

A Genetic Algorithm for Channel Estimation in Switch-based Hybrid Analog/Digital mmWave Massive MIMO Systems

Alec Poulin



Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

April 2020

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master.

© 2020 Alec Poulin

Abstract

Massive multiple-input multiple-output (MIMO) is one of the main technologies proposed to meet the stringent requirements of the upcoming fifth generation (5G) standard for wireless communications. However, the cost and the energy consumption of the radio-frequency (RF) chains in massive MIMO systems precludes the use of a traditional scheme where each antenna is equipped with one RF chain. This has led to the introduction of hybrid analog/digital systems, where an analog module between the antennas and the digital baseband processor allows a reduced number of RF chains.

This thesis addresses the problem of pilot-based channel estimation in hybrid analog/digital massive MIMO systems for future millimetre wave (mmWave) communications. To further reduce system cost and implementation complexity of the analog module, we consider an alternative architecture derived from RF switches as opposed to the phase shifters in the conventional literature. The estimation problem is modelled as a combinatorial optimization problem where the aim is to minimize the mean square error (MSE) between the real channel and the estimated channel over a finite set of allowed values for the switches. To solve the estimation problem, a genetic algorithm (GA) is developed for the novel switch-based hybrid analog/digital massive MIMO architecture. Simulations of MIMO transmission over realistic mmWave channel models show that the proposed GA is able to estimate channels as accurately as, if not more than, an existing solution using phase shifters.

Sommaire

Les systèmes à entrées-multiples-sorties-multiples (MIMO) massifs sont une des principales technologies mises de l'avant pour atteindre les demandes exigeantes de l'imminent standard de communication sans fil de cinquième génération (5G). Cependant, le coût et la consommation d'énergie des chaînes de fréquence radio (RF) rendent irréalisable la structure traditionnelle où chaque antenne est équipée d'une chaîne RF. C'est ce qui a poussé à l'introduction des systèmes hybrides analogues/digitaux, où un module analogue ajouté entre les antennes et le processeur digital de bande de base permet un nombre réduit de chaînes RF.

Ce mémoire aborde le problème d'estimation de canal utilisant des pilotes dans les systèmes hybrides analogues/digitaux MIMO massifs pour les futures communications à ondes millimétriques (mmWave). Pour réduire davantage les coûts de système et la complexité d'implémentation du module analogue, nous considérons une architecture alternative dérivée des commutateurs RF, par opposition aux déphaseurs retrouvés dans la littérature conventionnelle. Le problème d'estimation est modelé comme un problème d'optimisation combinatoire dont le but est de minimiser l'erreur quadratique moyenne (MSE) entre le vrai canal et le canal estimé par rapport à un ensemble fini de valeurs admises pour les commutateurs. Pour résoudre le problème d'estimation, un algorithme génétique (GA) est développé pour la nouvelle architecture de système MIMO hybride analogue/digital à commutateurs. Des simulations de transmission MIMO sur des modèles réalistes de canaux mmWave démontrent que le GA proposé est capable d'estimer les canaux aussi précisément, sinon plus, qu'une solution existante utilisant des déphaseurs.

Acknowledgments

I would never have been able to complete this thesis project without the help, guidance, and support of many people. Firstly, I would like to thank my supervisor Pr. Benoit Champagne for his intuition and his insightful advice, but also for his positive attitude and encouragement. He gave me a golden opportunity thanks to which I learned a lot, and to this I am very grateful. I am also indebted to his financial support through research contracts with InterDigital Canada Ltée. and grants with the Natural Sciences and Engineering Research Council of Canada (NSERC). My esteemed colleague Alireza Morsali was also always present to discuss ideas and answer questions. Without his passion, his guidance, and his knowledge, this project would not have culminated. I would also like to thank my other colleagues, especially Seyyed Saleh Hosseini and Ali Baghaki for their help and knowledge on various subjects.

The most important person I would like to thank is my wife, Shannon. She offered me her continuous love and understanding during the whole venture and her support never faltered. She truly is a source of inspiration for me. Likewise, I am grateful of my parents, Johanne and Daniel, without whom none of what I ever accomplished would have been possible, and my mother-in-law, Kathleen, for her moral support, her empathy, and her help with many things.

Contents

1	Introduction	1
1.1	Adapting MIMO for 5G	1
1.2	Literature Review	5
1.3	Thesis Objectives and Contributions	8
1.4	Thesis Organization	11
2	System Model and Problem Formulation	12
2.1	System Model	12
2.2	Channel Model	15
2.3	Problem Formulation	17
3	Review of Existing Solutions	21
3.1	Preliminary Steps	21
3.2	Unconstrained Optimization Methods	23
3.2.1	Single-Pilot-Symbol Unconstrained Optimization	23
3.2.2	Block Selection	24
3.2.3	Sequential Optimization	24
3.2.4	Alternating Optimization	26
3.3	Projection onto the Set of Constrained Combiners and Quantization	28
4	Proposed Switch-Based Architecture for Channel Estimation	29
4.1	Structural Elements	29
4.2	Problem Formulation with Switches	32
4.3	Channel Estimation using Genetic Algorithm	34

5	Simulation Results	39
5.1	Methodology	39
5.2	Algorithm Convergence	42
5.3	Comparison with Exhaustive Search	42
5.4	Results for Massive MIMO Systems	46
6	Conclusion and Future Prospects	56
6.1	Thesis Summary	56
6.2	Future Prospects	58
A	Derivation of the Channel Covariance Matrix and its Properties	60
A.1	Derivation for an Arbitrary Antenna Array	60
A.2	Derivation for a Uniform Linear Array	61
B	Derivation of the Optimal Baseband Processor Matrix	63
C	Derivation of the Bounds on the Objective Function	66
	References	68

List of Figures

1.1	Block diagram of a massive MIMO transceiver built with a hybrid analog/digital architecture. The analog unit acts as an interface between a large number N of antennas and a small number L of RF chains. The arrows indicate the travel direction of signals during a reception.	4
1.2	Illustration of the effects of pilot contamination on beamforming. The green arrow represents the pilot sequence sent by the user inside the cell, whereas the red arrow represents the interfering pilot of the user with the same pilot sequence in an adjacent cell.	7
1.3	Four hybrid-architecture configurations for massive MIMO transceivers that use switches instead of phase shifters.	9
2.1	Block diagram of the system model, consisting of a hybrid mmWave massive MIMO base station and K single-antenna mobile stations.	13
2.2	Geometrical representation of a plane wave signal impinging on a ULA at an angle θ . The small circles represent the antennas and the blue lines represent the signal propagation rays, which are approximated as parallel to each other (i.e. forming a plane wave) because the distance between the BS and the MS is much greater than the antenna aperture, given by $(N - 1)d$	16
3.1	Effect of diagonal loading on the normalized MSE of the channel estimation performed by the SO method as a function of the parameter ε and for different SNRs. The parameters $L = 8$, $N = 64$, $T = 1$, and $P = 6$ are used.	25
4.1	Different types of switches that can be used in an analog combiner to implement different sets \mathcal{B} of allowed values in the matrices \mathbf{S}_t	31

4.2	Switches based on the in-phase and quadrature signal decomposition as described in Section 4.1. These switches allow more possibilities to be implemented for the sets \mathcal{B} of values allowed in the matrices \mathbf{S}_t	33
5.1	MSE of the population versus iteration number of the GA. The blue dots are the MSEs of the individuals and the red crosses are the population averages at each iteration.	43
5.2	NMSE of channel estimate versus SNR for the GA, the SO method, and the exhaustive search. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $N = 6, L = 2, T = 1, P = 2$ are used.	44
5.3	NMSE of channel estimate versus SNR for the GA, the SO method, and the exhaustive search. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, j, -j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $N = 6, L = 2, T = 1, P = 2$ are used.	45
5.4	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers N of antennas. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $L = 8, T = 1, P = 6$ are used.	47
5.5	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers N of antennas. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $L = 8, T = 1, P = 6$ are used.	48
5.6	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers P of propagation paths. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $L = 8, N = 32, T = 1$ are used.	49
5.7	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers P of propagation paths. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $L = 8, N = 32, T = 1$ are used.	50

5.8	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers L of RF chains at the BS. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $N = 32, T = 1, P = 6$ are used. . .	51
5.9	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers L of RF chains at the BS. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $N = 32, T = 1, P = 6$ are used.	52
5.10	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers T of pilot symbols. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $L = 8, N = 32, P = 6$ are used. . .	53
5.11	NMSE of channel estimate versus SNR for the GA and the SO method with different numbers T of pilot symbols. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $L = 8, N = 32, P = 6$ are used. . .	54
5.12	NMSE of channel estimate versus SNR for the GA with different sets of allowed values. The parameters $L = 8, N = 64, T = 1, P = 6$ are used.	55

List of Algorithms

1	Alternating optimization to design optimal unconstrained combiners for multiple trainings	27
2	Genetic algorithm applied to finding analog combiners for switching networks	38

List of Acronyms

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
4G	Fourth Generation
5G	Fifth Generation
ADC	Analog-to-Digital Converter
AO	Alternating Optimization
AoA	Angle of Arrival
AWGN	Additive White Gaussian Noise
BS	Base Station
CNN	Convolutional Neural Network
C-RAN	Cloud Radio Access Network
CSI	Channel State Information
dB	Decibel
DNN	Deep Neural Network
EVD	Eigenvalue Decomposition
FDD	Frequency-Division Duplexing
GA	Genetic Algorithm
Gbps	Gigabit per Second
GHz	Gigahertz
i.i.d.	Independent and Identically Distributed
IoT	Internet of Things
M2M	Machine-to-Machine
Mbps	Megabit per Second

MHz	Megahertz
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean Square Error
mmWave	Millimetre Wave
MS	Mobile Station
MSE	Mean Square Error
MU	Multi-User
NMSE	Normalized Mean Square Error
OFDM	Orthogonal Frequency-Division Multiplexing
RF	Radio Frequency
SNR	Signal-to-Noise Ratio
SO	Sequential Optimization
SVD	Singular Value Decomposition
TDD	Time-Division Duplexing
ULA	Uniform Linear Array
UPA	Uniform Planar Array

Notation

a	A	Non-bold symbols are scalars
\mathbf{a}		Bold lower case symbols are vectors
\mathbf{A}		Bold upper case symbols are matrices
\mathbf{A}^T		Transpose of matrix \mathbf{A}
\mathbf{A}^H		Conjugate transpose of matrix \mathbf{A}
\mathbf{A}^\dagger		Moore-Penrose pseudoinverse of matrix \mathbf{A}
$\ \mathbf{a}\ $		Euclidean 2-norm of vector \mathbf{a}
$\ \mathbf{A}\ _F$		Frobenius norm of matrix \mathbf{A}
$\text{tr}(\mathbf{A})$		Trace of matrix \mathbf{A}
$\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_N)$		Square block-diagonal matrix with the matrices $\mathbf{A}_1, \dots, \mathbf{A}_N$ on its main diagonal and zeros elsewhere
\mathbf{I}_N		$N \times N$ identity matrix
$\mathbf{I}_{M \times N}$		$M \times N$ matrix with ones on its main diagonal and zeros elsewhere
$[\mathbf{A}]_{(m,n)}$		$(m, n)^{\text{th}}$ entry of matrix \mathbf{A}
$\mathbf{A}_{[n_1, \dots, n_K]}$		Matrix made of the $n_1^{\text{th}}, \dots, n_K^{\text{th}}$ columns of matrix \mathbf{A}
$\mathbf{A}_{[m:n]}$		Matrix made of the columns of matrix \mathbf{A} with indices m through n , inclusively
$\mathbb{E}[\mathbf{X}]$		Entrywise expectation of random matrix \mathbf{X}
$\mathcal{CN}(\mathbf{m}, \mathbf{R})$		Complex Gaussian random vector with mean \mathbf{m} and covariance matrix \mathbf{R}
$\mathcal{U}(a, b)$		Random variable uniformly distributed on the interval $[a, b]$
\mathbb{N}		Set of natural numbers, $\mathbb{N} = \{0, 1, 2, \dots\}$
\mathbb{R}		Set of real numbers
\mathbb{R}_+		Set of non-negative real numbers, $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$

\mathbb{C}	Set of complex numbers
$ \mathcal{B} $	Cardinality of discrete set \mathcal{B}
$[x]$	Real number x rounded to the nearest integer
j	Imaginary unit, $j = \sqrt{-1}$

Chapter 1

Introduction

This chapter begins with an overview of the 5G standard with an emphasis on massive MIMO, mmWave communications, and hybrid analog/digital architecture. Then, a survey of the literature on channel estimation within the context of these technologies follows. A summary of the objectives and contributions of this thesis comes after. Finally, the organization of the thesis is presented.

1.1 Adapting MIMO for 5G

Some reports have predicted that the global mobile traffic will increase sevenfold between 2017 and 2022 [1]. The load on the cellular network will thus build up tremendously. This tendency is mainly due to the growing popularity of video calls and high definition streaming services on smartphones. Added to these trends is the emergence of the internet of things (IoT) and machine-to-machine (M2M) communications. All-in-all, statistics indicate that there will be roughly 1.5 mobile-connected devices for each person on Earth by 2022 [2]. These numbers tell us that tomorrow's telecommunication networks will need to be rethought and redesigned, with the requirement that their entry points be able to manage more connections with higher data rates than it is currently possible.

