbol{\beta}\|^2$ of the rays. To better emphasize its behavior, let us consider a simple example with independent $|\beta_p|^2$ of mean μ and variance σ^2. The resulting large scale fading term is then

$$\frac{\mathbb{V}\mathrm{ar}\left\{\|\boldsymbol{\beta}\|^2\right\}}{\mathbb{E}\left\{\|\boldsymbol{\beta}\|^2\right\}^2} = \frac{1}{P}\left(\frac{\sigma}{\mu}\right)^2.$$

It clearly appears that more rays lead to reduced large-scale variations. This stems from the fact that any shadowing phenomenon is well averaged over P independent rays, hence becoming almost deterministic in rich scattering environments. This result is consistent with the literature on correlated Rayleigh fading channels where high rank correlation matrices provide a stronger channel hardening effect than low rank ones [5] as previously discussed in Section 2.2.1.

Small-scale fading. We condition the model on $\|\boldsymbol{\beta}\|^2$, that is the power conveyed by the channel, to remove the large scale fading effect.

Theorem. *The coefficient of variation CV^2_{MIMO} conditioned upon the aggregated ray powers $\|\boldsymbol{\beta}\|^2$ can be expressed as*

$$\begin{aligned} CV^2_{\|\boldsymbol{\beta}\|^2} &= \mathcal{E}^2(\mathcal{A}_{tx},\mathcal{D}_{tx})\mathcal{E}^2(\mathcal{A}_{rx},\mathcal{D}_{rx})\alpha^2(\boldsymbol{\beta}) \\ \textit{where } \alpha^2(\boldsymbol{\beta}) &= 1 - \frac{\mathbb{E}_{\boldsymbol{\beta}|\|\boldsymbol{\beta}\|^2}\left\{\|\boldsymbol{\beta}\|_4^4\right\}}{\|\boldsymbol{\beta}\|^4}. \end{aligned} \tag{2.12}$$

The small-scale fading contribution to CV^2 thus consists of a product of the quantities defined in Equations (2.10) that depend only on the antenna array topologies ($\mathcal{A}_{tx}$, $\mathcal{A}_{rx}$)

and ray distributions ($\mathcal{D}_{tx}$, $\mathcal{D}_{rx}$) multiplied by a propagation conditions factor $\alpha^2(\boldsymbol{\beta})$ that depends only on the statistics of the ray powers $\boldsymbol{\beta}$.

Array topologies. This paragraph focuses on the influence of $\mathcal{A}_{tx}$ on the quantity $\mathcal{E}^2(\mathcal{A}_{tx},\ \mathcal{D}_{tx})$ (the study is done only at the emitter, the obtained results being also valid at the receiver). A uniform distribution of the rays over the unit sphere is considered ($\mathcal{D}_{tx} = \mathcal{U}_{\mathcal{S}_2}$). Equation (2.10) yields

$$\mathcal{E}^2(\mathcal{A}_{tx},\ \mathcal{D}_{tx}) = \frac{1}{N_t^2}\mathbb{E}\left\{\left|\sum_{i=1}^{N_t} \mathrm{e}^{2j\pi\frac{\vec{a}_{tx,i}\cdot(\vec{u}_{tx,p}-\vec{u}_{tx,p'})}{\lambda}}\right|^2\right\}.$$

A well-known situation is when the inner sum involves exponentials of independent uniformly distributed phases and hence corresponds to a random walk with N_t steps of unit length. The above expectation then consists in the second moment of a Rayleigh distribution and $\mathcal{E}^2(\mathcal{A}_{tx},\ \mathcal{D}_{tx}) = \frac{1}{N_t}$. A necessary condition to such a case is to have (at least) a half wavelength antenna spacing Δd to ensure that phases are spread over $[0,\ 2\pi]$. On the other hand, phase independences are expected to occur for asymptotically large Δd. It is however shown hereafter that such assumption turns out to be valid for much more reasonable value of Δd.

Numerical evaluations of $\mathcal{E}^2$ are performed versus Δd (Figure 2.1), and versus N_t (Figure 2.2). Uniformly distributed rays over the 3D unit sphere ($\mathcal{D}_{tx} = \mathcal{D}_{rx} = \mathcal{U}_{\mathcal{S}_2}$) and Uniform Linear, Circular and Planar Arrays (ULA, UCA and UPA) are considered. As a reminder, the smaller $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{D}_{tx})$ the better the channel hardening. In Figure 2.1, $\mathcal{E}^2$ reaches the asymptote $1/N_t$ for all array types with $\Delta d = \frac{\lambda}{2}$ and remains almost constant for larger Δd. Figure 2.2 shows that $\mathcal{E}^2$ merely follows the $1/N_t$ law whatever the array type. We thus conclude that the independent uniform phases situation discussed above is a sufficient model for any array topology given that $\Delta d \geq \frac{\lambda}{2}$. It is therefore assumed in the sequel that,

$$\mathcal{E}^2(\mathcal{A}_{tx}, \mathcal{U}_{\mathcal{S}_2}) \approx 1/N_t, \quad \mathcal{E}^2(\mathcal{A}_{rx}, \mathcal{U}_{\mathcal{S}_2}) \approx 1/N_r. \tag{2.13}$$

Propagation conditions. It is now interesting to point out that the propagation factor $\alpha(\boldsymbol{\beta})$ introduced in (2.12) is bounded by

$$0 \leq \alpha^2(\boldsymbol{\beta}) \leq 1 - 1/P. \tag{2.14}$$

Those bounds are deduced from the following inequality which holds for any vector of

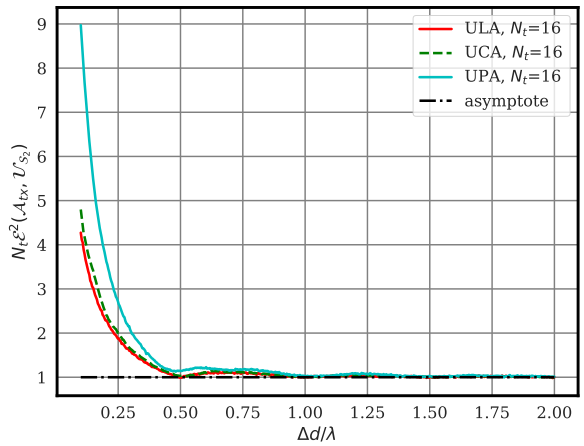

Figure 2.1 – Numerical evaluation of $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{U}_{\mathcal{S}_2})$ for various array types and increasing antenna spacing Δd. The values are normalized so the asymptote is 1.

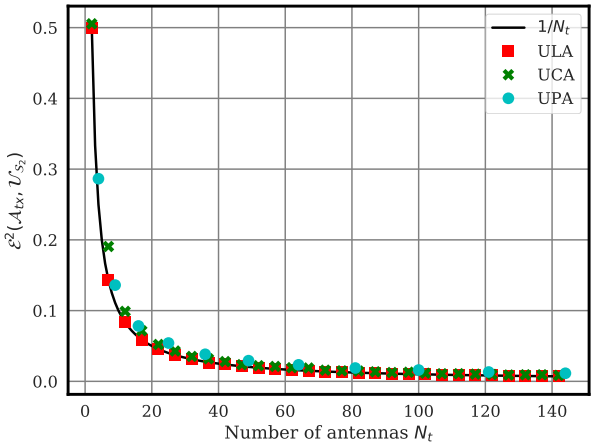

Figure 2.2 – Numerical evaluation of $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{U}_{\mathcal{S}_2})$ for various antenna arrays at the half wavelength. The lower, the better.

size P

$$\|\boldsymbol{\beta}\|_2^4/P \leq \|\boldsymbol{\beta}\|_4^4 \leq \|\boldsymbol{\beta}\|_2^4. \tag{2.15}$$

The right inequality comes from the convexity of the square function. Equality is achieved when there is only one contributing ray, i.e. no multipath occurs. In that case $CV^2_{\|\boldsymbol{\beta}\|^2} = 0$ and the MIMO channel power is deterministic. The left part of Equation (2.15) is given by Hölder's inequality. Equality is achieved when there are P rays of equal power. Then, taking the expectation on each member of Equation (2.15) yields (2.14).

In contrast to the large-scale fading, more rays lead to more small-scale fluctuations.

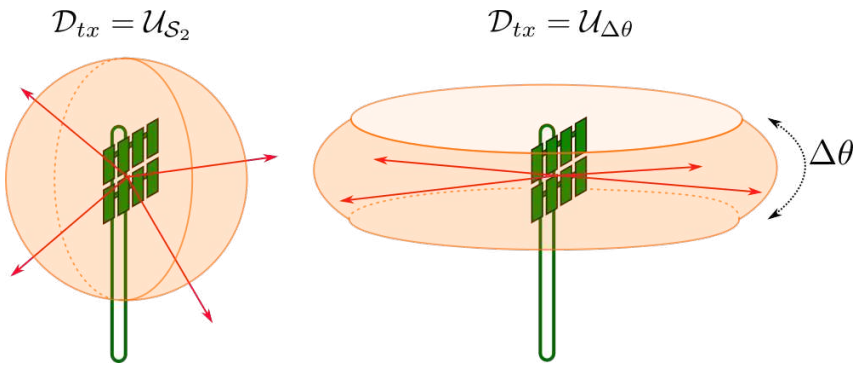

Figure 2.3 – On the leftmost side, uniform distribution of the rays on the unit sphere $\mathcal{U}_{\mathcal{S}_2}$. On the rightmost side, constrained distribution of the rays with solid angle $\mathcal{U}_{\Delta\theta}$.

It is indeed well known that a richer scattering environment increases small-scale fading.

Restrictions on ray distributions. In a mobility scenario, random translations and rotations might affect the receiver, leading to the previously used uniform distribution of the received rays over the unit sphere. However emitted rays at the base station may have much more constrained distributions (especially in the elevation plane). Equation (2.11) is also valid for those cases but the $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{D}_{tx})$ term shall be re-evaluated.

We study the ULA case using a uniform DOD distribution with constrained elevation in the interval $[-\Delta\theta/2, \Delta\theta/2]$ illustrated on Figure 2.3. This distribution is denoted $\mathcal{U}_{\Delta\theta}$ and we have the equality $\mathcal{U}_{\pi} = \mathcal{U}_{\mathcal{S}_2}$.

The behavior of $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{U}_{\Delta\theta})$ for increasing number of antennas is given in Figure 2.4. We can observe a clear degradation when $\Delta\theta$ decreases. This is coherent with literature on channel hardening, constrained rays induce correlated channel matrices, mitigating the channel hardening effect. This model leads to the pinhole channel for asymptotically small $\Delta\theta$. It corresponds to situations when rich scattering leads to a low correlation channel covariance matrix, where the channel is known to not harden [4]. Figure 2.5 shows the effect of antenna spacing on $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{U}_{\Delta\theta})$. It can be observed on this figure that the channel hardening loss due to restricted distribution of the rays by $\Delta\theta$ can be compensated by increasing antenna spacing above the half-wavelength. The influence of $\Delta\theta$ on channel hardening is depicted by Figure 2.6. A clear reduction of channel hardening can be observed when θ decreases.

Simulations. For illustration purposes, we evaluate the coefficient of variation through both Monte-Carlo simulations and Equation (2.11). We use uniform linear arrays (ULA) with inter-antenna spacing of $\frac{\lambda}{2}$ at both transmitter and receiver sides. A total of $P = 6$

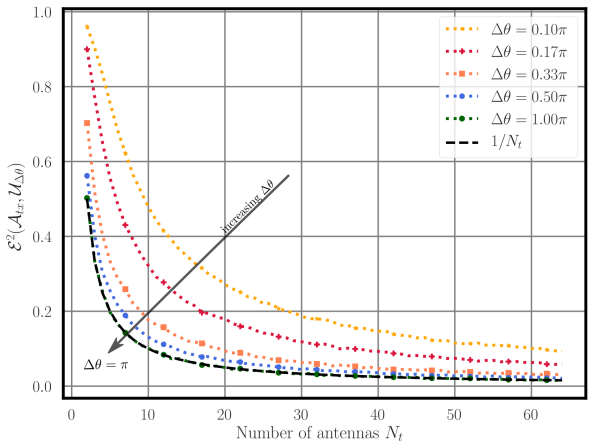

Figure 2.4 – Numerical evaluation of $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{U}_{\Delta\theta})$ for increasing $\Delta\theta$ with fixed half-wavelength antenna spacing. The lower, the better.

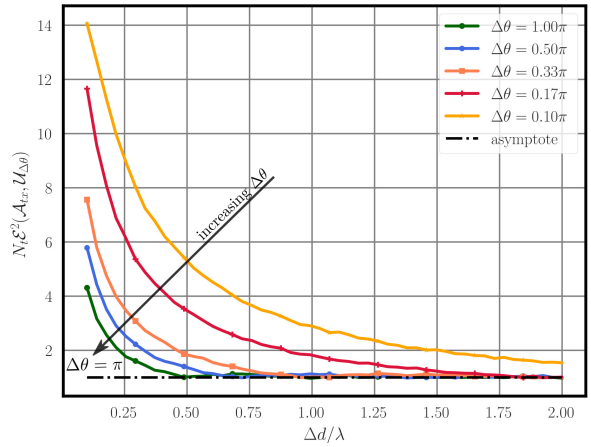

Figure 2.5 – Numerical evaluation of $\mathcal{E}(\mathcal{A}_{tx}, \mathcal{U}_{\Delta\theta})$ for increasing $\Delta\theta$ with fixed number of antennas $N_t = 16$. The lower, the better.

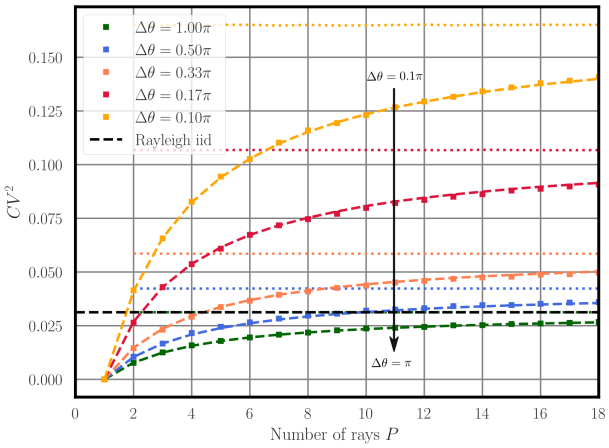

Figure 2.6 – Comparison between (2.11), simulated CV^2 and correlated Rayleigh channel hardening measure (2.7), the dashed, squared and dotted lines respectively. A 16×2 ULA setup is considered. Uniform distribution of DoDs and DoAs with constrained DoDs and complex Gaussian gains. The large-scale factor is normalized.

paths were randomly generated with Complex Gaussian gains $\beta_p \sim \mathcal{CN}(0,1)$, uniform DOD $\vec{u}_{tx,p} \sim \mathcal{U}_{\mathcal{S}_2}$ and DOA $\vec{u}_{rx,p} \sim \mathcal{U}_{\mathcal{S}_2}$.

Based on the general formula given in (2.11), on the interpretations and evaluations of its terms and on the geometrical terms approximations in Equation (2.13) we derive the expression of channel hardening for this illustrating example as

$$CV^2_{\text{approx}} = \frac{1}{N_t N_r}(1 - 1/P) + 1/P.$$

Simulation and approximated formula are compared in Figure 2.7 in which small-scale and large-scale contributions are easily evidenced.

Comparison with the Gaussian model. In this section we compare our channel hardening expression conditioned by $\|\boldsymbol{\beta}\|^2$ (local behavior) to the correlated Rayleigh fading channel in both strong scattering and low scattering situations.

In the high scattering regime, the correlated Rayleigh channels approach is known to perform well. Without any particular constraint on the DoD and DoA, the correlation matrix is given by $\mathbf{R} = \mathbf{Id}$ (the well-known Rayleigh i.i.d. channel model). Using this result with Equation (2.7) yields

$$CV^2_{iid} = \frac{1}{N_t N_r}.$$

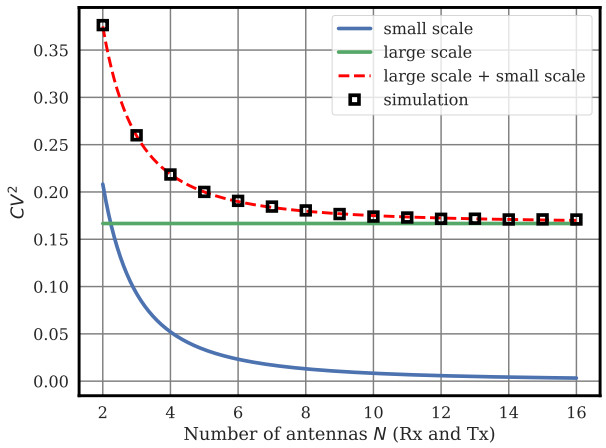

Figure 2.7 – Comparison between Equation (2.11) and numerically evaluated CV^2. Uniform distribution of DoDs and DoAs over the unit sphere and complex Gaussian gains.

Using the realistic model in a rich scattering environment, the propagation term $\alpha(\mathbf{c})$ of small-scale variations (2.12) reaches the upper bound of (2.14). This yields the limit

$$CV^2 \xrightarrow{P\to\infty} CV^2_{iid} \tag{2.16}$$

which is coherent with the interpretation of the model. This behavior can be observed on the lowest curve of Figure 2.6.

However in the low scattering regime (considering P rays), the Rayleigh modeling approach is not valid any more. The sparse channel is not modeled properly and formula (2.7) underestimates the channel hardening effect. Our approach introduces a propagation factor $\alpha(\mathbf{c})$ that models the scattering effect and the sparsity of the channel. According to results given by the ray-based model Figure 2.6, channel hardening is increased in low scattering scenarios and reaches the correlated Rayleigh hardening for $P \to \infty$, which is consistent with the previous physical interpretation.

2.2.4 Conclusion

In this section, we have extended previous studies on channel hardening using a propagation-based model. We have separated influences of antenna array topologies and propagation characteristics on the channel hardening phenomenon. Large-scale and small-scale contributions to channel variations have been evidenced. This contribution provides a general framework to study channel hardening using accurate propagation models.

To illustrate the overall behavior of channel hardening, this framework have been used with generic model parameters and hypotheses. The scaling laws evidenced for simpler channel models are conserved provided the antennas are spaced by at least half a wavelength. The results are consistent with state of the art and provide further insights on the influence of array topology and propagation on channel hardening. The proposed expression can easily be exploited with various propagation environments and array topologies to provide a more precise understanding of the phenomenon compared to classical channel descriptions based on Rayleigh fading models.

2.3 Multiplexing gain

The second aspect of the capacity is the multiplexing gain. In this section we propose to rewrite the optimal capacity using only a few variables, namely the transmission rank, the product of all used eigenvalues and the sum of their inverses. We show that this exact expression can be approximated using only two variables with minor capacity prediction error. In order to achieve the optimal capacity, the input covariance matrix $\mathbf{Q}$ introduced in Equation (2.4) has to be determined according to the propagation channel $\mathbf{H}$. First we introduce the calculation of $\mathbf{Q}$ that enables the optimal capacity. Using this result we introduce the new formula to describe the single user MIMO capacity in a later stage.

2.3.1 Optimal precoding

In the seminal paper [3] published by Emre Telatar in 1999, the expression of the optimal precoding matrix was derived. In this section we remind this important result of MIMO literature. At each time interval, the received signal $\mathbf{y}$ is given by the linear transform

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}.$$

In order to optimize the capacity (2.5), the input symbols $\mathbf{x}$ have to be shaped so that the covariance matrix $\bar{\mathbf{Q}} = \frac{\mathbf{x}\mathbf{x}^H}{\mathrm{Tr}(\mathbf{x}\mathbf{x}^H)}$ maximizes the capacity

$$SE_{MIMO} = \log_2\left(\det\left(\mathbf{I} + \rho\bar{\mathbf{H}}^H\bar{\mathbf{Q}}\bar{\mathbf{H}}\right)\right). \tag{2.17}$$

The covariance matrix $\bar{\mathbf{Q}}$ has to be chosen by the transmitter using the linear transformation

$$\mathbf{x} = \mathbf{P}\mathbf{s}$$

where $\mathbf{s}$ contains independent symbols and matrix $\mathbf{P}$ is called the *precoding matrix*. The relation with the input covariance matrix is given by $\bar{\mathbf{Q}} = \mathbf{P}\mathbf{P}^{\mathbf{H}}$. In order to maximize the capacity, Telatar showed that $\bar{\mathbf{Q}}_{opt}$ should take the form

$$\bar{\mathbf{Q}}_{opt} = \mathbf{V}\boldsymbol{\eta}\mathbf{V}^{\mathbf{H}}$$

where $\mathbf{V}$ is an unitary matrix that diagonalizes $\bar{\mathbf{H}}\bar{\mathbf{H}}^{H} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{H}}$ and $\boldsymbol{\eta}$ is the diagonal matrix containing the power allocated to each precoding vector. The power allocation weights $[\boldsymbol{\eta}]_{i,i}$ are calculated from the eigenvalues of the normalized channel $\boldsymbol{\Delta} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{\min(N_t,N_r)})$ using the water-filling algorithm [3]. The spectral efficiency when matrix $\mathbf{V}$ is used for precoding is given by

$$SE_{MIMO} = \sum_{i=1}^{\min(N_t,N_r)} \log_2(1 + \rho[\boldsymbol{\eta}]_{i,i}\lambda_i) \tag{2.18}$$

as the channel is diagonalized. The normalization of the matrices $\mathbf{H}$ and $\mathbf{Q}$ imposes that $\sum_{i=1}^{\min(N_t,N_r)} \lambda_i = 1$ and $\sum_{i=1}^{\min(N_t,N_r)} [\boldsymbol{\eta}]_{i,i} = 1$. From [9], the result of the waterfilling algorithm can be written analytically as

$$\begin{cases} [\boldsymbol{\eta}]_{i,i} & = \frac{1+\sum_{j=1}^{L} \frac{1}{\rho\lambda_j}}{L} - \frac{1}{\rho\lambda_i} \text{ if } i \leq L \\ [\boldsymbol{\eta}]_{i,i} & = 0 \text{ otherwise} \end{cases} \tag{2.19}$$

where L denote the *transmission rank*, that is the number of channel eigenvalues used to transmit data. The optimal transmission rank is given as the largest integer L that verifies the condition

$$\frac{1 + \sum_{j=1}^{L} \frac{1}{\rho\lambda_j}}{L} > \frac{1}{\rho\lambda_{L+1}}$$

2.3.2 Derivation of the new formula

In this PhD, we provide a new analytical formula for the capacity, obtained by injecting the power allocation formula (2.19) within the capacity formula (2.18). This formula is useful to quantify the behavior of the capacity in a particular environment only based on

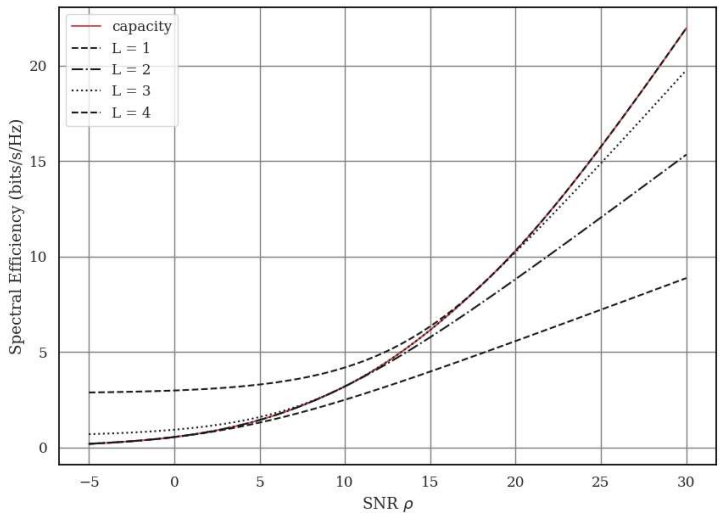

Figure 2.8 – Spectral efficiency of a MIMO channel with maximum transmission rank $L = 4$.

two channel-dependent variables.

Theorem. *From Equations (2.19) and (2.18), the spectral efficiency can be rewritten as*

$$SE_{MIMO} = \log_2\left(\left(\frac{\rho + R_L}{L}\right)^L D_L\right) \quad bps\ /\ Hz \tag{2.20}$$

where $R_L = \sum_{j=1}^{L} \frac{1}{\lambda_j}$ *and* $D_L = \prod_{j=1}^{L} \lambda_j$. *$L$ denotes the optimal transmission rank.*

Beyond the classical variables ρ, the SNR, and L the transmission rank, this new expression only depend on two channel parameters which are the sum of the inverses and the product of the L largest eigenvalues, namely

$$\begin{cases} R_L = \sum_{j=1}^{L} \frac{1}{\lambda_j} \\ D_L = \prod_{j=1}^{L} \lambda_j. \end{cases}$$

This formula enables a precise evaluation of the spectral efficiency based on few parameters. Note that the optimal transmission rank L changes with the SNR. Equation (2.20) is illustrated on Figure 2.8. The channel spectral efficiency follows the curve $L = 1$ at low SNR. As the signal to noise ratio improves the achievable transmission rank increase and the spectral efficiency switches to the higher order curve. For full-rank transmis-

sion, variables $R_{L_{max}}$ and $D_{L_{max}}$ can be rewritten using Vieta's formulas applied to the characteristic polynomial of $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ as

$$\begin{cases} R_L = \sum_{j=1}^{L_{max}} \frac{1}{\lambda_j} = -\frac{a_1}{a_0} = |\frac{a_1}{\det(\bar{\mathbf{H}}\bar{\mathbf{H}}^H)}| \\ D_L = \prod_{j=1}^{L_{max}} \lambda_j = a_0 = \det(\bar{\mathbf{H}}\bar{\mathbf{H}}^H) \end{cases}$$

where the constants a_1 and a_0 are extracted from the expression of the characteristic polynomial

$$\chi_{\mathbf{H}\mathbf{H}^H}(\lambda) = \det(\mathbf{H}\mathbf{H}^H - \lambda\mathbf{I}) = \sum_{i=0}^{L} a_i \lambda^i$$

2.3.3 Approximations of the spectral efficiency

In the high SNR regime, the spectral efficiency increase linearly with $\log(\rho)$ as evidenced on Figure 2.8. The transition from a transmission rank L to $L+1$ actually happens at the beginning of this high SNR regime. The approximation of the spectral efficiency by the high SNR asymptote is tight except at the vicinity of transmission rank transitions. In the high SNR regime, the SE can be approximated by the formula

$$SE_{MIMO} \xrightarrow{\rho \to \infty} L \log_2(\rho) - L \log_2(L) + \log_2(D_L) \tag{2.21}$$

The asymptotes of the SE for a transmission rank $L = 2$ are displayed on Figure 2.9. In Equation (2.21) the SE increases linearly with the logarithm of D_L and the SNR ρ. For consistency we measure the SNR and both parameters R_L and D_L in dB.

On Figure 2.10 we show the gap between the true SE and the approximation using only the asymptotic behavior (Equation (2.21)) for transmission rank $L > 1$. For $L = 1$ the Spectral Efficiency only depends on λ_1 and approximations are not required. The approximated curve is then piecewise affine for $L \geq 2$ and only depend on the R_L parameter, the transmission rank L and the SNR ρ. Such expression is useful to predict the capacity using only a handful of parameters. This technique can also be used to provide an approximate SNR value for the rank transition based on the asymptotes' intersection, that is

$$\rho_L = \frac{1}{\lambda_{L+1}} \frac{(L+1)^{L+1}}{L^L}.$$

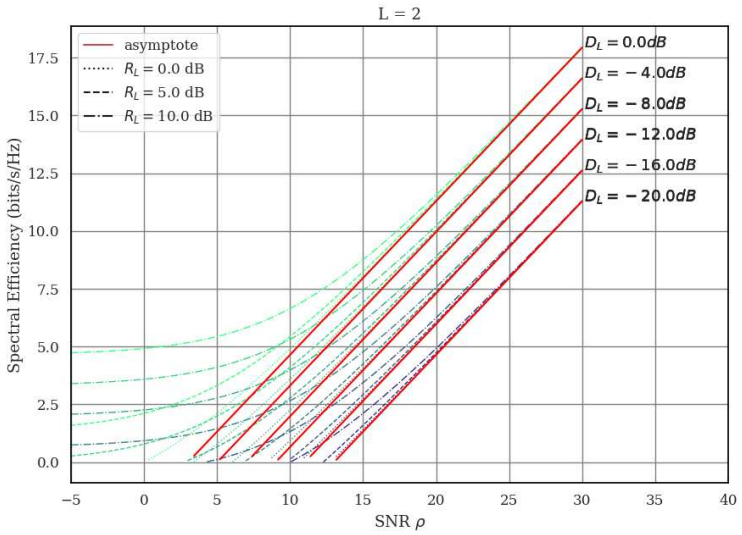

Figure 2.9 – Asymptotes of the spectral efficiency and influence of the parameters R_L and D_L for a transmission rank $L = 2$.

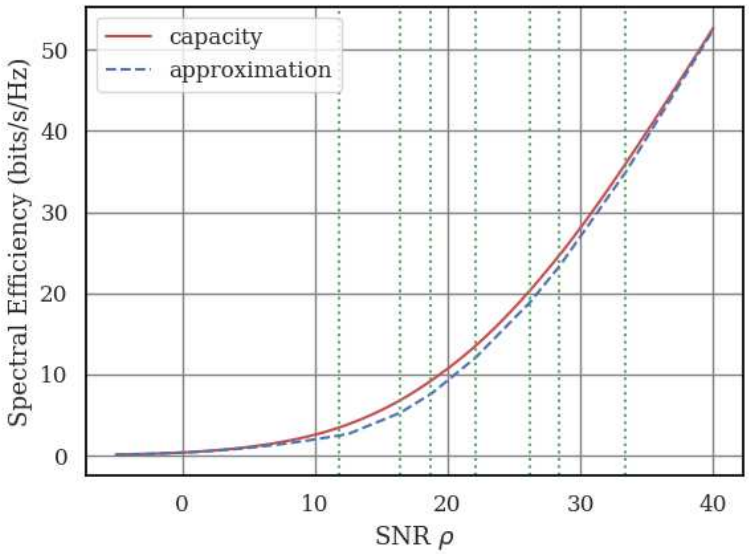

Figure 2.10 – Approximation of the SE where the maximum transmission rank is $L_{max} = 8$. The asymptotes' intersection points, that denote transitions to the next transmission rank $L + 1$, are displayed as vertical green dashed lines.

2.4 Conclusion

In this first chapter we studied analytically the spectral efficiency gain of MIMO channels. In the literature, the benefits of MIMO systems have been classified into two main categories: small scale fading reduction (also called channel hardening or diversity gain) and spatial multiplexing.

We proposed an original analysis of the channel hardening phenomenon using a ray-based model that enables a more in-dept understanding of the contributions from antenna array topologies and propagation conditions to the reduction of small-scale fading. This analysis is complementary to previous studies involving the well-known Correlated Gaussian model and the conclusions using both methods are identical in the high scattering regime.

Then we proposed a novel expression for the spectral efficiency that depends on fewer parameters than the classical formula. We evaluated the effect of the parameters on the spectral efficiency and proposed a simplified approximate formula using a piecewise affine function. This expression is useful to predict the capacity of the channel by measuring only few parameters and can be useful to predict rank transitions.

BIBLIOGRAPHY

[1] Matthieu Roy et al., « MIMO Channel Hardening for Ray-based Models », *in*: *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Limassol: IEEE, Oct. 2018.

[2] C. E. Shannon, « A Mathematical Theory of Communication », *in*: *Bell System Technical Journal* (July 1948).

[3] Emre Telatar, « Capacity of Multi-antenna Gaussian Channels », *in*: *European Transactions on Telecommunications* (Nov. 1999).

[4] Hien Quoc Ngo and Erik G. Larsson, « No downlink pilots are needed in TDD massive MIMO », *in*: *IEEE Transactions on Wireless Communications* (2017).

[5] Emil Björnson, Jakob Hoydis, and Luca Sanguinetti, « Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency », *in*: *Foundations and Trends® in Signal Processing* (2017).

[6] Sara Gunnarsson et al., « Channel Hardening in Massive MIMO - A Measurement Based Analysis », *in*: *arXiv:1804.01690 [cs, math]* (Apr. 2018).

[7] Àlex Oliveras Martínez, Elisabeth De Carvalho, and Jesper Ødum Nielsen, « Massive MIMO properties based on measured channels: Channel hardening, user decorrelation and channel sparsity », *in*: *Signals, Systems and Computers, 2016 50th Asilomar Conference on*, IEEE, 2016.

[8] Sara Gunnarsson et al., « Channel Hardening in Massive MIMO: Model Parameters and Experimental Assessment », *in*: *IEEE Open J. Commun. Soc.* (2020).

[9] Eitan Altman, Konstantin Avrachenkov, and Andrey Garnaev, « Closed form solutions for water-filling problems in optimization and game frameworks », *in*: *Telecommun Syst* (June 2011).

Chapter 3

Efficient Wideband MIMO Channel Representation

In this chapter we analyze the statistical properties of the Saleh-Valenzuela channel model. In particular we derive the first and second order statistics of both Saleh-Valenzuela separated clusters and whole channel. Those statistics are fundamental in the organization of the manuscript as they are utilized in the next two sections of this chapter as well as in the two following chapters, by providing a simpler formulation of the Saleh-Valenzuela cluster model.

In the second section we search for an efficient representation basis for Saleh-Valenzuela clusters. We compare the optimal eigenvector basis with the well-known Fourier basis, in both the delay and angular domains.

Then we calculate how much coefficients are required to model a particular channel given its Saleh-Valenzuela parameters. We compare two modeling approaches, one suitable for massive MIMO and one typically used for SISO and small-scale MIMO and we draw the boundary between both regimes.

3.1 Second order statistics

The multi-carrier angular Saleh-Valenzuela model (Section 1.4.2) with ULA antenna arrays at both ends can be rewritten using the formalism of Equation (1.15) as

$$\mathbf{h} = \sqrt{N_f N_r N_t} \sum_{q=1}^{Q} \sum_{p=1}^{P} \beta_{p,q} \mathbf{e_f}(\tau_q + \tau_{p,q}) \otimes \mathbf{e_r}(\theta_{rx,q} + \Delta\theta_{rx,p,q}) \otimes \mathbf{e_t}^*(\theta_{tx,q} + \Delta\theta_{tx,p,q}) \quad (3.1)$$

The channel is made of several multipath components, each of them is characterized by a delay of arrival τ_q and main directions of arrival and departure $\theta_{rx,q}$ and $\theta_{tx,q}$ (defined by a simple scalar angle as we consider ULA arrays at both ends). Equation (3.1) can be

rewritten as a sum of clusters

$$\mathbf{h} = \sum_{q=1}^{Q} \mathbf{c_q} \tag{3.2}$$

In this section we study the statistical properties of those Saleh-Valenzuela clusters conditioned on τ_q, $\theta_{rx,q}$ and $\theta_{tx,q}$. From the Saleh-Valenzuela model definition proposed in Section 1.4.2, the expectation of a cluster is $\mathbb{E}\{\mathbf{c_q}\} = 0$ (it consists in a sum of gains).

The phase independence property ensures that $\mathbb{E}\{\beta_p \beta_{p'}^*\} = 0$, thus removing cross-correlation terms between rays. The second order channel statistics are given by

$$\begin{aligned}
\mathbb{E}\{\mathbf{c_q}\mathbf{c_q}^H\} =& \mathbb{E}\{N_f \sum_{p=1}^{P} |\beta_{p,q}|^2 \mathbf{e_f}(\tau_q + \tau_{p,q})\mathbf{e_f}^H(\tau_q + \tau_{p,q})\} \\
&\otimes \mathbb{E}\{N_r \mathbf{e_r}(\theta_{rx,q} + \Delta\theta_{rx,p,q})\mathbf{e_r}^H(\theta_{rx,q} + \Delta\theta_{rx,p,q})\} \\
&\otimes \mathbb{E}\{N_t \mathbf{e_t}(\theta_{tx,q} + \Delta\theta_{tx,p,q})\mathbf{e_t}^H(\theta_{tx,q} + \Delta\theta_{tx,p,q})\}^H \\
=& \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q) \otimes \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{r}}(\theta_{rx,q}) \otimes \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}\,H}(\theta_{tx,q})
\end{aligned} \tag{3.3}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q)$ denote the frequency domain cluster covariance and $\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{r}}(\theta_{rx,q})$ and $\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}}(\theta_{tx,q})$ the angular domain receive and transmit covariances. We evaluate those covariance matrices in this section.

In particular, we evaluate the second order statistics of both Saleh-Valenzuela clusters and the whole channel in the frequency domain. They have already been calculated in the papers [1, 2], however the starting point of their demonstration is the delay-domain channel representation, given by Equation (1.9). This approach faces technical difficulties due to the dirac functions, that are overcome by using advanced point process theory. The starting point of our demonstration is the frequency domain representation of the channel, which avoid this technical problem and leads to a much simpler proof. This work has been published at the international conference GLOBECOM 2019 [3].

3.1.1 Single cluster statistics

In this section we derive the expression of the covariance and Power Delay Profile (PDP) of a single cluster of the Saleh-Valenzuela channel model. Without loss of generality, we set $\tau_q = 0$. The frequency domain complex gain of a cluster is given by

$$c_q(f) = \sum_{p=1}^{P} \beta_{p,q}\, \mathrm{e}^{-2\pi j f \tau_{p,q}}\,. \tag{3.4}$$

Due to the exponentially decaying intra-cluster powers with increasing delay, only a finite number of dominant rays are distinguishable. P is usually chosen high enough to embrace the whole multi-path power while keeping computational complexity low for simulation purposes. For analytical computation convenience, we will consider that $P \to \infty$. To simplify the expressions, we normalize the power of the first ray as $\mathbb{E}\{|\beta_{1,1}|^2\} = 1$.

Theorem. *The cluster covariance function is evaluated as*

$$R_{c_q}(\Delta f) = \mathbb{E}\left\{|\beta_{1,q}|^2\right\}\left(1 + \frac{\lambda\gamma}{1 - 2\pi j\Delta f\gamma}\right). \tag{3.5}$$

The details of the derivations are given in appendix B.1. From the frequency domain expression of the Saleh-Valenzuela channel, we derive the second order statistics for a fixed number of rays P using a recurrence on P. The cluster covariance is obtained by taking the limit of this expression as P tends to infinity.

We find that the random process c_q is Wide Sense Stationary (WSS), that is the covariance only depends on the frequency difference Δf. It is noticeable that this expression consists of two terms, the former being constant and the latter vanishing as the frequency difference increases. This feature is also observed in [4]. We obtain the PDP by applying the Fourier Transform on the autocorrelation,

$$S_{c_q}(\tau) = \mathbb{E}\left\{|\beta_{1,q}|^2\right\}\left(\delta(\tau) + \lambda\,\mathrm{e}^{-\tau/\gamma}\,u(\tau)\right) \tag{3.6}$$

where $u(\tau)$ is the Heaviside step function. The average power of the cluster for any frequency f is given by

$$P_{c_q} = \mathbb{E}\left\{|c_q(f)|^2\right\} = \mathbb{E}\left\{|\beta_{1,q}|^2\right\}(1 + \lambda\gamma)$$

where $\mathbb{E}\{|\beta_{1,q}|^2\}$ is the average power of the first ray of cluster q. It can be evaluated using a recurrence as

$$\mathbb{E}\left\{|\beta_{1,q}|^2\right\} = \mathbb{E}\left\{|\beta_{1,q-1}|^2\right\}\left(\frac{1}{1 + \frac{1}{\Lambda\Gamma}}\right) = \mathbb{E}\left\{|\beta_{1,1}|^2\right\}\left(\frac{1}{1 + \frac{1}{\Lambda\Gamma}}\right)^{q-1}. \tag{3.7}$$

where Γ and Λ are the cluster power decay constant and rate of arrival of clusters as defined in the Saleh-Valenzuela model (Section 1.4). $\mathbb{E}\{|\beta_{1,q-1}|^2\}$ is the average power of the very first ray, which is normalized to 1 (see above) to simplify the expressions.