Increasing data rates is one of the objectives of the 5G standard for wireless telecommunication, which defines the upcoming fifth generation of mobile networks. Its requirements specify, among other things, increased target values for the aggregate data rate, or area capacity, as well as for the edge rate:

- The area capacity is the total amount of data the network can support and is measured in bits per unit time and unit area. This capacity will need to be 1000 times higher in the fifth generation (5G) than it was in the fourth generation (4G) [3].
- The edge rate is the worst data rate a user can expect when in range of a base station (BS). The goal of 5G for this metric is between 100 Mbps and 1 Gbps, which represents an increase by a factor of 100 from 4G [3].

Massive multiple-input multiple-output (MIMO) and millimetre wave (mmWave) signals are two technologies proposed to reach these exacting objectives. Massive MIMO, as opposed to traditional MIMO, is a regime where each BS is equipped with a number of antennas on the order of hundreds or more [4]. This allows spatial multiplexing to be fully used and therefore a higher information capacity and thus higher data rates to be reached. Because of the very large number of antennas, however, the fundamental properties of massive MIMO are different than those of MIMO and more research must be done to better understand its operation and design.

Different antenna array configurations can be used for different massive MIMO applications: linear, rectangular, cylindrical, and distributed, for example [5]. Even arrays with a full three-dimensional (3D) structure can be considered. However, according to Gauss's laws of electromagnetism, the electric and magnetic fields in a volume are fully described by the fields at the surface of the volume. This limits the usefulness of 3D arrays, because the antennas inside the structure are superfluous and do not contribute to the information capacity of the array. Only the antennas at the surface do [6].

With the use of mmWave signals, i.e. signals with wavelengths on the order of millimetres¹, the new antenna arrays will be smaller and include more antennas than they would if they used signals with larger wavelengths, as it is the case today with operation in the microwave band. This will allow tens of antennas to fit in a single smartphone and hundreds or thousands in a BS [7]. Also, almost all current commercial radio communications, which include mobile phones, AM/FM radio, and Wi-Fi, use frequencies between 300 MHz and 3 GHz, leaving the frequencies above 3 GHz unexploited. If the oxygen and the water vapour absorption bands are excluded, a span of approximately 250 GHz is available between 3 GHz and 300 GHz [7]. Exploiting these frequencies would allow a much larger volume of data to be exchanged. This possibility, however, comes along with a few drawbacks: mmWave

¹That is, with frequencies in the range from approximately 30 GHz to 300 GHz.

signals do not travel very well through most solid materials and undergo significant attenuations in heavy rain [7]. Nevertheless, mmWave frequencies open the door for high-throughput small-range communications.

A serious concern that comes with providing a significantly enhanced performance with 5G is that the energy consumption of many components and sub-systems in the network will likely increase. Along with plans to extend the mobile coverage to less populated areas in various regions of the world, the total energy consumption of the global network could increase unacceptably if no special measures are taken. Thus, another important objective for future networks is to improve the energy efficiency of their components in order to keep the total energy consumption of the network at least constant, if not to reduce it.

To achieve this goal while allowing increased data usage, the energy per bit would need to be reduced by a factor of at least 100. Different technologies and schemes developed for 5G could potentially help reach this goal, including the following:

- The cloud radio access network (C-RAN) architecture, which allows a more fluid management of computing resources [8].
- The use of smaller and heterogeneous cells, which are far more efficient than the currently used larger cells [9].
- The use of hybrid analog/digital architecture for massive MIMO, which allows fewer radio-frequency (RF) chains to be employed in transceivers [7].
- The use of high-resolution beamforming with large-scale antenna arrays, which allows to focus the transmitted signals and to receive weaker signals [4].

Beamforming is achieved through interaction between electromagnetic waves. On one hand, when the waves emanating from the antennas during the transmission of a signal are in phase with each other, they interact constructively and more power is transmitted. The angular region where this happens is referred to as a beam. Conversely, outside of the beam, the transmitted waves are out of phase and interact destructively. Hence only a small amount of power, if any, is emitted in these directions. Adjusting the phase of the signals sent by each antenna allows the direction of the beam to be chosen. When the number of antennas employed is larger, the beam is easier to direct and can be made sharper. From the perspective of energy efficiency in 5G, beamforming allows signals to be focused in smaller areas and hence wasteful spills of power to be reduced.

On the other hand, during the reception of a signal, the waves received from a particular direction at the antennas can be delayed, or phase-shifted, electronically so that they be in phase with each other and interact constructively when summed. Also, because the electrical noise picked up by antennas tends to be spatially uncorrelated, adding together phase-aligned signals from many antennas increases the signal-to-noise ratio (SNR). This SNR gain, which is higher when there are more antennas, allows a reduction of the source transmit power.

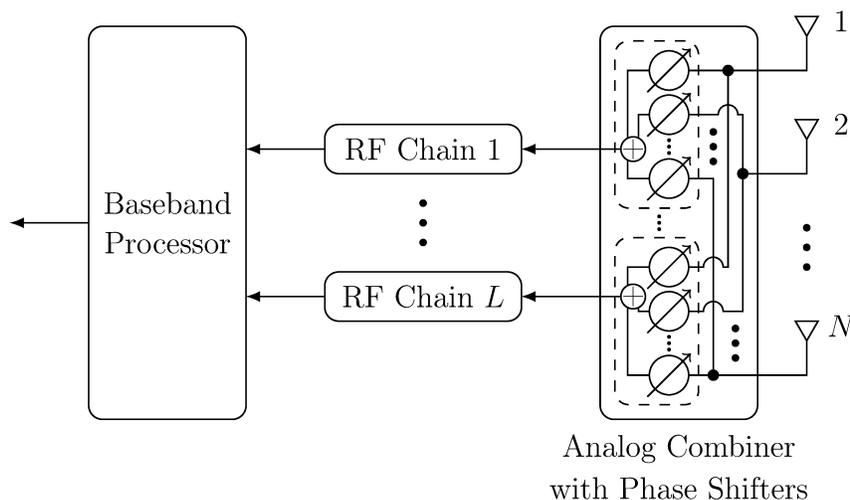


Fig. 1.1 Block diagram of a massive MIMO transceiver built with a hybrid analog/digital architecture. The analog unit acts as an interface between a large number N of antennas and a small number L of RF chains. The arrows indicate the travel direction of signals during a reception.

As previously mentioned, the hybrid analog/digital architecture is also proposed to help reduce material and energy costs of future massive MIMO systems. Using more antennas requires more RF chains, which in turn consume more power. RF chains are units that convert the analog signals received by the antennas into digital signals that are suitable for their treatment by the baseband processor during the reception, and vice versa during the transmission. Their large number, as anticipated for massive MIMO, is problematic because they are expensive and not power-efficient [5, 7]. One approach to cope with this limitation is to use an analog processor (called analog combiner, or decoder, for the reception and analog beamformer, or precoder, for the transmission) to convert the signal between a large number N of antennas and a much smaller number $L \ll N$ of RF chains, resulting in what is called a hybrid analog/digital architecture, as depicted in Fig. 1.1. This solution, however,

introduces a new layer of difficulty, since the baseband processing unit no longer has direct access to the antennas, which usually leads to decreased performance. The minimum number of RF chains required to achieve an analog precoder with the same performance as a fully digital one is investigated in [10–12].

1.2 Literature Review

The focus of this thesis is on channel estimation in the context of hybrid analog/digital mmWave massive MIMO. As exposed in the previous section, these combined technologies hold the potential to significantly boost the performance of future wireless networks, but need to be better understood before they can be most efficiently used in practice. Different approaches have been explored in recent works in order to develop channel estimation algorithms that are specifically designed for this architecture. Typically, these approaches consider either a point-to-point setup, where only two transceivers communicate with each other, or a multi-user (MU) setup, where one BS communicates with many mobile stations (MSs). The former situation is common in the backhaul, for example when two BSs are in a chain of relays between two users of the network, whereas the latter situation usually happens at an entry point of the network. Furthermore, some methods are designed to work specifically with frequency-division duplexing (FDD), whereas others are designed to work with time-division duplexing (TDD). One advantage of TDD over FDD is that in TDD, under slowly time-varying conditions, the uplink (MS to BS) and the downlink (BS to MS) channel matrices are transposes of one another, which means that the channels estimated in one direction can be used as well in the other direction. In contrast, in FDD, both channels must be individually estimated due to the loss of coherence across different frequency bands.

A method exploiting the angular sparsity of the channel is presented in [13], where the focus is on the estimation of a point-to-point narrow-band flat-fading mmWave channel with both transceivers having a hybrid architecture. Both the single-path case and the multiple-path case are considered, and the estimation is performed by going through a hierarchical multi-resolution codebook. At the first iteration of the algorithm in the single-path case, the entire angular domain around the receiver is separated into K equally sized sectors. For each sector, a beamformer is designed to extract the received signal power and the sector with the highest received power is kept. At the next iteration, this sector is separated into K smaller sectors and again, the one with the highest received power is kept. The process

is repeated until the desired resolution is attained. The multi-path scenario is handled by means of a similar iterative procedure, but with the contribution of each path subtracted from the received power pattern prior to the next iteration.

Various extensions of the above beam search approach to the broad-band frequency-selective fading channel model, which is more representative of the conditions of operation of mmWave massive MIMO systems, have been proposed in the literature. In particular, the works in [14,15] exploit sparse formulation and compressed sensing concepts to solve the wideband channel estimation problem for the hybrid MIMO architecture. In [16], the channel estimation problem for MIMO orthogonal frequency-division multiplexing (OFDM) systems after beam searching is formulated as an harmonic retrieval problem, where the channel paths are associated to a discrete set of frequencies and gains. Two different methods are then proposed for the estimation of these parameters.

In [17], the problem of uplink channel estimation of a massive MU MIMO system is investigated. The BS is equipped with a hybrid analog/digital architecture and each MS is equipped with a single antenna. In contrast to the above works based on the concept of beam search, [17] proposes an efficient pilot-based estimation approach that exploits the spatial correlation matrix of the massive MIMO channel. Both the single and the multiple-pilot cases are visited, leading to different algorithms with varying complexity and performance levels. Moreover, the channel correlation matrix is also used in [18], where an efficient least-squares approach based on subspace tracking is presented to estimate its dominant eigenvectors.

All of the aforementioned works use pilot sequences to estimate channels. Pilot sequences are signals that are known beforehand by the BS and the MS. The knowledge of the pilot sequence beforehand along with the received signal is what allows the channel to be estimated. In order for the BS to distinguish between the different users, their attributed pilot sequences need to be orthogonal to each other. However, the number of orthogonal pilot sequences is limited, where the limit is estimated to be about 200 in a typical scenario [5]. For this reason, neighbouring cells usually have to reuse the same set of pilot sequences, which leads to the problem of *pilot contamination*. This problem arises when the transmission of a pilot sequence in a cell interferes with the transmission the same sequence in another cell, as illustrated in Fig. 1.2a. A beamformer designed with pilot contamination will tend to leak power towards a user in another cell, as shown in Fig. 1.2b, thus increasing the amount of interference power for this user.

To lessen this detrimental effect, [19] develops a channel estimation algorithm that

works under the assumption of pilot contamination. Other approaches explore pilot allocations [20–22] and pilot designs [23, 24] that aim to minimize the interference. In [25], two methods are presented where adjacent BSs cooperate to reduce the contamination. Alternatively, other works develop ways of estimating channels without using pilot sequences or using them to a lesser extent, leading to so-called blind and semi-blind channel estimation methods. Both [26] and [27] present blind estimation techniques, the former being based on the maximum likelihood principle and the latter being based on adaptive independent component analysis. In [28], a blind technique and a semi-blind technique are presented for time-invariant channels, but both perform well even in the time-varying scenario. Finally, a low-complexity semi-blind estimation approach is proposed in [29] for massive MU MIMO systems.

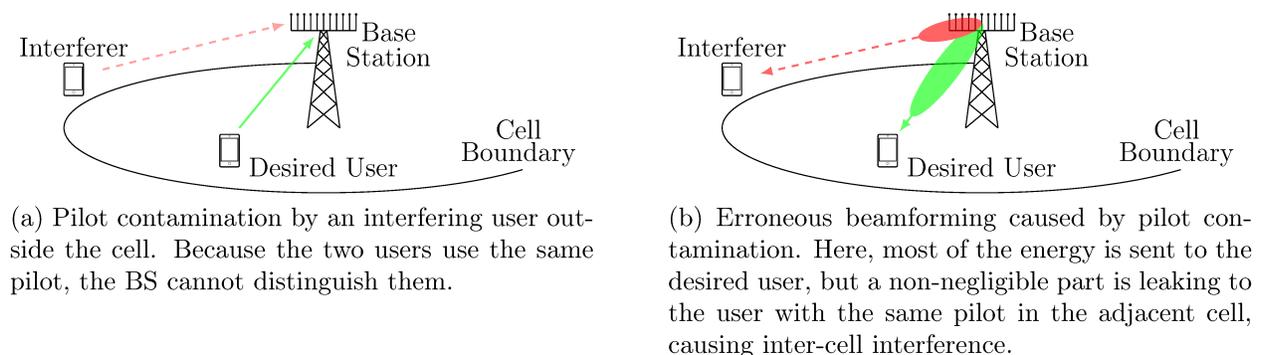


Fig. 1.2 Illustration of the effects of pilot contamination on beamforming. The green arrow represents the pilot sequence sent by the user inside the cell, whereas the red arrow represents the interfering pilot of the user with the same pilot sequence in an adjacent cell.

In recent years, attention has turned towards artificial intelligence, opening a new realm of possibilities for mmWave channel estimation. Several approaches based on deep neural networks (DNNs) have been investigated. For instance, [30] investigates the use of deep convolutional neural networks (CNNs) for channel estimation in mmWave massive MIMO systems with hybrid architecture. Three different CNN configurations are proposed for this purpose: one exploiting the spatial and frequency correlation of time-invariant channels; one exploiting the spatial, frequency, and time correlation of time-varying channels; and a last one designed to reduce the pilot overhead. DNNs have also been proposed for super-resolution channel and angle of arrival (AoA) estimation in massive MIMO systems [31], as well as

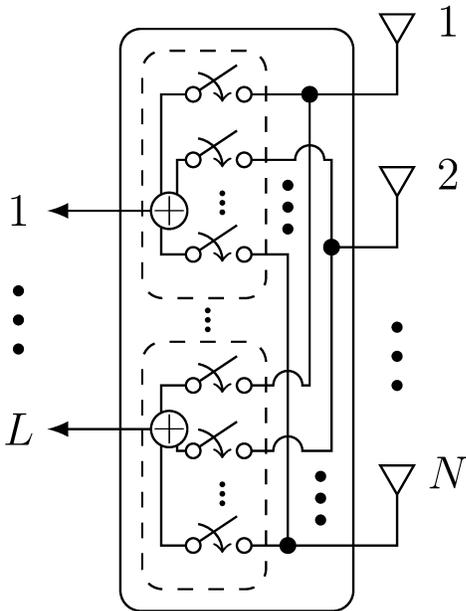
for channel estimation in such systems when implemented with mixed resolution analog-to-digital converters (ADCs) [32]. Deep learning along with pilot contamination is considered in [33] for a convolutional blind denoising network and in [34] to develop a new pilot allocation scheme. Bayesian learning has also been explored as an alternative solution to DNNs for mmWave channel estimation under various scenarios, including two-dimensional (2D) arrays of arbitrary geometry [35], off-grid models that characterize spatial sample mismatching [36], and again super-resolution channel estimation [37].

Other works consider realistic limitations that are often neglected in the usual models. For example, carrier frequency offset, which provides a measure of the imperfection of the time-frequency synchronization across devices, is taken into account in the angular sparsity channel estimation algorithm proposed in [38]. Also, hardware and channel state information (CSI) imperfections are dealt with in the beamforming schemes presented in [39] and [40].

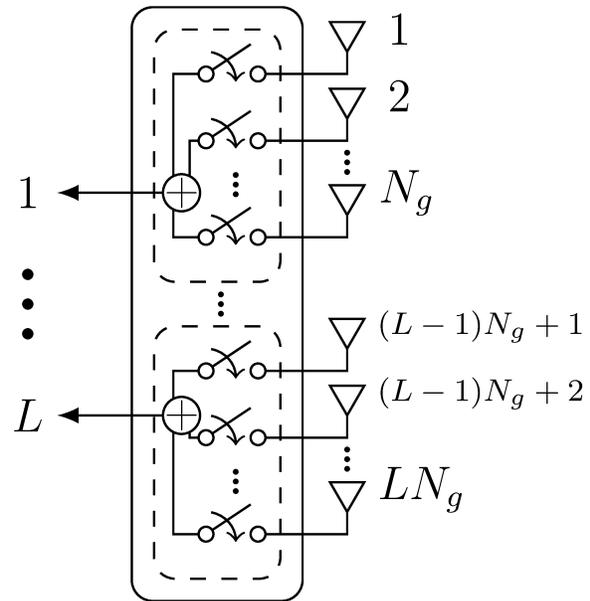
Finally, a new type of hybrid architecture where switches are used instead of phase shifters was recently proposed in [41]. Specifically, four different switching network configurations, reproduced in Fig. 1.3 for reference, and two phase shifter configurations are presented in [41], together with a compressed-sensing-based channel estimation method and a power consumption model. According to this model, all of the switching network configurations consume less power than a fully digital system, and most of them use less power than phase shifter networks. The configuration in Fig. 1.3a is also used in [42], where a hybrid precoder is designed for a mmWave MIMO system. In [43], the switching network in Fig. 1.3d is used with a matrix completion algorithm to estimate mmWave massive MIMO channels. Alternatively, it is possible to use switches in combination with phase shifters. This is the case in [44], where switches are coupled with constant, non-tunable phase shifters. The performance of this solution is almost as good as that of a system equipped with regular phase shifters. However, this setup can only be pre-designed for a restricted range of frequencies and is thus not suitable for broadband applications.

1.3 Thesis Objectives and Contributions

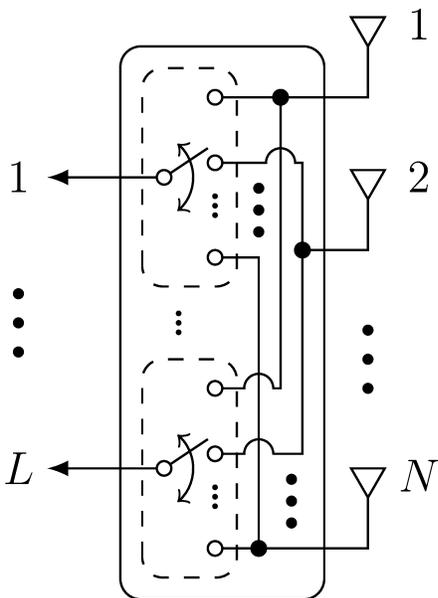
The channel estimation methods designed for massive MIMO systems with hybrid architecture discussed in the previous sections almost all present analog units equipped with phase shifters, as in Fig. 1.1. This framework enables the number of antennas to be increased while keeping a reduced number of RF chains, which is desirable since these are costly and



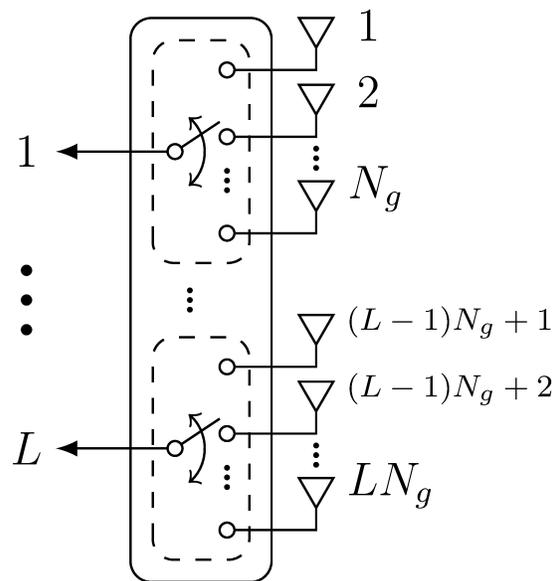
(a) In this architecture, every RF chain is fed a sum of signals coming from any antenna.



(b) In this architecture, every RF chain is fed a sum of signals coming from a subset of $N_g = N/L$ antennas.



(c) In this architecture, every RF chain is fed the signal of any one antenna.



(d) In this architecture, every RF chain is fed the signal of one antenna in a subset of $N_g = N/L$ antennas.

Fig. 1.3 Four hybrid-architecture configurations for massive MIMO transceivers that use switches instead of phase shifters.

not power-efficient. However, phase shifters exhibit some drawbacks, most notably their relatively slow speed of operation. Other types of hardware are more attractive than phase shifters in some aspects. This is the case with RF switches, like the ones presented in Fig. 1.3 and discussed in the previous section. In principle, hybrid architecture based on RF switches are cheaper, faster, and more power-efficient than those based on phase shifters [41]. However, more research needs to be conducted in order to design and better understand practical massive MIMO systems based on them.

The objective of this thesis is to develop and investigate a pilot-based channel estimation technique for mmWave massive MIMO channels with a switch-based hybrid architecture. Specifically, we consider the problem of uplink channel estimation of a massive MU MIMO system where the BS is equipped with a switch-based hybrid architecture and each MS is equipped with a single antenna. The main contributions can be listed in four points, as follows:

- New switch architectures are introduced to offer more control and flexibility on the signal computation while still retaining the advantages of switches described earlier.
- The uplink channel estimation problem for the switch-based MIMO system is formulated by extending the mean-square estimation framework presented in [17].
- A genetic algorithm (GA) is developed to solve the discrete optimization problem for the switch states and obtain the desired channel estimates.
- The performance of the proposed GA for channel estimation with switch-based hybrid architecture is compared to that of selected algorithms presented in [17] for channel estimation with phase shifter networks.

Our result show that in spite of their inherent simplicity, hybrid networks based on RF switches can estimate channels as well as those based on phase shifters, if not better than them in some conditions. Hence, considering their advantageous features, as mentioned above, RF switches hold the potential for a future class of mmWave massive MIMO transceivers with hybrid analog/digital architecture.

1.4 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 introduces the massive MIMO system and mmWave channel model used in this thesis and formulates the channel estimation as an optimization problem over a matrix space. Chapter 3 briefly reviews the solutions proposed in [17] for this problem. Chapter 4 presents different switch configurations, introduces new ones, reformulates the problem in terms of them, and presents the GA used to estimate channels with switching networks. Chapter 5 present the simulation results along with performance comparisons. Finally, Chapter 6 summarizes and concludes the thesis, and proposes future research prospects.

Chapter 2

System Model and Problem Formulation

In this chapter, the mathematical formalism used in this thesis and other related works to model mmWave massive MIMO systems and channels is presented, together with a few of their properties. Then, the pilot-based channel estimation problem for massive MIMO systems with hybrid architecture is formulated as a constrained optimization problem.

2.1 System Model

The present work investigates pilot-based channel estimation in a single-cell massive MIMO system for the case of uplink transmission from single-antenna MSs to a multi-antenna BS. In practice, provided the coherence time of the wireless channels is sufficiently large, TDD is used because the knowledge of the uplink channel then allows the downlink channel to be easily computed [4]. The presentation of the system model and the problem formulation will remain general in this chapter. Particular aspects of the model and problem that are specific to the use of switches will be presented in Chapter 4.

As shown in Fig. 2.1, the BS is equipped with N antennas and L RF chains. The antenna outputs are fed to a hybrid analog/digital processing system, consisting of an analog combiner, L individual RF chains, and a baseband processor, in that order. The analog combiner can be implemented in different ways. If it is made of phase shifters, as in most of the literature on hybrid massive MIMO, the narrow-band signals coming from the antennas are time-delayed, or phase-shifted, by different amounts, added together, and fed to the RF chains. If the analog combiner is built with switches instead, the signals can either go through different paths or be blocked, as will be explained in further details in Section 4.1.

In both cases, the analog combiner is used to convert a signal vector with N components to a lower-dimensional signal with L components, where $L < N$. In the context of massive MIMO, N can be quite large, e.g. on the order of 100, while typical values of L range from $N/10$ to $N/2$. The lower-dimensional signal vector at the analog combiner output is then fed to the RF chains, where its components are down-converted to baseband and digitally sampled by ADCs. The digital output signals from the RF chains are then fed to a baseband processor, which in turn completes the channel estimation.

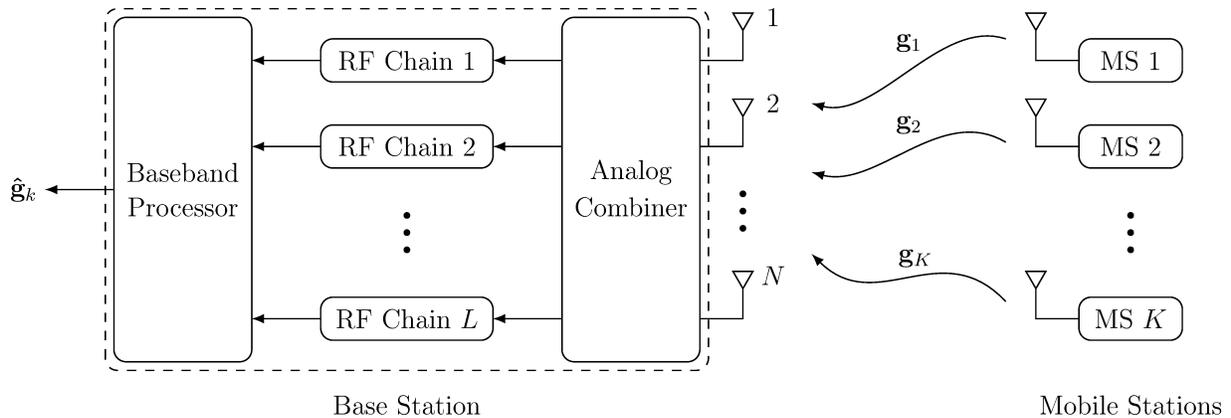


Fig. 2.1 Block diagram of the system model, consisting of a hybrid mmWave massive MIMO base station and K single-antenna mobile stations.