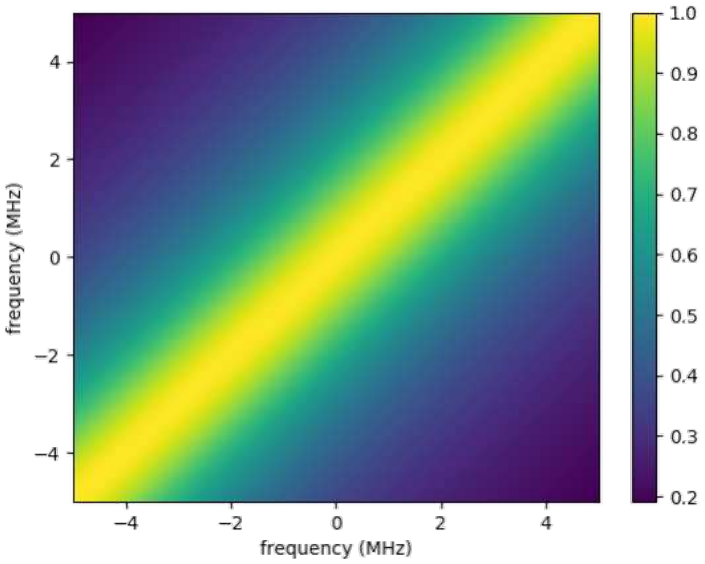

Figure 3.1 – Normalized modulus of the frequency domain correlation matrix for a single cluster.

We are interested in the cluster second order statistics covariance matrix between subcarriers $\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q)$. This introduces a sampling of the continuous function $c_q(f)$ corresponding to the locations of the OFDM subcarriers. The frequency domain covariance matrix is obtained from Equation (3.5) as

$$[\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q)]_{k,k'} = P_q \left(1 + \frac{\lambda\gamma}{1 - 2\pi j \frac{k'-k}{N_f} B\gamma}\right) \mathrm{e}^{2\pi j \frac{k'-k}{N_f} B\tau_q} . \tag{3.8}$$

where τ_q denotes the delay of cluster q and $P_q = \mathbb{E}\{|\beta_{1,q}|^2\}$ denotes the cluster power. The modulus of this covariance matrix (normalized) is given in Figure 3.1. The covariance between adjacent subcarriers (coefficients close to the diagonal of the matrix) is much more stronger than between spaced out subcarriers (coefficients away from the diagonal). The width of the diagonal high correlation region is linked to the coherence bandwidth of the channel [5]. Such covariance matrix will be a central piece of Chapters 4 and 5.

3.1.2 Multiple clusters channel analysis

In this section we derive the expression of the covariance and PDP of the full Saleh-Valenzuela channel. The complex frequency gain of the channel is given by

$$H(f) = \sum_{q=1}^{Q} \sum_{p=1}^{P} \beta_{p,q} \, \mathrm{e}^{-2\pi j f(\tau_{p,q}+\tau_q)} \, . \tag{3.9}$$

As previously done, we will consider that $P \to \infty$ and $Q \to \infty$ and we normalize the power of the first ray $\mathbb{E}\left\{|\beta_{1,1}|^2\right\} = 1$.

Theorem. *The covariance function of the whole Saleh-Valenzuela channel is given as*

$$R_H(\Delta f) = \left(1 + \frac{\Lambda\Gamma}{1 - 2\pi j \Delta f \Gamma}\right)\left(1 + \frac{\lambda\gamma}{1 - 2\pi j \Delta f \gamma}\right). \tag{3.10}$$

The covariance function calculation details are given in appendix B.2. The random process H is also WSS. This expression is advantageously stated as the product of the covariances of two single-exponentially decaying channels. Then we apply the Fourier Transform to the covariance function to get the PDP

$$S_H(\tau) = \delta(\tau) + \Lambda \, \mathrm{e}^{-\tau/\Gamma} \, u(\tau) + \lambda \, \mathrm{e}^{-\tau/\gamma} \, u(\tau) + \Lambda\lambda \frac{\Gamma\gamma}{\Gamma - \gamma} \left(\mathrm{e}^{-\tau/\gamma} - \mathrm{e}^{-\tau/\Gamma}\right) u(\tau)$$

where $u(\tau)$ is the Heaviside step function. The average power of the Saleh-Valenzuela channel at any frequency f is finally given by

$$P_H = \mathbb{E}\left\{|H(f)|^2\right\} = (1 + \lambda\gamma)\,(1 + \Lambda\Gamma)\,.$$

3.1.3 Angular domain cluster statistics

In this section we evaluate the Saleh-Valenzuela cluster covariance matrix in the angular domain $\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}}(\theta_{tx,q})$. For the sake of brevity we only treat the transmit side covariance matrix. As both ends are identical, the results can immediately be applied to the receive side. The expression of the covariance matrix is

$$\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}}(\theta_{tx,q}) = \mathbb{E}\{N_t \mathbf{e_t}(\theta_{tx,q} + \Delta\theta_{tx,p,q}) \mathbf{e_t}^H(\theta_{tx,q} + \Delta\theta_{tx,p,q})\}$$

where the expectation is taken over $\Delta\theta_{tx,p,q}$. From this expression and from the ULA steering vector formula (1.8) we derive the angular domain cluster covariance matrix as

$$[\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}}(\theta_{tx,q})]_{l,l'} = \int_{-\infty}^{\infty} p_{\Delta\theta}(\Delta\theta)\, \mathrm{e}^{-2\pi j(l'-l)\frac{\|\vec{a}_{tx}\|^2}{\lambda}\cos(\theta_{tx,q}+\Delta\theta)}\, d\Delta\theta$$

As a reminder, clusters are modeled as a group of rays with common direction of departure and arrival. In the extended Saleh-Valenzuela model (Section 1.4.2), let us remind that the angle offset is Laplace distributed as

$$p(\theta_{tx}) = \frac{1}{\sqrt{2}\sigma_\theta}\, \mathrm{e}^{-\sqrt{2}\frac{|\theta_{tx}-\bar{\theta}_{tx}|}{\sigma_\theta}}$$

where σ_θ denotes the angular spread.

3.2 Sparse Channel Representation

From the previous second order statistics derivation, it is now possible to represent each cluster as a vector of random variables $\mathbf{c_q}$ which have similar weights but are correlated according to the covariance matrices previously obtained. In this section we investigate in which extend we can establish a representation of the same cluster with fewer stochastic parameters such as

$$\mathbf{c_q} \approx \sum_{l=1}^{L} z_l \mathbf{e_l} \tag{3.11}$$

where z_l are random coefficients and $\mathbf{e_l}$ are deterministic vectors. The basis vectors $\mathbf{e_l}$ have to be optimized to yield the smallest possible number L of components while mitigating the representation error, quantified by the Mean Square Error (MSE). In other words, our goal is to find the smallest set of vectors $\{\mathbf{e_1}, \ldots, \mathbf{e_L}\}$ that can represent the channel while keeping the MSE $\mathbb{E}\left\{\|\mathbf{c_q} - \sum_{l=1}^{L} z_l \mathbf{e_l}\|^2\right\}$ below a certain predefined threshold η as

$$\{\mathbf{e_1}, \ldots, \mathbf{e_L}\} = \underset{\{\mathbf{u_1},\ldots,\mathbf{u_L}\}}{\operatorname{argmin}} L$$
$$wrt.\ \ \mathbb{E}\left\{\|\mathbf{c_q} - \sum_{l=1}^{L} z_l \mathbf{u_l}\|^2\right\} < \eta.$$

In this section, we compare the optimal eigencomponent decomposition (optimality means that it yields the lowest MSE for a given number of representation vectors L) with

the commonly used Fourier decomposition in both frequency and angular domains.

3.2.1 Optimal sparse representation

The optimal (that is it yields the smallest MSE for a fixed number of coefficients) solution of the problem above is to set vectors $\mathbf{e_l}$ as the eigenvectors basis [6] (normalized $\|\mathbf{e_l}\|^2 = 1$) of the cluster covariance matrix $\mathbf{\Sigma_q}$. Note that this matrix is Toeplitz symmetric positive semi-definite, thus the eigencomponents are well defined, eigenvalues are all positive and eigenvectors are orthogonal. This decomposition contains $N_f N_t N_r$ terms, however most of them have negligible impact on the representation quality and can be removed. The error induced when removing those terms from the sum (3.11) is given by

$$\mathbb{E}\left\{\|\mathbf{c_q} - \sum_{l\in\mathcal{N}} z_l \mathbf{e_l}\|^2\right\} = \sum_{l\notin\mathcal{N}} \mathbb{V}ar\left\{z_l\right\}. \tag{3.12}$$

The MSE induced by not including some coefficients in the sum (3.11) is equal to the sum of their variances. The coefficients variances are obtained from the eigenvalues λ_l of the eigenvectors as

$$\mathbb{V}ar\left\{z_l\right\} = \mathbb{E}\left\{|z_l|^2\right\} = \mathbb{E}\left\{|\mathbf{c_q}\mathbf{e_l}^H|^2\right\} = \mathbb{E}\left\{\mathbf{e_l}\mathbf{c_q}^H\mathbf{c_q}\mathbf{e_l}^H\right\} = \mathbf{e_l}\mathbf{\Sigma_q}\mathbf{e_l}^H = \lambda_l$$

We showed in the previous section that the covariance matrix $\mathbf{\Sigma_q}$ can be decomposed into $\mathbf{\Sigma_q^f}(\tau_q) \otimes \mathbf{\Sigma_q^r}(\theta_{rx,q}) \otimes \mathbf{\Sigma_q^t}^H(\theta_{tx,q})$. Using Kronecker product ($\otimes$) properties, the eigenvectors of $\mathbf{\Sigma_q}$ can hence be written as

$$\mathbf{e_l} = \mathbf{e_{l_f}^f} \otimes \mathbf{e_{l_r}^r} \otimes \mathbf{e_{l_t}^t}^*$$

where $\mathbf{e_{l_f}^f}$, $\mathbf{e_{l_r}^r}$ and $\mathbf{e_{l_t}^t}$ are eigenvectors of $\mathbf{\Sigma_q^f}(\tau_q)$, $\mathbf{\Sigma_q^r}(\theta_{rx,q})$ and $\mathbf{\Sigma_q^t}^H(\theta_{tx,q})$. Likewise the eigenvalues λ_l can be written as

$$\lambda_l = \lambda_{l_f} \otimes \lambda_{l_r} \otimes \lambda_{l_t}$$

where λ_{l_f}, λ_{l_r} and λ_{l_t} are the corresponding eigenvalues of $\mathbf{\Sigma_q^f}(\tau_q)$, $\mathbf{\Sigma_q^r}(\theta_{rx,q})$ and $\mathbf{\Sigma_q^t}^H(\theta_{tx,q})$.

Thus frequency domain and angular domain decomposition bases can be studied and compared separately. Those are the most efficient decomposition bases. However the vectors $\mathbf{e_l}$ depend on the parameters that govern the cluster, namely the angle of departure,

angle of arrival and delay, and on the statistical Saleh-Valenzuela parameters λ (intra-cluster ray rate of arrival), γ (intra-cluster ray power decrease time constant) and σ_θ (intra-cluster angular spread).

In the next section we test the efficiency of the widely used sub-optimal Fourier projection basis against the optimal basis. This decomposition is motivated by the Toeplitz structure of covariance matrices $\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q)$, $\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{r}}(\theta_{rx,q})$ and $\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}\,H}(\theta_{tx,q})$. Indeed in the asymptotic case when the size of a Toeplitz matrix rises to infinity, its eigenvectors approach Fourier basis vectors [7].

The coefficients z_l that represent the channel in the Fourier domain are usually called *taps* in what is called a tap delay line model. Taps are then denoted by the lower case variable h. In the following we characterize the variances of the channel taps h in both the delay and angular domains.

3.2.2 Frequency domain representation

For any channel $H[k]$ represented in the frequency domain, the Fourier decomposition coefficients are usually called *time domain taps*, denoted by lower case symbols h_n defined as

$$h_n = \frac{1}{\sqrt{N_f}} \sum_{k=0}^{N_f-1} H[k]\, \mathrm{e}^{2\pi jnk/N_f}, n \in [\![0, N_f - 1]\!]. \tag{3.13}$$

The frequency domain channel can be rewritten as

$$H[k] = \frac{1}{\sqrt{N_f}} \sum_{n=0}^{N_f-1} h_n\, \mathrm{e}^{-2\pi jnk/N_f}\,.$$

Using the linear combination notation introduced above, the frequency domain channel $\mathbf{h}$ can be rewritten

$$\mathbf{h} = \sum_{n=0}^{N_f-1} h_n \mathbf{e_n} \tag{3.14}$$

where the coefficients of the Fourier basis vectors are given by

$$[\mathbf{e_n}]_k = \frac{1}{\sqrt{N_f}}\, \mathrm{e}^{-2\pi jnk/N_f} \tag{3.15}$$

We can recognize the frequency domain characteristic vectors defined in Equation (1.11). Those basis vectors are orthonormal thus the MSE introduced by discarding non-significant

time domain coefficients is given by

$$\mathbb{E}\left\{\|\mathbf{h}-\sum_{n\in\mathcal{N}} h_n\mathbf{e_n}\|^2\right\} = \sum_{n\notin\mathcal{N}} \mathbb{V}ar\left\{h_n\right\}. \tag{3.16}$$

Theorem. *The time domain taps are zero mean complex random variables characterized by their variance*

$$\mathrm{Var}(h_n) = \int_{-\infty}^{\infty} S_H(\tau) D_{N_f}^2\left(\pi\frac{n-\tau B}{N_f}\right)\mathrm{d}\,\tau \tag{3.17}$$

where S_H denotes the PDP of H and $D_N(x)$ denotes the normalized Dirichlet kernel

$$D_N(x) = \frac{\sin\left(Nx\right)}{N\sin\left(x\right)}. \tag{3.18}$$

The proof for this formula is given in Annex B.3. Using Equations (3.16) and (3.17), we can evaluate the MSE induced by the truncation without relying on Monte-Carlo simulations. We evaluated the representation efficiency for both the optimal eigenvector basis and the Fourier decomposition on a single cluster model ($\mathbf{h} = \mathbf{c_q}$) using Equation (3.6) as the cluster PDP. Results are shown on Figure 3.2 in millimeter wave scenarios and Figure 3.3 for centimeter wave scenarios. We measured analytically the relative MSE induced by truncating the decomposition, only keeping the N more significant coefficients (N in abscissa). Note that for the same number of parameters, the truncation error is always lower with the eigendecomposition than with the Fourier basis. Broadly speaking, we can observe that as the bandwidth increase, more coefficients are required to represent the clusters. The gap between the optimal eigenbasis and the Fourier basis is small enough for typical configurations to confirm the use of the Fourier basis.

3.2.3 Angular domain Representation

We perform the same analysis in the angular domain by comparing the optimal eigendecomposition and a Fourier (steering vector) decomposition basis. As the main direction of cluster θ_q is already known, we use a shifted Fourier basis designed to minimize the number of significant coefficients. The first vector of the decomposition is shifted to ensure that it matches the main cluster direction $\mathbf{e_0}(\theta_q) = \mathbf{e_t}(\theta_q)$. This approach is also used in

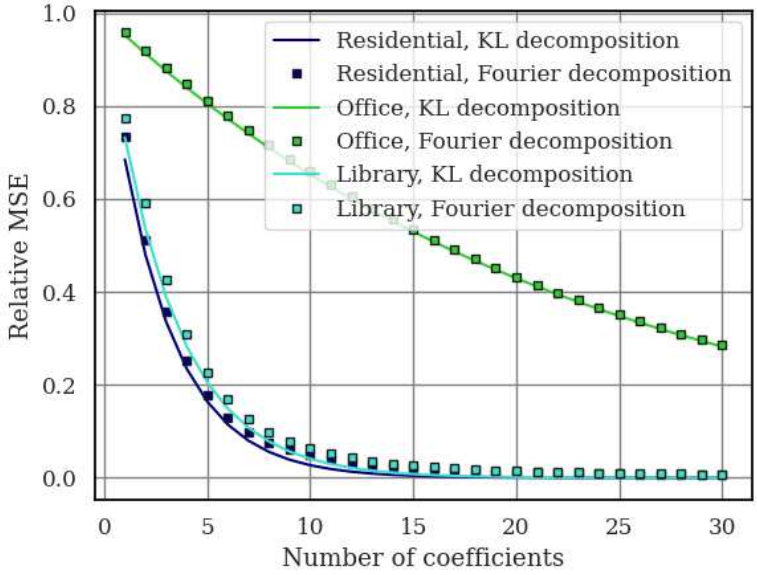

Figure 3.2 – Relative MSE when truncating the representation. In this evaluation we used the Numerology 3 defined in the 5G NR standard ($120kHz$ subcarrier spacing, $400MHz$ bandwidth).

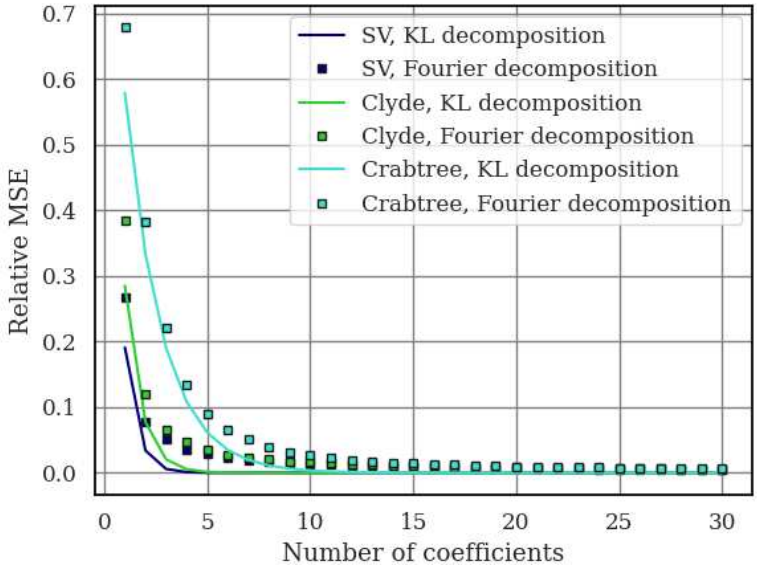

Figure 3.3 – Relative MSE when truncating the representation. In this evaluation we used the standard LTE numerology ($15kHz$ subcarrier spacing, $20MHz$ bandwidth).

Chapter 5 and in the research paper [8]. The shifted Fourier basis vectors are

$$[\mathbf{e_l}(\theta_q)]_m = \frac{1}{\sqrt{N_t}} \mathrm{e}^{-2\pi j\left(m-\frac{N_t-1}{2}\right)\left(\frac{l}{N_t}+\frac{\|\vec{a}_{tx}\|}{\lambda}\cos(\theta_q)\right)}.$$

In this shifted Fourier basis the angular domain cluster can be rewritten as

$$\mathbf{c_q} = \sum_{p=1}^{P} \beta_{p,q}\mathbf{e_t}(\theta_q+\theta_{p,q}) = \sum_{l=0}^{N_t-1} h_l \mathbf{e_l}(\theta_q)$$

The basis vectors $\mathbf{e_l}(\theta_q)$ remains orthonormal to each others. Thus we have the same MSE expression as in the frequency domain when discarding non-significant angular-domain taps, as

$$\mathbb{E}\left\{\|\mathbf{c_q} - \sum_{l\in\mathcal{N}} h_l \mathbf{e_l}(\theta_q)\|^2\right\} = \sum_{l\notin\mathcal{N}} \mathbb{V}ar\left\{h_l\right\}. \tag{3.19}$$

The angular taps variances are given by

$$\mathbb{V}ar\left\{h_l\right\} = \mathbb{E}\left\{\sum_{p=1}^{P}|\beta_{p,q}|^2\right\}\mathbb{E}\{|\langle \mathbf{e_t}(\theta_q+\Delta\theta_{p,q}), \mathbf{e_l}(\theta_q)\rangle|^2\}. \tag{3.20}$$

where the first term is a channel normalization constant and we are mainly interested in the latter term.

Theorem. *The second factor of Equation (3.20) is given by*

$$\mathbb{E}\{|\langle \mathbf{e_t}(\theta_q+\Delta\theta_{p,q}), \mathbf{e_l}(\theta_q)\rangle|^2\} = \int_0^{2\pi} f_{\Delta\theta}(\Delta\theta) D_{N_t}\left(x_l(\Delta\theta)\right) D_{N_t}\left(x_l(\Delta\theta)\right) d\Delta\theta \tag{3.21}$$

where $x_l(\Delta\theta) = \frac{l}{N_t} - \frac{\|\vec{a}_{tx}\|(\cos(\theta_q)-\cos(\theta_q+\Delta\theta))}{\lambda}$ *and* $D_N(x)$ *denotes the Dirichlet kernel (defined in Equation (3.18)) and* $f_{\Delta\theta}(\Delta\theta)$ *is the angular distribution of the intra-cluster rays (typically a Laplace distribution).*

From Equation (3.20), and the above theorem, we can now compare the efficiency of the angular Fourier decomposition with the optimal eigendecomposition. In this section we adapted the Fourier basis so that the first Fourier vector corresponds to the cluster main direction. As a comparison, the Fourier basis only depends on the cluster main direction θ_q and the number of antennas while the optimal eigenvector basis also depends on the angular spread $\sigma_{\Delta\theta}$. We tested various configurations on Figure 3.4. On sub-figure 3.4.a, several cluster main angles are tested while other parameters are fixed. As the

angular resolution is higher at array broadside ($\theta_q = 90°$) than endfire, more coefficients are required to model the cluster. Several angular spreads are tested on sub-figure 3.4.b while keeping the cluster main direction at broadside. More coefficients are required to model clusters with larger angular spreads. Finally we also tested on sub-figure 3.4.c the less commonly used Gaussian distribution against the classical Laplacian distribution.

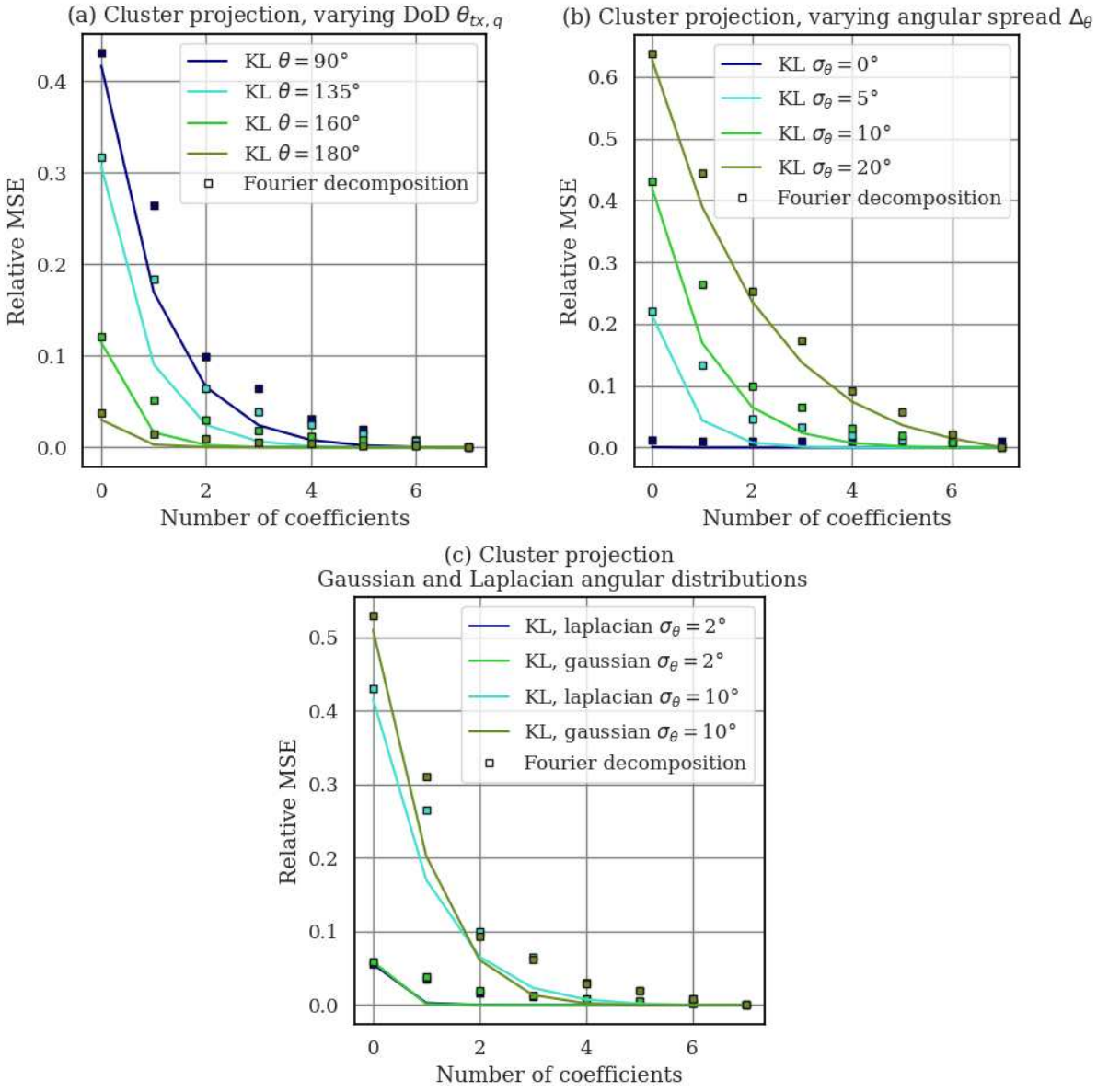

Figure 3.4 – Relative MSE when truncating the representation. In this evaluation we used $N = 8$ antennas. In Figure (a), the angular spread is fixed $\sigma_{\Delta\theta} = 10°$. In Figures (b) and (c), the cluster main direction is fixed to broadside ($\theta_q = 90°$).

3.3 Degrees of Freedom of MIMO Channels

Those coefficients are the degrees of freedom that describe the MIMO channel. We consider two modeling approaches. First we analyze each frequency-dependent MIMO channel matrix coefficient $H_{i,j}[k],\ k \in [0, N_f - 1]$. Each one of those coefficients $H_{i,j}[k]$ represents the equivalent SISO channel from an emitting antenna to a receiving antenna on subcarrier k. The first modeling approach consists in representing the MIMO channel as a collection of SISO channels (Figure 3.5). This is the usual working approach in small size MIMO systems. We count how much *time domain taps* are required to model each one of the $N_r \times N_t$ antenna to antenna radio links $H_{i,j}[k]$, over the frequency band B and according to the modeling error constraint ε. The truncation error is given by the sum of the taps variances as in Equation (3.16). The variances are calculated using Equation (3.17) where the PDP of the whole Saleh-Valenzuela channel is given by Equation (3.17). As all radio links experience the same propagation conditions, we multiply this number of taps by the amount of radio links $N_r \times N_t$. We called this method the *antenna based modeling*.

On one other hand, the channel consists in a sum of clusters $\mathbf{h} = \sum_{q=1}^{Q} \mathbf{c_q}$. Each cluster can be modeled as a linear combination of basis vectors (see Section 3.2.1). To simplify the evaluations we approximate in the angular domain all clusters by their main directions of departure and arrival. This is equivalent to truncating the angular domain Fourier decomposition of Section 3.2.3 to its first, more significant coefficient. This hypothesis simplifies further evaluations by eliminating the influences of angular spread $\sigma_{\Delta\theta}$ and cluster direction θ_q at both transmit and receive sides. However it limits the scope to small angular spreads and slightly underestimates the number of coefficients required to model each cluster.

We model each cluster in the frequency domain $\mathbf{c_q}$ using the same methodology used when modeling radio links. Each cluster is represented by a set of Fourier domain taps $h_{q,n}$ as

$$\mathbf{c_q} = \sum_{n=0}^{N_f-1} h_{q,n}\mathbf{e_n}$$

where $\mathbf{e_n}$ are Fourier vectors as defined in Equation (3.14). This representation is evidenced on Figure 3.6

We determine the minimal amount of coefficients required to model the full channel using a cluster decomposition. Note that this approach usually requires much more co-

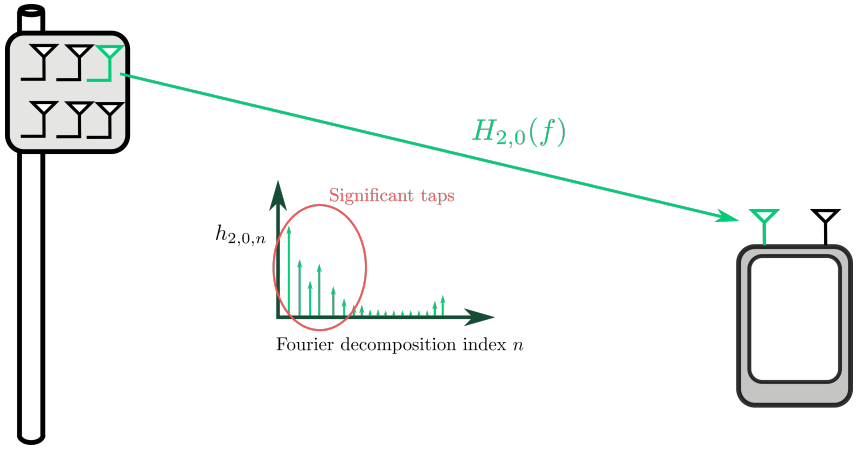

Figure 3.5 – *Antenna based modeling* : each radio link between a transmitting antenna and a receiving antenna is modeled as a *tapped delay line*.

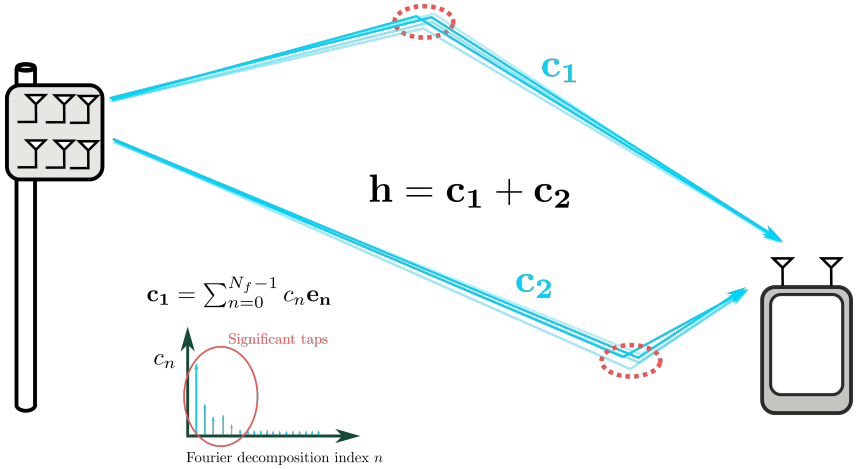

Figure 3.6 – *Propagation based modeling* : each cluster constituting the channel is modeled as a *tapped delay line*.

efficients than the first modeling technique. However the number of coefficients doesn't scale with the array sizes. We called this method the *propagation based modeling*.

We evaluated the two approaches in both centimeter (Figure 3.7) and millimeter wave (Figure 3.8) scenarios (see Section 1.1). In the first case we used the standard LTE numerology ($B = 20MHz$, $N_f = 1200$, $15kHz$ subcarrier spacing) for sub $6GHz$ bands. In the second case we used the new 5G NR numerology $\mu = 3$ ($B = 400MHz$, $N_f = 3300$, $120kHz$ subcarrier spacing). The *antenna based approach* is as expected more efficient for small number of antennas whereas the *propagation based approach* is more efficient for large-scale MIMO. The transition from the former to the later approaches happens surprisingly fast at $N_r \times N_t > 10$ at most. This configuration is easily attained if we consider multiple receive antennas.

We neglected the cluster angular spread. Wider arrays could be considered by splitting the clusters into multiple subclusters depending on $N_r \times N_t$. The flat curves of Fig. 6 would be slightly increasing and the previous observations and conclusions wouldn't change.

This is useful to have some intuition to understand when dedicated signal processing leveraging the angular sparsity is required to mitigate the complexity overhead when scaling up the array sizes. We draw in red the boundary between both regimes relative error between 5% and 40%.

3.4 A simplified cluster model

On a signal processing standpoint, working with Saleh-Valenzuela clusters is challenging. To prepare the follow-up of the manuscript, a simplified Saleh-Valenzuela cluster model is introduced in this section. It will be used later on for theoretical analyses and to derive practical algorithms. According to this simplified model, each cluster $\mathbf{c_q}$ is a realization of a complex centered Gaussian distribution as

$$\mathbf{c_q} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma_q}(\tau_q, P_q, \gamma_q, \lambda_q, \theta_{tx,q}, \theta_{rx,q}, \sigma_{\Delta\theta,tx}, \sigma_{\Delta\theta,rx})\right) \tag{3.22}$$

with covariance matrix $\mathbf{\Sigma_q}$, given by Equation (3.3). In this model the cluster angles of departure $\theta_{tx,q}$, arrival $\theta_{rx,q}$, delay τ_q; the statistical parameters λ, γ and $\sigma_{\Delta\theta}$ are fixed. In particular they rule the shape of the covariance matrix $\mathbf{\Sigma_q}$. This covariance matrix is divided into a frequency covariance matrix $\mathbf{\Sigma_q^f}$ and transmit and receive angular covariance

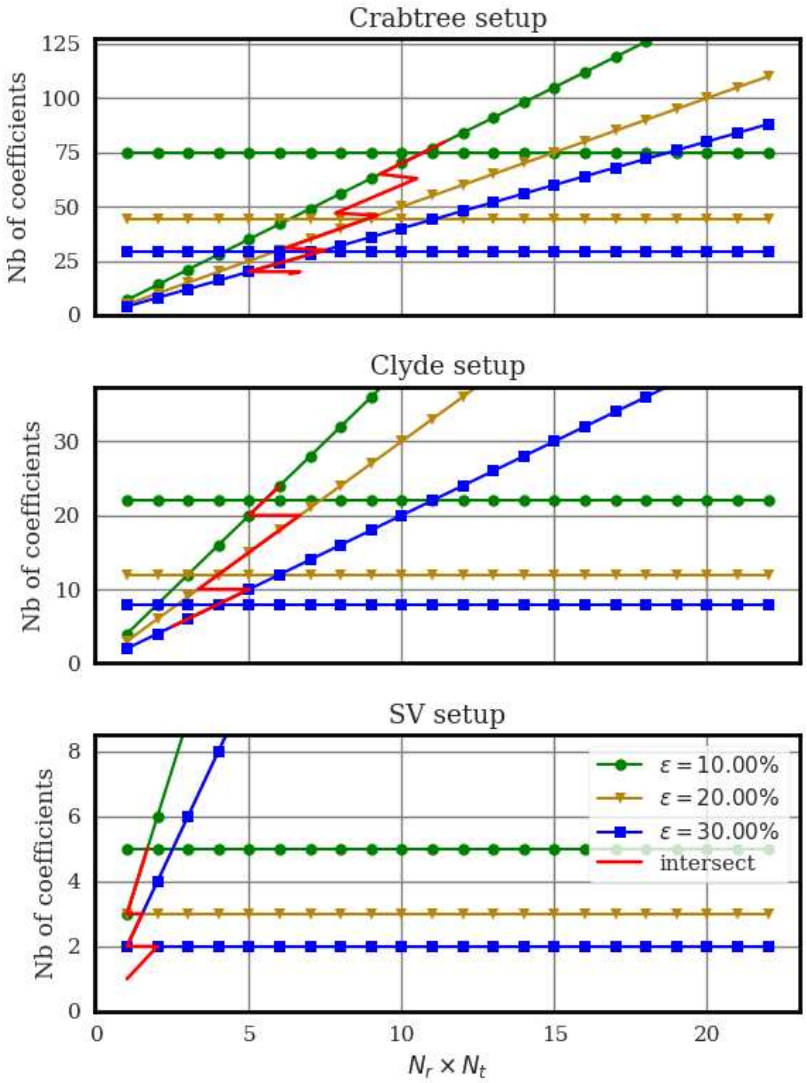

Figure 3.7 – Number of taps required to achieve the relative MSE ε for both the term by term MIMO channel and the cluster-based representation. Evaluations performed in centimeter wave scenarios with a classical LTE $20MHz$ numerology.

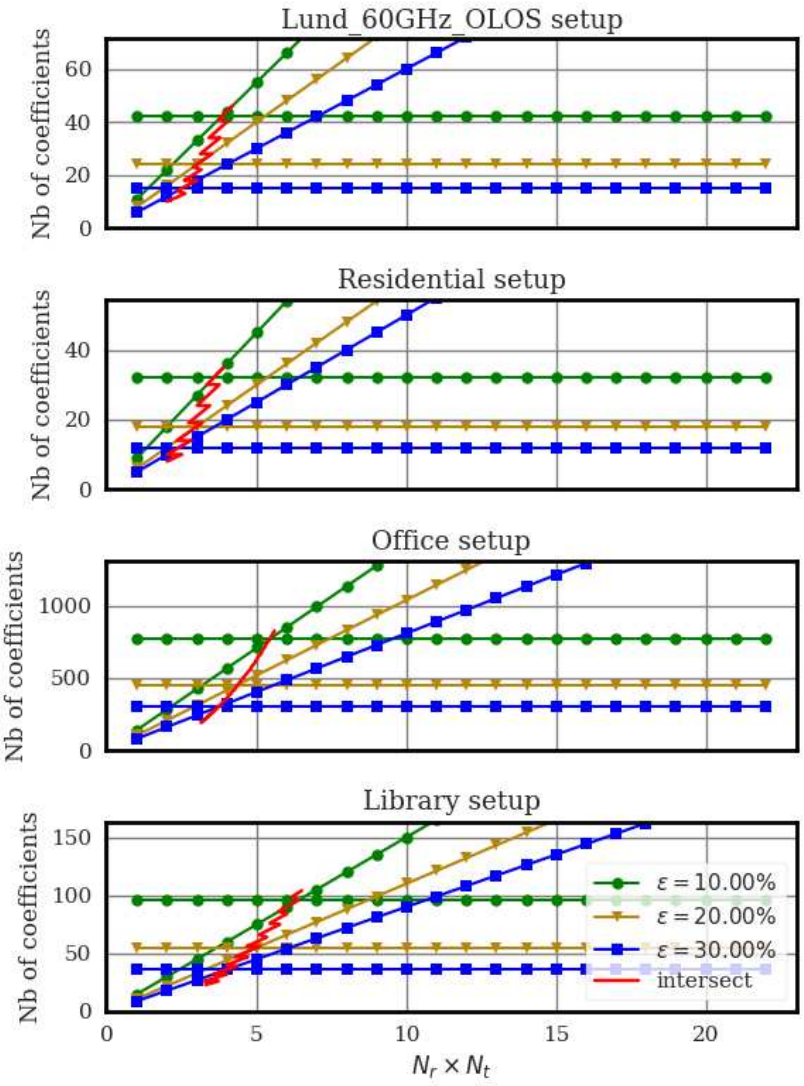

Figure 3.8 – Number of taps required to achieve the relative MSE ε for both the term by term MIMO channel and the cluster-based representation. Evaluations performed in millimeter wave scenarios with the 5G NR numerology.

matrices $\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}}$ and $\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{r}}$ as

$$\mathbf{\Sigma_q} = \mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q, P_q, \lambda_q, \gamma_q) \otimes \mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{r}}(\theta_{rx,q}, \sigma_{\Delta\theta,rx}) \otimes \mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{t}\,H}(\theta_{tx,q}, \sigma_{\Delta\theta,tx}) \tag{3.23}$$

It is then no longer necessary to use a ray-based approach to describe a Saleh-Valenzuela cluster. This expression is more convenient as random vectors are simpler to handle and generate. In the following chapters, theoretical analyses and algorithms will be derived based on this simplified model. However simulations will still be performed using the ray generation technique to verify the adequacy with the simplified model.

3.5 Conclusion

In this chapter we have studied the statistical properties of Saleh-Valenzuela channels. First we have derived the second order statistics of both clusters and the whole channel. Those statistics will be reused in the following chapters.

Based on those second order statistics, we have performed a decomposition of Saleh-Valenzuela clusters into a basis expansion model. We have studied two classical bases, the optimal eigenbasis and the widely used Fourier basis. We have shown that the gap between both bases is not significant in most cases, thus motivating the use of the simpler Fourier basis.