For the sake of simplicity, the pilot sequences used in this work are simple time-orthogonal sequences: when one MS sends a pilot, all the other MSs remain mute. This design greatly simplifies the problem formulation and analysis, allowing a single MS to be considered at any given symbol time, while still providing the same performance as other pilot designs [17]. Hence, the user indices are not used in the rest of this thesis.

The uplink training process is mathematically modelled as follows. The MS sends to the BS T pilot symbols $\varphi_t \in \mathbb{C}$, $t \in \{1, \dots, T\}$, known beforehand and all satisfying $|\varphi_t| = 1$. At time t , the BS receives a signal

$$\mathbf{r}_t = \sqrt{\rho} \mathbf{g} \varphi_t + \tilde{\mathbf{n}}_t, \quad (2.1)$$

where $\rho \in \mathbb{R}_+$ is the power of the pilot, $\mathbf{g} \in \mathbb{C}^{N \times 1}$ represents the uplink channel vector between the considered MS and the BS, and $\tilde{\mathbf{n}}_t \in \mathbb{C}^{N \times 1}$ is an additive white Gaussian noise (AWGN) vector with zero mean, unit variance, and complex circular Gaussian distribution,

i.e. $\tilde{\mathbf{n}}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ ¹. After being processed by the analog combiner and multiplied by the conjugate of the pilot symbol, the signal received at the baseband processor is

$$\mathbf{y}_t = \mathbf{F}_t(\sqrt{\rho}\mathbf{g}\varphi_t + \tilde{\mathbf{n}}_t)\bar{\varphi}_t = \mathbf{F}_t(\sqrt{\rho}\mathbf{g} + \mathbf{n}_t), \quad (2.2)$$

where $\mathbf{F}_t \in \mathbb{C}^{L \times N}$ represents the analog combiner at time t and $\mathbf{n}_t = \bar{\varphi}_t \tilde{\mathbf{n}}_t \in \mathbb{C}^{N \times 1}$ is an AWGN vector with the same statistics as $\tilde{\mathbf{n}}_t$, i.e. $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$.

The T signals successively received at the BS can be combined together if the following vectors and matrices are defined:

$$\mathbf{y}_c = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix}, \quad \mathbf{n}_c = \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_T \end{bmatrix}, \quad (2.3)$$

$$\mathbf{F}_c = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \vdots \\ \mathbf{F}_T \end{bmatrix}, \quad \mathbf{F}_d = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}_T \end{bmatrix}, \quad (2.4)$$

where $\mathbf{y}_c \in \mathbb{C}^{LT \times 1}$, $\mathbf{n}_c \in \mathbb{C}^{NT \times 1}$, $\mathbf{F}_c \in \mathbb{C}^{LT \times N}$, and $\mathbf{F}_d \in \mathbb{C}^{LT \times NT}$. These quantities then satisfy the following relation:

$$\mathbf{y}_c = \sqrt{\rho}\mathbf{F}_c\mathbf{g} + \mathbf{F}_d\mathbf{n}_c. \quad (2.5)$$

In practice, the channel vector \mathbf{g} exhibits a random character. Here, a second order characterization with zero mean, i.e. $\mathbb{E}[\mathbf{g}] = \mathbf{0}$, and covariance matrix $\mathbf{R} := \mathbb{E}[\mathbf{g}\mathbf{g}^H]$ are considered. Without loss of generality, the channel vector \mathbf{g} is assumed to be normalized as $\mathbb{E}[\mathbf{g}^H\mathbf{g}] = \text{tr}(\mathbf{R}) = N$. We also assume that the channel vector \mathbf{g} and the additive noise terms \mathbf{n}_t are uncorrelated, i.e. $\mathbb{E}[\mathbf{g}\mathbf{n}_t^H] = \mathbf{0}$ for all t .

¹In this work, without loss of generality, we set the noise variance to 1. Any desired SNR level can be achieved by adjusting the pilot power ρ accordingly.

2.2 Channel Model

The model used in this thesis to represent the low-scattering mmWave channel between the MS and the multi-antenna BS [13] is a linear combination of array response vectors:

$$\mathbf{g} = \frac{1}{\sqrt{\sigma_\alpha^2 P}} \sum_{p=1}^P \alpha_p \mathbf{a}(\theta_p), \quad (2.6)$$

where P is the number of propagation paths, $\alpha_p \sim \mathcal{CN}(0, \sigma_\alpha^2)$ and $\theta_p \sim \mathcal{U}(0, 2\pi)$ are respectively the complex gain, with variance σ_α^2 , and the AoA of the p^{th} path, and the function $\mathbf{a} : \mathbb{R} \rightarrow \mathbb{C}^{N \times 1}$ is the array response of the antenna array used at the BS, as further described below. In (2.6), all the path gains α_p and AoAs θ_p are statistically independent from each other.

While the channel estimation algorithm developed in this thesis works with antenna arrays of arbitrary shape, for simplicity, only uniform linear arrays (ULAs) are considered in our presentation. For ULAs, the array response returns a column vector expressed as

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 \\ e^{j\frac{2\pi d}{\lambda} \sin \theta} \\ \vdots \\ e^{j(N-1)\frac{2\pi d}{\lambda} \sin \theta} \end{bmatrix}, \quad (2.7)$$

where the AoA θ is measured from a line that is perpendicular to the axis of the array, d is the distance between the antennas, and λ is the wavelength of the signal at the operating frequency. The form of the array response vector in (2.7) can be derived from basic physical principles by assuming plane wave propagation along with the ULA geometry, as represented in Fig. 2.2. Specifically, the difference between the complex phases of two entries in the vector $\mathbf{a}(\theta)$ is defined to be equal to the difference between the phases of the signals received at the two antennas corresponding to the entries. If the phase difference between two adjacent antennas is denoted by $\Delta\phi$, then from the problem geometry

$$\Delta\phi = \frac{2\pi d}{\lambda} \sin \theta. \quad (2.8)$$

This yields expression (2.7), where by convention, the first entry of an array response vector is set to 1.

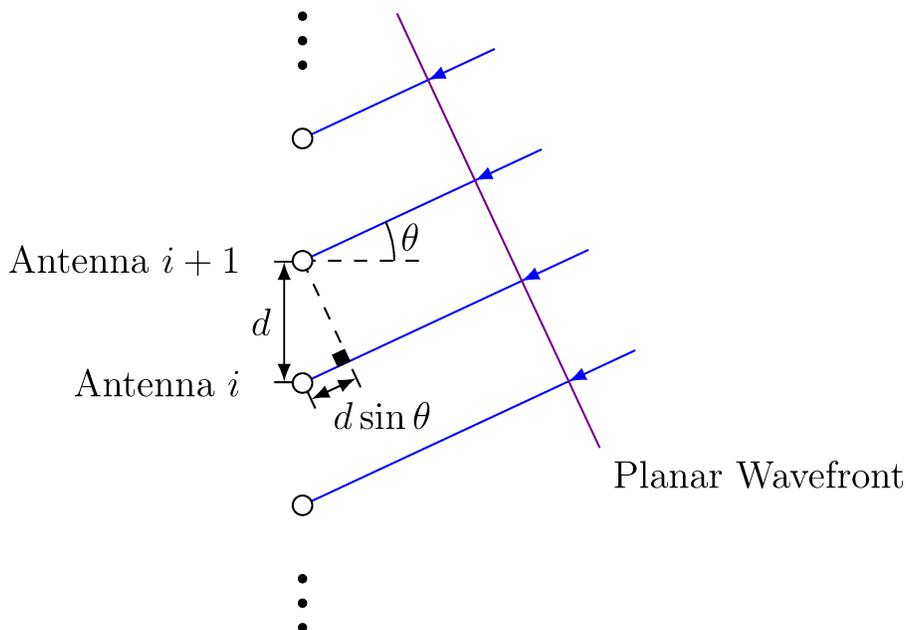


Fig. 2.2 Geometrical representation of a plane wave signal impinging on a ULA at an angle θ . The small circles represent the antennas and the blue lines represent the signal propagation rays, which are approximated as parallel to each other (i.e. forming a plane wave) because the distance between the BS and the MS is much greater than the antenna aperture, given by $(N - 1)d$.

ULAs are simple antenna arrays that allow easy mathematical modelling. However, more sophisticated array configurations can exhibit a better performance in some situations. One important limitation of ULAs is that they cannot generate one-dimensional (1D), pencil-like beam patterns. If the antennas are approximated as isotropic radio receivers, the rotational symmetry around the axis of the array implies that the main beam in the beamforming pattern has a conical shape. Two signals whose direction of arrival form the same angle with the axis of the array are thus indistinguishable by a ULA.

An example of a simple array configuration that allows control over two degrees of freedom is a uniform planar array (UPA). Since the antennas in a UPA are located on a planar grid in two dimensions, the array can steer a pencil-shaped beam with adjustable elevation and azimuth angles [3]. However, this kind of antenna array, when placed in free space, actually steers two beams that are symmetric to each other with respect to the plane of the array. In practice, this is avoided by blocking one of the two beams from one UPA and by having different UPAs serving specific sectors around an antenna. Overall, the ability to tailor a

narrow beam along elevation and azimuth has the clear advantage of increasing the received power from the desired source and reducing the interference from the neighbouring users and cells.

In practice however, antennas are not isotropic receivers and although their number and density are quite large in the case of mmWave communications, the beamforming patterns sometimes have defects. Other types of antenna arrays designed to perform within these limitations include cylindrical and distributed arrays [5]. Cylindrical arrays can beamform as well as UPAs, but have the added advantage of requiring less space for the same number of antennas. Different kinds of sparse cylindrical arrays are studied and compared in [45]. In contrast, distributed arrays can be used to take advantage of the geometry of a specific location.

Finally, the channel estimation techniques presented in this work make use of the spatial channel covariance matrix, $\mathbf{R} = \mathbb{E}[\mathbf{g}\mathbf{g}^H]$. It is shown in Appendix A that using the definition \mathbf{g} in (2.6), \mathbf{R} can be expressed as

$$\mathbf{R} = \frac{1}{P} \sum_{p=1}^P \mathbf{a}(\theta_p)\mathbf{a}(\theta_p)^H \quad (2.9)$$

and has rank at most P . In the special case of a ULA, with the array response vector given by (2.7), the $(m, n)^{\text{th}}$ entry of matrix \mathbf{R} takes the form

$$[\mathbf{R}]_{(m,n)} = \frac{1}{P} \sum_{p=1}^P e^{(m-n)j\frac{2\pi d}{\lambda} \sin \theta_p}. \quad (2.10)$$

It can further be noted that the diagonal entries $[\mathbf{R}]_{(n,n)}$ are all 1, so that \mathbf{R} admits the normalization $\text{tr}(\mathbf{R}) = N$, as desired.

2.3 Problem Formulation

The goal of this thesis is to estimate the channel \mathbf{g} from the MS to the BS by producing an estimated channel $\hat{\mathbf{g}}$ that is as close as possible to the real channel in the minimum mean square error (MSE) sense. In the context of the hybrid analog/digital massive MIMO system model presented in Section 2.1, this can be formulated as the following minimization

problem:

$$\begin{aligned}
& \underset{\mathbf{F}_1, \dots, \mathbf{F}_T, \mathbf{W}}{\text{minimize}} \quad \mathbb{E}[\|\mathbf{g} - \hat{\mathbf{g}}\|^2] & (2.11) \\
& \text{subject to} \quad \hat{\mathbf{g}} = \mathbf{W}\mathbf{y}_c \\
& \quad \mathbf{y}_c = \sqrt{\rho}\mathbf{F}_c\mathbf{g} + \mathbf{F}_d\mathbf{n}_c \\
& \quad \mathbf{F}_t \in \mathcal{F} \quad \forall t \in \{1, \dots, T\}
\end{aligned}$$

where $\mathbf{W} \in \mathbb{C}^{N \times LT}$ represents the baseband processor matrix and \mathcal{F} is the set of feasible matrices for the analog component, which depends on the specific hardware architecture used at the BS. For example, the entries of matrices representing phase shifters are restrained to unit-modulus complex numbers and those of matrices representing switching networks are taken from a finite set, as specified in Chapter 4.

As a first step in the constrained minimization of the MSE in (2.11), it is possible to find an expression for the optimal \mathbf{W} in terms of \mathbf{F}_t , which is given by

$$\mathbf{W}_{\text{opt}} = \sqrt{\rho}\mathbf{R}\mathbf{F}_c^H (\rho\mathbf{F}_c\mathbf{R}\mathbf{F}_c^H + \mathbf{F}_d\mathbf{F}_d^H)^{-1}. \quad (2.12)$$

The proof for the optimality of this expression is presented in detail in Appendix B. Using this expression and defining

$$\begin{aligned}
\mathcal{M}(\mathcal{F}) &= \mathbb{E}[\|\mathbf{g} - \hat{\mathbf{g}}\|^2] \\
&= \mathbb{E}[\|\mathbf{g} - \mathbf{W}_{\text{opt}}\mathbf{y}_c\|^2],
\end{aligned} \quad (2.13)$$

where $\mathcal{F} := (\mathbf{F}_1, \dots, \mathbf{F}_T)$, the objective function in (2.11) can be rewritten as

$$\mathcal{M}(\mathcal{F}) = \text{tr}\left(\mathbf{R} - \rho\mathbf{R}\mathbf{F}_c^H (\rho\mathbf{F}_c\mathbf{R}\mathbf{F}_c^H + \mathbf{F}_d\mathbf{F}_d^H)^{-1} \mathbf{F}_c\mathbf{R}\right). \quad (2.14)$$

It is interesting to note that using the optimal baseband processor places an upper bound on the objective function. As shown in Appendix C, for any $\mathcal{F} \in \mathcal{F}^T$ and any $\mathcal{F} \subseteq \mathbb{C}^{L \times N}$, the following inequality holds:

$$0 \leq \mathcal{M}(\mathcal{F}) \leq N. \quad (2.15)$$

In the special case where only one pilot symbol is used, the cost function (2.14) can be simplified. This special case is also explored in Section 3.2.1, but the approach there assumes that \mathbf{R} is invertible, which is not necessary here. If $T = 1$, it can be seen from (2.4) that the

matrices $\mathbf{F}_1 = \mathbf{F}_c = \mathbf{F}_d =: \mathbf{F}$ become one and the same. This implies that

$$\mathcal{M}(\mathcal{F}) = \text{tr}\left(\mathbf{R} - \rho\mathbf{R}\mathbf{F}^{\text{H}}(\rho\mathbf{F}\mathbf{R}\mathbf{F}^{\text{H}} + \mathbf{F}\mathbf{F}^{\text{H}})^{-1}\mathbf{F}\mathbf{R}\right) \quad (2.16)$$

$$= \text{tr}\left(\mathbf{R} - \rho\mathbf{R}\mathbf{F}^{\text{H}}(\mathbf{F}(\rho\mathbf{R} + \mathbf{I})\mathbf{F}^{\text{H}})^{-1}\mathbf{F}\mathbf{R}\right). \quad (2.17)$$

One of the requirements that appears from the above expression is that the $L \times N$ matrix \mathbf{F} (here, $L < N$) needs to be full-rank, otherwise the matrix $\mathbf{F}(\rho\mathbf{R} + \mathbf{I})\mathbf{F}^{\text{H}}$ would not be invertible. From this point, it is assumed that this condition is satisfied, i.e. that the rows of \mathbf{F}_t are linearly independent for all t . Besides, we note that the term $\rho\mathbf{R} + \mathbf{I}$ is Hermitian positive-definite because for any deterministic $\mathbf{z} \in \mathbb{C} \setminus \{\mathbf{0}\}$,

$$\begin{aligned} \mathbf{z}^{\text{H}}(\rho\mathbf{R} + \mathbf{I})\mathbf{z} &= \rho\mathbf{z}^{\text{H}}\mathbb{E}[\mathbf{g}\mathbf{g}^{\text{H}}]\mathbf{z} + \mathbf{z}^{\text{H}}\mathbf{z} \\ &= \rho\mathbb{E}\left[(\mathbf{g}^{\text{H}}\mathbf{z})^{\text{H}}(\mathbf{g}^{\text{H}}\mathbf{z})\right] + \mathbf{z}^{\text{H}}\mathbf{z} \\ &= \rho\underbrace{\mathbb{E}\left[\|\mathbf{g}^{\text{H}}\mathbf{z}\|^2\right]}_{\geq 0} + \underbrace{\|\mathbf{z}\|^2}_{> 0} > 0. \end{aligned} \quad (2.18)$$

The matrix $\rho\mathbf{R} + \mathbf{I}$ thus has a unique Hermitian positive-definite square root \mathbf{M} , such that $\mathbf{M}^{\text{H}} = \mathbf{M}$ and $\rho\mathbf{R} + \mathbf{I} = \mathbf{M}^2 = \mathbf{M}\mathbf{M}^{\text{H}}$. Next, using the fact that if \mathbf{A}^{-1} exists, then $\mathbf{A}^{-1} = \mathbf{A}^{\dagger}$, along with the general properties of the pseudo-inverse that if $\mathbf{A}^{\text{H}} = \mathbf{B}$, then $(\mathbf{A}\mathbf{B})^{\dagger} = \mathbf{B}^{\dagger}\mathbf{A}^{\dagger}$, and that $(\mathbf{A}^{\dagger})^{\text{H}} = (\mathbf{A}^{\text{H}})^{\dagger}$, the term $(\mathbf{F}(\rho\mathbf{R} + \mathbf{I})\mathbf{F}^{\text{H}})^{-1}$ can be rewritten as

$$\begin{aligned} (\mathbf{F}(\rho\mathbf{R} + \mathbf{I})\mathbf{F}^{\text{H}})^{-1} &= (\mathbf{F}\mathbf{M}\mathbf{M}^{\text{H}}\mathbf{F}^{\text{H}})^{\dagger} \\ &= \left((\mathbf{F}\mathbf{M})(\mathbf{F}\mathbf{M})^{\text{H}}\right)^{\dagger} \\ &= (\mathbf{F}\mathbf{M})^{\dagger\text{H}}(\mathbf{F}\mathbf{M})^{\dagger}. \end{aligned} \quad (2.19)$$

The objective function is therefore

$$\begin{aligned} \mathcal{M}(\mathcal{F}) &= \text{tr}\left(\mathbf{R} - \rho\mathbf{R}\mathbf{F}^{\text{H}}(\mathbf{F}\mathbf{M})^{\dagger\text{H}}(\mathbf{F}\mathbf{M})^{\dagger}\mathbf{F}\mathbf{R}\right) \\ &= \text{tr}(\mathbf{R}) - \rho\text{tr}\left(\left((\mathbf{F}\mathbf{M})^{\dagger}\mathbf{F}\mathbf{R}\right)^{\text{H}}(\mathbf{F}\mathbf{M})^{\dagger}\mathbf{F}\mathbf{R}\right) \\ &= N - \rho\left\|\left(\mathbf{F}\mathbf{M}\right)^{\dagger}\mathbf{F}\mathbf{R}\right\|_{\text{F}}^2, \end{aligned} \quad (2.20)$$

where the last line follows from the definition of the Frobenius norm, $\|\mathbf{A}\|_{\text{F}} := \sqrt{\text{tr}(\mathbf{A}^{\text{H}}\mathbf{A})}$,

and the normalization condition previously imposed on the covariance matrix \mathbf{R} , i.e. $\text{tr}(\mathbf{R}) = N$.

The above expression of the MSE objective function is useful because it greatly simplifies (2.14) when only one pilot symbol is used. In addition to this use, its derivation allowed to find important requirements for the analog combiner matrices, which are useful for the practical implementation of the algorithm presented in Chapter 4.

Chapter 3

Review of Existing Solutions

In this chapter, existing solutions for the channel estimation problem defined in the previous chapter are presented. Following an overview of common preliminary steps, four methods are described. The first one works only in the special case where one pilot symbol is used, whereas the other three work with more pilot symbols. The idea behind each of these techniques is to first find an unconstrained solution for the optimal analog combiners and then to project the solution on the feasible set. The projection is detailed after the presentation of the four methods along with a way to quantize the entries of the combiners.

3.1 Preliminary Steps

The optimization problem (2.11) is formulated in [17] for a hybrid system with phase shifters as the analog components, that is, where the feasible set for the analog combiner matrices is

$$\mathcal{F} = \{e^{j\theta} \mid \theta \in [0, 2\pi[\}^{L \times N}. \quad (3.1)$$

The solution approach presented in [17] begins by temporarily discarding the constant-magnitude constraint imposed on the entries of on the analog combiners \mathbf{F}_t , which is equivalent to replacing the constraint set \mathcal{F} by $\mathbb{C}^{L \times N}$. Then, the optimal unconstrained combiners are computed using one of the four methods reviewed in the following section. Finally, these combiners are projected onto the original feasible set \mathcal{F} .

As a preliminary step before the development of any of these four methods, the Woodbury matrix identity [46] is used under the assumption that \mathbf{R} is full-rank to simplify the MSE

expression (2.14) to

$$\mathcal{M}(\mathcal{F}) = \text{tr} \left(\left(\mathbf{R}^{-1} + \rho \mathbf{F}_c^H (\mathbf{F}_d \mathbf{F}_d^H)^{-1} \mathbf{F}_c \right)^{-1} \right). \quad (3.2)$$

In applying this identity, it is implicitly assumed that the matrices \mathbf{R} and $\mathbf{F}_d \mathbf{F}_d^H$ are non-singular.

The singular value decomposition (SVD) of the matrix representing the analog combiner is then computed as $\mathbf{F}_t = \mathbf{U}_t \mathbf{\Sigma}_t \mathbf{V}_t^H$, where $\mathbf{U}_t \in \mathbb{C}^{L \times L}$ is the left singular vector matrix, $\mathbf{\Sigma}_t \in \mathbb{R}^{L \times N}$ is the singular value matrix, and $\mathbf{V}_t \in \mathbb{C}^{N \times N}$ is the right singular vector matrix, and where \mathbf{U}_t and \mathbf{V}_t are unitary, i.e. $\mathbf{U}_t^H \mathbf{U}_t = \mathbf{I}_L$ and $\mathbf{V}_t^H \mathbf{V}_t = \mathbf{I}_N$, and $\mathbf{\Sigma}_t$ is diagonal with non-negative entries. Also denote by $\mathbf{V}_{t,L}$ the matrix composed of the L first columns of \mathbf{V}_t . Then, it is shown in [17] that replacing \mathbf{F}_t by $\mathbf{V}_{t,L}^H$ for every t in (3.2) yields the same MSE. It is thus possible to use \mathbf{F}_t and $\mathbf{V}_{t,L}^H$ interchangeably in terms of estimation performance.

This fact can be used to further simplify the MSE expression as

$$\mathcal{M}(\mathcal{F}) = \text{tr} \left(\left(\mathbf{\Lambda}^{-1} + \rho \sum_{t=1}^T \tilde{\mathbf{V}}_t \tilde{\mathbf{V}}_t^H \right)^{-1} \right), \quad (3.3)$$

where $\tilde{\mathbf{V}}_t := \mathbf{U}^H \mathbf{V}_{t,L}$, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$ and \mathbf{U} are respectively the eigenvalue matrix and the unitary eigenvector matrix of $\mathbf{R} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$. Accordingly, $\mathbf{U}^H \mathbf{U} = \mathbf{I}_N$ while $\mathbf{\Lambda}$ is defined to have its diagonal entries ordered non-increasingly: $\lambda_1 \geq \dots \geq \lambda_N \geq 0$. Note that the entries of $\mathbf{\Lambda}$ are real and non-negative because \mathbf{R} is Hermitian and positive semi-definite.

It should be noted that the assumption that \mathbf{R} be full-rank, i.e. invertible or positive-definite in this case, is not always a valid approximation for mmWave channels, because they are characterized by a small amount of scattering [13]. The poor scattering reduces the number P of propagation paths between the wireless devices, and thus the rank of \mathbf{R} . However, diagonal loading techniques [47] can be employed to overcome this issue without significantly affecting the proposed methods.

3.2 Unconstrained Optimization Methods

Herein, four methods developed in [17] to find the optimal unconstrained combiners are reviewed briefly: one that works only with one pilot symbol, and three more general ones that work with an arbitrary number of pilot symbols. These last three methods are respectively called Block Selection, Sequential Optimization (SO), and Alternating Optimization (AO).