We quantified how much coefficients are required to model a MIMO channel according to two modeling approaches. This comparison is based on the second order channel statistics that were previously derived. The matrix coefficient (resp. multipath) based approach is suitable for small-scale (resp. large-scale) MIMO. We performed extensive numerical evaluations using calibrated Saleh-Valenzuela models available in the literature and found out that the transition between those two representations arises at surprisingly small number of antennas.

We finally proposed a simplified cluster model based on a Gaussian distribution parameterized by the cluster power, the cluster delay of arrival, the shape parameters λ and γ, and the angular parameters. This model is simpler to generate and study as it does not requires to throw rays characterized by gains and directions of departure and arrival. As it is based on well-known Gaussian distributions, it can be easily handled in analytical theoretical studies.

BIBLIOGRAPHY

[1] Arjan Meijerink and Andreas F. Molisch, « On the Physical Interpretation of the Saleh–Valenzuela Model and the Definition of Its Power Delay Profiles », *in*: *IEEE Transactions on Antennas and Propagation* (Sept. 2014).

[2] Morten Lomholt Jakobsen, Troels Pedersen, and Bernard Henri Fleury, « Analysis of the stochastic channel model by Saleh-Valenzuela via the theory of point processes », *in*: (2012).

[3] Matthieu Roy, Stephane Paquelet, and Matthieu Crussiere, « Degrees of Freedom of Ray-Based Models for mm-Wave Wideband MIMO-OFDM », *in*: *GLOBECOM 2019 - 2019 IEEE Global Communications Conference*, Waikiloa Village, 2019.

[4] Milan S. Derpich and Rodolfo Feick, « Second-Order Spectral Statistics for the Power Gain of Wideband Wireless Channels », *in*: *IEEE Transactions on Vehicular Technology* (Mar. 2014).

[5] Wout Debaenst et al., « RMS Delay Spread vs. Coherence Bandwidth from 5G Indoor Radio Channel Measurements at 3.5 GHz Band », *in*: *Sensors* (Jan. 2020).

[6] Bruce Hajek, *Random Processes for Engineers*, 1st ed., Cambridge University Press, Mar. 2015.

[7] Nikolai K Nikolski, Danièle Gibbons, and Greg Gibbons, *Toeplitz matrices and operators*, 2020.

[8] Hongxiang Xie et al., « A Unified Transmission Strategy for TDD/FDD Massive MIMO Systems With Spatial Basis Expansion Model », *in*: *IEEE Trans. Veh. Technol.* (Apr. 2017).

Chapter 4

Cluster estimation and separation

As introduced in Chapter 1, the propagation channel consists of multiple paths resulting from reflections on the environment. Channel measurement campaigns found that those paths are grouped into clusters that have a particular, exponentially decreasing power delay profile, which is accurately modeled by the well-known Saleh-Valenzuela model [1].

In this chapter, we address the recovery of the individual clusters that constitute the channel based on received symbols. The cluster separation problem statement is presented in Section 4.1. State of the art techniques are presented in Section 4.2. We present our novel approach for cluster separation in Section 4.3. In its current form, this technique is limited to delay domain cluster separation. Moreover further work is required to lift the remaining conditioning. Leads to a more generalized algorithm, that also take into account the spatial dimensions are given in Section 4.4.5.

4.1 Problem statement

Cluster separation is an old problem encountered in radio communications and channel characterization. As a reminder from previous sections, the multipath channel is made of a sum of clusters $\mathbf{c_q}$ as

$$\mathbf{h} = \sum_{q=1}^{Q} \mathbf{c_q} = \mathbf{Ac} \tag{4.1}$$

where $\mathbf{A} = [\mathbf{I}, \mathbf{I}, \cdots, \mathbf{I}]$ and $\mathbf{c} = [\mathbf{c_1}, \mathbf{c_2}, \cdots, \mathbf{c_Q}]$ are concatenations of identity matrices and cluster vectors. The cluster separation problem can then be seen as the resolution of the linear problem

$$\mathbf{y} = \mathbf{Xh} + \mathbf{n} = \mathbf{XAc} + \mathbf{n} \tag{4.2}$$

with unknown $\mathbf{c}$. Note that the number of cluster Q, and consequently the size of $\mathbf{c}$ is also generally not known. This inverse problem is ill-posed as there are $2QN_f$ real unknowns

(the complex vector $\mathbf{c}$) and only $2N_f$ real equations. The problem has to be regularized using additional assumptions on the cluster structure. Indeed Equation (4.2) does not take into account the characteristics of clusters. It has to be reformulated by incorporating a proper *cluster definition*, that describes the features of the structure that we want to extract.

We first present an overview of existing cluster separation techniques, that all share the same pattern. First a (linear or non-linear) transformation is applied to the samples to concentrate channel information into a few variables, from which clusters can easily be separated in a second stage, either by visual inspection or by using a heuristic function, that actually represents the cluster definition.

Then we present a novel algorithm for pilot-aided cluster estimation and identification. The procedure is based on an Expectation-Maximization (EM) algorithm that consists of two main steps (*expectation* and *maximization* steps) that are iteratively repeated until all clusters have been extracted from the received symbols. This novel algorithm is based on the cluster definition proposed at the end of Section 3.4.

In the following, we propose to solve the problem from frequency domain measurements. This is due to two main reasons:

— The impact of the cluster shape is more significant in the frequency domain than in the angular domain.

— The hardware developed at b<>com can operate on a very large bandwidth, up to 2 GHz. It provides a massive resolution in the frequency domain.

As a consequence we simplify the study by only considering the frequency domain parameters. The modifications required to extend the algorithm to the angular (transmit and receive) dimensions are described at the end of this chapter.

4.2 State of the art in cluster separation

Existing clustering algorithms always follow the same pattern. A sparse representation of the channel is first calculated using either linear (Fourier Transform) or non-linear (high resolution schemes) techniques. This step concentrate the information on a reduced set of parameters, from which clusters can be easily evidenced. The high resolution schemes typically used in the literature are Multiple Signal Classification (MUSIC) [2], Estimation of Signal Parameter Via Rotational Invariance Technique (ESPRIT) [3], Space Alternating Generalized Expectation-maximization (SAGE) [4] and CLEAN [5]. Those algorithms

decompose the channel into a sum of specular rays. The clusters are then estimated by grouping those rays either visually (cluster definition based on operator knowledge) or based on a heuristic function. In the latter case the heuristic defines what is a cluster. The literature on cluster separation is broad and abundant and we provide in Table 4.1 a short overview of the state of the art.

Reference	Sparse representation	Cluster separation
[6]	CLEAN	visual inspection
[7]	Fourier Transform	visual inspection
[8]	SAGE	KMeans
[9]	SAGE	KPowerMeans
[10]	SAGE	KMeans
[5]	CLEAN	Heuristic function
[11]	Fourier Transform	Heuristic function

Table 4.1 – Non-exhaustive state of the art in cluster separation.

The KMeans algorithm is a standard application-independent point clustering algorithm from the unsupervised machine learning field. It aims at partitioning points defined by coordinates into groups based on their distance. The main hyper-parameter of this heuristic is K, the number of groups. The KPowerMeans uses power-weighted distances based on estimated ray powers. This algorithm does not take into account the characteristic exponentially decreasing power delay profile of clusters.

To overcome those limitations, a clustering algorithm should take into account the characteristics of clusters. Papers [5] and [11] achieve this goal by introducing more complex heuristics inspired by the visual appearance of clusters.

In this thesis we propose a novel approach that leverages the cluster model presented in the previous chapter. We show that the first step, which consists in a sparse decomposition of the channel, is not necessary to separate clusters but still useful for illustration purposes and complexity reduction.

4.3 Separating Saleh-Valenzuela clusters

In this section we remind the clustered channel model already presented in the previous chapters. The channel is made of multipath components $\mathbf{c_q}$. As previously mentioned in the introduction, we focus on the frequency domain where the received symbols are given

as

$$\mathbf{y} = \mathbf{\Delta_x} \sum_{q=1}^{Q} \mathbf{c_q} + \mathbf{n} \tag{4.3}$$

where the diagonal coefficients of $\mathbf{\Delta_x}$ denote the known pilot symbols sent on each subcarrier at the transmitter. This is a particular case of sensing matrix $\mathbf{X}$ (Equation (4.2)) whose structure is constrained by the model at hand, previously stated in Section 1.5. We also make the assumption that the coefficients of $\mathbf{\Delta_x}$ have constant unitary amplitude, which means that $\mathbf{\Delta_x}\mathbf{\Delta_x}^H = \mathbf{I}$. Using this property, the division of the received samples by the pilots to recover the propagation channel can be achieved cheaply. Pilot schemes employed in modern communication systems feature this property [12].

As already mentioned in the introduction, recovering the clusters $\mathbf{c_q}$ from the received symbols $\mathbf{y}$ is an ill-posed inverse problem. We need to add prior information on the vectors $\mathbf{c_q}$ to regularize it. We use the cluster model introduced in Section 3.4 and recalled in Section 4.3.1 to regularize the inverse problem. This model is based on the Saleh-Valenzuela cluster second-order moment characterization performed in Chapter 3. It is a function of a set of parameters $\boldsymbol{\theta_q}$, from which we will only consider the frequency domain parameters P, τ_q, γ_q and λ_q.

The cluster separation problem actually consists of two main sub-problems. The separation itself requires the knowledge of the cluster prior information, that depends on $\boldsymbol{\theta_q}$. Estimating this prior $\boldsymbol{\theta_q}$ is easily achieved when clusters have already been separated. However, solving both sub-problems simultaneously is more tricky.

In Section 4.3.2, we evaluate the Cramér-Rao Lower Bound (CRLB) on the covariance of the parameters estimates that rule this model. This analysis is useful to check the well-posedness of the estimation problem. Then we solve two preliminary sub-problems, namely the estimation of the parameters of a single cluster embedded in noise and the separation of clusters with known parameters, in Sections 4.3.3 and 4.3.4 respectively. This brings together all the pieces that are required to address the main problem, that is the separation of clusters with unknown characteristics, solved in Section 4.4.

4.3.1 Cluster model

We use the cluster model already introduced in Section 4.3.1. It is represented by a vector $\mathbf{c_q}$ that follows a complex centered Gaussian vector as

$$\mathbf{c_q} = \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma_q}(\boldsymbol{\theta_q})\right). \tag{4.4}$$

The covariance matrix $\mathbf{\Sigma_q}$ of cluster q is parametrized by the cluster angles of departure $\theta_{tx,q}$, arrival $\theta_{rx,q}$, delay τ_q; the statistical parameters λ, γ and $\sigma_{\Delta\theta}$. As we focus on the frequency dimension of the channel, the cluster covariance matrix is simply the cluster frequency domain covariance

$$\mathbf{\Sigma_q} = \mathbf{\Sigma_q^f}(\tau_q, P_q, \gamma_q, \lambda_q). \tag{4.5}$$

We assumed that each cluster would be characterized by its own Saleh-Valenzuela cluster parameters γ_q, λ_q as this enables more flexibility in the cluster shapes. As a reminder, this matrix has already been evaluated in Equation (3.8) as

$$[\mathbf{\Sigma_q^f}(\tau_q, P_q, \gamma_q, \lambda_q)]_{k,k'} = P_q \left(1 + \frac{\lambda_q \gamma_q}{1 - 2\pi j \frac{k'-k}{N_f} B\gamma_q}\right) \mathrm{e}^{2\pi j \frac{k'-k}{N_f} B\tau_q}.$$

This model constitutes the prior information to be inserted in our estimation algorithm.

4.3.2 Bounds on parameters estimates

As cluster parameters $\boldsymbol{\theta_q}$ are not known beforehand, at some point they have to be estimated. In this section we derive a lower bound on the variance of the estimates. Based on the analysis of this bound, we make sure that the estimation problem is well-posed and well-conditioned. The lower bound is also a useful benchmark to evaluate how well a given practical estimator operates and how much room there is for improvement.

Among the various existing bounds, we choose to evaluate the CRLB which is, by comparison, easy to determine and well-suited for this particular estimation problem. As a reminder the received samples are given by Equation (4.3) as

$$\mathbf{y} = \mathbf{\Delta_x} \sum_{q=1}^{Q} \mathbf{c_q} + \mathbf{n}$$

We model clusters as Gaussian vectors with covariance matrix $\mathbf{\Sigma_q^f}(\tau_q, P_q, \gamma_q, \lambda_q)$. Moreover, the thermal noise vector $\mathbf{n}$ is also modeled as a zero-mean i.i.d. Gaussian vector $\mathbf{n} \sim$

$\mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$. Thus $\mathbf{y}$ is a sum of Gaussian vectors, therefore it is Gaussian distributed as

$$\mathbf{y} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Delta_x}\left(\sum_{q=1}^{Q} \mathbf{\Sigma_q^f}(\tau_q, P_q, \gamma_q, \lambda_q)\right)\mathbf{\Delta_x}^H + \sigma_n^2\mathbf{I}\right).$$

The CRLB is obtained by inversion of the Fisher Information matrix. When considering complex zero-mean Gaussian distributed samples, the Fisher Information matrix between parameter θ_i and parameter θ_j is given in [13, Equation (15.52) p. 525] as

$$\mathbf{I}_{\theta_i,\theta_j} = \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_i}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\right\} \tag{4.6}$$

where $\mathbf{\Sigma} = \mathbf{\Delta_x}\left(\sum_{q=1}^{Q} \mathbf{\Sigma_q^f}(\tau_q, P_q, \gamma_q, \lambda_q)\right)\mathbf{\Delta_x}^H + \sigma_n^2\mathbf{I}$.

We first verify that the cluster parameter estimation problem is well-conditioned. To this end we evaluate the CRLB for the parameters of a single cluster embedded in noise, namely $P_q, \tau_q, \gamma_q, \lambda_q$.

Theorem. *The structure of the Fisher Information matrix for the set of parameters* $[\tau_q, P_q, \gamma_q, \lambda_q]$ *is given as*

$$\mathbf{I}(\tau_q, P_q, \gamma_q, \lambda_q) = \begin{bmatrix} I_{\tau_q} & 0 & I_{\tau_q,\gamma_q} & I_{\tau_q,\lambda_q} \\ 0 & I_{P_q} & I_{P_q,\gamma_q} & I_{P_q,\lambda_q} \\ I_{\gamma_q,\tau_q} & I_{\gamma_q,P_q} & I_{\gamma_q} & I_{\gamma_q,\lambda_q} \\ I_{\lambda_q,\tau_q} & I_{\lambda_q,P_q} & I_{\lambda_q,\gamma_q} & I_{\lambda_q} \end{bmatrix} \tag{4.7}$$

Some matrix coefficients can be easily evaluated. For instance $I_{P_q} = \sum_{k=1}^{N_f} \frac{\nu_k^2}{(P\nu_k+\sigma_n^2)^2}$ *where* ν_k *are the eigenvalues of* $\mathbf{\Sigma_q^f}$*. The mixed terms between* P_q *and* τ_q *can also be analytically computed as* $\mathbf{I}_{P_q,\tau_q} = \mathbf{I}_{\tau_q,P_q} = 0$.

This is proven in Annex C.1. Note that when extending this analysis to the angular domain, the Fisher Information matrix coefficients between DOA, DOD, delay τ_q and power P_q are also null, as evidenced in Chapter 6 and in the paper [14]. The Cramér-Rao matrix is obtained by inverting the Fisher Information matrix as

$$\mathbf{C_{CR}}(\tau_q, P_q, \gamma_q, \lambda_q) = [\mathbf{I}(\tau_q, P_q, \gamma_q, \lambda_q)]^{-1} \tag{4.8}$$

Theorem. *(Reminder) In this multivariate estimation problem, the covariance matrix*

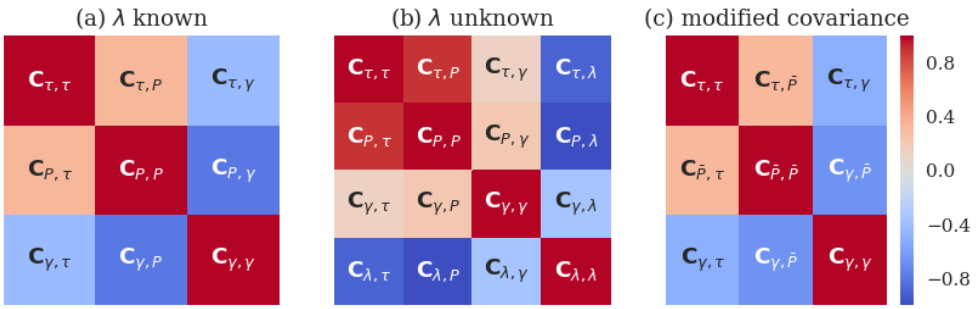

Figure 4.1 – Numerical evaluation of the normalized Cramér-Rao bound on parameters estimates. On subfigure (a) we consider that parameter λ is known beforehand. On subfigure (b) λ is unknown and has to be estimated. The Cramér-Rao bound for the modified covariance is given of Figure (c). Original Saleh-Valenzuela parameters, $20MHz$ LTE numerology.

$\mathbf{C_{est}}$ *of any unbiased estimator verifies [13, Equation (3.24) p. 44]*

$$\mathbf{C_{est}} - \mathbf{C_{CR}}(\tau_q, P_q, \gamma_q, \lambda_q) \geq \mathbf{0} \tag{4.9}$$

where $\geq \mathbf{0}$ *means that the matrix is positive semidefinite. This implies that the variance of any practical estimator* $\hat{\theta}_i$ *of parameter* θ_i*, given by* $\mathbb{V}ar\{\hat{\theta}_i\} = [\mathbf{C_{est}}]_{i,i}$ *for each* $\theta_i \in \{\tau_q, P_q, \gamma_q, \lambda_q\}$*, verifies*

$$\mathbb{V}ar\{\hat{\theta}_i\} \geq [\mathbf{C_{CR}}(\tau_q, P_q, \gamma_q, \lambda_q)]_{i,i} \tag{4.10}$$

We evaluate the Cramér-Rao bound for a single cluster in a $20MHz$ LTE setup ($N_f = 1200$, $15kHz$ subcarrier spacing) over a Saleh-Valenzuela channel (original set of parameters) on Figure 4.1. We normalized the matrices by the diagonal elements to obtain a correlation matrix instead of a covariance matrix. This operation highlights the correlations between the coefficients. In a general manner, the closer the off-diagonal coefficients are to 0, the better. It means that the estimation problem is well-conditioned and the parameters can be estimated jointly. We notice that when λ_q is known (Subfigure (a)), the problem is well conditioned. However when λ_q is unknown and should be estimated along with other parameters, strong off-diagonal correlation coefficients appear. The problem gets ill-conditioned. Indeed, when $\lambda\gamma \gg 1$ (condition which is usually achieved), the covariance matrix can be approximated by

$$[\mathbf{\Sigma}^{\boldsymbol{f}}_{\boldsymbol{q}}(\tau_q, P_q, \gamma_q, \lambda_q)]_{k,k'} \approx P_q \frac{\lambda_q \gamma_q}{1 - 2\pi j \frac{k'-k}{N_f} B\gamma_q} \mathrm{e}^{2\pi j \frac{k'-k}{N_f} B\tau_q}.$$

Parameters P_q and λ_q actually play the same role in this equation and can't be jointly estimated. To improve the estimation problem conditioning, we choose to discard parameter λ and aggregate the total cluster power into $\bar{P}_q = \lambda_q \gamma_q P_q$. This simplifies the estimation problem by removing one parameter while improving the conditioning of the system (Figure 4.1 (c)). However it can lead to biased estimates in certain circumstances when $\lambda\gamma \approx 1$. The novel covariance formula, used in the following developments, is given by

$$[\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q, P_q, \gamma_q, \lambda_q)]_{k,k'} \approx \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}(\tau_q, \bar{P}_q, \gamma_q)]_{k,k'} = \frac{\bar{P}_q}{1 - 2\pi j \frac{k'-k}{N_f} B\gamma_q} \mathrm{e}^{2\pi j \frac{k'-k}{N_f} B\tau_q} . \tag{4.11}$$

4.3.3 Single cluster parameters estimation

Estimating cluster parameters τ_q, $\bar{P}_q$, γ_q of a single cluster can be easily performed using standard solvers. In this section we take advantage of this simple estimation problem to introduce the representation method that we use for the Cramér-Rao bound. We also verify that the hypotheses made beforehand (Gaussian vector hypothesis in Section 4.3.1 and modified covariance matrix in Equation (4.11)) does not degrade the performance (bias and covariance) of the estimators.

The distribution of a single cluster $\mathbf{c_q}$ using the modified covariance matrix is given by

$$\mathbf{c_q} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\tau_q, \bar{P}_q, \gamma_q)\right). \tag{4.12}$$

We assume that this cluster is not known perfectly. Thus it is polluted by some zero-mean Gaussian distributed noise $\mathbf{n}$ parametrized by its covariance matrix $\sigma_n^2\mathbf{I}$. The distribution of the measured cluster $\tilde{\mathbf{c}}_\mathbf{q} = \mathbf{c_q} + \mathbf{n}$ is then given as

$$\tilde{\mathbf{c}}_\mathbf{q} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\tau_q, \bar{P}_q, \gamma_q) + \sigma_n^2\mathbf{I}) \tag{4.13}$$

The maximum likelihood estimate of the cluster parameters $\boldsymbol{\theta_q} = \{\tau_q, \bar{P}_q, \gamma_q\}$ is given by

$$\hat{\boldsymbol{\theta}}_\mathbf{q} = \underset{\boldsymbol{\theta_q}}{\operatorname{argmax}}\ p(\tilde{\mathbf{c}}_\mathbf{q}|\boldsymbol{\theta_q}) \tag{4.14}$$

where the probability density function of $\tilde{\mathbf{c}}_\mathbf{q}$ is given as

$$p(\tilde{\mathbf{c}}_\mathbf{q}|\boldsymbol{\theta_q}) = \frac{1}{(2\pi)^N |\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q}) + \sigma_n^2\mathbf{I}|} \mathrm{e}^{-\mathbf{y}^H \left(\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q}) + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{y}} \tag{4.15}$$

The estimators for the set of parameters $\boldsymbol{\theta_q}$ are then obtained using gradient descent techniques or grid search. For illustration purposes, we estimated the cluster parameters $\boldsymbol{\theta_q}$ using a maximum likelihood approach. We assumed that the cluster was embedded into thermal noise with known covariance matrix $\sigma_n^2\mathbf{I}$. We also used the modified covariance function (4.11), where parameter λ_q is no longer required. The estimation has been performed by maximizing Equation (4.15) using the general purpose *BFGS* (Broyden, Fletcher, Goldfarb and Shanno) algorithm [15] implementation proposed in the *scipy.optimize* Python package. The optimization process is rather slow but could be widely improved by designing a dedicated maximization algorithm for this problem. The performance has been evaluated using Monte-Carlo simulations and are compared with the CRLB calculated using Equation (4.6) in Figure 4.3.

According to the properties of the Cramér-Rao Lower Bound, the covariance of any estimator verifies

$$\mathbf{C_{est}} - \mathbf{C_{CR}}(\tau_q, P_q, \gamma_q, \lambda_q) \geq \mathbf{0} \tag{4.16}$$

where $\geq \mathbf{0}$ means definite positive (all eigenvalues are positive) and $\mathbf{C_{est}}$ denotes the empirical estimator covariance matrix, given as

$$\mathbf{C_{est}} = \mathbb{E}\left\{(\hat{\boldsymbol{\theta}}_\mathbf{q} - \mathbb{E}\left\{\hat{\boldsymbol{\theta}}_\mathbf{q}\right\})(\hat{\boldsymbol{\theta}}_\mathbf{q} - \mathbb{E}\left\{\hat{\boldsymbol{\theta}}_\mathbf{q}\right\})^T\right\}. \tag{4.17}$$

The multivariate relation (4.16) is not trivial to understand. To provide a more straightforward interpretation of the bound, we have chosen a representation method that fits well to the scatterplots of Figure 4.3. This method is further explained in Appendix C.2. The covariance matrix of two random variables is a 2×2 matrix, where diagonal terms are the variances of each variable and the off-diagonal term is the covariance between them, as

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2 \end{bmatrix}. \tag{4.18}$$

Because this matrix is symmetric positive-definite, it is diagonalizable with eigenvalues $\lambda_1 \geq \lambda_2 > 0$ in an orthonormal basis, which actually represents a rotation by an angle θ. The curve represented by the quadratic norm equation $\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x} = 1$ is an ellipse in the x_1, x_2 plane with semi-major axis λ_1, semi-minor axis λ_2 and rotation angle θ as depicted on Figure 4.2.

We study the pairwise relations between the parameters of $\boldsymbol{\theta_q}$. The covariance matrix

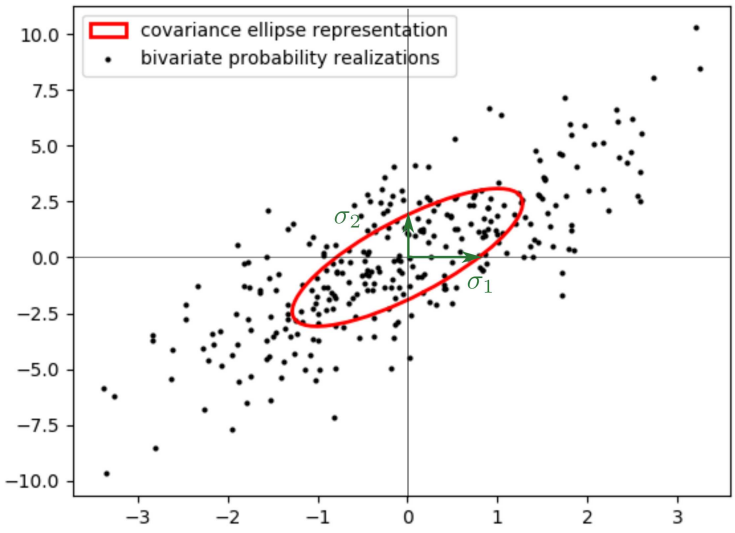

Figure 4.2 – Ellipse representation of the covariance matrix of a distribution.

of parameters θ_i and θ_j is given by :

$$\mathbf{R_{i,j}} = \begin{bmatrix} [\mathbf{C}]_{i,i} & [\mathbf{C}]_{i,j} \\ [\mathbf{C}]_{j,i} & [\mathbf{C}]_{j,j} \end{bmatrix} \tag{4.19}$$

where $\mathbf{C}$ can be either the optimal CRLB or the estimated covariance matrix, $\mathbf{C_{CR}}$ or $\mathbf{C_{est}}$. $\mathbf{R_{i,j}}$ is called a *principal submatrix* of $\mathbf{C}$. We draw on Figure 4.3 the corresponding ellipses that represent the optimal lower bound CRLB covariance (in red) and the measured covariance (in black) for each pair of parameters.

Theorem. *The ellipse that represents the CRLB is always inside any unbiased practical estimator covariance ellipse. Hence the red CRLB ellipse represents a geometric lower bound on the achievable covariance of any unbiased practical estimator.*

This theorem is proven in Annex C.2.

This property is showcased on Figure 4.3. We estimated the parameters of a single Saleh-Valenzuela (original set of parameters, see Table 1.1) cluster parametrized by delay $\tau_q = 100ns$, $\gamma_q = 20ns$ and $\bar{P} = (1 + \lambda_q\gamma_q) = 5$. We displayed the practical estimator covariance (in black) and the Cramér-Rao Lower Bound (in red) using the ellipse representation. The ellipses have been translated by the true parameter values

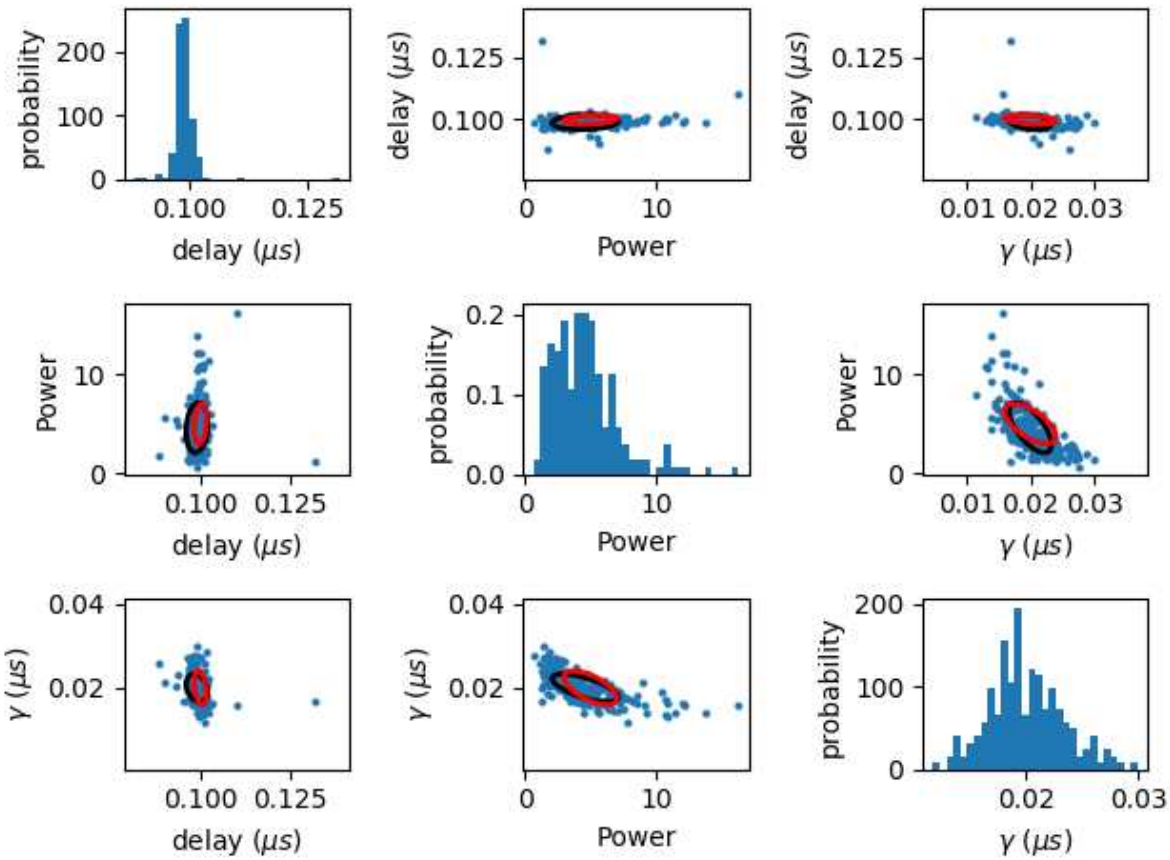

Figure 4.3 – Cluster parameters $\boldsymbol{\theta_q} = [\tau_q, \bar{P}_q, \gamma_q]$ estimation performance calculated using Monte-Carlo simulations and compared to the theoretical CRLB.

$[\tau_q = 100ns, \bar{P}_q = 5, \gamma_q = 20ns]$ for the red CRLB ellipse and the mean estimated parameter values $\mathbb{E}\left\{\hat{\boldsymbol{\theta}}_\mathbf{q}\right\}$ for the black ellipse. This enables a geometrical analysis of both bias and variance of the estimator.

Despite modeling the channel (which is a sum of rays) as a Gaussian vector and having simplified its covariance matrix to remove one parameter, the CRLB is still very close to the actual Maximum Likelihood estimator. The estimates are a bit biased, in particular the delay estimate as both ellipses are not concentric. Broadly speaking, empirical results follow the theoretical bounds; the Gaussian vector and modified covariance matrix hypotheses can be accepted in further calculations.

4.3.4 Separation of clusters with known characteristics

In this section we assume that the sets of parameters $\boldsymbol{\theta_q} = \{\bar{P}_q, \tau_q, \gamma_q\}$ that defines the clusters are known. We denote by $\boldsymbol{\theta}$ the vector that contains all cluster parameters $\boldsymbol{\theta} = [\boldsymbol{\theta_1}, \cdots, \boldsymbol{\theta_Q}]$. From Equation (4.3), the received symbols are given as

$$\mathbf{y} = \boldsymbol{\Delta}_\mathbf{x}\mathbf{A}\mathbf{c} + \mathbf{n} = \mathbf{A}'\mathbf{z} \tag{4.20}$$

where $\mathbf{z} = \begin{bmatrix}\mathbf{c}\\ \mathbf{n}\end{bmatrix}$ is obtained by concatenation of $\mathbf{c}$ and $\mathbf{n}$ and $\mathbf{A}'$ is defined as

$$\mathbf{A}' = \begin{bmatrix}\boldsymbol{\Delta}_\mathbf{x} & \boldsymbol{\Delta}_\mathbf{x} & \cdots & \boldsymbol{\Delta}_\mathbf{x} & \mathbf{I}\end{bmatrix}. \tag{4.21}$$

We wish to determine the distribution of $\mathbf{z}$ (the clusters) given the received symbols $\mathbf{y}$, $p(\mathbf{z}|\mathbf{y})$. This is an ill-posed inverse problem, as there are $2(Q+1)N_f$ unknowns and only $2N_f$ equations. We leverage the cluster model presented beforehand to regularize the separation problem. In this subproblem, model parameters $\tau_q, \bar{P}_q, \gamma_q$ are perfectly known for each cluster q. We have access to the distribution of the received symbols $p(\mathbf{y})$ and the distribution of the clusters $p(\mathbf{z})$. The latter is complex zero-mean Gaussian distributed with covariance

$$\boldsymbol{\Sigma}_\mathbf{z}(\boldsymbol{\theta}) = \begin{bmatrix}\boldsymbol{\Sigma_1}(\tau_1, \bar{P}_1, \gamma_1) & \cdots & 0 & 0\\ \vdots & \cdots & 0 & 0\\ 0 & \cdots & \boldsymbol{\Sigma_Q}(\tau_Q, \bar{P}_Q, \gamma_Q) & 0\\ 0 & \cdots & 0 & \sigma_n^2\mathbf{I}\end{bmatrix}$$

where $\boldsymbol{\theta}$ denotes the set of all parameters.

Theorem. *The conditional distribution of* $\mathbf{z}$ *given* $\mathbf{y}$ *and the set of parameters* $\boldsymbol{\theta}$ *is complex Gaussian distributed with mean and covariance matrix*

$$\begin{cases}\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}} = \boldsymbol{\Sigma}_\mathbf{z}(\boldsymbol{\theta})\mathbf{A}'^H\left(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{z}(\boldsymbol{\theta})\mathbf{A}'^H\right)^{-1}\mathbf{y}\\ \boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}} = \boldsymbol{\Sigma}_\mathbf{z}(\boldsymbol{\theta})\mathbf{A}'^H\left(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{w}(\boldsymbol{\theta})\mathbf{A}'^H\right)^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{z}(\boldsymbol{\theta})\end{cases} \tag{4.22}$$

This result is proven in Annex C.3. From this theorem, we can deduce that the mean

$\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}$ provides the maximum likelihood estimator for the clusters $\mathbf{c_q}$ as

$$\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}} = \hat{\mathbf{z}} = \begin{bmatrix} \hat{\mathbf{c}}_\mathbf{1} \\ \vdots \\ \hat{\mathbf{c}}_\mathbf{Q} \\ \hat{\mathbf{n}} \end{bmatrix}, \tag{4.23}$$

and that the covariance matrix $\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}$ of the distribution provides the variance of the estimator. This matrix can be rewritten in a cluster covariance block-diagonal fashion as

$$\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}} = \begin{bmatrix} [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{1,1} & [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{1,2} & \cdots & [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{1,Q+1} \\ \vdots & \vdots & \vdots & \vdots \\ [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{Q,1} & [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{Q,2} & \cdots & [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{Q,Q+1} \\ [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{Q+1,1} & [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{Q+1,2} & \cdots & [\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{Q+1,Q+1} \end{bmatrix} \tag{4.24}$$

where each element $[\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{q,q'}$ is a complex $N_f \times N_f$ matrix. The matrices on the diagonal of $\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}$, denoted as $[\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{q,q}$ are the error covariance matrices of cluster q. They measure the uncertainty on the estimated clusters on each subcarrier. Cross-diagonal elements $[\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}]_{q,q'}$ with $q \neq q'$ denote the cross-covariance error between cluster q and cluster q'.

We have evaluated this cluster separation method using Saleh-Valenzuela channel realizations. To be able to separate clusters without the angular dimension, we need to set the channel bandwidth high enough ($B = 200MHz$) so that clusters are separable. In practice clusters that overlap in the delay domain could be separated by adding the angular dimension to the model. Such generalization is discussed in Section 4.4.5 at the end of this Chapter. On Figures 4.5 and 4.6 the SNR is set to 30dB (high SNR regime). As the tail of the first cluster overlaps with the head of the second one, the second cluster is slightly noisier than the first one. As a result, the overlapping part of each cluster in the frequency domain features a higher error as depicted on Figure 4.6. On Figures 4.7 and 4.8, a single cluster is featured in a low SNR (10dB) scenario. The error from the cluster separation process is due to the strong thermal noise polluting the received samples.

The two examples represent respectively situations where the separation process works great and a situation where it performs poorly. The process can yield poor results either because clusters are impinging on each other (Clusters 1 and 2 in Figure 4.7) or because of thermal noise (Cluster 3 in Figure 4.7). Note that the error bounds on the estimate given by the covariance matrix $\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}}$ of (4.22) provides a valuable hint on the reliability

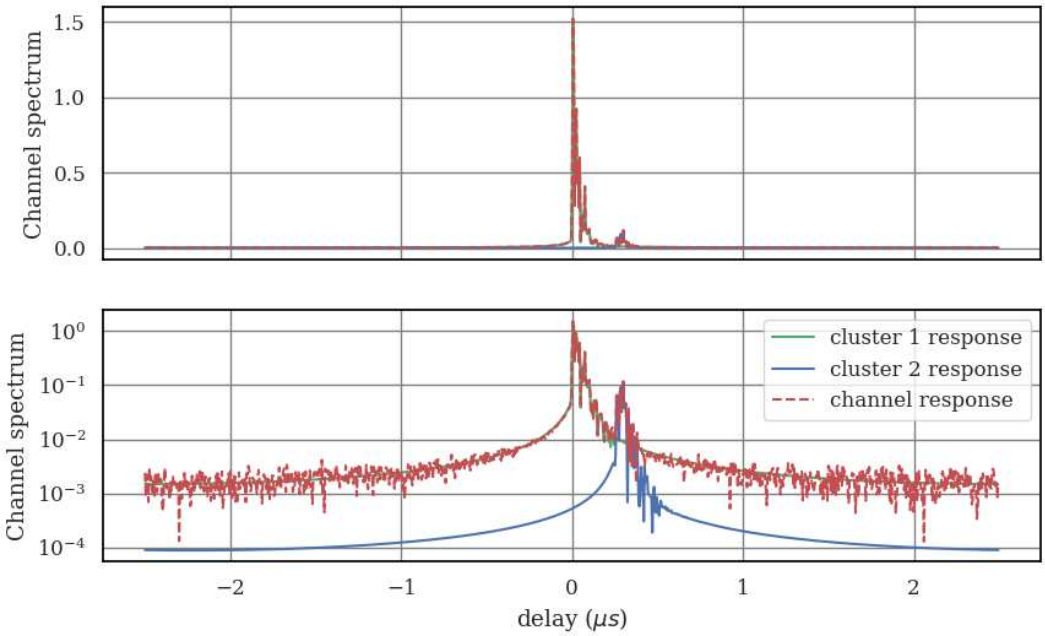

Figure 4.4 – Delay-domain channel realization with 2 main clusters, linear and logarithmic scale representation. $B = 200$MHz.