3.2.1 Single-Pilot-Symbol Unconstrained Optimization

The first estimation method developed in [17] works only when one pilot symbol is used, i.e. when $T = 1$. The MSE in (3.3) reduces in this case to

$$\mathcal{M}(\mathcal{F}) = \text{tr} \left(\left(\mathbf{\Lambda}^{-1} + \rho \tilde{\mathbf{V}}_1 \tilde{\mathbf{V}}_1^H \right)^{-1} \right). \quad (3.4)$$

It is then possible to use the Woodbury identity again together with basic properties of the trace operator and the fact that $\tilde{\mathbf{V}}_t^H \tilde{\mathbf{V}}_t = \mathbf{I}_L$ to write the MSE as

$$\mathcal{M}(\mathcal{F}) = \text{tr}(\mathbf{\Lambda}) - \text{tr} \left(\tilde{\mathbf{V}}_1^H \mathbf{\Lambda}^2 \tilde{\mathbf{V}}_1 \left(\tilde{\mathbf{V}}_1^H (\mathbf{\Lambda} + \rho^{-1} \mathbf{I}_N) \tilde{\mathbf{V}}_1 \right)^{-1} \right). \quad (3.5)$$

The problem of minimizing the MSE is thus equivalent in this case to maximizing the second term on the right hand side in (3.5) over $\tilde{\mathbf{V}}_1 \in \mathbb{C}^{N \times L}$ under the constraint that $\tilde{\mathbf{V}}_1^H \tilde{\mathbf{V}}_1 = \mathbf{I}_L$. However, the objective function of this new problem is simply the Block Generalized Rayleigh Quotient of $\tilde{\mathbf{V}}_1$ with respect to the pair of matrices, or pencil, $(\mathbf{\Lambda}^2, \mathbf{\Lambda} + \rho^{-1} \mathbf{I}_N)$. As explained in [17], the solution of this problem is a matrix composed of the eigenvectors corresponding to the L greatest generalized eigenvalues of the pencil. This solution turns out to be the L first columns of \mathbf{I}_N : $\tilde{\mathbf{V}}_1^{\text{opt}} = \mathbf{I}_{N \times L}$. Using the definition of $\tilde{\mathbf{V}}_t = \mathbf{U}^H \mathbf{V}_{t,L}$ and the fact that \mathbf{U} is unitary, the associated analog combiner matrix is found to be

$$\mathbf{F}_1^{\text{opt}} = (\mathbf{V}_{t,L}^{\text{opt}})^H = (\mathbf{U}_{[1:L]})^H. \quad (3.6)$$

The corresponding minimum value of the MSE is given by

$$\mathcal{M}(\mathcal{F}) = N - \sum_{l=1}^L \frac{\lambda_l^2}{\lambda_l + 1/\rho} \quad (3.7)$$

3.2.2 Block Selection

The authors of [17] then proceed with three techniques that use multiple pilot symbols to estimate the channel. Doing so allows better performances in general than the above single-pilot-symbol method.

The first multiple-pilot-symbol method, called Block Selection, is inspired by the result of the previous subsection and extends the idea of the method therein by taking the eigenvectors associated with the LT largest generalized eigenvalues of the pencil $(\mathbf{\Lambda}^2, \mathbf{\Lambda} + \rho^{-1}\mathbf{I}_N)$ to produce T solutions $\tilde{\mathbf{V}}_1^{\text{opt}}, \dots, \tilde{\mathbf{V}}_T^{\text{opt}}$. The T analog combiner matrices $\mathbf{F}_1, \dots, \mathbf{F}_T$ can then be computed from \mathbf{U} for all $t \in \{1, \dots, T\}$ as

$$\mathbf{F}_t^{\text{opt}} = (\mathbf{U}_{[(t-1)L+1:tL]})^{\text{H}}. \quad (3.8)$$

This method has a very low complexity, but is less effective than the following ones due to its heuristic nature.

3.2.3 Sequential Optimization

The second method to take advantage of more than one pilot symbol is called Sequential Optimization (SO) and is the one used as a comparison benchmark in our simulation experiments reported in Chapter 5. Like the previous methods, it aims at minimizing the MSE in (3.3) under the constraint that $\tilde{\mathbf{V}}_t^{\text{H}}\tilde{\mathbf{V}}_t = \mathbf{I}_L \forall t \in \{1, \dots, T\}$.

For this method, the MSE is rewritten as $\mathcal{M}(\mathcal{F}) = \mathcal{M}_T(\mathcal{F})$, where for $t \in \{1, \dots, T\}$, we define

$$\mathcal{M}_t(\mathcal{F}) := \text{tr} \left(\left(\mathbf{\Gamma}_t^{-1} + \rho \tilde{\mathbf{V}}_t \tilde{\mathbf{V}}_t^{\text{H}} \right)^{-1} \right), \quad (3.9)$$

and

$$\mathbf{\Gamma}_t^{-1} := \begin{cases} \mathbf{\Gamma}_{t-1}^{-1} + \rho \tilde{\mathbf{V}}_{t-1} \tilde{\mathbf{V}}_{t-1}^{\text{H}}, & t > 1 \\ \mathbf{\Lambda}^{-1}, & t = 1. \end{cases} \quad (3.10)$$

It can be easily verified that the matrix $\mathbf{\Gamma}_t$ so defined remains positive definite (and hence invertible) as long as \mathbf{R} is also positive definite (i.e. $\lambda_i > 0$ for all $i \in \{1, \dots, N\}$).

The basic idea behind the SO method is to minimize the intermediate MSE in (3.9) sequentially, i.e. starting with $t = 1$, finding the minimizer $\tilde{\mathbf{V}}_1^{\text{opt}}$ of $\mathcal{M}_1(\mathcal{F})$, substituting this minimizer in $\mathcal{M}_2(\mathcal{F})$, and then proceeding to $t = 2$, and so on, up to $t = T$. For each of these T steps, the same procedure as for the single-pilot-symbol optimization is followed to

find $\tilde{\mathbf{V}}_t^{\text{opt}}$, which is the matrix composed of the eigenvectors corresponding to the L largest generalized eigenvalues of the pencil $(\mathbf{\Gamma}_t^2, \mathbf{\Gamma}_t + \rho^{-1}\mathbf{I}_N)$. Interestingly, it can be shown that the matrices $\mathbf{\Gamma}_t$ at each iterations are diagonal matrices, although their diagonal entries are not necessarily in decreasing order. Furthermore, the optimal $\tilde{\mathbf{V}}_t^{\text{opt}}$ turns out to be a subset of the columns of the identity matrix in a certain order. If $\{\tilde{\gamma}_1, \dots, \tilde{\gamma}_N\}$ are the eigenvalues of the pencil and $j_1, \dots, j_N \in \{1, \dots, N\}$ are a set of indices such that $\tilde{\gamma}_{j_1} \geq \dots \geq \tilde{\gamma}_{j_N}$, then $\tilde{\mathbf{V}}_t^{\text{opt}} = (\mathbf{I}_N)_{[j_1, \dots, j_L]}$. Using the definition of $\tilde{\mathbf{V}}_t$ in (3.3), the analog combiner is found to be

$$\mathbf{F}_t^{\text{opt}} = (\mathbf{U}_{[j_1, \dots, j_L]})^H. \quad (3.11)$$

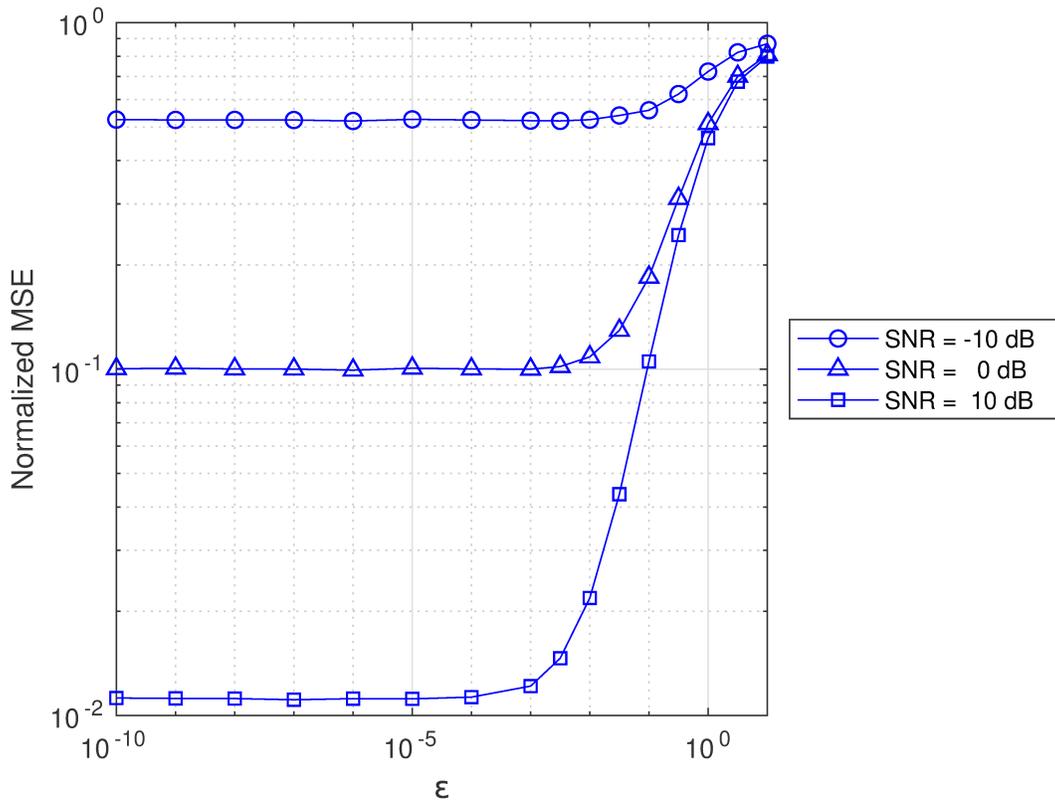


Fig. 3.1 Effect of diagonal loading on the normalized MSE of the channel estimation performed by the SO method as a function of the parameter ε and for different SNRs. The parameters $L = 8$, $N = 64$, $T = 1$, and $P = 6$ are used.

While this method is not globally optimal, it tends to produce better channel estimates than the previous one. Also, it is as effective as the following one, but it has a lower complexity. The SO methods assume that matrix \mathbf{R} is full-rank, and more precisely positive-

definite. In the case where \mathbf{R} is not full-rank, diagonal loading as discussed in [48] can be used as a workaround solution, which in practice amounts to substitute \mathbf{R} for $\mathbf{R} + \varepsilon \mathbf{I}_N$ for some small $\varepsilon > 0$. To illustrate the effect of diagonal loading, simulation results for the SO method with a rank-deficient covariance matrix¹ are shown in Fig. 3.1. These and other similar results indicate that using values smaller than approximately 10^{-4} for ε works well, while larger values worsen the results gradually as ε increases, because the modified matrix $\mathbf{R} + \varepsilon \mathbf{I}_N$ gets further away from the true covariance matrix \mathbf{R} when this is the case.

3.2.4 Alternating Optimization

The fourth and final method, called Alternating Optimization (AO), is an iterative one. During each iteration, each matrix $\tilde{\mathbf{V}}_j$, $j \in \{1, \dots, T\}$, is optimized while the other matrices $\{\tilde{\mathbf{V}}_t | t \neq j\}$ are fixed. The iterative process stops when the MSE stabilizes within a given tolerance.

To achieve this, the MSE in (3.3) is rewritten as

$$\mathcal{M}(\mathcal{F}) = \text{tr} \left(\left(\mathbf{Q}_j^{-1} + \rho \tilde{\mathbf{V}}_j \tilde{\mathbf{V}}_j^H \right)^{-1} \right), \quad (3.12)$$

where for $j \in \{1, \dots, T\}$, we define

$$\mathbf{Q}_j^{-1} = \mathbf{\Lambda}^{-1} + \rho \sum_{t=1, t \neq j}^T \tilde{\mathbf{V}}_t \tilde{\mathbf{V}}_t^H. \quad (3.13)$$

Note that the MSE as expressed in (3.12) does not actually depend on j , but its form allows the same procedure as for the single-pilot-symbol optimization to be directly used². This procedure is therefore employed to optimize over the matrix $\tilde{\mathbf{V}}_j$ during each iteration. This AO approach is presented explicitly in Algorithm 1, where the relative decrease in MSE is used as a stopping criterion. During our simulations, a value of $\varepsilon_{\text{tol}} = 10^{-3}$ for the algorithm provided good results. This method exhibits the same performance as the previous one, but has a higher complexity.

¹The rank-deficient covariance matrix is obtained based on the model in (2.6) where the number of paths, $P = 6$, is chosen to be less than the number of antennas, $N = 64$. The simulation methodology used in our work will be presented in details in Section 5.1

²This remains true as long as the matrices $\tilde{\mathbf{V}}_t$ are properly initialized in the AO algorithm, i.e. the product $\tilde{\mathbf{V}}_t \tilde{\mathbf{V}}_t^H$ must be diagonal.