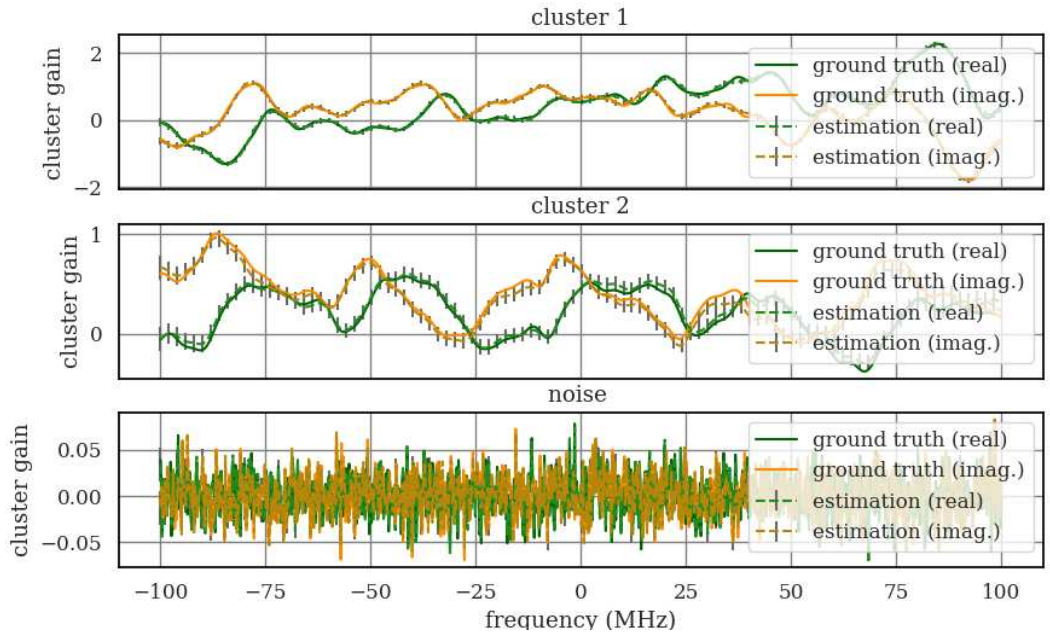

Figure 4.5 – Frequency domain cluster separation results with known cluster parameters. The uncertainty on the estimated channel is measured by the standard deviation (4.22) and denoted by error bars. $SNR = 30dB$.

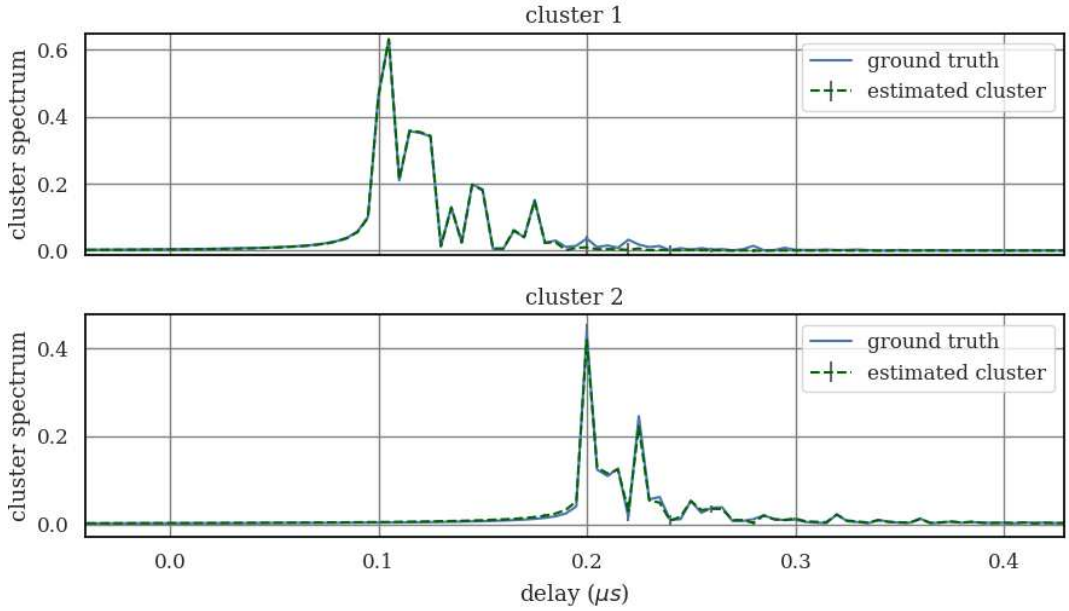

Figure 4.6 – Delay domain cluster separation results with known cluster parameters. The estimator variance (4.22) translated to the delay domain is denoted by error bars. $SNR = 30dB$.

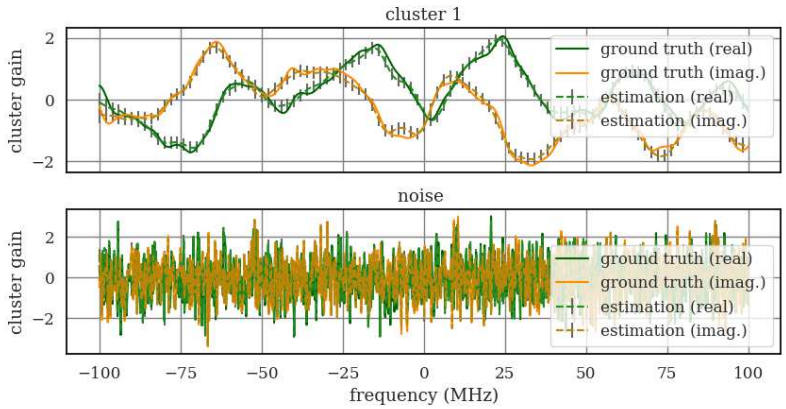

Figure 4.7 – Frequency domain cluster separation results with known cluster parameters. The uncertainty on the estimated channel is measured by the standard deviation (4.22) and denoted by error bars. $SNR = 10dB$.

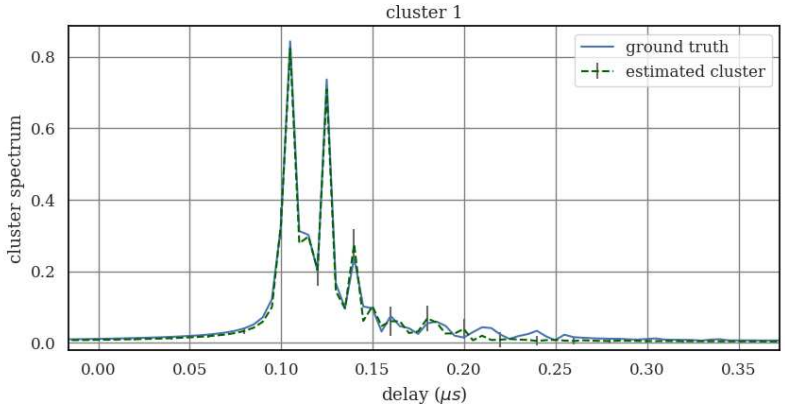

Figure 4.8 – Delay domain cluster separation results with known cluster parameters. The estimator standard deviation (4.22) translated to the delay domain is denoted by error bars. $SNR = 10dB$.

of the resulting estimated clusters. This covariance matrix can be translated to the delay domain using a Fourier change of base as

$$\mathbf{\Sigma}^{\mathbf{F}}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}} = \mathbf{F}^H \mathbf{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta}} \mathbf{F} \tag{4.25}$$

where $\mathbf{F}$ is a Fourier transformation matrix. Using this change of basis we display the delay-domain estimated clusters on Figures 4.6 and 4.8 alongside with the variance on the estimates. On Figure 4.8 the variance is particularly high where the two first clusters overlap each others.

4.4 Separation of clusters with unknown characteristics

The full cluster separation problem is actually made of two main sub-problems. The separation itself requires the knowledge of cluster parameters, which themselves can be perfectly estimated only if the clusters have been separated beforehand. This joint estimation-separation problem can be solved efficiently by using an *Expectation-Maximization* algorithm [16].

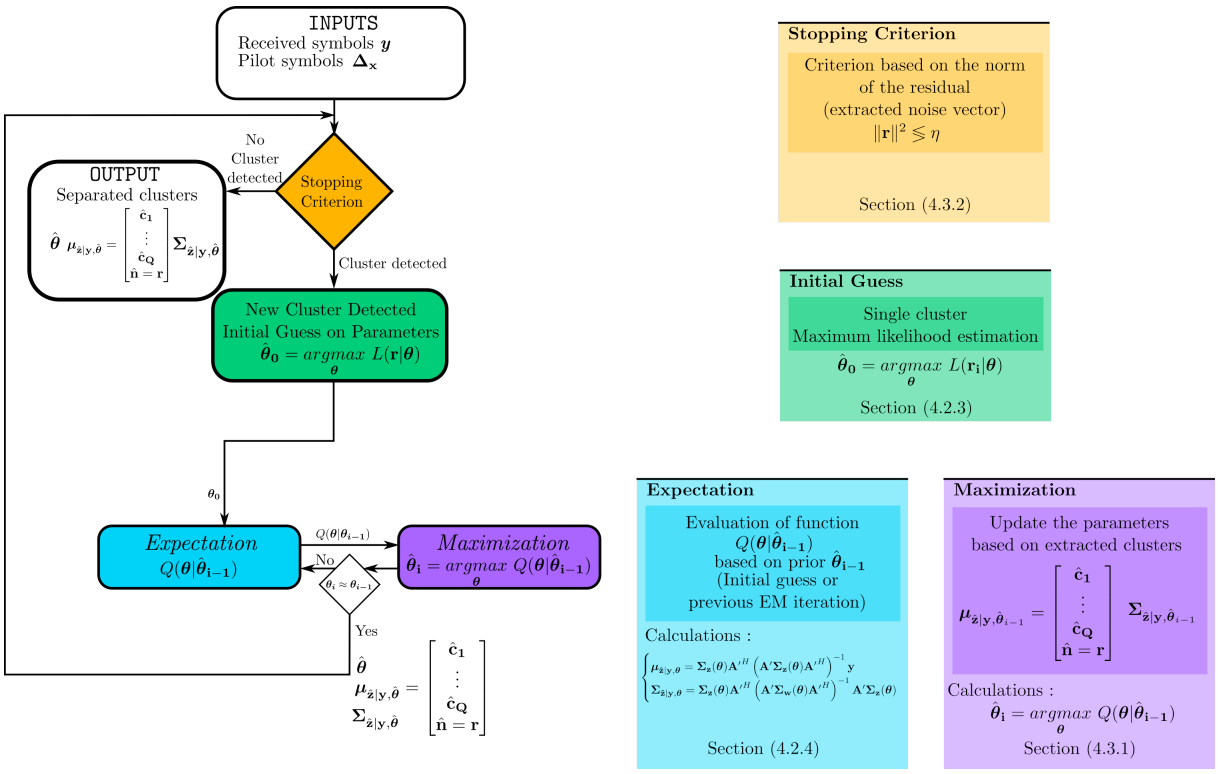

Figure 4.9 – Flowchart of the cluster separation and estimation procedure.

4.4.1 Derivation of the Expectation-Maximization algorithm

In this section we walk through all the steps required to correctly setup the Expectation-Maximization algorithm. Assuming that we would like to estimate all cluster parameters $\boldsymbol{\theta}$ jointly based on the received samples $\mathbf{y}$, we would have to maximize the log-likelihood

$$L(\mathbf{y}|\boldsymbol{\theta}) = \ln\left(p(\mathbf{y}|\boldsymbol{\theta})\right). \tag{4.26}$$

Optimizing this log-likelihood is very complex as it requires a costly $3Q$ parameters joint optimization. Moreover Q is not known beforehand. It can be rewritten using rules on conditional probabilities as

$$L(\mathbf{y}|\boldsymbol{\theta}) = \ln\left(p(\mathbf{y},\mathbf{z}|\boldsymbol{\theta})\right) - \ln\left(p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta})\right) \tag{4.27}$$

where $\mathbf{z}$ has been defined in Section 4.3.4 as the concatenation of the cluster gains $\mathbf{c}$ and the noise vector $\mathbf{n}$. Note that $\mathbf{y}$ is entirely defined by $\mathbf{z}$ from $\mathbf{y} = \mathbf{A}'\mathbf{z}$, thus the former term $\ln\left(p(\mathbf{y},\mathbf{z}|\boldsymbol{\theta})\right)$ can be rewritten as $\ln\left(p(\mathbf{z}|\boldsymbol{\theta})\right)$.

Note that $L(\mathbf{y}|\boldsymbol{\theta})$ does not depend on $\mathbf{z}$. Then the expectation taken over $\mathbf{z}$ of the log-likelihood assuming that the received samples $\mathbf{y}$ and an approximation of the cluster parameters $\boldsymbol{\theta_c}$ are known verifies $L(\mathbf{y}|\boldsymbol{\theta}) = \mathbb{E}\left\{L(\mathbf{y}|\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta_c}\right\}$ where $\boldsymbol{\theta_c}$ is an arbitrary set of cluster parameters. By applying this expectation to (4.27) we obtain

$$L(\mathbf{y}|\boldsymbol{\theta}) = \mathbb{E}\left\{\ln\left(p(\mathbf{z}|\boldsymbol{\theta})\right)|\mathbf{y}, \boldsymbol{\theta_c}\right\} - \mathbb{E}\left\{\ln\left(p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})\right)|\mathbf{y}, \boldsymbol{\theta_c}\right\} \tag{4.28}$$

Those two terms are commonly denoted in EM literature as

$$\begin{cases} Q(\boldsymbol{\theta}|\boldsymbol{\theta_c}) = \mathbb{E}\left\{\ln\left(p(\mathbf{z}|\boldsymbol{\theta})\right)|\mathbf{y}, \boldsymbol{\theta_c}\right\} \\ H(\boldsymbol{\theta}|\boldsymbol{\theta_c}) = \mathbb{E}\left\{\ln\left(p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})\right)|\mathbf{y}, \boldsymbol{\theta_c}\right\}. \end{cases} \tag{4.29}$$

Theorem. *(Reminder) For any set of parameters $\boldsymbol{\theta_c}$, function H verifies*

$$H(\boldsymbol{\theta}|\boldsymbol{\theta_c}) \leq H(\boldsymbol{\theta_c}|\boldsymbol{\theta_c}). \tag{4.30}$$

This comes from

$$H(\boldsymbol{\theta}|\boldsymbol{\theta_c}) - H(\boldsymbol{\theta_c}|\boldsymbol{\theta_c}) = \mathbb{E}\left\{\ln\left(\frac{p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta_c})}\right)|\mathbf{y}, \boldsymbol{\theta_c}\right\} \leq \ln\left(\mathbb{E}\left\{\frac{p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta_c})}|\mathbf{y}, \boldsymbol{\theta_c}\right\}\right) = 0 \tag{4.31}$$

where the first equality comes from the linearity of the expectation and the rightmost inequality from Jensen's inequality on the concave function ln.

From this result, we obtain the fundamental theorem that enables the estimation of cluster parameters.

Theorem. *The sequence $L(\mathbf{y}|\boldsymbol{\theta}_i)$ converges to a local maxima of function $L(\mathbf{y}|\boldsymbol{\theta})$ when $i \to \infty$ where the sequence $\boldsymbol{\theta}_i$ is defined as*

$$\boldsymbol{\theta}_i = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{i-1}) \tag{4.32}$$

Indeed $Q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i-1}) \geq Q(\boldsymbol{\theta}_{i-1}|\boldsymbol{\theta}_{i-1})$ (by definition of $\boldsymbol{\theta}_i$) and $H(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i-1}) \leq H(\boldsymbol{\theta}_{i-1}|\boldsymbol{\theta}_{i-1})$ (from Equation (4.30)). Thus the log-likelihood improves at each iteration $L(\mathbf{y}|\boldsymbol{\theta}_i) \geq L(\mathbf{y}|\boldsymbol{\theta}_{i-1})$.

At the core of the algorithm resides the function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_i)$ which has to be first evaluated for the set of parameters $\boldsymbol{\theta}_i$ obtained from prior knowledge (initial guess or previous

iteration). The function can be detailed as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta_c}) &= -\ln((2\pi)^{N_f}) - \ln(|\boldsymbol{\Sigma_z}(\boldsymbol{\theta})|) - \mathbb{E}\left\{\mathbf{z}^H {\boldsymbol{\Sigma_z}}^{-1}(\boldsymbol{\theta})\mathbf{z}|\mathbf{y}, \boldsymbol{\theta_c}\right\} \\ &= -\ln((2\pi)^{N_f}) - \ln(|\boldsymbol{\Sigma_z}(\boldsymbol{\theta})|) - \mathrm{Tr}({\boldsymbol{\Sigma_z}}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}) - \boldsymbol{\mu}^H_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}} {\boldsymbol{\Sigma_z}}^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}} \end{aligned} \tag{4.33}$$

where $\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}$ and $\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}$ are the mean and covariance matrix of the Gaussian vector $\mathbf{z}$ given the parameters $\boldsymbol{\theta_c}$ and the received symbols $\mathbf{y}$. Those first and second statistics have already been evaluated and are given by Equation (4.22). Equation (4.33) can be split into a sum of clusters-related components as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta_c}) = &-\ln((2\pi)^{N_f}) \\ &-\sum_{q=1}^{Q} Q_q(\boldsymbol{\theta_q}|\boldsymbol{\theta_c}) \\ &-\ln((\sigma_n^2)^{N_f}) - \frac{1}{\sigma_n^2}\mathrm{Tr}([\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]_{Q+1,Q+1}) - \frac{1}{\sigma_n^2}[\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]^H_{Q+1}[\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]_{Q+1} \end{aligned} \tag{4.34}$$

where the quantities $Q_q(\boldsymbol{\theta_q}|\boldsymbol{\theta_c})$ are given by

$$Q_q(\boldsymbol{\theta_q}|\boldsymbol{\theta_c}) = \ln(|\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})|) + \mathrm{Tr}({\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}}^{-1}(\boldsymbol{\theta_q})[\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]_{q,q}) + [\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]^H_q {\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}}^{-1}(\boldsymbol{\theta_q})[\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]_q$$

This last equation is of particular interest because it enables a sequential estimation of cluster parameters.

The final algorithm, illustrated on Figure 4.9, can be decomposed into three main steps :

- *Initial guess :* this step provides a coarse estimation of the parameters $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ of a particular cluster. This estimate is used to initialize the EM algorithm.
- *Expectation :* we evaluate the function $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{\mathbf{i-1}})$, based on prior knowledge on the cluster parameters $\hat{\boldsymbol{\theta}}_{\mathbf{i-1}}$ obtained either by an initial guess or the previous iteration. This step consists in the calculation of $\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\hat{\boldsymbol{\theta}}_{\mathbf{i-1}}}$ and $\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\hat{\boldsymbol{\theta}}_{\mathbf{i-1}}}$.
- *Maximization :* we find the optimal parameters $\boldsymbol{\theta}_{\mathbf{q}}$ for each cluster sequentially that maximize the function $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{i-1})$. We obtain the refined cluster parameters $\hat{\boldsymbol{\theta}}_i$.

From the received samples, the parameters of the most powerful cluster are extracted in the *initial guess* step. Using this coarse knowledge, the cluster is extracted from the signal and separated from thermal noise and other clusters in the *expectation* step. The

first guess on the parameters is then refined in the *maximization* step. Those two steps are repeated iteratively until the estimated parameters stop improving, forming an EM algorithm. We repeat this process on the residual of the *expectation* step, that contains noise and other clusters.

4.4.2 Maximization Step

In the maximization step, the functions

$$Q_q(\boldsymbol{\theta_q}|\boldsymbol{\theta_c}) = \ln(|\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})|) + \mathrm{Tr}({\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}}^{-1}(\boldsymbol{\theta_q})[\boldsymbol{\Sigma}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]_{q,q}) + [\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]_q^H {\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}}^{-1}(\boldsymbol{\theta_q})[\boldsymbol{\mu}_{\hat{\mathbf{z}}|\mathbf{y},\boldsymbol{\theta_c}}]_q$$

are optimized sequentially. However this require the inversion of matrix ${\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}}^{-1}(\boldsymbol{\theta_q})$ which is ill-conditioned as most eigenvalues are non-significant (see Chapter 3). This requires some kind of regularization to be performed. We propose two kinds of regularizations :

— Replace matrix $\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})$ by the regularized matrix $\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q}) + \mu\mathbf{I}$ where μ is the regularization factor. This regularization works for delay τ_q and γ_q parameters estimations but yield heavily biased power estimates $\hat{P}_q$.

— Remove the null space of $\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})$ by projecting the matrix into its image space $\mathbf{P}\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})\mathbf{P}^H$ where $\mathbf{P}$ is the set of all eigenvectors of $\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})$ associated to eigenvalues above threshold μ (the regularization factor). This regularization works great for power estimation $\hat{P}_q$.

4.4.3 Stopping criterion

This algorithm requires a stopping criterion, otherwise at some point it will start extracting clusters from thermal noise. The stopping criterion can be either:

— A predefined number of iterations L. This requires to know the number of clusters Q lying within the channel to set $L = Q$.

— A criterion based on the norm of the residual. This is the solution proposed here. We will see in Chapter 6 that this approach is sub-optimal compared to other advanced detection techniques, for the specific case of specular rays.

In this section we study the second stopping criterion technique. If the residual $\mathbf{r}$ only contains noise, then its squared norm follows a χ^2 distribution.

Theorem. *Assuming that* $\mathbf{r} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$*, the squared norm* $\|\mathbf{r}\|^2$ *follows the* χ^2 *distri-*

bution

$$\|\mathbf{r}\|^2 \sim \chi^2(2N_f, \sigma_n^2/2)$$

parametrized by $2N_f$ *degrees of freedom and* $\sigma_n^2/2$ *scale parameter. The factor* 2 *comes from the complex values that contains both a real and an imaginary part, each weighted by a variance* $\sigma_n^2/2$.

We wish to define a threshold such as $\|\mathbf{r}\|^2 \lessgtr \eta$ to test the presence of a cluster within the residual. If the norm falls below a certain threshold it is considered as containing only noise. Otherwise it contains a cluster. The threshold of the detector is calibrated based on a *false alarm* probability, which represent the eventuality of $\|\mathbf{r}\|^2$ exceeding the threshold η even if it contains only noise. The false alarm probability for the threshold η is given by

$$P_{FA} = 1 - F_{\chi^2(2N_f, \sigma_n^2/2)}(\eta) \tag{4.35}$$

where $F_{\chi^2(2N_f, \sigma_n^2/2)}$ denotes the cumulative probability function of the chi-square distribution $\chi^2(2N_f, \sigma_n^2/2)$. The threshold, calibrated using the desired false alarm probability P_{FA}, is given by

$$\eta = F^{-1}_{\chi^2(2N_f, \sigma_n^2/2)}(1 - P_{FA}) \tag{4.36}$$

where $F^{-1}_{\chi^2(2N_f, \sigma_n^2/2)}$ denotes the reciprocal of the cumulative probability function. It is also called the percentile function. This threshold is used to stop the iterative algorithm before it starts extracting clusters from thermal noise.

4.4.4 Illustrating example

An illustrating example of the cluster separation algorithm is presented in this section. It is applied to the two-clusters Saleh-Valenzuela channel realization presented on Figure 4.10.

In order to make the algorithm work, a former knowledge of the γ parameter is assumed in this section. This is a current limitation of the algorithm, required to prevent the algorithm first iteration from considering the whole channel as a single cluster parametrized by a very large γ constant. This constraint could be relaxed by introducing a proper regularization to the problem.

As the channel contains two clusters, only two iterations are required by the algorithm to fully extract all the multipath components. During the first pass, the algorithm tries to extract the most powerful cluster, as demonstrated on Figure 4.11. As the two clusters that

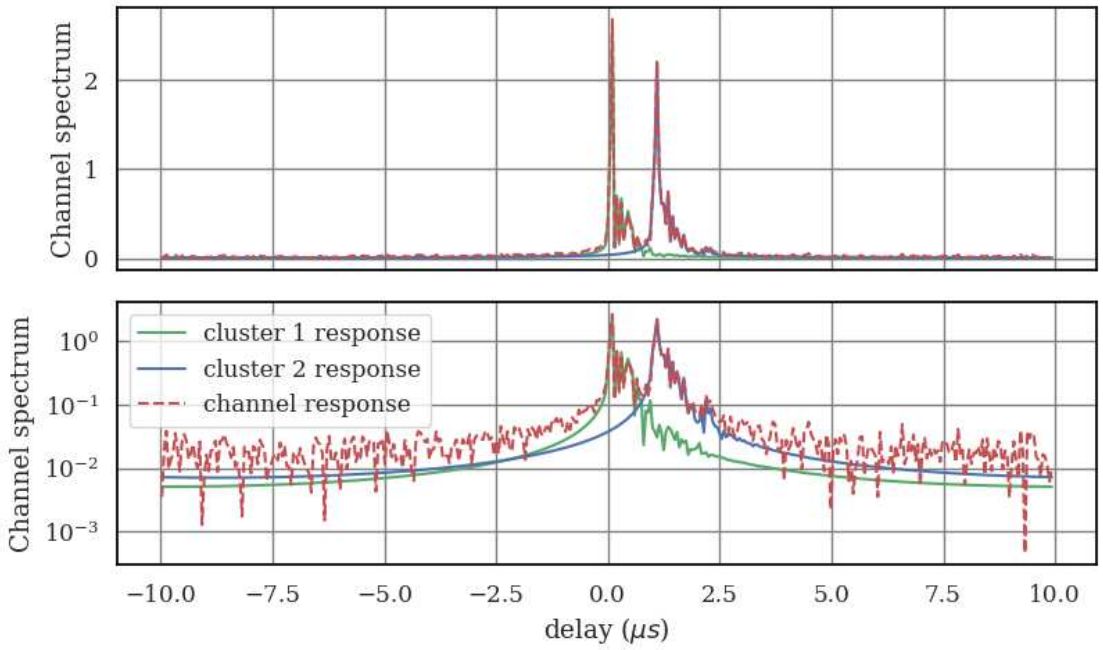

Figure 4.10 – The two-clusters Saleh-Valenzuela channel realization used to highlight the cluster separation technique.

constitute the channel are a bit overlapping each other, the estimated cluster contains all of the most powerful cluster as well as part of the second cluster. Its remains are embedded in the rejected thermal noise within the residue.

A second pass of the algorithm on the residual fully extracts the second cluster, while correcting the first estimated cluster, as depicted on Figure 4.12. There is only some small remaining aliases that comes from the EM algorithm not converging quickly enough. The frequency domain responses of both clusters are presented on Figure 4.13.

4.4.5 Generalization to the angular domain

We have presented a novel approach for the estimation and separation of clusters constituting a propagation channel. As it is, the process is limited to the frequency domain. It can be extended to the angular domain by modifying the channel covariance matrix and adding the corresponding parameters. There are however some challenges to overcome for a practical implementation

— The frequency domain covariance matrix $\mathbf{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{f}}$ admit an analytical expression that facilitates the optimization process in the *initial guess* and *maximization* steps.

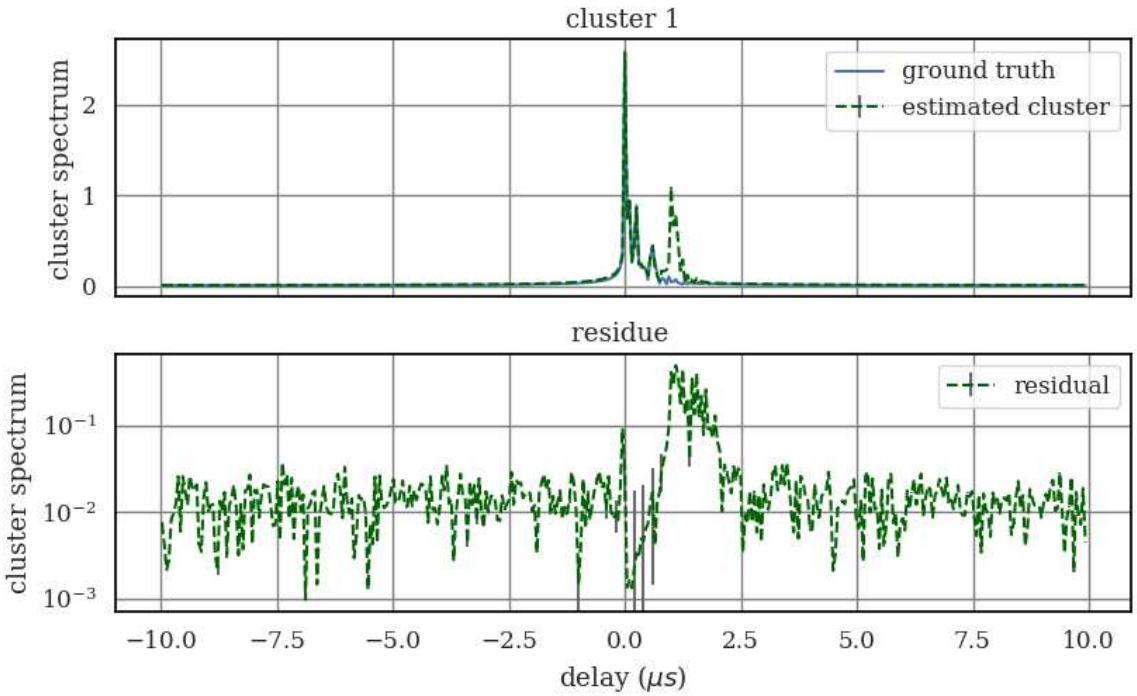

Figure 4.11 – First pass of the clustering algorithm. The most powerful cluster is extracted from the channel samples. It is mixed with a small part of the second cluster. The residual contains thermal noise plus the remaining non-extracted clusters.

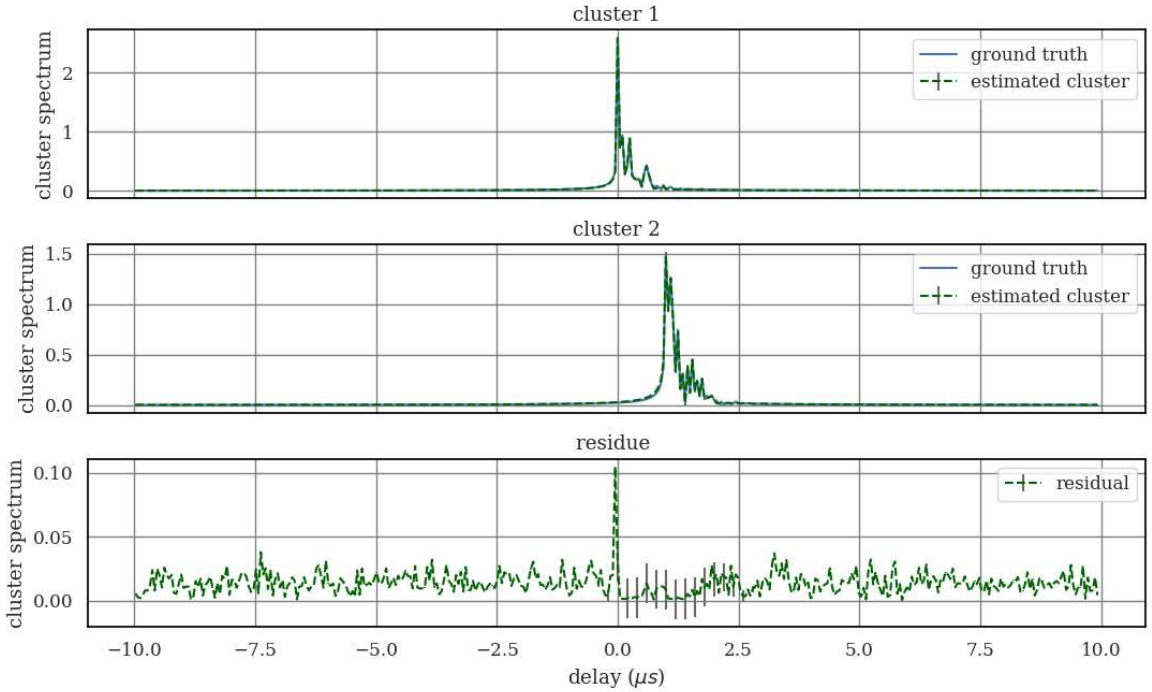

Figure 4.12 – Second pass of the clustering algorithm on the residual. This time both clusters are extracted.

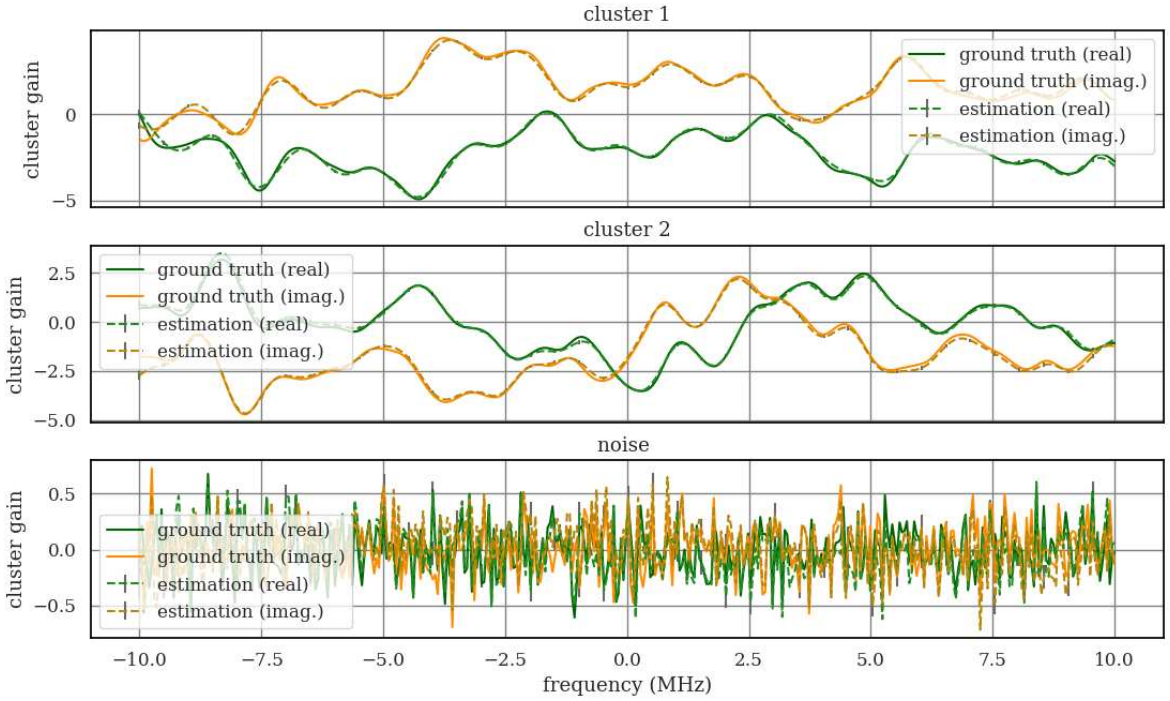

Figure 4.13 – Second pass of the clustering algorithm on the residual. This time both clusters are extracted and displayed in the frequency domain.

In contrast, the angular domain covariance matrices $\mathbf{\Sigma}_q^t$ and $\mathbf{\Sigma}_q^r$ does not admit an analytical expression. A simplified model should be designed and used in the optimization process.

— This extension involves optimizations over a larger number of variables. This would require to improve the optimization process, which has been overlooked in this chapter.

4.5 Conclusion

In this chapter we propose a novel cluster separation technique, designed for the frequency domain. This algorithm could benefit from a pre-processing sparse channel projection step (typically a Fourier Transform) that would reduce the complexity, though such step is not necessary. In all previous state of the art techniques, this preprocessing step is always required. We propose a proof of concept that could we improved significantly by

— using solvers optimized for the specific optimization problems encountered

— adjusting the regularizations required when dealing with low rank matrix inverses

— adding the angular domain dimension required to separate clusters when not considering Ultra Wide Band (UWB) channels

— designing a more efficient detector that can find clusters below the noise floor

— including prior information on cluster parameters to avoid falling into local maxima of the likelihood function (cluster with null power or cluster with large γ_q that encompass the whole channel for instance)

We deal with some of those optimizations in Chapter 6 for the specific case of specular ray detection and estimation.

The results featured in this chapter contribute to the understanding of the cluster separation problem. The Cramér-Rao bound calculations indicate the circumstances when parameter estimation is feasible and provides the ultimate performance of any unbiased estimator. Also, contrary to most state of the art cluster separation techniques, our approach provides uncertainty bounds on the resulting estimated clusters.

BIBLIOGRAPHY

[1] Adel AM Saleh and Reinaldo Valenzuela, « A statistical model for indoor multipath propagation », *in*: *IEEE Journal on selected areas in communications* (1987).

[2] R. Schmidt, « Multiple emitter location and signal parameter estimation », *in*: *IEEE Trans. Antennas Propagat.* (Mar. 1986).

[3] R. Roy and T. Kailath, « ESPRIT-estimation of signal parameters via rotational invariance techniques », *in*: *IEEE Trans. Acoust., Speech, Signal Processing* (July 1989).

[4] B.H. Fleury, P. Jourdan, and A. Stucki, « High-resolution channel parameter estimation for MIMO applications using the SAGE algorithm », *in*: *2002 International Zurich Seminar on Broadband Communications Access - Transmission - Networking (Cat. No.02TH8599)*, Zurich, Switzerland: IEEE, 2002.

[5] Michael Corrigan et al., « Automatic UWB clusters identification », *in*: *2009 IEEE Radio and Wireless Symposium*, San Diego, CA, USA: IEEE, Jan. 2009.

[6] R.J.-M. Cramer, R.A. Scholtz, and M.Z. Win, « Evaluation of an ultra-wide-band propagation channel », *in*: *IEEE Trans. Antennas Propagat.* (May 2002).

[7] Johan Karedal et al., « A Measurement-Based Statistical Model for Industrial Ultra-Wideband Channels », *in*: *IEEE Trans. Wireless Commun.* (Aug. 2007).

[8] N. Czink et al., « Cluster Angular Spreads in a MIMO Indoor Propagation Environment », *in*: *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, Berlin, Germany: IEEE, 2005.

[9] Nicolai Czink et al., « A Framework for Automatic Clustering of Parametric MIMO Channel Data Including Path Powers », *in*: *IEEE Vehicular Technology Conference*, Hyatt Regency Montreal, Montreal, QC, Canada: IEEE, Sept. 2006.

[10] Carl Gustafson et al., « On mm-Wave Multipath Clustering and Channel Modeling », *in*: *IEEE Transactions on Antennas and Propagation* (Mar. 2014).

[11] Ruisi He et al., « On the Clustering of Radio Channel Impulse Responses Using Sparsity-Based Methods », *in*: *IEEE Transactions on Antennas and Propagation* (June 2016).

[12] *3GPP TS 38.211: "NR; Physical channels and modulation"*, 2020.

[13] Steven M. Kay, *Fundamentals of statistical signal processing*, Prentice Hall signal processing series, Englewood Cliffs, N.J: Prentice-Hall PTR, 1993.

[14] Luc Le Magoarou and Stéphane Paquelet, « Performance of MIMO channel estimation with a physical model », *in*: *arXiv:1902.07031 [cs, eess]* (Feb. 2019).

[15] Jorge Nocedal and Stephen J. Wright, *Numerical optimization*, 2nd ed, Springer series in operations research, New York: Springer, 2006.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin Reviewed work(s): « Maximum Likelihood from Incomplete Data via the EM Algorithm », *in*: *Journal of the Royal Statistical Society. Series B (Methodological)* (1977).

Part II

Model Exploitation for Massive MIMO Digital Processing

Chapter 5

Using uplink Channel Information for downlink precoding in FDD

Acquiring the Channel State Information (CSI) at the transmitter to calculate the precoder that maximizes the capacity (see Chapter 2) is a major challenge in MIMO systems. The CSI, measured at the receiver side, has to be fed back to the transmitter. This is particularly challenging in Frequency Division Duplex (FDD) as uplink and downlink channels are different, since they stand on adjacent frequency bands.

In Section 5.1 we describe the massive MIMO CSI acquisition problem. The solutions to provide the CSI at the transmitter vary heavily depending on the chosen duplex scheme (TDD or FDD). In this chapter we focus on the more challenging FDD case. To this end, we analyzed the performance of a downlink CSI estimation method based on a linear extrapolation scheme of the measured uplink channel. Our analysis leverages the results obtained in Chapters 3 (second order statistics of the channel) and 4 (cluster separation) to quantify the achievable extrapolation range. The error induced by the process has been measured both in terms of MSE and Reduction of Beamforming Gain (RBG), the latter having a direct influence on the system capacity.

5.1 Context

As a reminder, we showed in Chapter 2 that the Spectral Efficiency of a MIMO system is given by the equation

$$SE_{MIMO} = \log_2 \left(\det \left(\mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{H}^H(f) \mathbf{Q} \mathbf{H}(f) \right) \right) \tag{5.1}$$

where the precoding matrix $\mathbf{Q}$ is actually designed at the transmitter, based on the CSI, to maximize the SE. The actual process used to derive the precoder based on the channel $\mathbf{H}(f)$ is given in Section 2.3.1.

In mobile broadband communications, the radio access network consists in a central base station that serves a set of User Equipments (UE) (cell phones, cars, computers, ...) located in its close vicinity (the area served by a base station is called the *cell*). The base station can rely on large antenna arrays while the UE only have a handful of antennas. The data traffic is also most of the time asymmetrical : the UE send short requests to the network which replies with the corresponding heavier online content. Thus the downlink channel is busier than its uplink counterpart and is most of the time optimized for high data rates while uplink is optimized for UE power consumption.

5.1.1 FDD and TDD

There are two main modes of operation in wireless communications : TDD and FDD. In TDD (described in Figure 5.1), the same frequency band is used for both uplink and downlink. The transmit and receive channels are the same. Thus the channel measured at the receiver can be reused on the transmit side to perform the precoding operation. This property is called *channel reciprocity* [1]. The remaining adjustments and calibrations required stem from the mismatch between hardware impairments of the transmitter and receiver radio frequency chains.

In FDD (described in Figure 5.2), uplink and downlink channels use different frequency bands. Thus the uplink channel $\mathbf{h_U}$ (measured on the receive side of the base station) can't be directly used to design the downlink precoder. In this case, researchers came up with two main sets of solutions.

The simplest solution is to feed back the CSI, determined at the receive side of the UE, to the base station, through dedicated control messages on the uplink channel. Today, this solution is implemented in the 4G LTE and 5G NR industry standards. The feedback consists of an integer called the Rank Indicator (RI) that defines the achievable transmission rank of the downlink channel and the Prefered Matrix Indicator (PMI), which is the set of indexes that identify a particular predefined precoding matrix.

However the feedback overhead increase with the number of transmit antennas, which render its use impractical in massive MIMO systems. *Channel reciprocity* is an inherent feature of TDD systems. To solve this fundamental problem, researchers recently investigated the potential approaches that would enable *partial channel reciprocity* to mitigate the feedback required to design the precoder. In particular, the angles of departure (DOD) and arrival (DOA) of the clusters that constitute the propagation channel depend on geometrical features of the environment that are frequency-independent. This property, called

angular reciprocity, has been tested and verified in channel measurement campaigns by both academia [2, 3] and industry [4]. The frequency-independent directional parameters can be estimated on the uplink and reused to precode on the downlink transmit side in a process called *channel extrapolation*. Eventually, it paves the way to smart precoding strategies in FDD mode that may not rely on a feedback channel. This is what is investigated in the following.

5.1.2 State of the Art on Channel Extrapolation

This strategy exploits some key frequency independent features of the propagation channel. First estimated on the uplink channel, those reciprocal characteristics can be used as a prior information to reduce the overhead. For instance the main directions of departure exhibit this reciprocity property. Those have been exploited in [5, 6, 7, 8] to obtain an approximate CSI used either to refine the feedback scheme or to perform simple angular precoding.

Recently, papers [9, 10] proposed to fully extrapolate the downlink channel from the uplink measurements. The principle consists in retrieving all the reciprocal parameters of a ray-based model (complex path gains, delays, directions of departure) using high resolution estimation techniques. The results were promising in the Line of Sight (LOS) scenario that features a strong specular ray well-defined in the space-frequency domain. However it performed badly in Non Line Of Sight (NLOS) as the algorithm struggles to separate intra-cluster rays out of the richer multipath environment.

Indeed the multipath components that make up the propagation channel feature closely spaced rays in the space-frequency domain. Due to the finite resolution of practical antenna arrays, those intra-cluster rays remain unresolvable. This issue could be mitigated by first separating the clusters constituting the channel (for instance using high resolution techniques [9, 10] or using the separation process proposed in Chapter 4) then using frequency domain extrapolation of the frequency-dependent cluster gains.

5.2 Limits of Channel Extrapolation

As the authors of [4] pointed out, only the angular characteristics of the channel are reciprocal. The full reciprocity from uplink to downlink does not hold. An interesting question is then: depending on the propagation characteristics, how far, in terms of

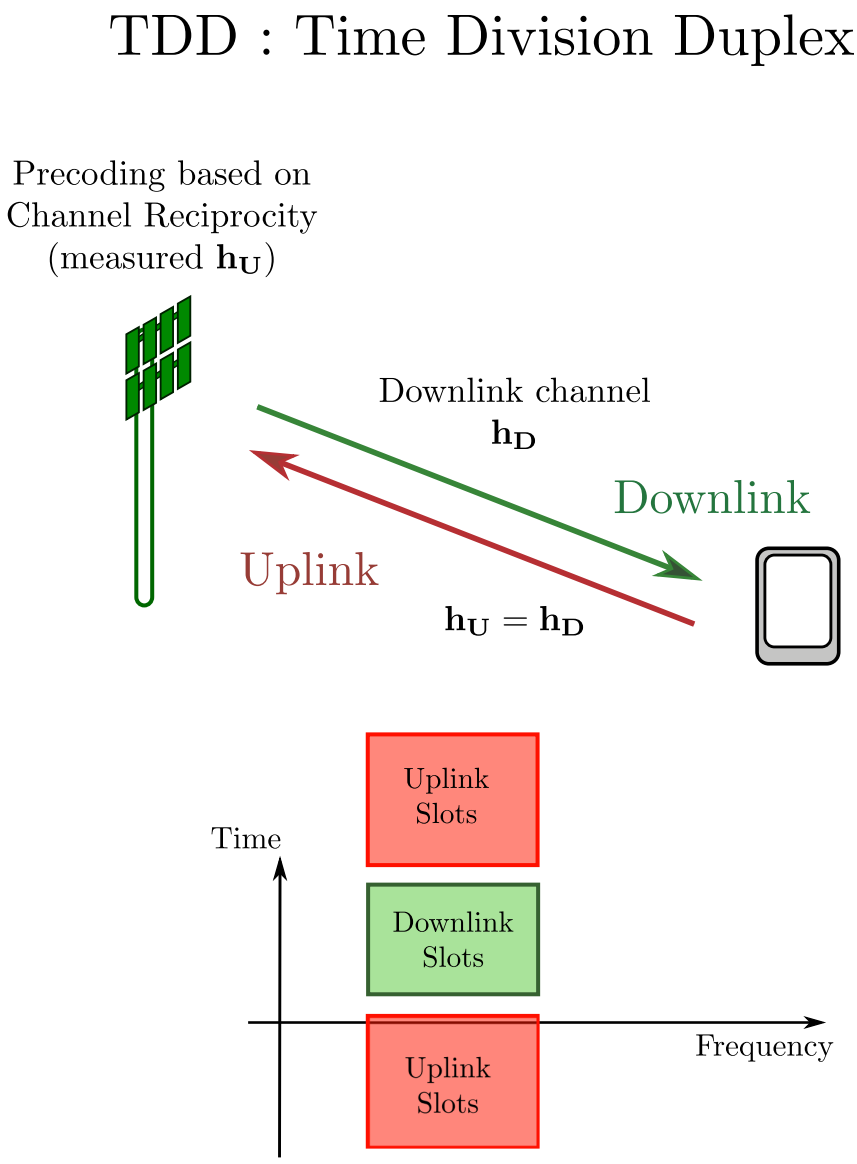

Figure 5.1 – Working principle of a MIMO system operating in TDD. The same frequency band is used for both uplink and downlink, which are separated in the time domain with dedicated time slots. The uplink channel can be reused for downlink precoding.

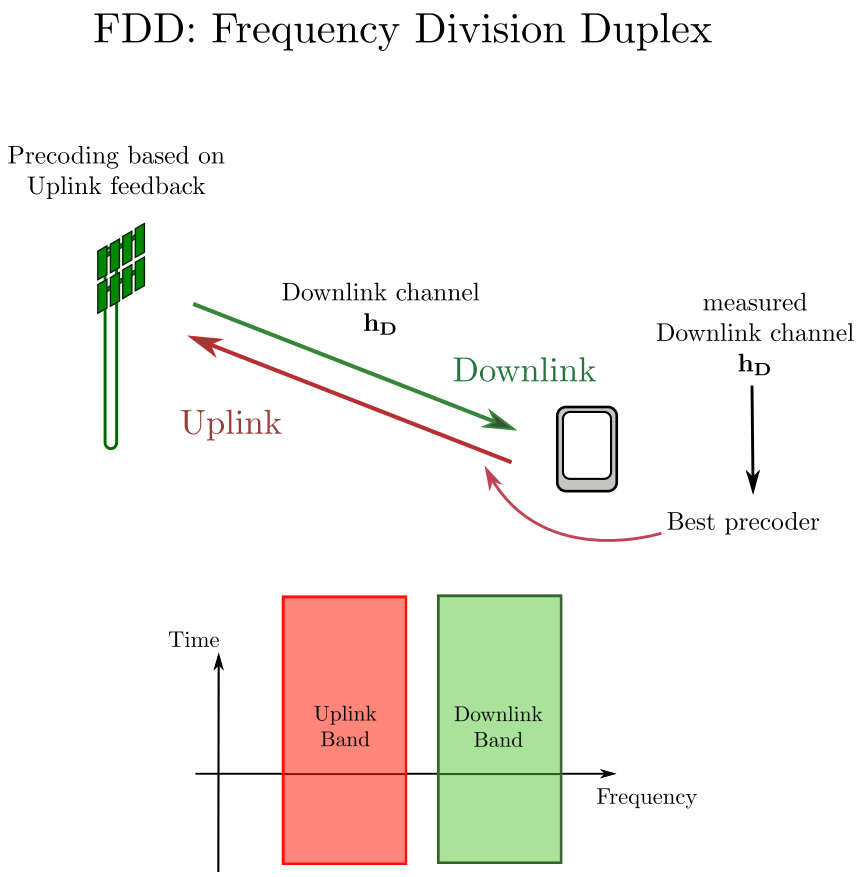

Figure 5.2 – Working principle of a MIMO system operating in FDD. uplink and downlink are on adjacent frequency bands. The base station can't rely directly on uplink measurements to design the precoder.

frequency spacing between uplink and downlink, can we extrapolate the *non-reciprocal* properties of the channel? In this section, we answer this question by studying how far a noisy Saleh-Valenzuela cluster can be extrapolated outside its measurement range. To this end, the clusters that constitute the propagation channel are modeled as complex Gaussian vectors $\mathbf{c_q}$ characterized by their covariance matrix as previously described in Chapters 3 and 4. The extrapolation range is quantified by how far the downlink channel $\mathbf{h_d}$ can be estimated from its uplink counterpart $\mathbf{h_u}$, as

$$\hat{\mathbf{h}}_\mathbf{d} = f(\mathbf{h_u}) \tag{5.2}$$

The corresponding extrapolation error is quantified on each subcarrier by the standard MSE (Mean Square Error) criterion $\varepsilon = \mathbb{E}\left\{|[\mathbf{h_d}]_k - [\hat{\mathbf{h}}_\mathbf{d}]_k|^2\right\}$ and by the RBG, which corresponds to the SNR loss assuming Maximum Ratio Transmission (MRT) precoding. The RBG is thus directly linked to the channel capacity loss.

5.2.1 Angular domain projection

To simplify the equations we study a standard clustered multipath Multi-Input-Single-Output (MISO) channel. Without loss of generality, the transmit side antenna array is set as an ULA. As in many aspects of our work, the extension to other arrays is achievable through the Kronecker expression model. Using again the vectorized Kronecker notations of Equation (1.15), the channel can be written as a sum of clusters

$$\mathbf{h} = \sum_{q=1}^{Q} \mathbf{c_q},$$

where the clusters $\mathbf{c_q}$ are given by

$$\mathbf{c_q} = \sqrt{N_f N_t} \sum_{p=1}^{P} \beta_{p,q} \mathbf{e_f}(\tau_{p,q} + \tau_q) \otimes \mathbf{e_t}^*(\vec{u}_{tx,p,q}).$$

It is assumed that the clusters that constitute the propagation have already been separated beforehand, using either a state of the art clustering algorithm, as proposed in Section 5.1.2, or the process described in Chapter 4. In either case, we now have access to cluster parameters, such as the main angle of departure $\theta_{tx,q}$ and the cluster delay τ_q. We also assume that the cluster shape parameters λ and γ are known perfectly. As intra-cluster rays are not resolvable, we use a Basis Expansion Model carefully chosen to

provide a sparse representation of the channel over the angular dimension.

We then characterize the first and second order moments of the basis coefficients on which the uplink to downlink extrapolation process relies on. We assume that the cluster delay τ_q is known. Without loss of generality we can set $\tau_q = 0$ to simplify the formulas.

The main direction of cluster θ_q is already known so we can use a shifted Fourier basis (already studied in Chapter 3) designed to minimize the number of significant coefficients. More details on this approach can be found in paper [7]. The shifted Fourier basis vectors are

$$\mathbf{e_l}(\theta_q)[m] = \frac{1}{\sqrt{N_t}} \, \mathrm{e}^{-2\pi j\left(m-\frac{N_t-1}{2}\right)\left(\frac{l}{N_t}+\frac{\|\vec{a}_{tx}\|}{\lambda}\cos(\theta_q)\right)}.$$

In this basis the cluster tensor can be rewritten as

$$\mathbf{c_q} = \sqrt{N_f N_t} \sum_{l=0}^{N_t-1} \sum_{p=1}^{P} \alpha_{p,l,q} \mathbf{e_f}(\tau_{p,q}) \otimes \mathbf{e_l}^*(\theta_q)$$

where the coefficients $\alpha_{p,l,q}$ are given by

$$\alpha_{p,l,q} = \beta_{p,q} \langle \mathbf{e_t}(\theta_q + \Delta\theta_{p,q}), \mathbf{e_l}(\theta_q) \rangle.$$

Cluster q is decomposed into N_t *sub-clusters* defined by fixed orthogonal steering vectors. The gain of each sub-cluster at subcarrier k, denoted $\mathbf{c_{q,l}}[k]$, can be recovered as

$$\mathbf{c_{q,l}} = \sqrt{N_f N_t} \sum_{p=1}^{P} \alpha_{p,l,q} \mathbf{e_f}(\tau_{p,q}).$$

Note that $\mathbb{E}\{\mathbf{c_{q,l}}\} = 0$ since the ray gains $\beta_{p,q}$ are centered. The cross-covariance matrix between two sub-clusters l and l' is given by

$$\mathbf{\Sigma_{q,l,l'}} = \mathbb{E}\left\{\mathbf{c_{q,l}} \mathbf{c_{q,l'}^H}\right\} = \left[\mathbf{\Sigma_q^t}(\theta_q)\right]_{l,l'} \mathbf{\Sigma_q^f}(\tau_q) \tag{5.3}$$

where $\mathbf{\Sigma_{q,f}}$ is the covariance matrix of the equivalent SISO cluster (evaluated in Chapter 3 as Equation (3.8)) and $\mathbf{\Sigma_{q,tx}}$ is defined as

$$[\mathbf{\Sigma_{q,tx}}]_{l,l'} = \mathbb{E}_{\Delta\theta}\left\{N_t \langle \mathbf{e_t}(\theta_q + \Delta\theta), \mathbf{e_l} \rangle \langle \mathbf{e_t}(\theta_q + \Delta\theta), \mathbf{e_{l'}} \rangle^*\right\}.$$

For a Saleh-Valenzuela cluster with parameters λ and γ the frequency covariance

matrix have already been evaluated in Chapter 3 as

$$[\mathbf{\Sigma_f}]_{k,k'} = P_q \left(1 + \frac{\lambda\gamma}{1 - 2\pi j \Delta B \gamma (k' - k)}\right).$$

As a reminder, the proof can be found in Annex B.1. Similarly, the coefficients of $\mathbf{\Sigma_{q,tx}}$ are given by

$$\begin{aligned}[\mathbf{\Sigma_{q,tx}}]_{l,l'} = \int_0^{2\pi} & f_{\Delta\theta}(\Delta\theta) \\ & D_{N_t}\left(x_l(\Delta\theta)\right) D_{N_t}\left(x_{l'}(\Delta\theta)\right) d\Delta\theta\end{aligned} \tag{5.4}$$

where $x_l(\Delta\theta) = \frac{l}{N_t} - \frac{\|\vec{a}_{tx}\|(\cos(\theta_q) - \cos(\theta_q + \Delta\theta))}{\lambda}$ and $D_N(x)$ denotes the Dirichlet kernel and $f_{\Delta\theta}(\Delta\theta)$ is the angular distribution of the intra-cluster rays (typically a Laplace distribution). The diagonal coefficients ($l = l'$) have already been derived in Equation (3.20).

5.2.2 Cluster Extrapolation

The off-diagonal elements of matrix $\mathbf{\Sigma_{q,tx}}$ are Dirichlet kernel side lobe products that are quickly negligible as N_t increase. Its diagonal elements have already been evaluated in Equation (3.20). We use the approximation that sub-clusters are uncorrelated and can be studied separately. Each sub-cluster is then fully characterized by its frequency covariance matrix $\mathbf{\Sigma_f}$ weighted by a fraction of the cluster power P_q and by $[\mathbf{\Sigma_{q,tx}}]_{l,l}$. Examples of cluster power partitioning (diagonal elements of $\mathbf{\Sigma_{q,tx}}$) over the sub-clusters are given in Fig 5.3. Because of the angular shift introduced in Section 5.2.1, the first bin always represents the main direction of the cluster. As expected, more subclusters are required to model wider angular-spreads ($\sigma_{\Delta\theta}$) and broadside clusters ($\theta = 90°$).

As previously mentioned, subclusters can be studied separately. For ease of notation we denote any sub-cluster and its covariance matrix as

$$\begin{cases}\mathbf{c} = \mathbf{c_{q,l}} \\ \mathbf{\Sigma} = \mathbf{\Sigma_{q,l,l}}.\end{cases}$$

The subcarriers can be separated into two sets: uplink $\mathcal{U}$ and downlink $\mathcal{D}$ subcarriers. We

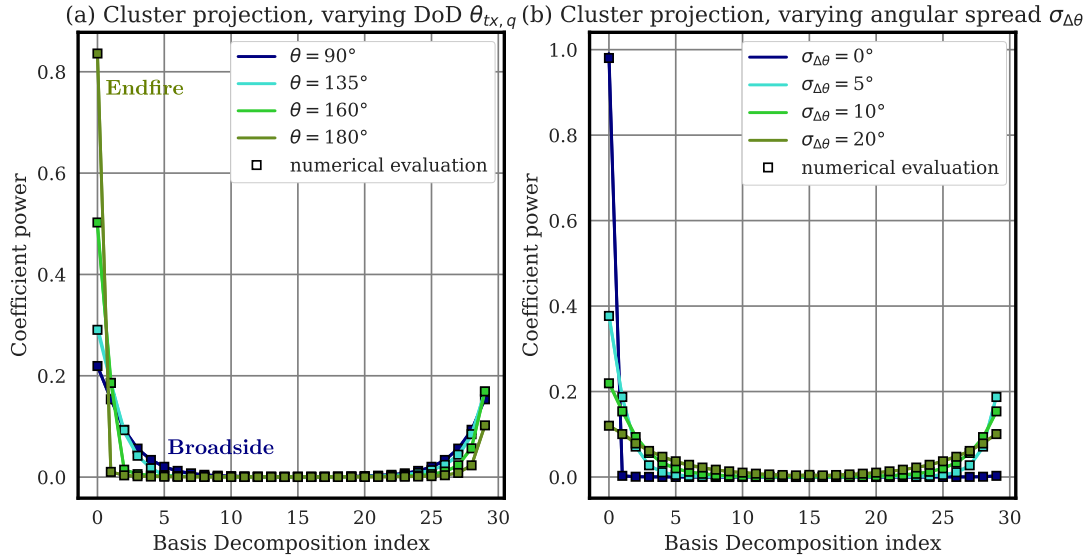

Figure 5.3 – Diagonal coefficients of the matrix $\mathbf{\Sigma_{q,tx}}$ for an ULA with $N_t = 30$, analytical expression (5.4) and numerical evaluation. On the left (a) the cluster angular spread is 10° and the cluster angle varies from 90° (broadside) to 180° (endfire). On the right (b) the angular spread of a broadside cluster varies from 0° (specular ray) to 20°.

separate the sub-cluster frequency gain vector into uplink and downlink vectors as

$$\begin{cases} \mathbf{c_u} = \mathbf{c}[k], \ k \in \mathcal{U} \\ \mathbf{c_d} = \mathbf{c}[k], \ k \in \mathcal{D}. \end{cases}$$

Our goal is to find the distribution of the downlink gains $\mathbf{c_d}$ given uplink gains measurements $\mathbf{c_u}$.

5.2.3 Sub-cluster gain modeling

We model a sub-cluster gain as a centered Wide Sense Stationary Gaussian random process with autocorrelation $\mathbf{\Sigma}$. Thermal noise and interference deteriorate the received uplink signal. Hence the measured sub-clusters are mixed with noise

$$\tilde{\mathbf{c}}_\mathbf{u} = \mathbf{c_u} + \mathbf{n}$$

with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. The covariance matrix of those noisy subclusters is denoted by

$$\tilde{\mathbf{\Sigma}}_\mathbf{u} = \mathbb{E}\left\{\tilde{\mathbf{c}}_\mathbf{u}\tilde{\mathbf{c}}_\mathbf{u}^H\right\} = \mathbf{\Sigma_u} + \sigma_n^2\mathbf{I}$$

where $\mathbf{\Sigma_u}$ is the noiseless uplink covariance matrix (a submatrix of $\mathbf{\Sigma}$) and σ_n^2 denotes the noise power. We also introduce the partial covariance vectors between the uplink channel and the downlink subcarrier k denoted by

$$\boldsymbol{\sigma}_{\mathbf{u}\cup k} = \boldsymbol{\sigma}_{k\cup\mathbf{u}}^H = \mathbb{E}\left\{\mathbf{c_u}[\mathbf{c_d}]_k^*\right\}.$$

We define the downlink subcarrier variance $\sigma_k^2 = \mathbb{E}\left\{[\mathbf{c_d}]_k[\mathbf{c_d}]_k^*\right\}$. The sub-cluster gain is WSS so this variance does actually not depend on k.

The autocorrelation of the uplink channel concatenated with gain k of the downlink is given by

$$\tilde{\mathbf{\Sigma}}_{\mathbf{u}\cup k} = \begin{bmatrix} \tilde{\mathbf{\Sigma}}_{\mathbf{u}} & \boldsymbol{\sigma}_{\mathbf{u}\cup k} \\ \boldsymbol{\sigma}_{k\cup\mathbf{u}} & \sigma_k^2 \end{bmatrix}.$$

Such model is used here after to find the posterior distribution of the downlink cluster coefficients $[\mathbf{c_d}]_k$.

5.2.4 Cluster gain extrapolation

The posterior distribution of $[\mathbf{c_d}]_k$ from the noisy sub-clusters measured on the uplink $\tilde{\mathbf{c}}_\mathbf{u}$ is given after classical manipulations by

$$\begin{aligned} f_{[\mathbf{c_d}]_k|\tilde{\mathbf{c}}_\mathbf{u}}([\mathbf{c_d}]_k|\tilde{\mathbf{c}}_\mathbf{u}) &= \frac{f_{[\mathbf{c_d}]_k,\tilde{\mathbf{c}}_\mathbf{u}}([\mathbf{c_d}]_k,\tilde{\mathbf{c}}_\mathbf{u})}{f_{\tilde{\mathbf{c}}_\mathbf{u}}(\tilde{\mathbf{c}}_\mathbf{u})} \\ &= \frac{|\tilde{\mathbf{\Sigma}}_\mathbf{u}|}{\pi|\tilde{\mathbf{\Sigma}}_{\mathbf{u}\cup k}|}\,\mathrm{e}^{-\frac{1}{\varepsilon_k^2}\left([\mathbf{c_d}]_k-\boldsymbol{\sigma}_{\mathbf{u}\cup k}^H\tilde{\mathbf{\Sigma}}_\mathbf{u}^{-1}\tilde{\mathbf{c}}_\mathbf{u}\right)^2}. \end{aligned} \tag{5.5}$$

$[\mathbf{c_d}]_k$ follows a complex Gaussian law, the mean of which is the Maximum a Posteriori (MAP) estimator

$$\hat{[\mathbf{c_d}]_k} = \boldsymbol{\sigma}_{\mathbf{u}\cup k}^H\left(\mathbf{\Sigma_u} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{c_u}$$

which is a linear estimator yielding the downlink channel coefficients at each subcarrier k from the knowledge of the uplink channel and the cluster statistics. The variance (MSE) of such estimator is given by

$$\varepsilon^2[k] = \sigma_k^2 - \boldsymbol{\sigma}_{\mathbf{u}\cup k}^H\left(\mathbf{\Sigma_u} + \sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\sigma}_{\mathbf{u}\cup k}. \tag{5.6}$$

The MSE is divided into two parts, σ_k^2 is the error without prior information and

$$\boldsymbol{\sigma}_{\mathbf{u}\cup k}^H \left(\boldsymbol{\Sigma}_{\mathbf{u}} + \sigma_n^2 \mathbf{I}\right)^{-1} \boldsymbol{\sigma}_{\mathbf{u}\cup k}$$

is the accuracy gain from uplink measurements.

5.2.5 Mean Square Error

In this section we provide analytical extrapolation MSE formulas for a single subcluster, verified by numerical evaluations. Those error formulas are subsequently extended to a full Saleh-Valenzuela channel. The single cluster formula is used to derive the channel MSE in a latter stage.

The Saleh-Valenzuela sub-cluster is parametrized by the ray arrival rate λ, the ray decay constant γ, the uplink bandwidth B_{ul}, the sub-cluster power P_q and the noise power σ_n^2. In this section we consider the normalized MSE derived from Equation (5.6) as

$$\frac{\mathbb{E}\left\{|[\hat{\mathbf{c}_{\mathbf{d}}}]_k - [\mathbf{c}_{\mathbf{d}}]_k|^2\right\}}{\mathbb{E}\left\{|[\mathbf{c}_{\mathbf{d}}]_k|^2\right\}} = 1 - \frac{\boldsymbol{\sigma}_{\mathbf{u}\cup k}^H \left(\boldsymbol{\Sigma}_{\mathbf{u}} + \sigma_n^2 \mathbf{I}\right)^{-1} \boldsymbol{\sigma}_{\mathbf{u}\cup k}}{\sigma_k^2}. \tag{5.7}$$

This expression only depends on λ, γ, B_{ul} and $\rho = \frac{P_q}{\sigma_n^2}$. We introduce the dimensionless parameters $\bar{\lambda} = \lambda / B_{ul}, \bar{\gamma} = \gamma B_{ul}$. Those notations are sufficient to fully describe the cluster.

We drew the mean square error obtained for two Saleh-Valenzuela environments with $B_{ul} = 9MHz$ and 200 pilots (matches a $10MHz$ LTE frame structure). The extrapolation process works well in moderately scattered environments, such as the original Saleh-Valenzuela model [11] (Fig. 5.4). However the extrapolation range is much shorter in rich scattering channels such as the one described in [12] (Fig. 5.5) by the Crabtree Building parameters. As expected this process works better for reduced intra-cluster scattering. The extrapolation potential of the channel also highly depends on the SNR. The process yields overall promising results at asymptotically high SNR but the performance decrease rapidly in presence of thermal noise.

Using the dimensionless parameters $\bar{\lambda}, \bar{\gamma}$ we can draw the 2D map of the achievable extrapolation range for any Saleh-Valenzuela cluster on Figure 5.6 at $\rho = 30dB$ for a 10% sub-cluster MSE. The extrapolation potential is given in logarithmic scale in percentages of the uplink bandwidth. We also computed the positions of characterized Saleh-Valenzuela channels of the literature [11, 12, 13] within this plane for $B_{ul} = 1MHz$

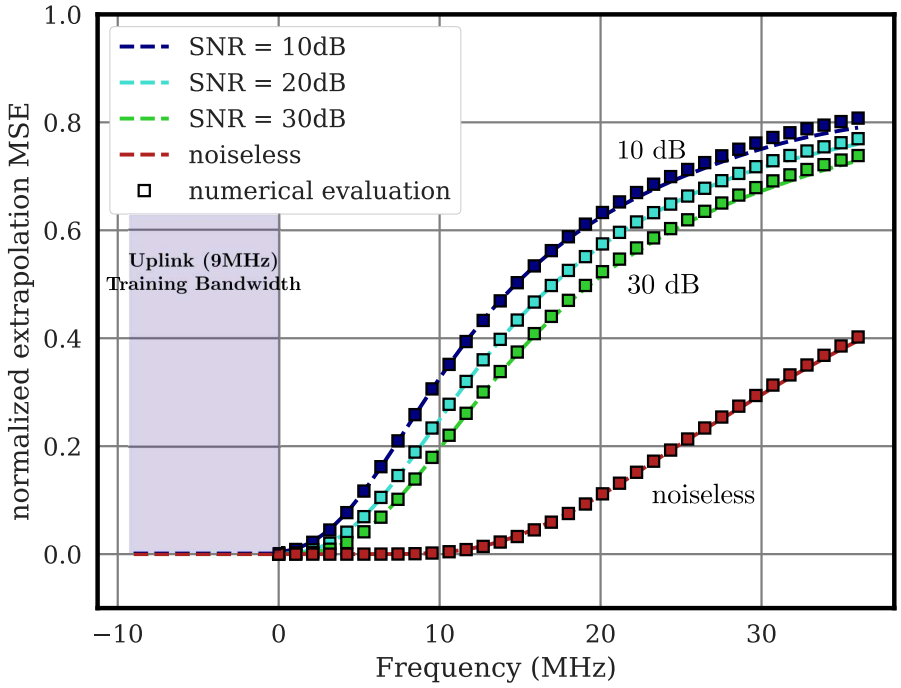

Figure 5.4 – Normalized MSE of the channel gain with $B_{UL} = 9MHz$ and $N_{UL} = 200$ uplink subcarriers, analytical (5.7) and numerical evaluation with original Saleh-Valenzuela parameters. The extrapolation scheme is very sensitive to noise.

(red points) and $B_{ul} = 10MHz$ (blue points). Channels with the same characteristics but different uplink bandwidths follow straight lines with -1 slope (in log scale). A $\times 10$ bandwidth increase means $\bar{\lambda}$ is a decade decreased and $\bar{\gamma}$ a decade increased. This abacus is then very useful to compare channels and interpret the effect of a bandwidth increase.

The extrapolation error is most of the time quantified by the Mean Square Error (MSE) criterion [9, 10]. We evaluate the extrapolation error at frequency k outside of the uplink training band. On subcarrier k, a cluster is given by the vector

$$\mathbf{c_{q,k}} = \sqrt{N_f N_t} \sum_{p=1}^{P} \beta_{p,q} \mathbf{e_f}(\tau_{p,q} + \tau_q)[k] \otimes \mathbf{e_t}^*(\vec{u}_{tx,p,q}). \tag{5.8}$$

Thus the channel on frequency k is given by

$$\mathbf{h_k} = \sum_{p=1}^{P} \mathbf{c_{q,k}}. \tag{5.9}$$

The linear extrapolation approach provides a simple expression for the MSE at sub-

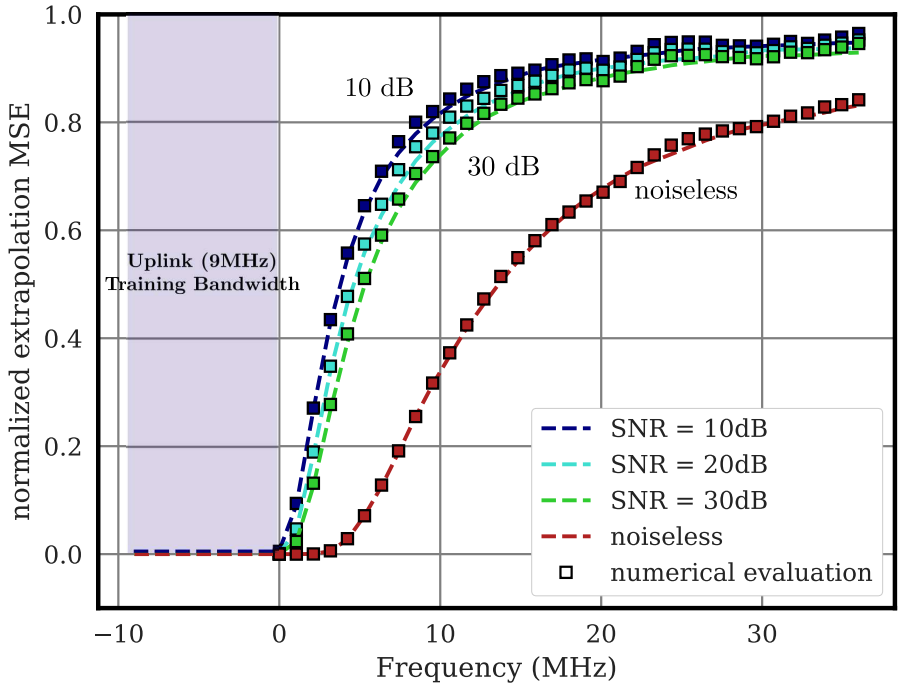

Figure 5.5 – Normalized MSE of the channel gain with $B_{UL} = 9MHz$ and $N_{UL} = 200$ uplink subcarriers, analytical (5.7) and numerical evaluation (linear scale). The extrapolation scheme is very sensitive to noise. Parameters are those of [12].

carrier k

$$\frac{\mathbb{E}\left\{\|\mathbf{h_k} - \hat{\mathbf{h}}_\mathbf{k}\|^2\right\}}{\mathbb{E}\left\{\|\mathbf{h_k}\|^2\right\}} = \frac{\sum_{q=1}^{Q}\sum_{l=0}^{N_t-1}\varepsilon_{q,l}^2[k]}{(1+\Lambda\Gamma)(1+\lambda\gamma)}$$

where $\varepsilon_{q,l}^2[k]$ is the variance of the estimated gain of subcluster l in cluster q at subcarrier k obtained from (5.6). The channel extrapolation error is the sum of the extrapolation errors from all its constituting clusters.

5.3 Reduction of Beamforming Gain

We have shown in the previous sections that the downlink channel extrapolated from uplink measurements follows a Complex Gaussian distribution whose mean and covariance have been characterized. In this section we will use this statistical structure to derive an approximate closed form formula of the expected Reduction of Beamforming gain at subcarrier k. To that end we introduce simplified notations. $\mathbf{h}$ denotes the $N_t \times 1$ MISO channel at subcarrier k ($\mathbf{h} = \mathbf{h_k}$). Similarly $\hat{\mathbf{h}}$ denotes the estimated MISO channel at subcarrier k ($\hat{\mathbf{h}} = \hat{\mathbf{h}}_\mathbf{k}$).

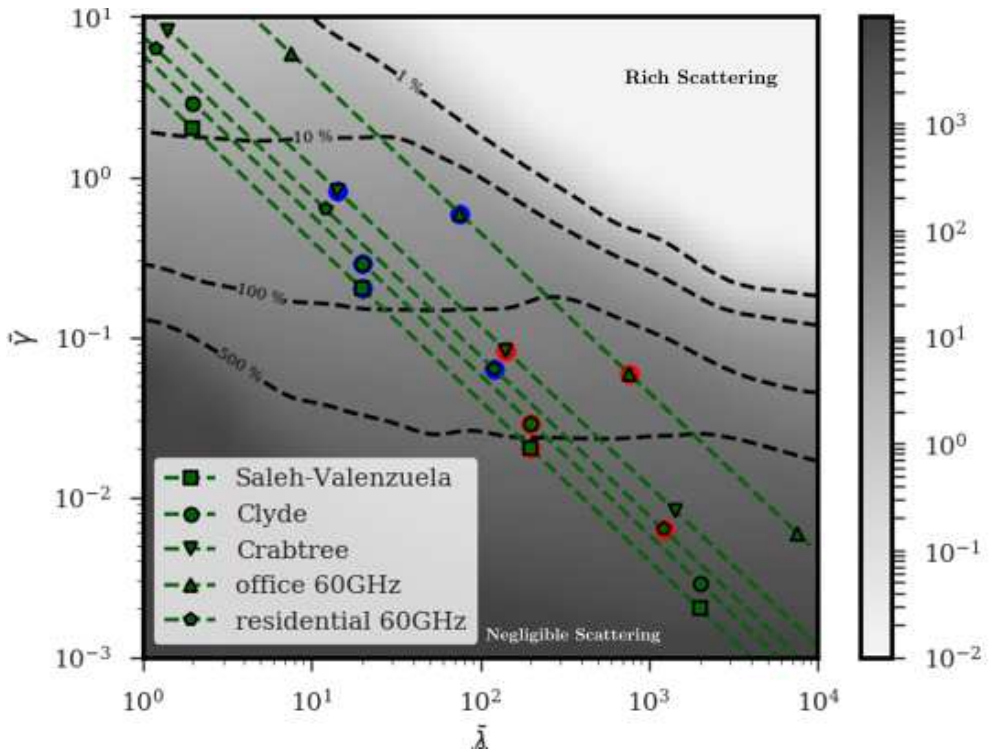

Figure 5.6 – Maximum achievable bandwidth extension for a single cluster (in % of the uplink bandwidth) as a function of the dimensionless cluster parameters $\bar{\lambda}$ and $\bar{\gamma}$, $\rho = 30dB$. Characterized channels have been placed onto the curve in dashed green lines (blue points: $B_{UL} = 10MHz$, red points: $B_{UL} = 1MHz$).

5.3.1 Analytical derivation of the RBG

The mean square error is not the best figure of merit of an extrapolation technique. Indeed the estimated downlink channel is then used to design a precoder $\mathbf{p}$ that shapes the transmitted signal in order to optimize the capacity. In this section we use a Maximum Ratio Transmission (MRT) precoder as it yields the optimal capacity in the single user case. In multi-user cases, it has been shown that the *favorable propagation* property of massive MIMO offers a natural interference reduction mechanism. The optimal precoder is denoted

$$\mathbf{p} = \frac{\mathbf{h}}{\|\mathbf{h}\|}.$$

The transmitter doesn't have access to the true channel $\mathbf{h}$. It uses the estimated channel $\hat{\mathbf{h}}$ to compute the precoder

$$\hat{\mathbf{p}} = \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}.$$

Capacity is the best figure of merit but is also hard to compute. Therefore most papers use intermediate metrics such as the Reduction of Beamforming Gain (RBG) [9] given by

$$RBG = \frac{|\mathbf{p}^H\mathbf{h}|^2}{|\hat{\mathbf{p}}^H\mathbf{h}|^2}.$$

This metric is directly linked to the single-user capacity. It can be rewritten in logarithmic scale

$$RBG_{dB} = 10\log\left(|\mathbf{p}^H\mathbf{h}|^2\right) - 10\log\left(|\hat{\mathbf{p}}^H\mathbf{h}|^2\right).$$

The optimal and sub-optimal beamforming gains $|\mathbf{p}^H\mathbf{h}|^2$ and $|\hat{\mathbf{p}}^H\mathbf{h}|^2$ have similar quadratic forms structures. We introduce the Gram matrix of all sub-clusters basis functions $\mathbf{G}$ so as $[\mathbf{G}]_{l+N_tq,l'+N_tq'} = \langle \mathbf{e_l}(\theta_q), \mathbf{e_{l'}}(\theta_{q'})\rangle$, the vector of all sub-clusters gains at subcarrier k $\mathbf{c}$ so as $\mathbf{c}_{l+N_tq} = \mathbf{c_{q,l}}[k]$ and the vector of all estimated sub-clusters gains at subcarrier k $\hat{\mathbf{c}}$ so as $\hat{\mathbf{c}}_{l+N_tq} = \hat{\mathbf{c}}_{\mathbf{q,l}}[k]$. $\boldsymbol{\mathcal{E}}$ is the diagonal covariance matrix of the channel gains $\mathbf{c}$. Taking the point of view of the transmitter, the true gains $\mathbf{c}$ are unknown. Only the distribution of $\mathbf{c} \sim \mathcal{CN}(\hat{\mathbf{c}}, \boldsymbol{\mathcal{E}})$ is known.