Algorithm 1: Alternating optimization to design optimal unconstrained combiners for multiple trainings

Input: $L, N, T, \mathbf{R} = \mathbf{U}\mathbf{A}\mathbf{U}^H, \rho, \varepsilon_{\text{tol}}$

```

1 for  $t \in \{1, \dots, T\}$  do
2   |  $\tilde{\mathbf{V}}_t^{(0)} = \mathbf{I}_{N \times L}$  // Initialization
3 end
4  $\mathcal{M}_0 := \mathcal{M}(\mathcal{F})$  // Using (3.12)
5  $n \leftarrow 1$ 
6 keepLooping  $\leftarrow$  True
7 while keepLooping is True do
8   | for  $j \in \{1, \dots, T\}$  do
9     | Compute  $\mathbf{Q}_j^{-1}$  as in (3.13)
10    | Solve for the optimal  $\tilde{\mathbf{V}}_j^{(n)}$  in (3.12) using the same method as for the
11    | single-pilot-symbol optimization, described in Section 3.2.1
12  end
13   $\mathcal{M}_n := \mathcal{M}(\mathcal{F})$  // Using (3.12)
14  if  $\frac{|\mathcal{M}_n - \mathcal{M}_{n-1}|}{\mathcal{M}_{n-1}} < \varepsilon_{\text{tol}}$  then
15    | keepLooping  $\leftarrow$  False
16  else
17    |  $n \leftarrow n + 1$ 
18  end
19 for  $t \in \{1, \dots, T\}$  do
20   |  $\mathbf{F}_t^{\text{opt}} = (\mathbf{U}\tilde{\mathbf{V}}_t^{(n)})^H$ 
21 end

```

Output: Optimal unconstrained RF combiners for multiple trainings,
 $\{\mathbf{F}_t^{\text{opt}} \mid t \in \{1, \dots, T\}\}$

3.3 Projection onto the Set of Constrained Combiners and Quantization

All the above optimization methods disregard the phase-only constraint of the original problem. The solutions must therefore be adapted to be used in an RF combiner. To do so, [17] simply projects the entries of $\mathbf{F}_t^{\text{opt}}$ onto the unit circle in the complex plane \mathbb{C} . That is, if the $(l, n)^{\text{th}}$ entry of the t^{th} optimal unconstrained combiner is $[\mathbf{F}_t^{\text{opt}}]_{(l,n)} = r_{l,n}e^{j\theta_{l,n}}$, where $r_{l,n} \in \mathbb{R}_+$ and $\theta_{l,n} \in [0, 2\pi[$, then the projected combiner is

$$[\mathbf{F}_t^{\text{proj}}]_{(l,n)} = e^{j\theta_{l,n}}. \quad (3.14)$$

Furthermore, if the phase shifters are implemented to provide only a finite number of possible phase delays, the entries of the RF combiners must be quantized. Quantization of numbers on the complex unit circle amounts to selecting the closest element of a finite subset of available points on the circle. The subset used in this work for the quantization is the one with $N_Q := 2^{N_B}$ points placed equidistantly along the circle, i.e. with angular separation of $\Delta\theta = 2\pi/N_Q$, and including the point $1 \in \mathbb{C}$, where N_B is the number of quantization bits. More precisely, this set is expressed as

$$\{e^{jn\Delta\theta} \mid n \in \{0, \dots, N_Q - 1\}\}. \quad (3.15)$$

When quantized to N_B bits, the entry $[\mathbf{F}_t^{\text{proj}}]_{(l,n)}$ becomes

$$[\mathbf{F}_t^{N_B \text{ bits}}]_{(l,n)} = e^{j\Delta\theta \lceil \frac{\theta_{l,n}}{\Delta\theta} \rceil}, \quad (3.16)$$

where $\lceil x \rceil$ is x rounded to the nearest integer.

Chapter 4

Proposed Switch-Based Architecture for Channel Estimation

This chapter presents a novel solution to the channel estimation problem defined in Chapter 2. First, new types of switches conceived specifically for this problem to replace phase shifters are introduced. A reformulation of the estimation problem in terms of these switches follows. Then, a genetic algorithm is developed to solve the resulting combinatorial optimization problem. A detailed description of the algorithmic procedure is given.

4.1 Structural Elements

When massive MIMO became a popular research subject, its feasibility became a concern because for transceivers with a hundred antennas or more, using as many RF chains as antennas would result in very expensive and power-hungry systems. This is why the design was adapted by adding phase shifters between the antennas and the RF chains, leading to the hybrid analog/digital processing structure, as explained in Chapter 1. The hybrid structure allows a smaller number of RF chains to be used with the same large number of antennas, while preserving the capacity to take advantage of spatial multiplexing.

Phase shifters, however, are still relatively onerous and complex pieces of hardware. Besides, another disadvantage of phase shifters is that they can delay signals by a precise phase only over a limited range of frequencies. This is because they control the delay indirectly by increasing or decreasing the length of the paths on which the signals travel [49]. To see this, let Δx be the change in path length controlled by the phase shifter and $\Delta\phi$ be the resulting

change in phase. These two quantities are related through

$$\frac{\Delta x}{\lambda'} = \frac{\Delta \phi}{2\pi} \quad (4.1)$$

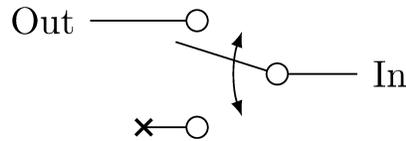
where λ' is the wavelength of the travelling signal inside the conducting material. This wavelength is determined by physical properties of the conductor and the frequency of the signal. Hence, if the frequency changes, the phase delay will change even if the path length remains constant. Since phase shifters are usually implemented with the ability to choose among only a finite number of path lengths, they can hardly compensate for this effect. Though this limitation is not of paramount importance when working with narrowband channels, as it is the case here, it should be taken into account when considering applications to broadband channels.

The solution proposed in this thesis to solve the channel estimation problem for massive MIMO systems with hybrid analog/digital architecture uses analog combiners that are implemented with switches instead of phase shifters. Switches can reduce the cost, the power consumption, and the building complexity of the analog module in the hybrid architecture [41]. Furthermore, the effect that the switches proposed here have on a signal is ideally independent of its frequency. This makes them particularly interesting for application in broadband communications. Specifically, referring to problem (2.11), the symbol \mathbf{S}_t will be used instead of \mathbf{F}_t to represent the switch-based analog combiner network used for the t^{th} pilot in our formulation.

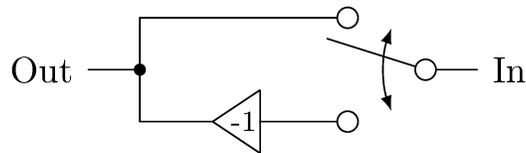
While the entries of the matrices \mathbf{F}_t take values in a continuous set $\{e^{j\theta} \mid \theta \in [0, 2\pi[\}$, the entries of the switching matrices \mathbf{S}_t can only take values in a discrete set, denoted as \mathcal{B} , which may contain 0 and selected points from the complex unit circle. The exact composition of \mathcal{B} depends on the type of switches used, as explained below.

Different types of switches are shown in Fig. 4.1. For simple on/off switches, the set of allowed values for the entries of \mathbf{S}_t is $\mathcal{B} = \{0, 1\}$, as in Fig. 4.1a. For switchable signal inverters, the set of allowed values is $\mathcal{B} = \{-1, 1\}$, as in Fig. 4.1b. For a combination of both, the set is $\mathcal{B} = \{0, -1, 1\}$, as in Fig. 4.1c. As described in Section 1.2, the antenna selection scheme is the most widely used architecture for hybrid massive MIMO systems that use switches [42, 43]. It is equivalent to the first type of switches, i.e. $\mathcal{B} = \{0, 1\}$, since signals between the antennas and the RF chains are either directly transmitted or blocked.

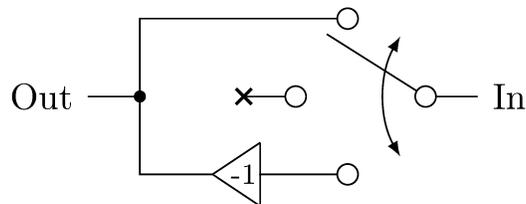
Here, to allow more possibilities in the design, we propose equipping the antennas with



(a) Switch used for $\mathcal{B} = \{0, 1\}$. The top branch corresponds to 1 and the bottom branch corresponds to 0.



(b) Switch used for $\mathcal{B} = \{-1, 1\}$. The top branch corresponds to 1 and the bottom branch corresponds to -1 .



(c) Switch used for $\mathcal{B} = \{0, -1, 1\}$. The top branch corresponds to 1, the middle branch corresponds to 0, and the bottom branch corresponds to -1 .

Fig. 4.1 Different types of switches that can be used in an analog combiner to implement different sets \mathcal{B} of allowed values in the matrices \mathbf{S}_t .

signal splitters that output the in-phase and the quadrature components separately. Using these devices with signal inverters allows a new type of switches that makes possible the inclusion of the imaginary unit and its inverse in the set of allowed values to be built. Two switches based on this architecture are depicted in Fig. 4.2. Fig. 4.2a shows a switch implementing $\mathcal{B} = \{-1, 1, -j, j\}$ while Fig. 4.2b shows a switch implementing $\mathcal{B} = \{0, -1, 1, -j, j\}$. It is important to note that in these figures, the switching elements for the in-phase and quadrature branches must be in the same respective position, as emphasized by the vertical dashed line.

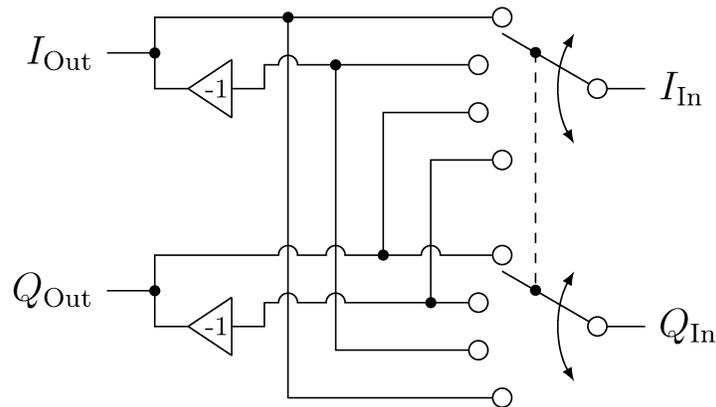
Another type of switch allowing more flexibility than the simple on/off or inverting switches shown in Fig. 4.1 was presented in [44]. However, since this design combines selection switches with constant, non-tunable phase shifters, it exhibits the same drawback as regular phase shifters, as described at the beginning of this section: In particular, it can only delay signals correctly in a restricted range of frequencies and is thus not suitable for broadband applications. This is not the case with the switches proposed in Fig. 4.1 and Fig. 4.2, which ideally perform the same function over a wide range of frequencies due to their discrete nature.

In this work, it is assumed that the switches are not partitioned among antenna elements and that each RF chain is equipped with a signal summer, that is, any group of antennas can be connected to any RF chain. Hence, our proposed architecture offers a generalization of the more flexible switching configuration in Fig. 1.3a. From a mathematical angle, this means that \mathbf{S}_t is not constrained to have a maximum number of non-zero entries per row or per column.

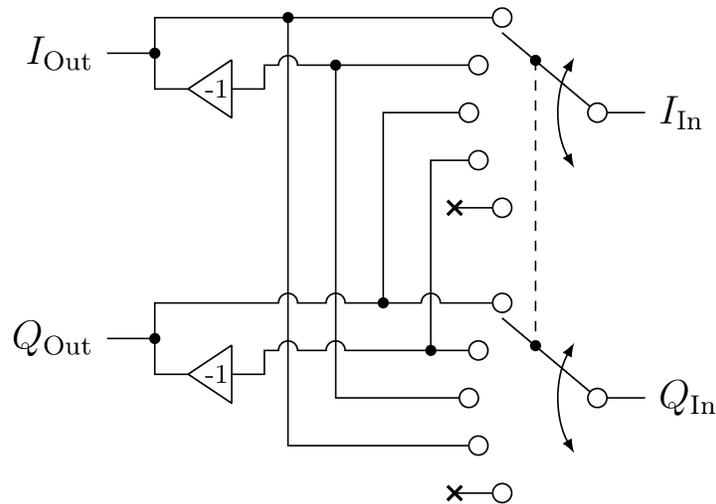
4.2 Problem Formulation with Switches

Using the hybrid analog/digital architecture with RF switches as introduced above, and defining the tuple of discrete combiner matrices $\mathcal{S} := (\mathbf{S}_1, \dots, \mathbf{S}_T)$, the optimization problem (2.11) can be rewritten as the following combinatorial minimum mean square error (MMSE) optimization problem:

$$\begin{aligned} & \underset{\mathcal{S}}{\text{minimize}} \mathcal{M}(\mathcal{S}) \\ & \text{subject to } \mathcal{S} \in \mathcal{B}^{L \times N \times T}, \end{aligned} \tag{4.2}$$



(a) Switch used for $\mathcal{B} = \{-1, 1, -j, j\}$. The first branch from the top corresponds to 1, the second branch corresponds to -1 , the third branch corresponds to j , and the last branch at the bottom corresponds to $-j$.



(b) Switch used for $\mathcal{B} = \{0, -1, 1, -j, j\}$. The first branch from the top corresponds to 1, the second branch corresponds to -1 , the third branch corresponds to j , the fourth branch corresponds to $-j$, and the last branch at the bottom corresponds to 0.