The inner products for optimal and sub-optimal precoding are

$$\mathbf{p}^H\mathbf{h} = \sqrt{N_t}\frac{\mathbf{c}^H\mathbf{G}\mathbf{c}}{\sqrt{\mathbf{c}^H\mathbf{G}\mathbf{c}}} = \sqrt{N_t}\sqrt{\mathbf{c}^H\mathbf{G}\mathbf{c}},\ \hat{\mathbf{p}}^H\mathbf{h} = \sqrt{N_t}\frac{\hat{\mathbf{c}}^H\mathbf{G}\mathbf{c}}{\sqrt{\hat{\mathbf{c}}^H\mathbf{G}\hat{\mathbf{c}}}}.$$

Their modulus squared yield the quadratic forms

$$|\hat{\mathbf{p}}^H\mathbf{h}|^2 = N_t\frac{\mathbf{c}^H\mathbf{G}\hat{\mathbf{c}}\hat{\mathbf{c}}^H\mathbf{G}\mathbf{c}}{\hat{\mathbf{c}}^H\mathbf{G}\hat{\mathbf{c}}} = N_t\mathbf{c}^H\mathbf{A}\mathbf{c},$$

$$|\mathbf{p}^H\mathbf{h}|^2 = N_t\mathbf{c}^H\mathbf{G}\mathbf{c}.$$

According to [14], the first and second order moments are

$$\begin{cases}\mathbb{E}\left\{|\hat{\mathbf{p}}^H\mathbf{h}|^2\right\}/N_t = \mathrm{Tr}\left(\mathbf{A}\boldsymbol{\mathcal{E}}\right) + \hat{\mathbf{c}}^H\mathbf{A}\hat{\mathbf{c}} \\ \mathbb{V}ar\left\{|\hat{\mathbf{p}}^H\mathbf{h}|^2\right\}/N_t^2 = \mathrm{Tr}\left((\mathbf{A}\boldsymbol{\mathcal{E}})^2\right) + 2\hat{\mathbf{c}}^H\mathbf{A}\boldsymbol{\mathcal{E}}\mathbf{A}\hat{\mathbf{c}} \\ \mathbb{E}\left\{|\mathbf{p}^H\mathbf{h}|^2\right\}/N_t = \mathrm{Tr}\left(\mathbf{G}\boldsymbol{\mathcal{E}}\right) + \hat{\mathbf{c}}^H\mathbf{G}\hat{\mathbf{c}} \\ \mathbb{V}ar\left\{|\mathbf{p}^H\mathbf{h}|^2\right\}/N_t^2 = \mathrm{Tr}\left((\mathbf{G}\boldsymbol{\mathcal{E}})^2\right) + 2\hat{\mathbf{c}}^H\mathbf{G}\boldsymbol{\mathcal{E}}\mathbf{G}\hat{\mathbf{c}}.\end{cases}$$

Note that $|\hat{\mathbf{p}}^H\mathbf{h}|^2$ follows the distribution

$$|\hat{\mathbf{p}}^H\mathbf{h}|^2/N_t \sim \tilde{\sigma}^2\chi_2^2\left(\frac{\tilde{\mu}^2}{\tilde{\sigma}^2}\right)$$

where $\tilde{\sigma}^2 = \frac{1}{2}\mathrm{Tr}\left(\mathbf{A}\boldsymbol{\mathcal{E}}\right)$, $\tilde{\mu}^2 = \hat{\mathbf{c}}^H\mathbf{A}\hat{\mathbf{c}}$ and $\chi_k^2(\lambda)$ represents the noncentral Chi-Squared distribution with non-centrality parameter λ and k degrees of freedom [15].

Using the following second order approximation

$$\mathbb{E}\left\{f(X)\right\} \approx f(\mathbb{E}\left\{X\right\}) + \frac{1}{2}f''(\mathbb{E}\left\{X\right\})\mathbb{V}ar\left\{X\right\}$$

and the previous formulas we can approximate very closely the Reduction of Beamforming Gain

$$\begin{aligned}\mathbb{E}\left\{RBG_{dB}\right\} \approx\ & 10\log\left(\mathbb{E}\left\{|\mathbf{p}^H\mathbf{h}|^2\right\}\right) - 10\log\left(\mathbb{E}\left\{|\hat{\mathbf{p}}^H\mathbf{h}|^2\right\}\right) \\ & - \frac{5\mathbb{V}ar\left\{|\mathbf{p}^H\mathbf{h}|^2\right\}}{\ln(10)\mathbb{E}\left\{|\mathbf{p}^H\mathbf{h}|^2\right\}^2} + \frac{5\mathbb{V}ar\left\{|\hat{\mathbf{p}}^H\mathbf{h}|^2\right\}}{\ln(10)\mathbb{E}\left\{|\hat{\mathbf{p}}^H\mathbf{h}|^2\right\}^2}.\end{aligned} \tag{5.10}$$

5.3.2 Numerical experiments

We illustrate those formulas with a simple 3-clusters model with equal parameters (λ, γ, P_q), $N_t = 30$ and null angular spread. The clusters main angles are uniformly distributed $\theta_q \sim [0, 2\pi]$. As equation (5.10) gives the expected downlink RBG for a specific uplink realisation, both numerically evaluated and expected RGB were averaged over multiple

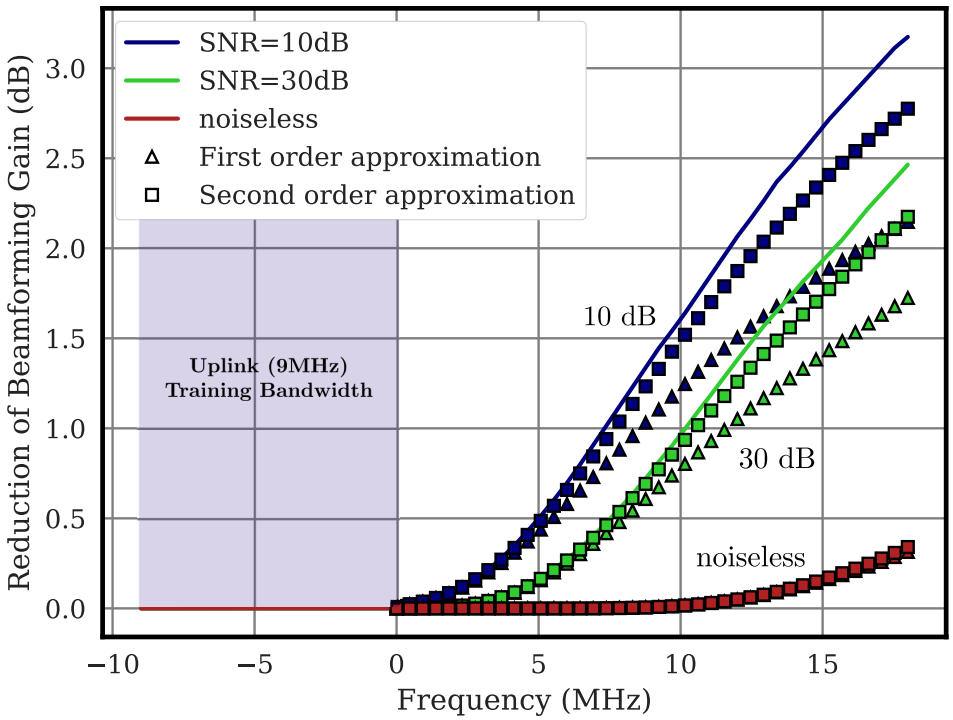

Figure 5.7 – Average Reduction of Beamforming Gain for 10dB, 30 dB and infinite SNR. Original Saleh-Valenzuela parameters.

realisations to obtain comparable results (Fig. 5.7). The approximation remains valid until the RGB reaches a 1.5 dB loss.

Comparing Figures 5.4 and 5.7, for a 10dB average SNR, the normalized cluster MSE reaches 60%. However, the RBG (SNR loss due to the extrapolation error) is only limited to 1.5dB, which is an acceptable loss. This illustrates how the RBG is overall a better measure of performance.

5.4 Conclusion

In this chapter we have investigated the limits of an extrapolation process on a wide-band MIMO system to infer the downlink channel from uplink measurements. As the clusters constituting the propagation are made of a large amount of densely distributed rays that are hardly separable, a specular representation of clusters in the angular-delay domain is not available. Clusters have been projected on a adequately chosen sparse BEM in the angular domain then extrapolated in the frequency domain. Using the posterior distribution of the downlink channel, we have obtained a closed form expression for the

MSE applied to Saleh-Valenzuela models. The extrapolation accuracy highly depends on both propagation channel characteristics and SNR. Cluster estimation errors for any parameters have then been summarized in a generic abacus, providing a convenient way to evaluate the extrapolation potential. We have also derived an approximate formula for the expected Reduction of Beamforming Gain.

We used the conventional MRT precoder, consistent with previous work in [9]. From the extrapolated channel distribution structure provided in this chapter, new precoding strategies that leverages this channel error structure might however outperform MRT. On the other hand, in the last section, we proposed estimates for the downlink channel gain. Those indicators could be extended to inter-user interference, providing inputs for massive MIMO resource allocation strategies.

BIBLIOGRAPHY

[1] Emil Björnson, Jakob Hoydis, and Luca Sanguinetti, « Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency », *in*: *Foundations and Trends® in Signal Processing* (2017).

[2] L. Bigler et al., « Experimental direction of arrival and spatial signature measurements at 900 MHz for smart antenna systems », *in*: *1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, Chicago, IL, USA: IEEE, 1995.

[3] K.I. Pedersen, P.E. Mogensen, and F. Frederiksen, « Joint directional properties of uplink and downlink channel in mobile communications », *in*: *Electron. Lett.* (1999).

[4] Zhimeng Zhong, Li Fan, and Shibin Ge, « FDD Massive MIMO Uplink and Downlink Channel Reciprocity Properties: Full or Partial Reciprocity? », *in*: *arXiv:1912.11221 [cs, eess, math]* (Dec. 2019).

[5] Yantao Han, Jiqing Ni, and Gaoke Du, « The potential approaches to achieve channel reciprocity in FDD system with frequency correction algorithms », *in*: *Proceedings of the 5th International ICST Conference on Communications and Networking in China*, Beijing, China: IEEE, 2010.

[6] Deepak Vasisht et al., « Eliminating Channel Feedback in Next-Generation Cellular Networks », *in*: *Proceedings of the 2016 conference on ACM SIGCOMM 2016 Conference - SIGCOMM '16*, Florianopolis, Brazil: ACM Press, 2016.

[7] Hongxiang Xie et al., « A Unified Transmission Strategy for TDD/FDD Massive MIMO Systems With Spatial Basis Expansion Model », *in*: *IEEE Trans. Veh. Technol.* (Apr. 2017).

[8] Hayder Almosa et al., « Downlink Channel Estimation with Limited Feedback for FDD Multi-User Massive MIMO with Spatial Channel Correlation », *in*: *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO: IEEE, May 2018.

[9] Thomas Choi et al., « Channel Extrapolation for FDD Massive MIMO: Procedure and Experimental Results », *in*: *arXiv:1907.11401 [cs, eess]* (July 2019).

[10] François Rottenberg et al., « Channel Extrapolation in FDD Massive MIMO: Theoretical Analysis and Numerical Validation », *in*: *arXiv:1902.06844 [cs, math]* (Feb. 2019).

[11] Adel AM Saleh and Reinaldo Valenzuela, « A statistical model for indoor multipath propagation », *in*: *IEEE Journal on selected areas in communications* (1987).

[12] Q.H. Spencer et al., « Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel », *in*: *IEEE Journal on Selected Areas in Communications* (Mar. 2000).

[13] Su-Khiong Yong et al., *G3c channel modeling sub-commitee final report*, Mar. 2007.

[14] A. M. Mathai, Serge B. Provost, and Takesi Hayakawa, *Bilinear Forms and Zonal Polynomials*, ed. by P. Diggle et al., Lecture Notes in Statistics, New York, NY: Springer New York, 1995.

[15] Marvin Kenneth Simon, *Probability distributions involving Gaussian random variables: a handbook for engineers and scientists*, New York: Springer, 2006.

Chapter 6

Improvements on Channel Estimation for Sparse MIMO channels

Channel estimation and denoising is challenging in massive MIMO systems as the number of coefficients to estimate increase with the array sizes. This scaling rule render channel estimation harder to perform. Novel channel estimation techniques dedicated to millimeter wave massive MIMO systems take advantage of the particular sparse structure of millimeter wave MIMO channels, thus reducing the number of unknowns in the problem.

The contributions of this chapter consist of two improvements related to massive MIMO OMP-based channel estimation: the stopping criterion and the ray estimation steps. In order to derive those contributions, the single specular ray detection and estimation sub-problem is fully characterized in Sections 6.3 and 6.4. The resulting advanced detection and estimation steps are then integrated into the OMP algorithm in Section 6.5 and compared on NYUSIM channel realizations (Section 6.1.1) against the state of the art approach (Section 6.2.1). In order to simplify the notations, we limit our analysis to the single carrier $N_t \times 1$ MISO channel. Generalization to wideband MIMO channels is discussed in Section 6.4.4.

This work is based on previous studies on OMP-based channel estimation performed by Luc Le Magoarou and Stéphane Paquelet at b<>com [1, 2].

6.1 Generalities

As the amount of remaining spectral resources shrinks, mobile broadband communication systems are moving towards higher frequencies, with a particular focus on millimeter waves. Those bands are currently barely exploited as their propagation characteristics are less attractive than those of lower frequency bands.

Indeed, the channel suffer from severe path-loss due to the lower wavelength. As a reminder the path-loss is given (in dB) by $PL = 10\log(\frac{4\pi d}{\lambda})^2$ where d denotes the distance between transmitter and receiver and λ the wavelength. Moreover, electromagnetic waves may suffer from high *atmospheric attenuation* in millimeter waves (in particular around $60GHz$) which forbids any long range use cases. For those reasons, millimeter waves are mostly considered as a viable solution for short-distance communications.

The combination of antenna arrays and millimeter waves is a particularly appealing bundle. As the wavelength λ is smaller, each individual antenna is more compact. This enables large antenna arrays to fit in smaller form factors while providing the massive array gain necessary to compensate for the severe path loss as well as the atmospheric attenuation.

As the frequency increase, multipath scattering is less significant. The propagation channel is thus usually dominated by the LOS path. The remaining scattering usually consists of a few specular rays. In this case, it is commonly stated that the channel is *sparse* in the angular domain [3, 4, 5]. This means that the channel can be modeled by a limited number of distinct specular paths, each characterized by a gain β_p, DOD $\vec{u}_{tx,p}$, DOA $\vec{u}_{rx,p}$ and delay τ_p, as in Equation (1.15), as

$$\mathbf{h}(f) = \sqrt{N_r N_t} \sum_{p=1}^{P} \beta_p \, \mathrm{e}^{-2\pi j f \tau_p} \, \mathbf{e_r}(\vec{u}_{rx,p}) \otimes \mathbf{e_t}^*(\vec{u}_{tx,p}). \tag{6.1}$$

where the dimension of $\mathbf{h}$ is $N_r \times N_t$.

6.1.1 The NYUSIM Channel Model

We used the publicly available channel model NYUSIM to generate realistic channel realizations. This model has been calibrated by the NYU Wireless research team at New York University [6, 7]. For illustration purposes, we display on Figures 6.1 and 6.2 the angles of departure and the corresponding Fourier domain angular channel response of a single carrier narrowband MISO system with 128 transmit antennas. In this chapter, NYUSIM channels are generated using the UMi (Urban Microcell, a typical use case of millimeter waves) parameters in NLOS conditions. As depicted on Figures 6.1 and 6.2, the channel only consists of a few specular rays and is thus sparse in the angular domain.

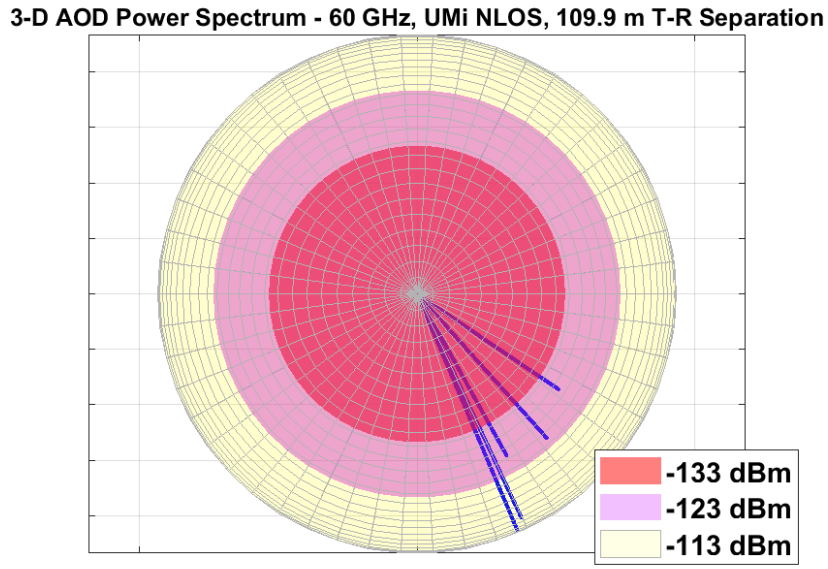

Figure 6.1 – Angular power spectrum of a NYUSIM channel realization. The channel only consists in a few specular paths.

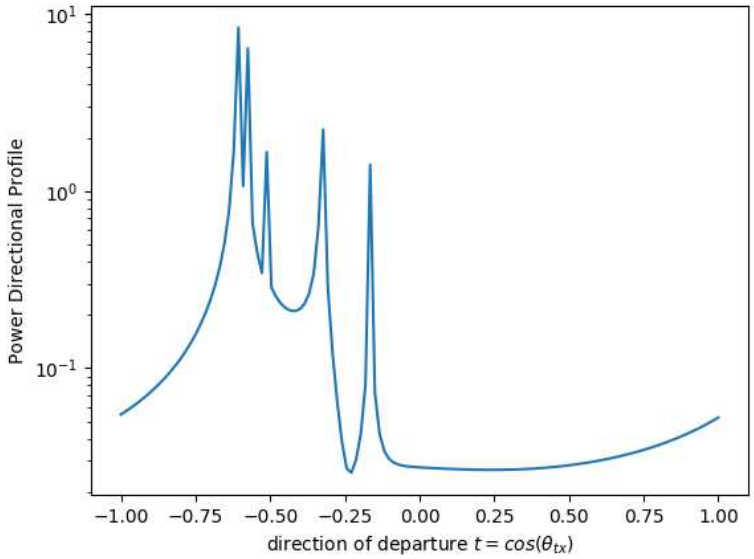

Figure 6.2 – Angular Fourier domain representation of the channel described on Figure 6.1 sampled by a 128 ULA antenna array.

6.2 Channel Estimation

In this section, we state the channel estimation problem. In massive MIMO systems, such mandatory step is challenging as the number of unknowns to estimate gets overwhelming. To cut down this burden, recent approaches use an OMP-based ray estimation that leverages the millimeter wave channel sparsity. This property of millimeter wave channels has been showcased in Section 6.1.1 on the NYUSIM channel simulator.

6.2.1 Problem Statement

The multipath propagation channel alters the signal traveling from the transmitter to the receiver. On the receive side, the symbols have to be cleared of the channel influence to be decoded efficiently. This operation is performed in two steps. First, the channel response is estimated based on known *pilot* signals sent by the transmitter. This step is called *channel estimation*. Then the channel influence is removed of the other data symbols based on this estimated response. In an OFDM system, the received symbols $\mathbf{y}$ are given by

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n}$$

where $\mathbf{X}$ are the known pilots (the *sensing matrix*), $\mathbf{n}$ denotes the thermal noise samples that pollute the received symbols and $\mathbf{h}$ is the propagation channel, previously described in Equation (1.15).

Channel estimation consists in estimating the propagation channel, characterized by the vector $\mathbf{h}$, from the received samples $\mathbf{y}$ and the known pilot signals $\mathbf{X}$. The unknown $\mathbf{h}$ consists in $2N_tN_rN_f$ real random variables, that have to be estimated to fully characterize the channel. Channel estimation is difficult in massive MIMO systems, because the number of parameters to estimate increase with $N_r \times N_t$. In order to cut down the number of parameters to estimate, researchers recently proposed to estimate the physical parameters β_p, τ_p, $\vec{u}_{tx,p}$ and $\vec{u}_{rx,p}$ of the rays that constitute the channel (depicted in Equation (6.1)), thus leveraging the angular domain sparsity.

6.2.2 Orthogonal Matching Pursuit

Several techniques can be employed to extract the ray parameters. In this chapter, we do not study the high resolution path extraction algorithms (MUSIC [8], ESPRIT [9] or SAGE [10]) which are also designed to extract the ray parameters but requires high

computation times. We rather consider the general purpose OMP algorithm [11], that extracts rays from the received samples sequentially. This tool, related to the field of *compressed sensing*, has recently been employed for massive MIMO channel estimation in multiple papers [1, 2, 12, 13, 14, 15].

OMP is a general purpose algorithm to solve inverse problems involving sparse representations. Assuming that the received signal consists of a linear combination of elements (the atoms) from a known dictionary, OMP greedily extracts the best atom, which is the most correlated with the residual, and updates the coefficients of all extracted atoms accordingly. However, this versatile algorithm has to be adapted to the particular problem at hand. Considering massive MIMO channel estimation, two main features have to be taken into account:

— **Feature 1.** The dictionary, which is the set of all possible *steering vectors*, spans over a continuum of vectors that represent all possible DODs (or DOAs or delays depending on the dimension).
— **Feature 2.** The received signal is polluted by i.i.d. complex Gaussian distributed thermal noise.

The OMP algorithms designed for massive MIMO channel estimation always follow the same pattern, described in paper [16] and summarized on Figure 6.3.

Specular ray estimation: First, the most powerful ray (characterized by its angle of arrival and angle of departure) is extracted from the received signal using a maximum likelihood estimation. This estimator is usually implemented using a grid search (exhaustive search over a quantized angular interval) which amounts to testing a quantized set of directions and picking the most likely. The likelihood is evaluated on each node of the estimation grid as described on Figure 6.4. We define

$$OS = N_{grid}/N_t \tag{6.2}$$

the oversampling factor (here in the DOD domain) as the ratio of the number of evaluation points over the number of transmit antennas. This important dimensioning factor (as well as its undersampling factor $US = 1/OS$ counterpart) will be reused later on when comparing different configurations. The node which yields the highest likelihood provides a confidence interval centered on this very node of width $\Delta_t = 2/(N_t OS)$ (described in Figure 6.5) where the DOD actually lies in. This uncertainty on the exact direction put

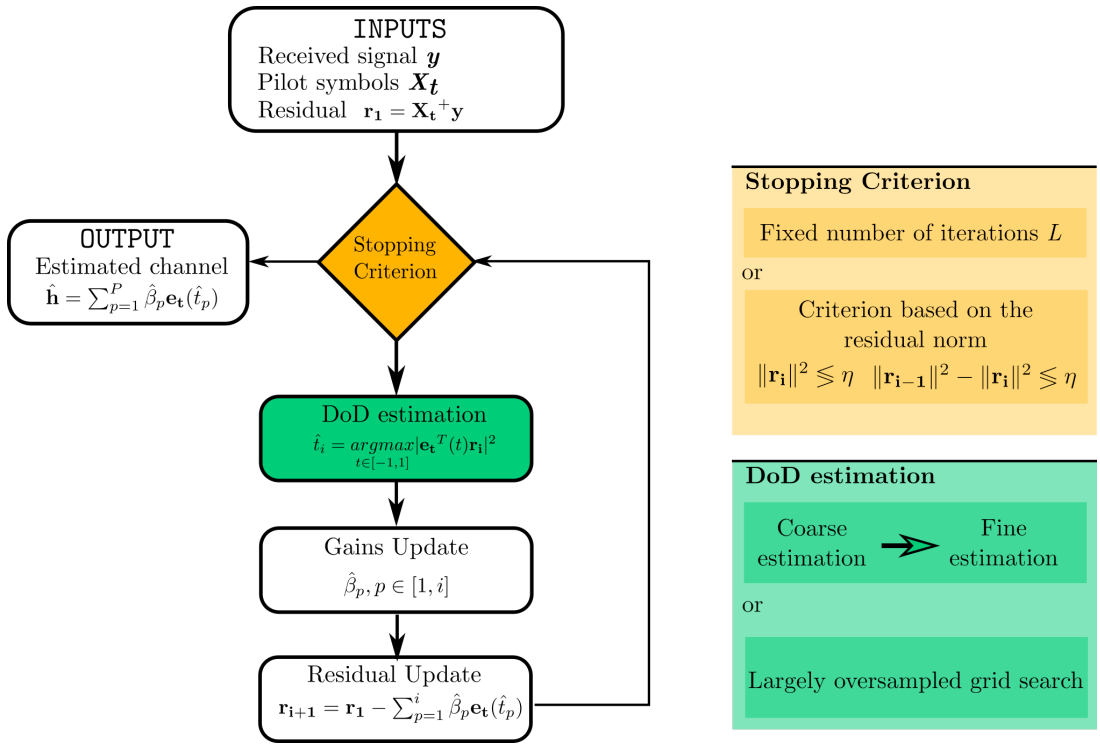

Figure 6.3 – Flowchart of a typical OMP algorithm for massive MIMO algorithm.

a bound on the variance of this simple estimator as

$$\mathbb{V}ar\left\{\hat{t}\right\} = \frac{\Delta_t^2}{12} = \frac{1}{3N_t^2OS^2} \tag{6.3}$$

which is the formula that characterize the variance of a uniform distribution of width $\Delta_t = 2/(N_tOS)$.

To ensure that the grid search estimation variance matches the CRLB (derived in Section 6.4.1), a conservative dimensioning based on a high oversampling factor $OS \gg 1$ would be required. This is a direct consequence of the continuous nature of the unknown θ_{tx} (*Feature 1*). To highlight this estimation performance saturation, both optimal standard deviation (square root of the CRLB) and grid search error floor are drawn on Figure 6.6 for $OS \in \{2, 4, 6, 8, 10\}$, for a 32 antennas narrowband MISO system. An analysis of the relation between oversampling and channel estimation error is provided in paper [2]. This step is a significant bottleneck for the OMP algorithm.

The Oversampling (OS) requirements can be mitigated by using a two-stage approach (coarse then fine estimation). Indeed, acquiring the approximate direction of the ray can

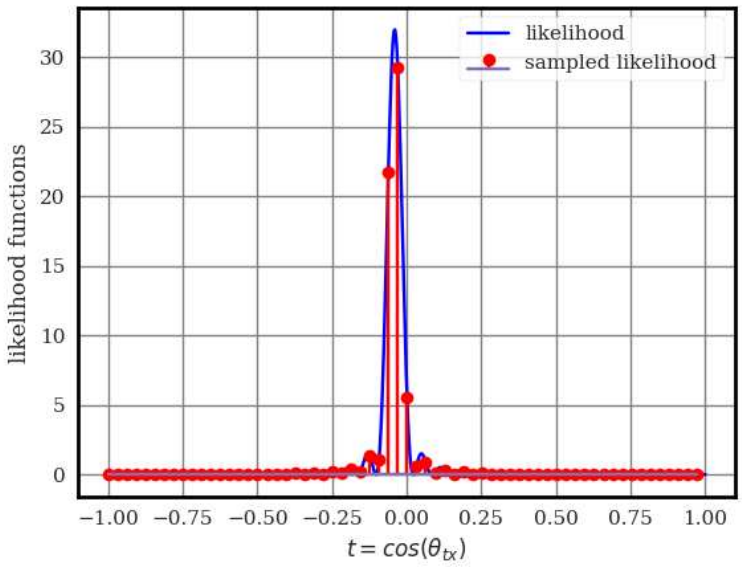

Figure 6.4 – Grid search for the most likely direction of departure with oversampling $OS = 2$.

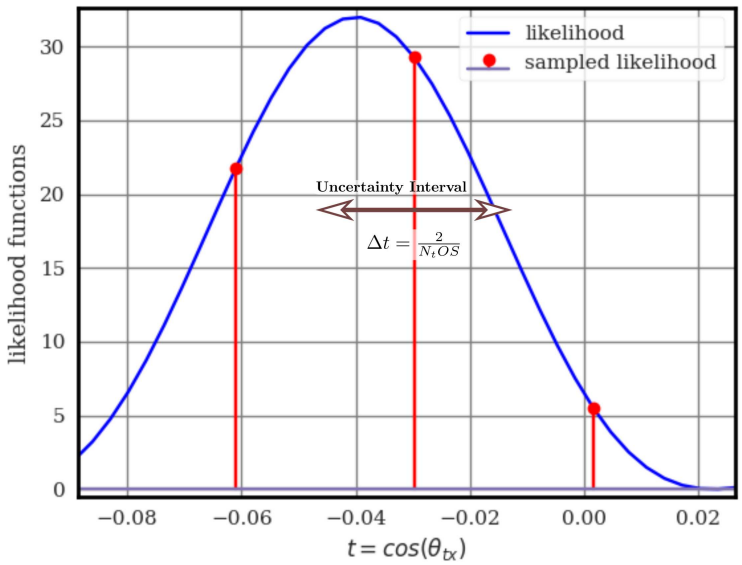

Figure 6.5 – Zoomed in version of Figure 6.4 that features the uncertainty interval on the estimated direction of departure.

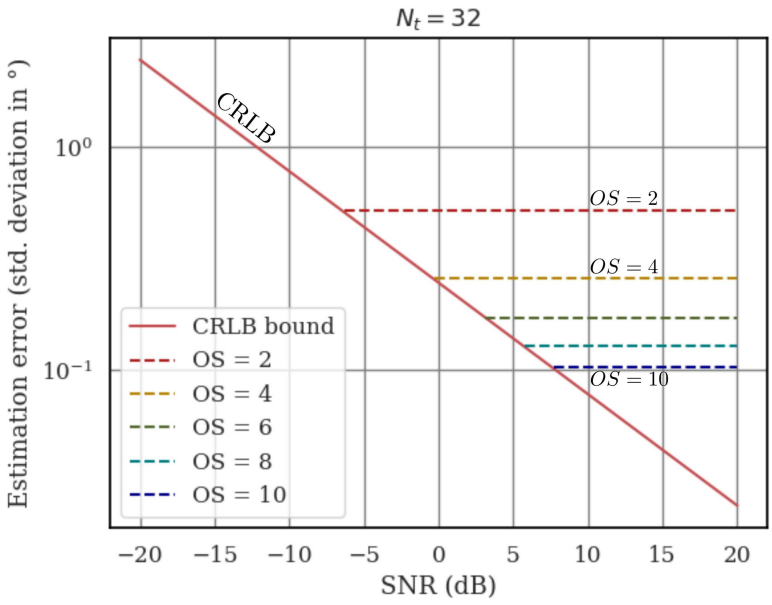

Figure 6.6 – Comparison between the Cramér-Rao bound on the angle of departure estimate and the minimum achievable variance using grid search with oversampling factor $OS \in \{2, 4, 6, 8, 10\}$.

be done at a lower cost, using only a $OS = 2$ oversampling factor [17, 16, 18]. We will analyze in Section 6.4 how this factor 2 arises, how we can reduce the oversampling requirement and what are the consequences on the performance.

This coarse estimate serves as a suitable starting point to attain a finer estimate in a second stage using a dedicated algorithm. This refinement stage usually consists in either a denser local grid search [17] or mathematical optimization (Gradient descent, Newton method, ...) [18].

Gains update: Once the best direction has been extracted from the residual, the complex coefficients $\hat{\beta}_p$ of all estimated rays are estimated using a Moore-Penrose inverse as

$$\hat{\boldsymbol{\beta}} = (\mathbf{E_t}^+)^* \mathbf{X_t}^+ \mathbf{y}$$

where $\hat{\boldsymbol{\beta}}$ denotes the vector of the gains associated with all currently estimated directions, $\mathbf{E_t}$ is the matrix that contains the stacked estimated steering vectors. $\mathbf{E_t}^+ = (\mathbf{E_t}^H \mathbf{E_t})^{-1} \mathbf{E_t}^H$ denotes the Moore-Penrose inverse of $\mathbf{E_t}$ and $\mathbf{X_t}^+ \mathbf{y}$ are the received symbols where the influence of the pilots have been removed.

The coefficients of all rays extracted since the beginning are updated at each iteration.

That is the difference between the Matching Pursuit algorithm that does not update the gains and the Orthogonal Matching Pursuit variants.

Residual update: The currently estimated channel is subtracted from the received symbols. The resulting residual is used as an input for the next iteration specular ray search.

Stopping criterion: OMP is an iterative algorithm that has to be stopped according to some criterion. This particular point is usually overlooked in the literature and is most of the time heuristically chosen. If the extraction procedure stops too early, the channel hasn't been fully extracted leading to a biased estimate. If it stops too late, it interprets noise as specular rays leading to a noisy estimate. Various stopping criteria have been proposed. In [18] the number of propagation paths is supposed to be known beforehand. However the optimal number of specular rays to be estimated is actually the result of a complicated trade-off that depends on the propagation environment and the Signal to Noise Ratio [2]. In most papers, the stopping criterion is based on an empirically set threshold on the norm of the residual [14, 17]. Other possible stopping criteria, mainly heuristics, are proposed in the paper [15].

However, the criteria based on the norm of the residual does not take into account *Feature 2*, that is the structure of the channel, which is concentrated on a few distinct directions, is radically different from the structure of the noise, which is spread out over all directions. *Feature 2* is taken into account in our analysis in this chapter.

6.3 The atomic ray detection problem

In this section we solve the single specular path detection problem. The classical detection strategies of the literature are presented in Section 6.3.1. Then we introduce a novel interval-based detector that can reduce the oversampling requirements in Section 6.3.2. The detection approaches are then characterized and compared in Section 6.3.3. The detection outcome will be used as the stopping criterion of the OMP algorithm.

In order to keep the document tractable, we consider only the single carrier MISO problem with a ULA array. It only involves the angle of departure θ_{tx}. As the steering vector is a function of the intermediate variable $t = \cos(\theta_{tx})$, all future derivations involves t instead of θ_{tx}. Generalization is discussed in Section 6.4.4.

6.3.1 Overview of classical detection approaches

The received samples $\mathbf{y}$ can contain either thermal noise plus the transmitted signal or only thermal noise. The two possible outcomes can be stated as a binary (H_{sig} against H_{noise}) hypothesis testing on the received samples $\mathbf{y}$ as

$$\begin{cases} H_{sig} : \mathbf{y} = \mathbf{X_t}\mathbf{h}(t) + \mathbf{n} \\ H_{noise} : \mathbf{y} = \mathbf{n} \end{cases}$$

Optimal approach: Assuming that the direction of departure is known, the probability density functions of the distribution of $\mathbf{y}$ under both hypotheses are given by

$$\begin{cases} H_{sig} : \mathbf{y} \sim \mathcal{CN}\left(\mathbf{X_t}\mathbf{h}(t), \sigma_n^2\mathbf{I}\right) \\ H_{noise} : \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I}) \end{cases}$$

Theorem. *The most powerful detection test for this problem is given by the matched filter*

$$|\mathbf{y}^H \mathbf{X_t} \mathbf{e_t}^*(t)|^2 \lessgtr \gamma \tag{6.4}$$

where γ is a predefined threshold calibrated according to a false alarm target α, given as

$$\alpha = \int_{\Re\{\mathbf{y}^H \mathbf{X_t} \mathbf{e_t}^*(t)\} < \gamma} p_{H_{noise}}(\mathbf{y}) d\mathbf{y}.$$

This theorem is proven in appendix D.1. This is the optimal detector. However it requires the knowledge of t, which is not yet available.

Naive approach: To detect whether a ray lies in the received samples $\mathbf{y}$ without prior knowledge on t, the most simple method is to test the likelihood of the null hypothesis $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$ using a *statistical significance test*. We reject the H_{noise} hypothesis if the test statistic falls behind a predefined threshold calibrated according to a false alarm rate α, given by

$$\alpha = P(\|\mathbf{y}\|^2 > \eta) = \int_{\eta}^{\infty} p_{\chi^2(2N_t, \sigma_n^2/2)}(x) dx = F_{\chi^2(2N_t, \sigma_n^2/2)}(\eta)$$

where η is the corresponding test threshold, p_{χ^2} and F_{χ^2} denote respectively the p.d.f. and the c.d.f. of a χ^2 distribution. The test threshold is set to achieve the chosen false error

rate α as $\eta = F^{-1}_{\chi^2(2N_t,\sigma_n^2/2)}(\alpha)$.

This test is largely sub-optimal as the structure of the signal under the H_{sig} hypothesis is not taken into account.

Generalized likelihood ratio test: As t is unknown, we don't have a perfect knowledge of the distribution of $\mathbf{y}$ under the signal plus noise hypothesis, that could allow the design of the optimal precoder. To overcome this problem, we used a *generalized likelihood ratio test* approach [19]. This technique consists in first estimating the DOD assuming H_{sig} using a maximum likelihood approach as

$$\hat{t} = \underset{t\in[-1,1]}{\operatorname{argmax}}\ p_{H_{sig}}(\mathbf{y}, t)$$

This amounts as finding t that maximizes

$$\hat{t} = \underset{t\in[-1,1]}{\operatorname{argmax}}\ |\mathbf{y}^H \mathbf{X_t} \mathbf{e_t}^*(t)|^2.$$

In the performance comparisons displayed in Section 6.3.3, the optimization step is performed using a grid search and is consistent with the state of the art OMP ray estimation step as described in Section 6.2.1. The main dimensioning factor is the oversampling factor of the grid OS.

This estimate is then injected into the test statistic (6.4) as

$$|\mathbf{y}^H \mathbf{X_t} \mathbf{e_t}^*(t)|^2 \lessgtr \gamma$$

which is the same function evaluated for the prior estimation, instead of Equation (6.4).

6.3.2 Interval-based detection

To reduce the number of nodes in the search grid without leaving gaps, we propose an alternative approach. We subdivide the search interval $[-1, 1]$ into K_t equal subsections of size $\Delta_t = 2/K_t$. By testing if a ray lies in one of the K_t intervals instead of testing the whole search space $[-1, 1]$, noise is filtered and curtailed thus increasing the test power. Each interval j has the shape $[t_j, t_{j+1}]$ with mean value $\bar{t}_j = (t_{j+1} - t_j)/2$ and width $\Delta_t = (t_{j+1} + t_j)/2$. For each interval $[t_j, t_{j+1}]$, we state the ratio likelihood as

$$r(\mathbf{y}, t_j < t < t_{j+1}) = \frac{p_{H_{sig}}(\mathbf{y}|t_j < t < t_{j+1})}{p_{H_{noise}}(\mathbf{y})} = \frac{\int_{t_j}^{t_{j+1}} p_{H_{sig}}(\mathbf{y}|t)dt}{p_{H_{noise}}(\mathbf{y})}.$$

The index of the interval which is the most likely to contain the direction of departure is given by

$$j_{opt} = \underset{j\in[0,K_t-1]}{\operatorname{argmax}}\ r(\mathbf{y}, t_j < t < t_{j+1}) \tag{6.5}$$

In annex D.2 we derive the following approximate first order sufficient statistic for this decision problem as a quadratic form

$$f(\mathbf{z}, j) = \mathbf{z}^H \mathbf{\Delta}^H(\bar{t}_j)\mathbf{R_0}(\Delta_t)\mathbf{\Delta}(\bar{t}_j)\mathbf{z} \tag{6.6}$$

where $\mathbf{z} = \mathbf{X_t}^H\mathbf{y}$, the interval mean $\bar{t}_j$ and size Δ_t are separated into two matrices $\mathbf{\Delta}(\bar{t}_j)$ and $\mathbf{R_0}(\Delta_t)$

$$\begin{cases} \mathbf{\Delta}(\bar{t}_i) = \operatorname{diag}(\mathbf{e_t}(\bar{t}_i)) \\ [\mathbf{R_0}]_{k,l} = \operatorname{sinc}(2\pi\Delta_t(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda}). \end{cases} \tag{6.7}$$

In particular $\mathbf{R_0}$ only depends on the constant interval size Δ_t chosen when dimensioning the system. Moreover this kernel matrix is real-valued symetric (diagonalizable in $\mathbb{R}$) and Toeplitz. $\mathbf{R_0}$ can also be decomposed off-line into eigencomponents $\mathbf{R_0} = \sum_{k=1}^{N_t} \lambda_k \mathbf{r_k}\mathbf{r_k}^T$. The sufficient statistic can be rewritten as a weighted sum of correlators

$$f(\mathbf{z}, j) = \sum_{k=1}^{N_t} \lambda_k |\mathbf{y}^H \mathbf{X_t}(\mathbf{r_k} \odot \mathbf{e_t}^*(\bar{t}_j))|^2 \tag{6.8}$$

We expect that most of the eigenvalues will have negligible power and the sum (6.8) can be truncated and reduced to a small number of correlators $\#\lambda << N_t$, enabling low-complexity implementation. In practice the number of required correlators is of the order of

$$\#\lambda = 1/OS + 1 \tag{6.9}$$

when $OS \leq 1$. Compared to the Generalized likelihood ratio test (GLRT) detection technique, this approach basically comes back to trading off $OS \times N_t$ different correlators (with large OS) for $OS \times N_t$ (with much smaller OS) correlators, each repeated $\#\lambda$ times with different windows, then summed. Overall, the number of correlators required is the same for both approaches. However, such architecture can be beneficial when it comes to pipelined hardware implementation. The effect of the four strongest windowed correlators depending on the actual direction of departure t is depicted on Figure 6.7, obtained on a $N_t = 16$ antenna array where the search space $[-1, 1]$ is divided into $K_t = 4$ intervals. At

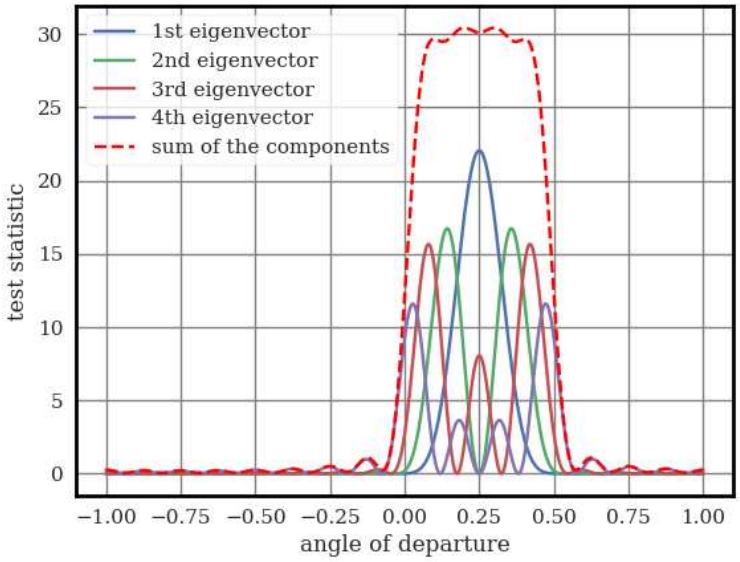

Figure 6.7 – Angular response of each individual weighted correlators of Equation (6.8) for the third interval detector out of four for any direction of departure $t \in [-1, 1]$. System with $N_t = 16$ antennas ($OS = 0.25$ or $US = 4$).

this point, it is more convenient to work with the *undersampling factor US* as

$$US = 1/OS = N_t/K_t$$

to avoid non-integer oversampling factors OS. It will be used in Section 6.3.3 when considering $OS < 1$. The setup presented on Figure 6.7 is characterized by $N_t = 16$ antennas and a search space divided into $K_t = 4$ subintervals, that corresponds to an undersampling factor of $US = 4$. Each windowed correlator is monitoring a subsection of the third interval. The sum of the responses is sketching the shape of a rectangle, providing a uniform response over the corresponding search interval.

The most likely interval j_{opt} is returned by solving (6.5). The value of this maximum can also be used as a test statistic to assess whether the estimated direction corresponds to an actual signal or noise. The false alarm rate is then given by the probability that $f(\mathbf{n}, j_{opt})$ exceed a threshold η when $\mathbf{n}$ is noise only

$$FA(\eta) = P(f(\mathbf{n}, j_{opt}) > \eta), \mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$$

In this section, the function $FA(\eta)$ is computed numerically using Monte-Carlo simulations. Then we invert this function to obtain the threshold η as a function of the chosen false alarm rate α

$$\eta = FA^{-1}(\alpha)$$

6.3.3 Analysis and comparison

In the previous sections, several detection techniques (*naive detector*, *Generalized likelihood ratio test* and the proposed *interval-based detector*) are presented. In this section, those detectors are evaluated and compared on a similar setup (ULA array with $N_t = 32$ antennas).

First, the raw detection performance is measured on Figure 6.8. The false alarm rate is set at $\alpha = 1\%$. The percentage of detected signals is evaluated relatively to the SNR. As expected, at a given SNR, the naive approach which does not account for the signal model is always the worst. In contrast, the GLRT with $OS = 2$ represents an upper bound (this is not proven, yet conjectured) on the detection performance. Note that this high oversampling factor is required to achieve acceptable performance. Reducing the oversampling requirements creates gaps in the search space and non-uniform detection performance. This is the reason why all state of the art coarse estimation techniques (presented in Section 6.2.1) rely on at least $OS = 2$.

The proposed interval-based approach is evaluated for different sampling factors, from $OS = 2$ to $US = 8$. Both $OS = 2$ and $OS = 1$ approaches the performance of the GLRT $OS = 2$ detector (the conjectured upper bound). As the oversampling factor is reduced, the detection performance decreases accordingly. As interval sizes get larger, thermal noise is less filtered. The detector ends up comparing more noise with the same amount of signal.

The characteristics of each detector are showcased using Receiver Operating Characteristic (ROC) curves for SNR=0dB (Figure 6.9) and SNR=-5dB (Figure 6.10) on the same $N_t = 32$ setup. A detection decision is actually the result of a compromise between false alarm rate (α) and test power (the percentage of detected signals). A ROC curve depicts the detector test power as a function of the false alarm. The best detection performance is achieved by the detector whose ROC curve lies above the others (best percentage of detected signals for a given false alarm rate). From Figures 6.9 and 6.10, the naive approach is the worst and the GLRT with $OS = 2$ put an upper bound, followed closely by the interval-based approaches with $OS = 2$ and $OS = 1$. Those conclusions are in line

with the interpretations of Figure 6.8.

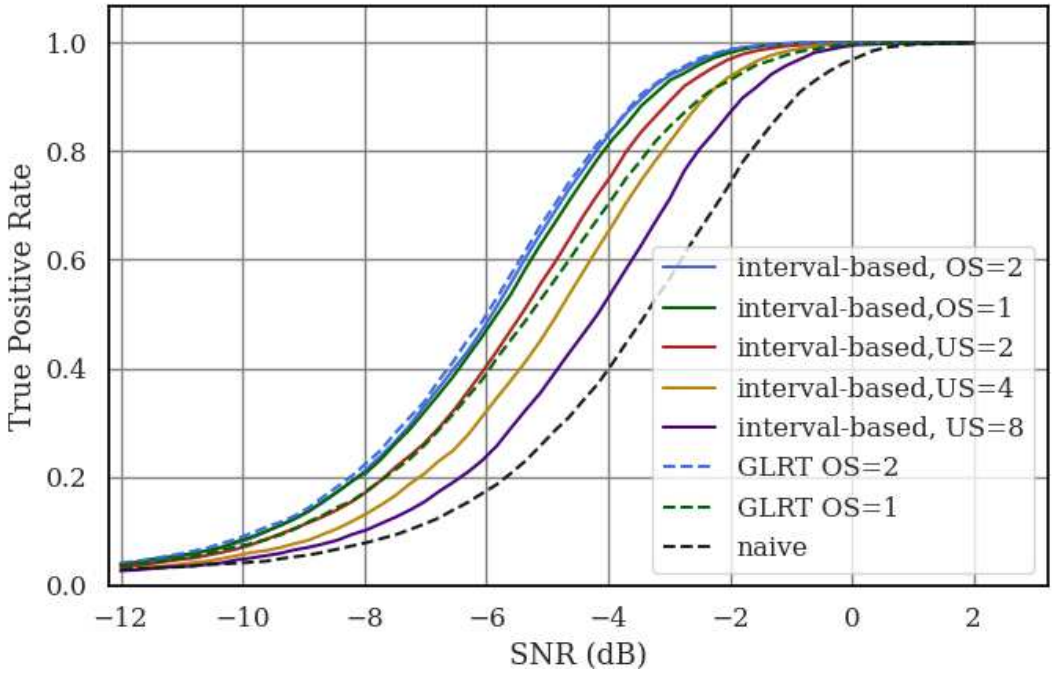

Figure 6.8 – Detection performance, 1% false alarm, 32 antennas.

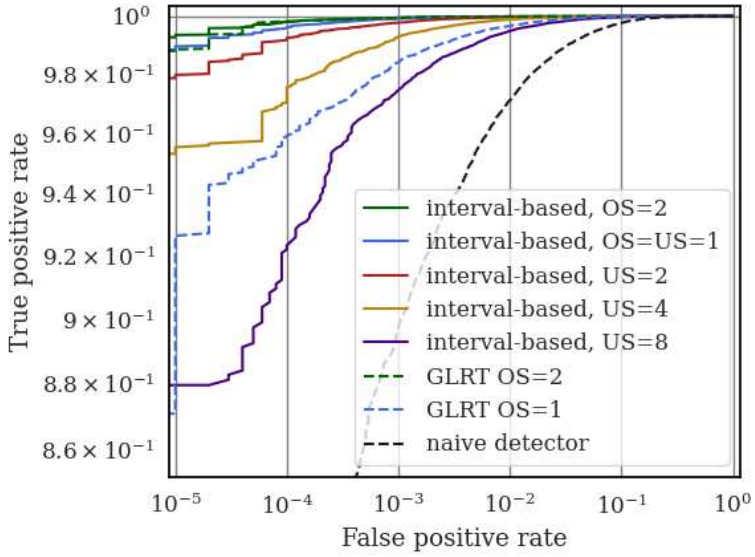

Figure 6.9 – ROC curves of the receivers encountered in the previous sections at a 0dB SNR.

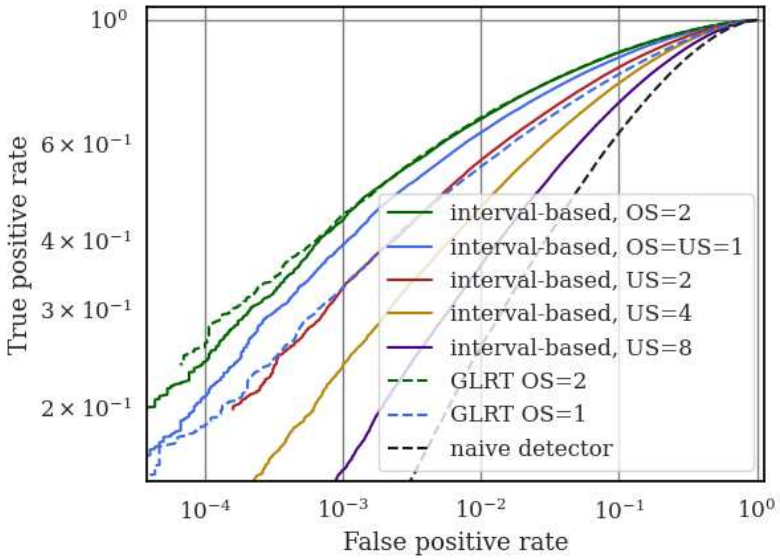

Figure 6.10 – ROC curves of the receivers encountered in the previous sections at a -5dB SNR.

6.4 The atomic ray estimation problem

In this section we solve the single specular path estimation problem. The specular ray channel model is described and analyzed in Section 6.4.1. As previously suggested in the literature [17, 16, 18], we propose a two-stage estimation technique. From the intermediate results of the detection decision presented in Section 6.3, a first coarse estimate is derived. It is used to initialize a numerical optimization technique that results in the fine estimate.

6.4.1 Analytical bounds

The singular specular ray channel is a particular simplified case of the multipath channel presented in Equation (1.7). It can be stated as

$$\mathbf{h}(\theta_{tx}) = \sqrt{N_t}\beta\mathbf{e_t}^*(\theta_{tx}) \tag{6.10}$$

where β is the ray gain and θ_{tx} is the ray DOD. In this section, the following steering vector notation is adopted

$$\mathbf{e_t}(\theta_{tx}) = \frac{1}{\sqrt{N_t}}\left[\mathrm{e}^{-2\pi j(-(N_t-1)/2)\frac{\|\vec{a}_{tx}\|\cos(\theta_{tx})}{\lambda}} \quad \dots \quad \mathrm{e}^{-2\pi j((N_t-1)/2)\frac{\|\vec{a}_{tx}\|\cos(\theta_{tx})}{\lambda}}\right].$$

It assumes that the centroid of the array is taken as origin point of the antenna positions in the array as featured in Section 1.3.1. This is done to simplify the CRLB evaluation, as in paper [1]. The focus is placed upon the angle of departure θ_{tx} as the goal is to optimize the ray estimation step. The ray gain β is considered as an interfering parameter. As the steering vector is a function of the intermediate variable $t = \cos(\theta_{tx})$, all future derivations involves t instead of θ_{tx}. This substitution leads to a non-uniform grid in the angle domain which has already been studied in paper [14].

The received symbols are given by

$$\mathbf{y} = \mathbf{X_t}\mathbf{h}(t) + \mathbf{n}$$

where $\mathbf{X_t}$ are the known transmitted pilot symbols, also called the *sensing matrix*. Then the received samples follow the complex Gaussian distribution

$$\mathbf{y} \sim \mathcal{CN}\left(\mathbf{X_t}\mathbf{h}(t), \sigma_n^2\mathbf{I}\right) \tag{6.11}$$

characterized by its mean $\mathbf{X_t h}(t)$ and the noise covariance identity matrix.

From the paper [1], the Fisher information for parameters β and t takes the form

$$\mathbf{I} = \begin{bmatrix} I_{|\beta|}0 & 0 & \\ 0 & I_{\arg(\beta)} & 0 \\ 0 & 0 & I_t \end{bmatrix}.$$

The CRLB is a lower bound on the variance of any practical estimator of an unknown parameter. Such bound is interesting as it enables theoretical studies (such as Figure 6.6) and offers a baseline for assessing the efficiency of any practical estimator. To this end, the derivation of the CRLB on parameter t, given as the inverse of I_t is required. In our case, we assume that $\mathbf{X_t}$ is unitary. In this case, the Fisher information I_t is given as

$$I_t = \frac{2N_t|\beta|^2}{\sigma_n^2} \left(\frac{\partial \mathbf{e_t}(t)}{\partial t} \right)^H \frac{\partial \mathbf{e_t}(t)}{\partial t}$$

where the derivative of $\mathbf{e_t}(t)$ is given by

$$\frac{\partial \mathbf{e_t}(t)}{\partial t} = \mathbf{D_t e_t(t)}$$

where $\mathbf{D_t}$ is a diagonal matrix that contains the exponents of the steering vector as

$$\mathbf{D_t} = \text{diag}([-2\pi j \left(-(N_t-1)/2\right) \frac{\|\vec{a}_{tx}\|}{\lambda}, \cdots, -2\pi j \left((N_t-1)/2\right) \frac{\|\vec{a}_{tx}\|}{\lambda}]).$$

A generalization of this bound for any antenna configuration is given in the paper [1].

6.4.2 Coarse estimation methods

The coarse estimate is obtained directly from the detection phase intermediate results 6.3. The index j_{opt} provides a convenient coarse estimator, namely the center of the j^{th} interval $\hat{t}^0 = \bar{t}_{j_{opt}}$. As previously stated in Equation (6.3), the variance of this simple estimator is given by

$$\mathbb{V}ar\left\{\hat{t}^0\right\} = \frac{\Delta_t^2}{12}$$

where Δ_t denotes the width of the search interval. This expression comes from the variance of a random variable uniformly distributed over an interval of length Δ_t. This estimator is comparable to the state of the art coarse grid search already presented in Section 6.2.2.

We can however derive a more accurate coarse estimator from the intermediary results of the detection phase. Indeed, using the interval-based detector, the responses of the individual windowed correlators (Figure 6.7) depend on t. We can leverage the intermediary results corresponding to the j_{opt} interval (where the ray was detected) to derive a novel estimation method. This technique works well and can achieve the CRLB when the channel consists only of one ray. However its performance decreases when considering multiple rays. To be integrated in OMP, a refinement step is required.

6.4.3 Fine estimation methods

The coarse estimate $\hat{t}^0$ obtained previously can be used as an initialization point for a numerical optimization procedure, to obtain a finer estimate. We focused our attention on the Newton-Raphson technique. However the raw likelihood function (displayed on Figure 6.4) features a dirac-like behavior at the vicinity of its maximum, which isn't a practical pattern for numerical optimization. That does not matter much if the initialization point is close enough to the actual direction of departure t. However when using $OS < 2$ and the first coarse estimation technique, $\hat{t}^0$ might fall too far from the peak leading to poor optimization conditioning and slow convergence time. To mitigate this potential issue, we introduce a modified likelihood function which is convolved with a kernel function p^{tx}

$$L(\mathbf{y}, \bar{t}) = \int_{-\infty}^{\infty} p^{tx}(t - \bar{t}) p(\mathbf{y}, t) dt. \tag{6.12}$$

We need to find a kernel function that smooths the likelihood and enables fast convergence towards the maximum to solve efficiently the optimization problem

$$\bar{t}_{opt} = \underset{\bar{t} \in [-1,1]}{\operatorname{argmax}} \; L(\mathbf{y}, \bar{t}). \tag{6.13}$$

The Newton-Raphson procedure is optimal for quadratic curves. However as the number of antennas increase the raw likelihood function $p(\mathbf{y}, t)$ behaves like a dirac at the DOD t. We choose an inverted parabola $p^{tx}(t)$ centered at 0 as the convolution kernel so that the modified likelihood behaves like a quadratic function that reaches its maximum at 0.

$$p^{tx}(t) = (t - \Delta_t/2)(t + \Delta_t/2) \mathbb{1}_{[-\Delta_t/2, \Delta_t/2]}(t)$$

Theorem. *Equation (6.13) can be approximated by the simpler optimization problem*

$$\bar{t}_{opt} = \underset{\bar{t}\in[-1,1]}{\operatorname{argmax}}\ f(\mathbf{y}, \bar{t}). \tag{6.14}$$

The novel function to optimize f is given as

$$f(\mathbf{z}, t) = \mathbf{z}^H \mathbf{\Delta}^H(t)\mathbf{R_0}(\Delta_t)\mathbf{\Delta}(t)\mathbf{z}. \tag{6.15}$$

where $\mathbf{z} = \mathbf{X_t}^H\mathbf{y}$ and the expression of $\mathbf{\Delta}(\bar{t})$ was already given in Equation (6.7) and the real-valued matrix $\mathbf{R_0}$ is given as

$$\begin{aligned}[\mathbf{R_0}]_{k,l}(\Delta_t) =& \frac{2}{(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})^3}\Bigg(\\ & \Delta_t(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})\cos\left(\frac{\Delta_t}{2}(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})\right). \\ & -2\sin\left(\frac{\Delta_t}{2}(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})\right)\Bigg)\end{aligned}$$

This result is proven in Annex D.2.

$\mathbf{R_0}$ only depends on the width Δ_t of the quadratic convolution kernel $p^{tx}(t)$. It is a real valued symmetric Toeplitz matrix. The first and second derivatives of Equation (6.15) are given by

$$\begin{cases}\frac{\partial f(\mathbf{z},t)}{\partial t} = \mathbf{z}^H\mathbf{\Delta}^H(t)(\mathbf{D}_{tx}\odot\mathbf{R_0})\mathbf{\Delta}(t)\mathbf{z} \\ \frac{\partial^2 f(\mathbf{z},t)}{\partial t^2} = \mathbf{z}^H\mathbf{\Delta}^H(t)(\mathbf{D}_{tx}\odot\mathbf{D}_{tx}\odot\mathbf{R_0})\mathbf{\Delta}(t)\mathbf{z}\end{cases}$$

The matrices $\mathbf{D}_{tx}\odot\mathbf{R_0}$ and $\mathbf{D}_{tx}\odot\mathbf{D}_{tx}\odot\mathbf{R_0}$ are diagonalizable. We denote λ'_k and λ''_k their respective eigenvalues and $\mathbf{r'_k}$ and $\mathbf{r''_k}$ the corresponding eigenvectors. The derivatives of Equation (6.15) are given as

$$f^{(1)}(\mathbf{z}, t) = \frac{\partial f(\mathbf{z}, t)}{\partial t} = \sum_{k=1}^{N_t}\lambda'_k|\mathbf{y}^H\mathbf{X_t}(\mathbf{r'_k}\odot\mathbf{e_t}^*(t))|^2,$$

$$f^{(2)}(\mathbf{z}, t) = \frac{\partial^2 f(\mathbf{z}, t)}{\partial t^2} = \sum_{k=1}^{N_t}\lambda''_k|\mathbf{y}^H\mathbf{X_t}(\mathbf{r''_k}\odot\mathbf{e_t}^*(t))|^2.$$

The coarse estimate $\hat{t}^0$ is refined by a Newton-Raphson optimization step to obtain

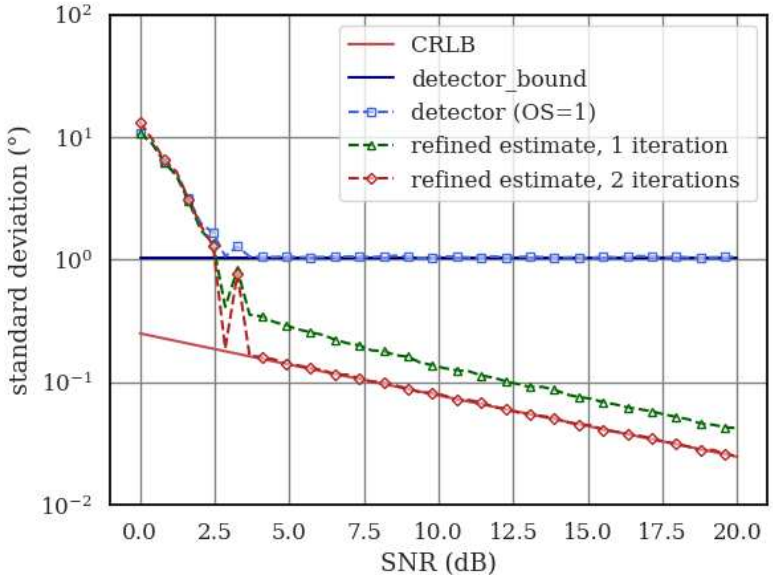

Figure 6.11 – Angle of departure estimation on a 8 antenna array ULA MISO channel. Comparison between the empirical estimator standard deviation and the Cramér-Rao Lower Bound.

the fine estimate $\hat{t}^1$

$$\hat{t}^1 = \hat{t}^0 - \frac{f^{(1)}(\mathbf{z}, \hat{t}^0)}{f^{(2)}(\mathbf{z}, \hat{t}^0)}$$

This refinement step can be iteratively repeated until the desired precision is obtained. It can eventually be replaced by a gradient descent or by a Levenberg-Marquardt optimization if function f is not concave.

We illustrate this two-stage estimation algorithm by a numerical evaluation of its performance on a MISO channel with $N_t = 8$ transmit antennas on Figure 6.11. The estimated angle standard deviation (square root of the variance) is compared to the optimal CRLB. As expected, the coarse estimate obtained from the detector (first coarse estimation technique, center of the most likely interval) is inaccurate compared to the theoretical CRLB. From this initialization point, only two iterations of the fine estimation technique are required to achieve the CRLB.

6.4.4 Generalization

In the previous section we presented a fast angle of departure estimation technique. Those techniques can be generalized up to the wideband MIMO case with joint angle of

arrival $r = \cos(\theta_{rx})$, angle of departure $t = \cos(\theta_{tx})$ and delay τ estimation. As previously showed, all dimensions can be treated using the same approach.

In the detection phase, we proposed a novel detector based on a subdivision of intervals. When adding other dimensions, the search space gets subdivided into rectangles of a $2D$ grid then cubic cells of a $3D$ mesh.

We proposed two techniques for coarse estimation. The foremost can be extended directly as the coarse estimates corresponds to the centroid of the most probable rectangle (or cubic depending on the dimension) cell. The extension of the second approach is a bit more complex to implement as it requires a multi-parameter function fit and has not been studied in this thesis for lack of time.

Finally the fine estimation method is based on numerical optimization tools, from which straightforward multivariate extensions already exists.

6.5 Integration into the OMP algorithm

We integrated this work within the classical OMP algorithm. In particular, from the flowchart presented in Figure 6.3, we modified the Specular Ray estimation step and the Stopping criterion. The latter step is replaced by the outcome of the detector presented in Section 6.3.2.

Most of current State of the art implementations of the OMP algorithm uses either heuristics or residual norm-based criteria [16, 17, 18]. Most of the performance gain of our approach comes from this new criterion.

We evaluated the performance of the modified algorithm on a communication system with 32 transmit antennas (ULA array). The channels where generated using the NYUSIM [6, 7] channel model (parameters UMi, Urban Microcell). The first performance indicator is the percentage of cases where at least one ray is extracted for both the residual norm-based and the advanced (ray-based) detectors. This channel detection performance is depicted on Figure 6.12. The novel approach offers a substantial 2dB detection gain at low SNR (at -5dB) that narrows as the SNR increases.

Then on Figure 6.13 we evaluate the MSE for both the original and the modified algorithms. We tested multiple oversampling factors $OS \in \{1, 2, 4, 6\}$ for the classical OMP algorithm. Then we evaluated the MSE of the modified algorithm assuming $OS = 1$. In the low SNR regime, the novel estimation method outperforms the classical one at low SNR. This gain is mainly due to the modified stopping criterion that enables the algorithm

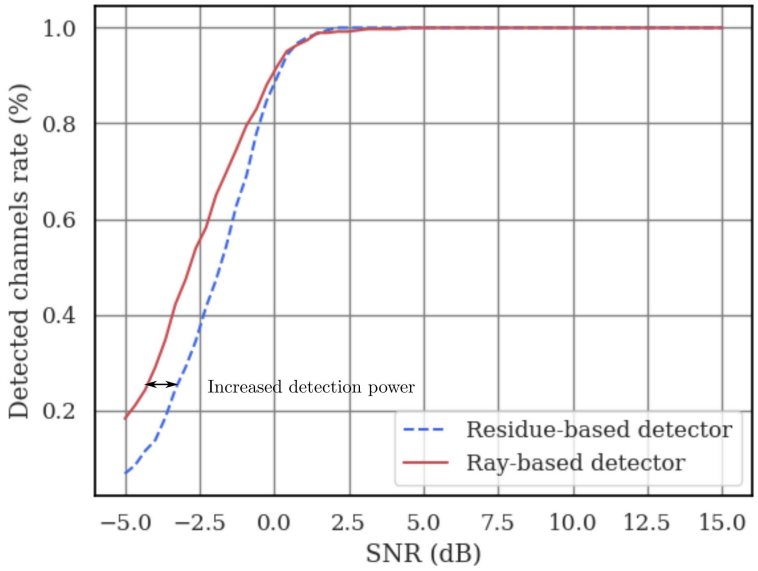

Figure 6.12 – Proportion of channels where at least one ray was detected, novel detector against state of the art residue based detector, $N_t = 32$.

to safely extract rays more deeply embedded in noise than the state of the art version.

Finally, note that at high SNR, the modified OMP algorithm achieve a performance similar to its $OS = 6$ original version counterpart while keeping a reasonable $OS = 1$ complexity. This is due to the fine estimation step proposed in Section 6.4.3.

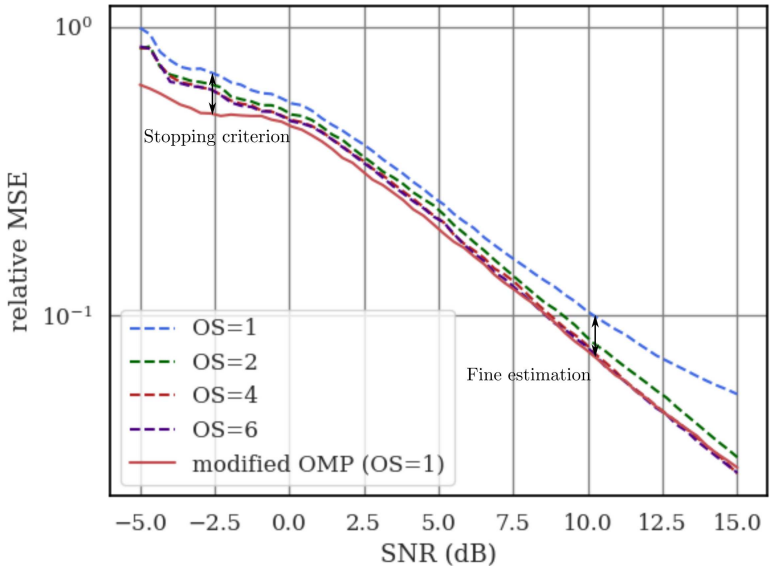

Figure 6.13 – Performance (MSE) of both state of the art OMP and the modified OMP, $N_t = 32$.

6.6 Conclusion

In this chapter we modified the classical OMP algorithm used for massive MIMO channel estimation to optimize two of its key steps : the ray estimation and the stopping criterion. We extracted and studied separately the detection and estimation of a single specular ray. We characterized the detector using ROC curves (fixed SNR) and the detection power at fixed false alarm rate. The estimation performance, evaluated using Monte-Carlo simulations, was compared to the theoretical CRLB. Finally this whole setup was integrated within the OMP algorithm, whose performance was characterized and compared with its original counterpart in the last section. The detection step provides a natural and efficient stopping criterion which was lacking to the algorithm. Our modifications can also mitigate the complexity of traditional methods relying on dense grid search by enabling sparser grids without creating gaps in the search space.

BIBLIOGRAPHY

[1] Luc Le Magoarou and Stéphane Paquelet, « Parametric channel estimation for massive MIMO », *in*: *arXiv preprint arXiv:1710.08214* (2017).

[2] Luc Le Magoarou and Stéphane Paquelet, « Performance of MIMO channel estimation with a physical model », *in*: *arXiv:1902.07031 [cs, eess]* (Feb. 2019).

[3] Mustafa Riza Akdeniz et al., « Millimeter Wave Channel Modeling and Cellular Capacity Evaluation », *in*: *IEEE J. Select. Areas Commun.* (June 2014).

[4] M. Shehata et al., « Angular Based Beamforming and Power Allocation Framework in a Multi-User Millimeter-Wave Massive MIMO System », *in*: *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, Porto: IEEE, June 2018.

[5] Z. Muhi-Eldeen, L.P. Ivrissimtzis, and M. Al-Nuaimi, « Modelling and measurements of millimetre wavelength propagation in urban environments », *in*: *IET Microw. Antennas Propag.* (2010).

[6] Mathew K. Samimi and Theodore S. Rappaport, « 3-D Millimeter-Wave Statistical Channel Model for 5G Wireless System Design », *in*: *IEEE Trans. Microwave Theory Techn.* (July 2016).

[7] Mathew K. Samimi et al., « 28 GHz Millimeter-Wave Ultrawideband Small-Scale Fading Models in Wireless Channels », *in*: IEEE, May 2016.

[8] M.D. Zoltowski and Kainam Thomas Wong, « Self-initiating MUSIC-based direction finding and polarization estimation in spatio-polarizational beamspace », *in*: *IEEE Trans. Antennas Propagat.* (Aug. 2000).

[9] A. Richter et al., « Joint estimation of DoD, time-delay, and DoA for high-resolution channel sounding », *in*: *VTC2000-Spring. 2000 IEEE 51st Vehicular Technology Conference Proceedings (Cat. No.00CH37026)*, Tokyo, Japan: IEEE, 2000.

[10] Susana Mota et al., « Estimation of the Radio Channel Parameters using the SAGE Algorithm », *in*: (2010).

[11] Liao Anping, Xie Jiaxin, and Yang Xaiobo, « The Exact Recovery of Sparse Signals Via Orthogonal Matching Pursuit », *in*: *JCM* (June 2016).

[12] Javier Rodriguez-Fernandez et al., « Exploiting Common Sparsity for Frequency-Domain Wideband Channel Estimation at mmWave », *in*: *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore: IEEE, Dec. 2017.

[13] Javier Rodríguez-Fernández et al., « Frequency-domain Compressive Channel Estimation for Frequency-Selective Hybrid mmWave MIMO Systems », *in*: *arXiv:1704.08572 [cs, math]* (Apr. 2017).

[14] Junho Lee, Gye-Tae Gil, and Yong H. Lee, « Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications », *in*: *IEEE Trans. Commun.* (June 2016).

[15] Die Hu and Lianghua He, « Channel Estimation for FDD Massive MIMO OFDM Systems », *in*: *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Toronto, ON: IEEE, Sept. 2017.

[16] Zhinus Marzi, Dinesh Ramasamy, and Upamanyu Madhow, « Compressive channel estimation and tracking for large arrays in mm wave picocells », *in*: *IEEE J. Sel. Top. Signal Process.* (Apr. 2016).

[17] Junho Lee, Gye-Tae Gil, and Yong H. Lee, « Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications », *in*: *2014 IEEE Global Communications Conference*, Austin, TX, USA: IEEE, Dec. 2014.

[18] Lorenz Weiland et al., « OMP with Grid-Less Refinement Steps for Compressive mmWave MIMO Channel Estimation », *in*: *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Sheffield: IEEE, July 2018.

[19] Steven M. Kay, *Fundamentals of statistical signal processing*, Prentice Hall signal processing series, Englewood Cliffs, N.J: Prentice-Hall PTR, 1993.

Conclusion and perspectives

The techniques used in SISO and small-scale MIMO communications do not scale well when increasing the number of antennas. As the envisioned sizes of antenna arrays used for wireless access networks increase, dedicated signal processing algorithms that relies on the angular properties of the channel have to be employed. To this end, we dedicated the first part of this manuscript to propagation channel modeling and characterization while the second part exploits those results for precoding and channel estimation.

Channel modeling and characterization

In this first part, we set up the ray-based (and cluster-based) channel model that is used in all the other chapters of this manuscript. In particular, we used extensively the well-known Saleh-Valenzuela stochastic channel model. We introduced the different dimensions of the propagation which are studied in the manuscript (direction of departure, direction of arrival, delay). Finally, we established the Kronecker product notations for the vectorized channel matrix which are used throughout the manuscript.

Then we introduced the concept of channel capacity, that is an upper bound of the achievable error-free data rate over a particular channel. In particular we studied the variations of the SNR of a MIMO channel in a multipath environment by characterizing analytically the *channel hardening* phenomenon. The originality of this contribution lies on the underlying physically motivated ray-based channel. It enables to separate the contributions from the antenna array shapes and the propagation channel properties.

Then we characterized by their first and second order statistics the properties of Saleh-Valenzuela clusters. Those results were not new, however our proofs were much simpler than the ones proposed in the state of the art. Then we search for an efficient way to represent those clusters. As previously mentioned, when the array sizes increase from small-scale MIMO up to the massive MIMO regime, dedicated signal processing have to be employed. In order to draw a boundary between both regimes, we evaluated how much coefficients are required to model the channel using a small-scale approach and a massive MIMO approach that leverages the angular sparsity. The channel switches from

the former to the latter regime at around $N_r N_t > 10$, which is quite small. Note that our analysis did not account for the angular spread of clusters, thus providing a lower bound on this shift. Such an analysis, including larger angular spreads, would be interesting to assess the boundary between both regimes for larger antenna arrays.

Finally we studied a cluster estimation and separation process. It differed from the state of the art on two main points. First it did not relied on heuristics to define the clusters. It leveraged the Saleh-Valenzuela cluster model introduced in the previous chapter. This enabled a theoretical analysis based on the Cramér-Rao bound of the cluster parameters. Most state of the art techniques rely on two stage algorithms (information concentration followed by cluster extraction). Our process did not necessarily relied on a first projection step, but such a preprocessing step could still be employed to mitigate the calculation complexity. The cluster estimation problem consisted into two subproblems (cluster parameter estimation and cluster separation with known parameters) that were linked to each other. The proposed technique employed an EM algorithm to estimate jointly cluster parameters and clusters themselves from received samples.

As it is, this technique is limited to the frequency domain and extension to the angular domains (transmit and receive) is discussed at the end of Chapter 4. Some other interesting perspectives would include defining dedicated and efficient estimators for the cluster parameters, finding the proper regularization for the EM maximization step and defining a tailored detection criterion. It would also be useful to separate the LOS ray from other clusters, as its characteristics (no delay or angular spread, very first ray in the delay domain) are clearly specific.

Model exploitation for massive MIMO digital processing

In order to design the precoder that maximizes the capacity, the channel state information (CSI) has to be recovered at the transmitter. In FDD, the CSI is usually estimated at the receiver and fed back to the transmitter on the uplink channel. Recently, researchers investigated extrapolating the CSI from the uplink band to the downlink transmission band. To this end, we investigated the efficiency of a frequency domain cluster extrapolation process based on its known second order statistics (derived in the previous part). We evaluated the extrapolation process efficiency with both the MSE and the RBG (directly linked to the capacity) criteria. This is useful to assess the estimated channel reliability

and make the decision to precode with this extrapolated CSI or fall back to a classical feedback strategy.

An interesting perspective would consist in testing this technique using the output of the cluster separation algorithm presented above. This could not be achieved in this thesis as the cluster separation process is limited to the frequency domain and does not include the angular domains.

Last, but not least, we studied some key aspects of the OMP-based millimeter wave massive MIMO channel estimation algorithm. In particular we provided a novel stopping criterion that works well at low SNR as well as a simplified specular ray estimator. Those enhancements were integrated within the OMP channel estimation algorithm framework developed at b<>com and compared with the former approach.

Testing this algorithm on real measured channels is the following step, to verify the channel properties and the simulated algorithm behavior. This paves the way to an efficient hardware implementation for real time processing within a millimeter wave MIMO demonstrator.

Appendix A

Channel Hardening measure for ray-based models

In this appendix we prove Formula (2.11). For the sake of simplicity, an intermediary matrix $\mathbf{A}$ is introduced. It is defined by

$$[\mathbf{A}]_{p,p'} = \begin{cases} 2|\gamma_{p,p'}|\cos(\phi_{p,p'}) & \text{if } p \neq p' \\ 1 & \text{if } p = p' \end{cases}$$

with $\phi_{p,p'} = \arg(\beta_p^* \beta_{p'} \gamma_{p,p'})$ the whole channel phase dependence and $\gamma_{p,p'}$ defined by Equation 2.2.2. $\|\mathbf{H}\|_F^2$ can be written using a quadratic form with vector $\boldsymbol{\beta} = [|\beta_1|, \cdots, |\beta_P|]^T$ and matrix $\mathbf{A}$, which can be decomposed into two terms $\mathbf{I}$ (identity) and $\mathbf{J}$

$$\frac{\|\mathbf{H}\|_F^2}{N_t N_r} = \boldsymbol{\beta}^T \mathbf{A} \boldsymbol{\beta} = \boldsymbol{\beta}^T \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{J} \boldsymbol{\beta}$$

where $\mathbf{J} = \mathbf{A} - \mathbf{I}$. $\mathbb{E}\{\mathbf{J}\} = \mathbf{0}$ so:

$$\frac{\mathbb{E}\{\|\mathbf{H}\|_F^4\}}{(N_t N_r)^2} = \mathbb{E}\left\{\|\boldsymbol{\beta}\|^4\right\} + \mathbb{E}\left\{(\boldsymbol{\beta}^T \mathbf{J} \boldsymbol{\beta})^2\right\}.$$

The ray independence properties yields the following weighted sum of coupled ray powers

$$\mathbb{E}\left\{(\boldsymbol{\beta}^T \mathbf{J} \boldsymbol{\beta})^2\right\} = \sum_{p \neq p'} \mathbb{E}\left\{|\beta_p|^2 |\beta_{p'}|^2\right\} \mathbb{E}\left\{[\mathbf{J}]_{p,p'}^2\right\}.$$

Considering i.i.d. rays, all the weights $\mathbb{E}\left\{[\mathbf{J}]_{p,p'}^2\right\}$ are identical. Using the weights notations introduced in (2.10) and the definition of the 4-norm yields the second order moment $\mathbb{E}\{\|\mathbf{H}\|_F^4\}$. We divide this result by the expectation (2.9) to derive the channel hardening measure (2.11).

Appendix B

Saleh-Valenzuela autocorrelation derivation

B.1 Cluster channel gain

In (3.4) the phases of the complex gains $\beta_{p,q}$ are all independent resulting in $\mathbb{E}\left\{\beta_{p_1,q}\beta^*_{p_2,q}\right\} = \delta_{p_1,p_2}\mathbb{E}\left\{|\beta_{p_1,q}\beta_{p_2,q}|\right\}$. The autocorrelation is then given by

$$R_{c_q}(f,f') = \sum_{p=1}^{\infty}\mathbb{E}\left\{|\beta_{p,q}|^2\,\mathrm{e}^{-2\pi j(f-f')\tau_{p,q}}\right\}.$$

The random process is Wide Sense Stationary (WSS). We set $\Delta f = f' - f$. Then using conditional expectation on $\tau_{p,q}$ and the exponential decay $\mathbb{E}\left\{|\beta_{p,q}|^2\middle|\tau_{p,q}\right\} = \mathbb{E}\left\{|\beta_{1,q}|^2\right\}\mathrm{e}^{-\tau_{p,q}/\gamma}$ we get

$$R_{c_q}(\Delta f) = \mathbb{E}\left\{|\beta_{1,q}|^2\right\}\sum_{p=1}^{\infty}\mathbb{E}_{\tau_{p,q}}\left\{\mathrm{e}^{(2\pi j\Delta f - 1/\gamma)\tau_{p,q}}\right\}. \tag{B.1}$$

Then using a recurrence over p we can prove that

$$\mathbb{E}_{\tau_{p,q}}\left\{\mathrm{e}^{(2\pi j\Delta f-1/\gamma)\tau_{p,q}}\right\} = \lambda^{p-1}\frac{1}{(\frac{1}{\gamma}+\lambda-2\pi j\Delta f)^{p-1}}.$$

Equation (B.1) is a geometric series whose ratio $r = \frac{1}{\frac{1}{\lambda\gamma}+1-2\pi j\Delta f/\lambda}$ always verify $|r|<1$ so that the series converges to

$$R_{cc}(\Delta f) = \mathbb{E}\left\{|\beta_{1,q}|^2\right\}\left(1+\frac{\lambda\gamma}{1-2\pi j\gamma\Delta f}\right).$$

B.2 Saleh-Valenzuela channel gain

Still using the independence of the phases of the complex gains in (3.9), we get a simplified expression. We find that this channel is also WSS. We obtain the expression

$$R_H(\Delta f) = \sum_{q=1}^{\infty}\sum_{p=1}^{\infty} \mathbb{E}\left\{|\beta_{p,q}|^2 \, \mathrm{e}^{2\pi j \Delta f(\tau_{p,q}+\tau_q)}\right\}. \tag{B.2}$$

Using a conditional expectation on the $\tau_{p,q}$ and the τ_q and using the double exponential decay formula of the Saleh-Valenzuela channel $\mathbb{E}\left\{|\beta_{p,q}|^2 \middle| \tau_{p,q}\right\} = \mathrm{e}^{-\tau_q/\Gamma} \, \mathrm{e}^{-\tau_{p,q}/\gamma}$, we get an expression similar to (B.1)

$$R_H(\Delta f) = \sum_{q=1}^{\infty}\sum_{p=1}^{\infty} \mathbb{E}\left\{\mathrm{e}^{-(1/\Gamma - 2\pi j \Delta f)\tau_q} \, \mathrm{e}^{-(1/\gamma - 2\pi j \Delta f)\tau_{p,q}}\right\}.$$

Because the random variables τ_q and $\tau_{p,q}$ are independent, we can split the expectation. Using the result of (B.1) on both parts, the equation can be rewritten as the product of two geometric series yielding the result (3.10).

B.3 Variances of time domain coefficients

The variance of the time domain coefficient h_n is

$$\mathbb{E}\left\{h_n h_n^*\right\} = \frac{1}{N_f^2} \sum_{k_1=0}^{N_f-1}\sum_{k_2=0}^{N_f-1} \mathbb{E}\left\{H[k_1]H[k_2]^*\right\} \mathrm{e}^{-2\pi j (k_2-k_1)n/N_f}.$$

Where we can recognize the autocorrelation $\mathbb{E}\left\{H[k_1]H[k_2]^*\right\} = R_{H,H}[k_2 - k_1]$. We use the autocorrelation expression from (B.2) with $\Delta f = (k_2 - k_1)\Delta B$

$$\mathbb{E}\left\{H[k_1]H[k_2]^*\right\} = \sum_{q=1}^{\infty}\sum_{p=1}^{\infty} \mathbb{E}\Big\{|\beta_{p,q}|^2 \, \mathrm{e}^{2\pi j (k_2-k_1)\Delta B(\tau_{p,q}+\tau_q)}\Big\}.$$

We insert this expression into the variance formula and get

$$\mathbb{E}\left\{h_n h_n^*\right\} = \frac{1}{N_f^2} \sum_{q=1}^{\infty} \sum_{p=1}^{\infty} \mathbb{E}\Bigg\{ |\beta_{p,q}|^2 \sum_{k_1=0}^{N_f-1} \sum_{k_2=0}^{N_f-1} \mathrm{e}^{2\pi j (k_2-k_1)\Delta B(\tau_{p,q}+\tau_q - n/(N_f \Delta B))} \Bigg\}$$

where the sums equals $D_{N_f}^2\left(\pi \Delta B(\tau_{p,q} + \tau_q - n/B)\right)$ resulting in

$$\mathbb{E}\left\{h_n h_n^*\right\} = \sum_{q=1}^{\infty} \sum_{p=1}^{\infty} \mathbb{E}\left\{ |\beta_{p,q}|^2 D_{N_f}^2(\pi \Delta B(\tau_{p,q} + \tau_q - n/B)) \right\}.$$

We introduce the continuous function defined for $x \in \mathbb{R}$

$$f(x) = \sum_{q=1}^{\infty} \sum_{p=1}^{\infty} \mathbb{E}\left\{ |\beta_{p,q}|^2 D_{N_f}^2(\pi \Delta B(\tau_{p,q} + \tau_q - x/B)) \right\}.$$

After a first Inverse Fourier Transform and substituting $y = B\tau_{p,q} + B\tau_q - x$

$$\sum_{q=1}^{\infty} \sum_{p=1}^{\infty} \mathbb{E}\left\{ |\beta_{p,q}|^2 \mathrm{e}^{2\pi j \nu (\tau_{p,q}+\tau_q)} \right\} \int_{-\infty}^{\infty} D_{N_f}^2(\pi y/N_f)\, \mathrm{e}^{-2\pi j \nu y/B}\, \mathrm{d}\, y.$$

The first line matches the autocorrelation function of the channel $R(\nu)$. The second part is the Fourier Transform of the squared Dirichlet kernel. We retrieve the function f using a FT:

$$f(x) = \int_{-\infty}^{\infty} S_H(\tau) D_{N_f}^2(\pi(\tau B - x)/N_f)\, \mathrm{d}\, \tau.$$

Applying this formula for $x = n$, we get the expression.

Appendix C

Cluster Separation

C.1 Cluster Fisher Information matrix

The structure of the Fisher Information matrix for a single cluster parametrized by $\boldsymbol{\theta_q} = [\tau_q, P_q, \gamma_q, \lambda_q]$ is given as

$$\mathbf{I}(\tau_q, P_q, \gamma_q, \lambda_q) = \begin{bmatrix} I_{\tau_q} & I_{\tau_q,P_q} & I_{\tau_q,\gamma_q} & I_{\tau_q,\lambda_q} \\ I_{P_q,\tau_q} & I_{P_q} & I_{P_q,\gamma_q} & I_{P_q,\lambda_q} \\ I_{\gamma_q,\tau_q} & I_{\gamma_q,P_q} & I_{\gamma_q} & I_{\gamma_q,\lambda_q} \\ I_{\lambda_q,\tau_q} & I_{\lambda_q,P_q} & I_{\lambda_q,\gamma_q} & I_{\lambda_q} \end{bmatrix}$$

where each coefficient is evaluated using the formula

$$\mathbf{I}_{\theta_i,\theta_j} = \mathrm{Tr}\left\{ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \right\}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Delta}_{\mathbf{x}} \boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\tau_q, P_q, \gamma_q, \lambda_q) \boldsymbol{\Delta}_{\mathbf{x}}{}^H + \sigma_n^2 \mathbf{I}$.

Assuming that $\boldsymbol{\Delta}_{\mathbf{x}} \boldsymbol{\Delta}_{\mathbf{x}}{}^H = \mathbf{I}$, the coefficients are evaluated as

$$\mathbf{I}_{\theta_i,\theta_j} = \mathrm{Tr}\left\{ \left(\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q}) + \sigma_n^2 \mathbf{I} \right)^{-1} \frac{\partial \boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})}{\partial \theta_i} \left(\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q}) + \sigma_n^2 \mathbf{I} \right)^{-1} \frac{\partial \boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})}{\partial \theta_j} \right\}.$$

We evaluate the terms I_{P_q} and I_{τ_q,P_q}.

Evaluation of I_{P_q}:

The derivative of matrix $\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})$ according to P_q is given by

$$\frac{\partial \boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\boldsymbol{\theta_q})}{\partial P_q} = \boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\tau_q, 1, \gamma_q, \lambda_q).$$

We set $\mathbf{U}\boldsymbol{\eta}\mathbf{U}^H$ the eigendecomposition of $\boldsymbol{\Sigma}^{\mathbf{f}}_{\mathbf{q}}(\tau_q, 1, \gamma_q, \lambda_q)$ (cluster with unit power). The Fisher Information I_{P_q} is given as

$$\begin{aligned}\mathbf{I}_{\theta_i,\theta_j} &= \mathrm{Tr}\left\{\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\right\}\\ &= \sum_{i=0}^{N_f-1}\frac{[\boldsymbol{\eta}]_i^2}{(P_q[\boldsymbol{\eta}]_i+\sigma_n^2)^2}\end{aligned}$$

Evaluation of I_{τ_q,P_q}: The cluster covariance matrix can be rewritten as

$$\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\tau_q,P_q,\gamma_q,\lambda_q)=\boldsymbol{\Delta}(\tau_q)\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(0,P_q,\gamma_q,\lambda_q)\boldsymbol{\Delta}^H(\tau_q)$$

where $[\boldsymbol{\Delta}(\tau_q)]_k=\mathrm{e}^{-2\pi jk\Delta_f\tau_q}$. The derivative of $\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})$ according to τ_q is given as

$$\begin{aligned}\frac{\partial\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})}{\partial\tau_q} &= \mathbf{D}_\tau\boldsymbol{\Delta}(\tau_q)\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(0,P_q,\gamma_q,\lambda_q)\boldsymbol{\Delta}^H(\tau_q)-\boldsymbol{\Delta}(\tau_q)\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(0,P_q,\gamma_q,\lambda_q)\boldsymbol{\Delta}^H(\tau_q)\mathbf{D}_\tau\\ &= \mathbf{D}_\tau\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})-\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})\mathbf{D}_\tau\end{aligned}$$

where $[\mathbf{D}(\tau)]_k=-2\pi jk\Delta_f$. The Fisher Information coefficient is given as

$$\begin{aligned}\mathbf{I}_{\tau_q,P_q} = &\ \mathrm{Tr}\left\{\left(\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})+\sigma_n^2\mathbf{I}\right)^{-1}\mathbf{D}_\tau\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})\left(\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\tau_q,1,\gamma_q,\lambda_q)\right\}\\ &-\mathrm{Tr}\left\{\left(\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})\mathbf{D}_\tau\left(\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\boldsymbol{\theta_q})+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{\mathbf{q}}^{\mathbf{f}}(\tau_q,1,\gamma_q,\lambda_q)\right\}\end{aligned}$$

Using the eigendecomposition, we have :

$$\begin{aligned}\mathbf{I}_{\tau_q,P_q} &= \mathrm{Tr}\left\{\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\mathbf{U}^H\mathbf{D}_\tau\mathbf{U}\boldsymbol{\eta}\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\right\}\\ &\quad-\mathrm{Tr}\left\{\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\mathbf{U}^H\mathbf{D}_\tau\mathbf{U}\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\right\}\\ &= \mathrm{Tr}\left\{\mathbf{U}^H\mathbf{D}_\tau\mathbf{U}\boldsymbol{\eta}\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\right\}\\ &\quad-\mathrm{Tr}\left\{\mathbf{U}^H\mathbf{D}_\tau\mathbf{U}\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\left(P_q\boldsymbol{\eta}+\sigma_n^2\mathbf{I}\right)^{-1}\boldsymbol{\eta}\right\}\end{aligned}$$

As diagonal matrices commute, $\mathbf{I}_{\tau_q,P_q}=0$. From the same calculation we get $\mathbf{I}_{P_q,\tau_q}=0$.

C.2 Cramér-Rao matrix properties

In this annex we present a way to interpret geometrically the Cramér-Rao bound. The vector parameter is denoted by $\boldsymbol{\theta}$. Using a practical estimator we obtain the estimate $\hat{\boldsymbol{\theta}}$.

The estimation error is characterized by its covariance matrix as

$$\mathbf{C_{est}} = \mathbb{E}\left\{(\hat{\boldsymbol{\theta}} - \mathbb{E}\left\{\hat{\boldsymbol{\theta}}\right\})(\hat{\boldsymbol{\theta}} - \mathbb{E}\left\{\hat{\boldsymbol{\theta}}\right\})^T\right\}$$

we assume that the estimator is unbiaised ($\mathbb{E}\left\{\hat{\boldsymbol{\theta}}\right\} = \boldsymbol{\theta}$). Then the Cramér-Rao Bound $\mathbf{C_{CR}}$ verifies for any estimator

$$\mathbf{C_{est}} - \mathbf{C_{CR}} \geq \mathbf{0} \tag{C.1}$$

Thus the eigenvalues of $\mathbf{C_{est}} - \mathbf{C_{CR}}$ are all positive. We interpret geometrically the relations between pairs of parameters (θ_i, θ_j). The 2×2 covariance matrix between estimated parameters $(\hat{\theta}_i, \hat{\theta}_j)$ is given by the following principal submatrix of $\mathbf{C_{est}}$ as

$$\mathbf{R_{est,i,j}} = \begin{bmatrix} [\mathbf{C_{est}}]_{i,i} & [\mathbf{C_{est}}]_{i,j} \\ [\mathbf{C_{est}}]_{j,i} & [\mathbf{C_{est}}]_{j,j} \end{bmatrix}$$

The Cramér-Rao corresponding covariance submatrix $\mathbf{R_{CR,i,j}}$ is obtained from $\mathbf{C_{CR}}$ as

$$\mathbf{R_{CR,i,j}} = \begin{bmatrix} [\mathbf{C_{CR}}]_{i,i} & [\mathbf{C_{CR}}]_{i,j} \\ [\mathbf{C_{CR}}]_{j,i} & [\mathbf{C_{CR}}]_{j,j} \end{bmatrix}$$

Using [7, (11), p. 73] and Equation (C.1), the eigenvalues of any principal submatrix $\mathbf{R_{est,i,j}} - \mathbf{R_{CR,i,j}}$ are always positive. Thus all those matrices are definite positive. This means that $\mathbf{R_{est,i,j}} \geq \mathbf{R_{CR,i,j}}$ where $\geq$ denotes the *partial semidefinite ordering* on symmetric matrices (see [8]). From [8, (9), p. 8.10] we get $\mathbf{R_{est,i,j}}^{-1} \leq \mathbf{R_{CR,i,j}}^{-1}$. For any 2D point $\mathbf{x}$ in the parameter space,

$$\mathbf{x}^T\mathbf{R_{est,i,j}}^{-1}\mathbf{x} \leq \mathbf{x}^T\mathbf{R_{CR,i,j}}^{-1}\mathbf{x} \tag{C.2}$$

with equality when $\mathbf{x} = \mathbf{0}$ or $\mathbf{R_{est,i,j}} = \mathbf{R_{CR,i,j}}$.

The parametric curve defined by $\mathbf{x}^T\mathbf{R_{est,i,j}}^{-1}\mathbf{x} = 1$ represents an ellipse in the $\mathbf{x}$ space as depicted in Figure C.1. We use this representation to interpret the Cramér-Rao Bound geometrically. On Figure 4.3 we represent the ellipses corresponding to both $\mathbf{R_{est,i,j}}$ (in black) and $\mathbf{R_{CR,i,j}}$ (in red).

The red ellipse $\mathcal{E}_{CR}$ in 4.3 that corresponds to the Cramér-Rao bound is given by the set of points $\mathbf{x}$ in the parameter space that verifies

$$\mathbf{x}^T\mathbf{R_{CR,i,j}}^{-1}\mathbf{x} = 1.$$

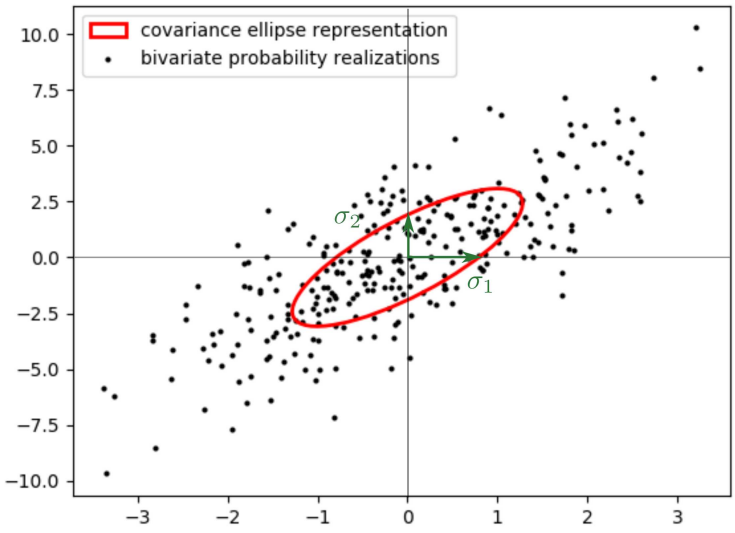

Figure C.1 – Ellipse representation of the covariance matrix of a distribution.

The black estimator ellipse $\mathcal{E}_{est}$ is given by the set of points $\mathbf{x}$ in the parameter space that verifies

$$\mathbf{x}^T {\mathbf{R}_{\mathbf{est,i,j}}}^{-1} \mathbf{x} = 1.$$

On the red ellipse $\mathcal{E}_{CR}$, the expression $\mathbf{x}^T {\mathbf{R}_{\mathbf{est,i,j}}}^{-1} \mathbf{x}$ is always inferior to 1 because of Equation (C.2). Thus the ellipse $\mathcal{E}_{CR}$ is inside $\mathcal{E}_{est}$.

C.3 Gaussian posterior distribution

Let $\mathbf{x}$ be a centered complex distributed Gaussian vector characterized by its covariance matrix $\mathbf{\Sigma_x}$. The probability density function of $\mathbf{x}$ is given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N_x} |\mathbf{\Sigma_x}|} \mathrm{e}^{-\mathbf{x}^H {\mathbf{\Sigma_x}}^{-1} \mathbf{x}}.$$

Let $\mathbf{y} = \mathbf{A}\mathbf{x}$. The vector $\mathbf{y}$ is then also a centered complex distributed Gaussian vector characterized by the covariance matrix $\mathbf{\Sigma_y} = \mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^H$. Its probability density function is given by

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{N_y} |\mathbf{\Sigma_y}|} \mathrm{e}^{-\mathbf{y}^H {\mathbf{\Sigma_y}}^{-1} \mathbf{y}}.$$

We evaluate the probability density distribution of the vector $\mathbf{x}$ given the Gaussian vector $\mathbf{y}$, denoted by $p(\mathbf{x}|\mathbf{y})$. From Bayes'law we have

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} \tag{C.3}$$

Note that the vector $\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}$ is given as

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\\ \mathbf{I}\end{bmatrix}\mathbf{x}$$

This vector is zero-mean Gaussian distributed and characterized by the block-diagonal covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix}\mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^H = \mathbf{\Sigma_y} & \mathbf{A}\mathbf{\Sigma_x}\\ \mathbf{\Sigma_x}\mathbf{A}^H & \mathbf{\Sigma_x}\end{bmatrix}$$

The covariance matrix $\mathbf{\Sigma}$ can be inverted blockwise as follows:

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix}\mathbf{\Sigma_y}^{-1} + \mathbf{\Sigma_y}^{-1}\mathbf{A}\mathbf{\Sigma_x}\mathbf{S}^{-1}\mathbf{\Sigma_x}\mathbf{A}^H\mathbf{\Sigma_y}^{-1} & -\mathbf{\Sigma_y}^{-1}\mathbf{A}\mathbf{\Sigma_x}\mathbf{S}^{-1}\\ -\mathbf{S}^{-1}\mathbf{\Sigma_x}\mathbf{A}^H\mathbf{\Sigma_y}^{-1} & \mathbf{S}^{-1}\end{bmatrix}$$

where $\mathbf{S} = \mathbf{\Sigma_x} - \mathbf{\Sigma_x}\mathbf{A}^H\mathbf{\Sigma_y}^{-1}\mathbf{A}\mathbf{\Sigma_x}$ denote the Schur complement of $\mathbf{\Sigma}$. We can rewrite the distribution of $\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}$ using the block diagonal decomposition of the inversed covariance matrix as

$$p(\mathbf{y},\mathbf{x}) = \frac{1}{(2\pi)^{N_x+N_y}|\mathbf{S}||\mathbf{\Sigma_y}|} \mathrm{e}^{-\mathbf{x}^H\mathbf{S}^{-1}\mathbf{x}+\mathbf{y}^H\mathbf{\Sigma_y}^{-1}\mathbf{A}\mathbf{\Sigma_x}\mathbf{S}^{-1}\mathbf{x}+\mathbf{x}^H\mathbf{S}^{-1}\mathbf{\Sigma_x}\mathbf{A}^H\mathbf{\Sigma_y}^{-1}\mathbf{y}-\mathbf{y}^H\left(\mathbf{\Sigma_y}^{-1}+\mathbf{\Sigma_y}^{-1}\mathbf{A}\mathbf{\Sigma_x}\mathbf{S}^{-1}\mathbf{\Sigma_x}\mathbf{A}^H\mathbf{\Sigma_y}^{-1}\right)\mathbf{y}} \tag{C.4}$$

In order to evaluate the posterior probability density of Equation (C.3), we divide the joint distribution given by Equation (C.4) by $p(\mathbf{y})$ which is recalled as

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{N_y}|\mathbf{\Sigma_y}|}\,\mathrm{e}^{-\mathbf{y}^H\mathbf{\Sigma_y}^{-1}\mathbf{y}}\,.$$

The posterior distribution of $\mathbf{x}$ given $\mathbf{y}$ is obtained as

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{(2\pi)^{N_x}|\mathbf{S}|}\, \mathrm{e}^{-(\mathbf{x}-\boldsymbol{\Sigma}_\mathbf{x}\mathbf{A}^H{\boldsymbol{\Sigma}_\mathbf{y}}^{-1}\mathbf{y})^H\mathbf{S}^{-1}(\mathbf{x}-\boldsymbol{\Sigma}_\mathbf{x}\mathbf{A}^H{\boldsymbol{\Sigma}_\mathbf{y}}^{-1}\mathbf{y})}$$

This is a Gaussian distribution parametrized by its mean vector and covariance matrix

$$\begin{cases} \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \boldsymbol{\Sigma}_\mathbf{x}\mathbf{A}^H(\mathbf{A}\boldsymbol{\Sigma}_\mathbf{x}\mathbf{A}^H)^{-1}\mathbf{y} \\ \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}} = \mathbf{S} = \boldsymbol{\Sigma}_\mathbf{x} - \boldsymbol{\Sigma}_\mathbf{x}\mathbf{A}^H(\mathbf{A}\boldsymbol{\Sigma}_\mathbf{x}\mathbf{A}^H)^{-1}\mathbf{A}\boldsymbol{\Sigma}_\mathbf{x} \end{cases}$$

Appendix D

Ray detection and estimation

D.1 Likelihood ratio

The detection problem consists in evaluating, based on the received samples $\mathbf{y}$ which hypothesis H_{sig} (the received samples contains the signal of interest) or H_{noise} (the received samples only contains noise) is the most probable. In each case, the model for $\mathbf{y}$ is given as

$$\begin{cases} H_{sig} : \mathbf{y} = \mathbf{X_t}\mathbf{h}(t) + \mathbf{n} \\ H_{noise} : \mathbf{y} = \mathbf{n} \end{cases}$$

In order to simplify the evaluation, we assume that the gain β is complex centered Gaussian distributed as $\beta \sim \mathcal{CN}(0, P)$. Otherwise dealing with the unknown phase $\arg(\beta)$ is difficult.

Thus the probability density function of $\mathbf{y}$ under hypothesis H_{sig} is

$$\begin{aligned} p(\mathbf{y}|t) &= \frac{1}{(2\pi)^{N_t}|\mathbf{X_t}\mathbf{\Sigma}(t){\mathbf{X_t}}^H + \sigma_n^2\mathbf{I}|}\, \mathrm{e}^{-\mathbf{y}^H(\mathbf{X_t}\mathbf{\Sigma}(t){\mathbf{X_t}}^H+\sigma_n^2\mathbf{I})^{-1}\mathbf{y}} \\ &= K\, \mathrm{e}^{-\mathbf{z}^H(\mathbf{\Sigma}(t)+\sigma_n^2\mathbf{I})^{-1}\mathbf{z}} \end{aligned}$$

where $\mathbf{\Sigma}(t) = N_t P \mathbf{e_t}(t){\mathbf{e_t}}^H(t)$ is the channel covariance matrix and $\mathbf{z} = {\mathbf{X_t}}^H\mathbf{y}$. The probability density function of $\mathbf{y}$ under hypothesis H_{noise} is

$$p_{H_{noise}}(\mathbf{y}|t) = \frac{1}{(2\pi\sigma_n^2)^{N_t}}\, \mathrm{e}^{-\frac{\|\mathbf{y}\|^2}{\sigma_n^2}}.$$

According to the Neyman-Pearson lemma, the most powerful test is given by the likelihood ratio r, given as

$$r(\mathbf{z}|t) = \frac{p_{H_{sig}}(\mathbf{y}|t)}{p_{H_{noise}}(\mathbf{y})} = \frac{(\sigma_n^2)^{N_t}}{|\mathbf{X_t}\mathbf{\Sigma}(t){\mathbf{X_t}}^H + \sigma_n^2\mathbf{I}|}\, \mathrm{e}^{-\mathbf{z}^H(\mathbf{\Sigma}(t)+\sigma_n^2\mathbf{I})^{-1}\mathbf{z}+\frac{\|\mathbf{z}\|^2}{\sigma_n^2}} \lessgtr \gamma$$

where γ is a predefined threshold. If $r(\mathbf{z}|t) > \gamma$ then H_{sig} is the most probable hypothesis. Otherwise if $r(\mathbf{z}|t) < \gamma$, H_{noise} is the most probable hypothesis. This detection criterion can be rewritten using the log-likelihood ratio as

$$\begin{aligned}\ln\left(r(\mathbf{z}|t)\right) &= -\mathbf{z}^H(\mathbf{\Sigma}(t)+\sigma_n^2\mathbf{I})^{-1}\mathbf{z} + \frac{\|\mathbf{z}\|^2}{\sigma_n^2}. \\ &= -\mathbf{z}^H(PN_t\mathbf{e_t}(t)\mathbf{e_t}^H(t)+\sigma_n^2\mathbf{I})^{-1}\mathbf{z} + \frac{\|\mathbf{z}\|^2}{\sigma_n^2}.\end{aligned}$$

Using the fact that $(PN_t\mathbf{e_t}(t)\mathbf{e_t}^H(t)+\sigma_n^2\mathbf{I})^{-1} = -\frac{PN_t}{(PN_t+\sigma_n^2)\sigma_n^2}\mathbf{e_t}(t)\mathbf{e_t}^H(t) + \frac{1}{\sigma_n^2}\mathbf{I}$, we obtain the matched filter detector expression

$$|\mathbf{y}^H\mathbf{X_t}\mathbf{e_t}^*(t)|^2 \lessgtr \gamma'$$

where γ' is a modified predefined threshold.

D.2 Convolution of the likelihood with a kernel function

We consider the same detection problem as in Annex D.1 that consists in evaluating which hypothesis H_{sig} or H_{noise} is valid based on the received samples $\mathbf{y}$. In each case, the model for $\mathbf{y}$ is given as

$$\begin{cases} H_{sig} : \mathbf{y} = \mathbf{X_t}\mathbf{h}(t) + \mathbf{n} \\ H_{noise} : \mathbf{y} = \mathbf{n} \end{cases}$$

In order to simplify the evaluation, we assume that the gain β is complex centered Gaussian distributed as $\beta \sim \mathcal{CN}(0, P)$. Otherwise dealing with the unknown phase $\arg(\beta)$ is difficult.

As well as in Annex D.1, the probability density function of $\mathbf{y}$ under hypothesis H_{sig} is

$$\begin{aligned}p(\mathbf{y}|t) &= \frac{1}{(2\pi)^{N_t}|\mathbf{X_t}\mathbf{\Sigma}(t)\mathbf{X_t}^H+\sigma_n^2\mathbf{I}|}\,\mathrm{e}^{-\mathbf{y}^H(\mathbf{X_t}\mathbf{\Sigma}(t)\mathbf{X_t}^H+\sigma_n^2\mathbf{I})^{-1}\mathbf{y}} \\ &= K\,\mathrm{e}^{-\mathbf{z}^H(\mathbf{\Sigma}(t)+\sigma_n^2\mathbf{I})^{-1}\mathbf{z}}\end{aligned}$$

where $\mathbf{\Sigma}(t) = N_tP\mathbf{e_t}(t)\mathbf{e_t}^H(t)$ is the channel covariance matrix and $\mathbf{z} = \mathbf{X_t}^H\mathbf{y}$. The

probability density function of $\mathbf{y}$ under hypothesis H_{noise} is

$$p_{H_{noise}}(\mathbf{y}|t) = \frac{1}{(2\pi\sigma_n^2)^{N_t}} \, \mathrm{e}^{-\frac{\|\mathbf{y}\|^2}{\sigma_n^2}}.$$

The difference with Annex D.1 resides in the now unknown direction of departure t that lies in the interval $[-1, 1]$. In order to mitigate this effect, we will perform several tests on a subdivision of $[-1, 1]$.

We wish assess whether the ray lies or not in the sub-interval $t_j < t < t_{i+j}$ centered on $\bar{t}_j$, of width Δ_t. The likelihood ratio is then given as

$$r(\mathbf{y}, t_j < t < t_{j+1}) = \frac{p_{H_{sig}}(\mathbf{y}|t_j < t < t_{j+1})}{p_{H_{noise}}(\mathbf{y})} = \frac{\int_{t_j}^{t_{j+1}} p_{H_{sig}}(\mathbf{y}|t)dt}{p_{H_{noise}}(\mathbf{y})} = \int_{t_j}^{t_{j+1}} r(\mathbf{z}|t)dt.$$

The likelihood ratio could be rewritten as

$$r(\mathbf{y}, t_j < t < t_{j+1}) = \frac{(\sigma_n^2)^{N_t}}{|\mathbf{X_t}\mathbf{\Sigma}(t)\mathbf{X_t}^H + \sigma_n^2\mathbf{I}|} \int_{t_j}^{t_{j+1}} \mathrm{e}^{-\mathbf{z}^H(\mathbf{\Sigma}(t)+\sigma_n^2\mathbf{I})^{-1}\mathbf{z}+\frac{\|\mathbf{z}\|^2}{\sigma_n^2}} \, dt \lessgtr \gamma$$

Using the fact that $(PN_t\mathbf{e_t}(t)\mathbf{e_t}^H(t) + \sigma_n^2\mathbf{I})^{-1} = -\frac{PN_t}{(PN_t+\sigma_n^2)\sigma_n^2}\mathbf{e_t}(t)\mathbf{e_t}^H(t) + \frac{1}{\sigma_n^2}\mathbf{I}$, we obtain the expression

$$r(\mathbf{y}, t_j < t < t_{j+1}) = K \int_{t_j}^{t_{j+1}} \mathrm{e}^{\frac{P}{(PN_t+\sigma_n^2)\sigma_n^2}\mathbf{z}^H\mathbf{e_t}(t)\mathbf{e_t}^H(t)\mathbf{z}} \, dt \lessgtr \gamma$$

As this reasoning will be useful later on, we state the following general integration problem

$$f(\mathbf{y}) = \int_{-\infty}^{\infty} p(t - \bar{t}) \, \mathrm{e}^{\frac{P}{(PN_t+\sigma_n^2)\sigma_n^2}\mathbf{z}^H\mathbf{e_t}(t)\mathbf{e_t}^H(t)\mathbf{z}} \, dt$$

The Taylor decomposition of the exponential function is given by

$$\mathrm{e}^{-x} = \sum_{k=0}^{\infty}(-1)^k\frac{x^k}{k!}.$$

We only use the first order approximation. The modified likelihood takes the form

$$f(\mathbf{y}) \approx K_0 + K' \sum_{k,l=1}^{N_t} [\mathbf{z}^H]_k[\mathbf{z}]_l \int_{-\infty}^{\infty} p(t - \bar{t}) \, \mathrm{e}^{2\pi j(l-k)\frac{\|\vec{a}_{tx}\|t}{\lambda}} \, dt.$$

where we can recognize the Fourier Transform of the kernel function p :

Table D.1 – Relation between the kernel functions and the Qaudratic form matrix $\mathbf{R}(p^{tx})$

Kernel function	**Kernel formula**	$\mathbf{R}(p^{tx})$
Rectangle	$p_{\Delta_t}(t) = \frac{\mathbb{1}_{[-\Delta_t/2,\Delta_t/2]}(t)}{\Delta_t}$	$[\mathbf{R}(p)]_{k,l} = N_t \operatorname{sinc}(2\pi\Delta_t(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})$
Piecewise Parabolic	$p_{\Delta_t}(t) = -(t-\Delta_t/2)(t+\Delta_t/2)$ $t \in [-\Delta_t/2, \Delta_t/2]$	$[\mathbf{R}(p)]_{k,l} = N_t \frac{2}{(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})^3}$ $\left(\Delta_t(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})\cos\left(\frac{\Delta_t}{2}(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})\right)\right.$ $\left.-2\sin\left(\frac{\Delta_t}{2}(\pi(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda})\right)\right)$

$$f(\mathbf{y}) \approx K_0 + K' \sum_{k,l=1}^{N_t} [\mathbf{z}^H]_k [\mathbf{z}]_l \phi_p\left((l-k)\frac{\|\vec{a}_{tx}\|}{\lambda}\right) \mathrm{e}^{2\pi j(l-k)\frac{\|\vec{a}_{tx}\|}{\lambda}\bar{t}}$$

The text statistic can be rewritten in a more compact way using a quadratic form as

$$\mathbf{z}^H \mathbf{\Delta}(\bar{t}) \mathbf{R}(p) \mathbf{\Delta}^*(\bar{t}) \mathbf{z} \lessgtr \gamma'$$

where γ' is another threshold value.

In Chapter 6, we consider two types of kernels, described in Table D.1. The rectangle one is used for the detection problem introduced before. The same calculations using the piecewise parabolic kernel are useful in Section 6.4.3. The function that is optimized is then $f_{opt}(t) = \mathbf{z}^H \mathbf{\Delta}(t) \mathbf{R}(p) \mathbf{\Delta}^*(t) \mathbf{z}$ where $\mathbf{R}(p)$ is defined using the piecewise parabolic kernel.

For implementation efficiency, the matrix $\mathbf{R}$ can be decomposed into eigencomponents as

$$\mathbf{R} = \sum_{k=1}^{N_t} \lambda_k \mathbf{r_k} {\mathbf{r_k}}^H$$

The test statistic function can then be rewritten as

$$\sum_{k=1}^{N_t} \lambda_k |\mathbf{y}^H \mathbf{X_t} (\mathbf{r_k} \odot {\mathbf{e_t}}^*(t_0))|^2 \lessgtr \gamma' \tag{D.1}$$

using an eigenvector decomposition of matrix $\mathbf{R}$. Most of the eigenvalues are non-significant and the sum (D.1) can be truncated. It actually consists in a weighted sum of correlators, windowed by the real vectors $\mathbf{r_k}$.

INSA de RENNES
Service des Formations

AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

Titre de la thèse:
A ray-based approach for Massive MIMO Systems: Statistical Analysis and Signal Processing

Nom Prénom de l'auteur : ROY MATTHIEU

Membres du jury :
- Monsieur LE RUYET Didier
- Monsieur CRUSSIERE Matthieu
- Madame COTTATELLUCCI Laura
- Monsieur DORE Jean-Baptiste
- Monsieur PAQUELET Stéphane
- Monsieur SIBILLE Alain
- Monsieur GORCE Jean-Marie

Président du jury : Alain Sibille

Date de la soutenance : 23 Novembre 2020

Reproduction de la these soutenue

☑ Thèse pouvant être reproduite en l'état
☐ Thèse pouvant être reproduite après corrections suggérées

Fait à Rennes, le 23 Novembre 2020

Signature du président de jury

Le Directeur,

M'hamed DRISSI

DOCTORAT BRETAGNE LOIRE / MATHSTIC

INSA RENNES

Titre : Utilisation de modèles par rayons pour les systèmes multi-antennes massifs : Analyse statistique et traitements numériques

Mots clés : MIMO, Modèles par rayons, Saleh-Valenzuela, Séparation de clusters

Résumé : Le passage des systèmes MIMO à faible nombre d'antennes vers les systèmes multi-antennes massifs entraîne un changement de paradigme dans les traitements numériques mis en œuvre. En effet la complexité des algorithmes de traitement du signal utilisés classiquement augmente avec les nombres d'antennes. D'autre part les larges réseaux d'antennes bénéficient d'une résolution angulaire plus précise. Il devient alors intéressant de mettre à profit la dimension angulaire de la propagation dans les traitements pour en maîtriser les performances et la complexité algorithmique. Dans cette thèse, nous avons dans un premier temps étudié et caractérisé les canaux de propagation multi-trajets, avec une attention particulière sur le modèle Saleh-Valenzuela. Cette étude préliminaire nous a permis de définir un modèle simplifié de cluster. A partir de ce modèle, nous avons étudié d'un point de vue théorique le problème de séparation de clusters et avons proposé un nouvel algorithme. Nous avons aussi utilisé ce modèle pour étudier les limites de l'extrapolation de clusters pour la détermination du précodage en FDD. Nous avons enfin proposé des améliorations à un algorithme d'estimation de canal MIMO massif, basé sur la technique OMP qui utilise la structure creuse des canaux de propagation en ondes millimétriques.

Title : A ray-based approach for massive MIMO systems: Statistical analysis and digital processing

Keywords : MIMO, Ray-based model, Saleh-Valenzuela, Cluster separation

Abstract : Moving from small-scale MIMO to large scale massive MIMO systems leads to a paradigm shift in digital processing techniques. Indeed, the complexity of classical signal processing algorithms scales with the antenna array sizes, so does the system angular resolution. It is then interesting to take advantage of the angular dimension properties of the propagation in the processing to constraint the algorithmic complexity while controlling the performance. In this thesis, we first studied and characterized multi-path propagation channels, with due consideration to the Saleh-Valenzuela model. This preliminary study provided us with a simplified cluster model. From this model, we studied theoretically the cluster separation problem and proposed a novel algorithm for this purpose. We also applied this model to assess the limits of cluster extrapolation for FDD precoding. We finally proposed enhancements to OMP-based Massive MIMO channel estimation that leverages the sparsity of millimeter wave channels.