Fig. 4.2 Switches based on the in-phase and quadrature signal decomposition as described in Section 4.1. These switches allow more possibilities to be implemented for the sets \mathcal{B} of values allowed in the matrices \mathbf{S}_t .

where the objective function is

$$\mathcal{M}(\mathcal{S}) = \text{tr}\left(\mathbf{R} - \rho\mathbf{R}\mathbf{S}_c^{\text{H}} (\rho\mathbf{S}_c\mathbf{R}\mathbf{S}_c^{\text{H}} + \mathbf{S}_d\mathbf{S}_d^{\text{H}})^{-1} \mathbf{S}_c\mathbf{R}\right) \quad (4.3)$$

and \mathbf{S}_c and \mathbf{S}_d are defined the same way as \mathbf{F}_c and \mathbf{F}_d , that is

$$\mathbf{S}_c = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_T \end{bmatrix}, \quad \mathbf{S}_d = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_T \end{bmatrix}. \quad (4.4)$$

This new problem is more challenging to solve than (2.11) because the feasible set is discrete rather than continuous, and the search complexity increases exponentially with the parameters L , N , and T . However, this complexity contrasts with the simplicity of the hardware used for implementing the solution. Switches are indeed simpler than the phase shifters used for the original problem, because they are faster, more affordable, and consume less power [41].

Once the matrices $\mathbf{S}_1, \dots, \mathbf{S}_T$ have been determined, they can be used to obtain the channel estimate in the same way as in (2.11) and (2.12), but where $\mathbf{F}_1, \dots, \mathbf{F}_T$ are now replaced by $\mathbf{S}_1, \dots, \mathbf{S}_T$. That is, $\hat{\mathbf{g}} = \mathbf{W}\mathbf{y}_c$ where \mathbf{W} is now given by

$$\mathbf{W}_{\text{opt}} = \sqrt{\rho}\mathbf{R}\mathbf{S}_c^{\text{H}} (\rho\mathbf{S}_c\mathbf{R}\mathbf{S}_c^{\text{H}} + \mathbf{S}_d\mathbf{S}_d^{\text{H}})^{-1}. \quad (4.5)$$

4.3 Channel Estimation using Genetic Algorithm

The solution proposed here to solve problem (4.2) is a genetic algorithm (GA). GAs are part of a larger class of algorithms called evolutionary algorithms. In general, an evolutionary algorithm generates a set, called population, of feasible points, called individuals. At the n^{th} iteration, the population \mathcal{P}_n evolves into a new one, \mathcal{P}_{n+1} , in which the individuals are ideally closer to an optimum of the objective function [50].

GAs are interesting tools for solving optimization problem because they do not require any assumption to be made on the solutions and can be adapted to solve virtually any type of problem. They have been used among other things to configure seismic dampers [51], to reduce the risks of terrorism and piracy [52], to optimize transportation [53, 54], and to generate new meteorological models [55]. Sometimes, the solutions obtained from this type

of algorithms, although effective, are unusual and different from what might be suggested by intuition [56, 57].

In GAs, a new population is produced from the previous one by means of selection, crossover, and mutation, as explained below:

- Selection, or elitism, consists in taking individuals of the previous, or old, population and inserting them directly into the new population. The individuals that are selected are usually the best ones in terms of a selected cost function.
- Crossovers consist in selecting pairs of individuals from the old population, called parents, and for each pair selecting and swapping parts of them to create two new individuals.
- Mutations consist in choosing individuals from the old population, changing some of their features, called genes, and inserting the modified individuals into the new population.

Let us now consider the application of a GA to solve the combinatorial problem (4.2). In the implementation presented here, the population consists in a collection of M individuals, where in turn each individual is a T -tuple $\mathcal{S} = (\mathbf{S}_1, \dots, \mathbf{S}_T)$ of matrices and is taken from the feasible set $\mathcal{B}^{L \times N \times T}$. At each iteration, indexed by the integer $n \in \mathbb{N}$, the current population \mathcal{P}_n is updated to \mathcal{P}_{n+1} by applying a combination of the mechanisms introduced above. Specifically, when a new population is produced, the individuals from the old population are first sorted according to their MSE, as defined in (4.3), and the best ones are used to produce the next population as follows:

- The M_S individuals with the lowest MSE are selected without modification, where $M_S \in \{0, \dots, M\}$ is an integer.
- The M_C individuals with the lowest MSE are used to perform the crossover, where $M_C \in \{0, \dots, M\}$ is an even integer. In the algorithm presented here, the best individuals are paired sequentially to form parents. For each pair, two new individuals, called children, are created by random crossover. In the first child, each entry of the matrices S_t is taken from one of the parents with probability $p = \frac{1}{2}$. Then for the second child, the corresponding entry is selected from the other parent. If M_C is an odd integer

instead, the $M_C + 1$ individuals with the lowest MSE are used to form parents, but the $(M_C + 1)^{\text{th}}$ child produced is discarded.

- The M_M individuals with the lowest MSE are chosen for the mutation process, where $M_M \in \{0, \dots, M\}$ is an integer. In the proposed solution, the mutation is carried on for each entry of the matrices S_t with probability μ . If the mutation occurs, it consists in replacing the corresponding entry by an equiprobable random element taken from the set of allowed values, i.e. an element of \mathcal{B} .

The iterative process continues until there is no improvement in the best or the average MSE for a given number c_{\max} of iterations. In our implementation of the GA, in order to reduce the overall run time, we used a time-varying population size M_n that decreases monotonically as a function of the iteration number n . Specifically, we let $M_n := M_0 \alpha^n$, where M_0 is the initial population size, $\alpha \in]0, 1]$ is a geometric factor that controls how fast the population size decreases, and $n \in \{0, 1, \dots\}$ is the iteration number. We also let the subpopulation sizes $M_S^{(n)}$, $M_C^{(n)}$, and $M_M^{(n)}$ for the three evolution mechanisms vary in a similar way with fixed proportions. Note that following this definition, these three sizes must be rounded to the closest integer values, i.e. $M_S^{(n)} = \lfloor m_S M_n \rfloor$, $M_C^{(n)} = \lfloor m_C M_n \rfloor$, and $M_M^{(n)} = \lfloor m_M M_n \rfloor$, where m_S , m_C , and m_M are the respective proportion factors satisfying $m_S + m_C + m_M = 1$.

The final GA, which is presented in Algorithm 2, takes as input the dimensions L and N of the matrices \mathbf{S}_t and their number T , and the set \mathcal{B} of allowed values. It also depends on execution parameters like the initial size M_0 of the population, the shrinking parameter α , the proportions $m_S, m_C, m_M \in [0, 1]$ of the population respectively created by each of the three evolution mechanisms, the probability $\mu \in [0, 1]$ of mutation for the matrix entries of the mutated individuals, and the number c_{\max} of iterations without improvement of the MSE before the algorithm stops.

The overall computational complexity of the GA can be obtained by evaluating the number of operations of each of its steps. To simplify this analysis, the fact that LT and N must be on the same order of magnitude in order to estimate all the parameters of the channel is used to approximate $LT \approx N$. First, generating an initial random population requires $k_1 M_0 N^2$ operations, where k_1 is the number of operations needed for the generation of a discrete random variable taken from \mathcal{B} . Sorting the population with respect to the MSE of the individuals at the n^{th} iteration requires $M_n t_{\mathcal{M}} + 2M_n \log_2 M_n + \mathcal{O}(M_n)$ operations, where

$t_{\mathcal{M}}$ is the complexity of the cost function (4.3) and is given by $t_{\mathcal{M}} = 10.3N^3 + \mathcal{O}(N^2)$ when only the diagonal matrix entries used for the trace operation are computed and the sparsity of \mathcal{S}_d is exploited. The selection, crossover, and mutation procedures respectively have complexities $k_S M_S^{(n)} N^2$, $k_C M_C^{(n)} N^2$, and $k_M M_M^{(n)} N^2$, where k_S is the number of operations required to copy an entry; k_C is the average number of operations required to copy an entry, to decide if it is swapped, and, if so, to swap it; and k_M is the average number of operations required to copy an entry, to decide if it is mutated, and, if so, to mutate it. All the other steps have negligible complexities compared to the ones mentioned above. Assuming that the algorithm stops at n_{iter} iterations and that $\alpha^{n_{\text{iter}}} \ll 1$, and taking into account the geometric expression for M_n given above, the total complexity of the algorithm is approximately given by

$$t_{\text{GA}} = \frac{M_0}{1 - \alpha} (10.3N^3 + 2 \log_2 M_0 + \mathcal{O}(N^2)). \quad (4.6)$$

In the case where $\alpha^{n_{\text{iter}}}$ cannot be neglected, for example when $\alpha = 1$ and the population size is constant, the complexity takes the form

$$t_{\text{GA}} = n_{\text{iter}} M_0 (10.3N^3 + 2 \log_2 M_0 + \mathcal{O}(N^2)). \quad (4.7)$$

For large numbers of antennas (e.g. $N = 32$) and typical sizes of initial population (e.g. $M_0 = 500$), both of these expressions are dominated by the first term in the parentheses, i.e. $N^3 \gg \log_2 M_0$.

Algorithm 2: Genetic algorithm applied to finding analog combiners for switching networks

Input: L, N, T, \mathcal{B}

GA Parameters: $M_0, \alpha, m_S, m_C, m_M, \mu, c_{\max}$

```

1  $n \leftarrow 0$ 
2 Generate initial population  $\mathcal{P}_n := \{\mathcal{S}_1^{(n)}, \dots, \mathcal{S}_{M_n}^{(n)}\}$  with each entry in each individual
   randomly picked from  $\mathcal{B}$  with uniform probability
3 Sort in increasing MSE order, so that  $\mathcal{M}(\mathcal{S}_1^{(n)}) \leq \dots \leq \mathcal{M}(\mathcal{S}_{M_n}^{(n)})$ 
4 Save  $\mathcal{S}_{\text{best}} \leftarrow \mathcal{S}_1^{(n)}$ ,  $\mathcal{M}_{\text{best}}^{\min} \leftarrow \mathcal{M}(\mathcal{S}_1^{(n)})$ , and  $\overline{\mathcal{M}}_{\text{best}} \leftarrow \frac{1}{M_n} \sum_{m=1}^{M_n} \mathcal{M}(\mathcal{S}_m^{(n)})$ 
5  $c \leftarrow 0$ 
6 while  $c < c_{\max}$  do
7    $n \leftarrow n + 1$ 
8    $M_n := M_0 \alpha^n$ 
9   // Selections:
10  Copy the  $\lfloor m_S M_n \rfloor$  best individuals from  $\mathcal{P}_{n-1}$  to  $\mathcal{P}_n$ 
11  // Crossovers:
12  Take the  $\lfloor m_C M_n \rfloor$  best individuals of  $\mathcal{P}_{n-1}$  to form parents
13  For each consecutive pair, produce children following the method in Section 4.3,
   and store them in  $\mathcal{P}_n$ 
14  // Mutations:
15  Copy the  $\lfloor m_M M_n \rfloor$  best individuals from  $\mathcal{P}_{n-1}$  to  $\mathcal{P}_n$  and randomly replace entries
   as described in Section 4.3
16  // Evaluation of the new population:
17  Sort in increasing MSE order, so that  $\mathcal{M}(\mathcal{S}_1^{(n)}) \leq \dots \leq \mathcal{M}(\mathcal{S}_{M_n}^{(n)})$ 
18  Save  $\mathcal{M}_n^{\min} := \mathcal{M}(\mathcal{S}_1^{(n)})$  and  $\overline{\mathcal{M}}_n := \frac{1}{M_n} \sum_{m=1}^{M_n} \mathcal{M}(\mathcal{S}_m^{(n)})$ 
19  if  $\mathcal{M}_n^{\min} < \mathcal{M}_{\text{best}}^{\min}$  or  $\overline{\mathcal{M}}_n < \overline{\mathcal{M}}_{\text{best}}$  then // If a new best MSE or a new best
   average MSE is found
20     $c \leftarrow 0$ 
21    if  $\mathcal{M}_n^{\min} < \mathcal{M}_{\text{best}}^{\min}$  then // If a new best MSE is found
22       $\mathcal{M}_{\text{best}}^{\min} \leftarrow \mathcal{M}_n^{\min}$ 
23       $\mathcal{S}_{\text{best}} \leftarrow \mathcal{S}_1^{(n)}$ 
24    end
25    if  $\overline{\mathcal{M}}_n < \overline{\mathcal{M}}_{\text{best}}$  then // If a new best average MSE is found
26       $\overline{\mathcal{M}}_{\text{best}} \leftarrow \overline{\mathcal{M}}_n$ 
27    end
28  else
29     $c \leftarrow c + 1$ 
30  end
31 end

```

Output: Analog combiners for switching networks $\mathcal{S}_{\text{best}}$

Chapter 5

Simulation Results

In this chapter, we evaluate the performance of the proposed GA for channel estimation in hybrid analog/digital massive MIMO systems with switch-based architecture. First, the methodology details how the simulations were performed and lists the values used for the different parameters. Then, the performance of this GA and that of the SO method presented in Chapter 3 for the hybrid architecture with phase shifters is compared to an exhaustive search for a small system setup. Comparisons between the GA and the SO method follow for different sets of parameters, covering a wide range of configurations. Finally, the different types of switches listed in 4.1 are compared in terms of their estimation performance.

5.1 Methodology

The massive MIMO system considered in the simulations is modelled after the one illustrated in Fig. 2.1. A ULA as shown in Fig. 2.2 is assumed, where the distance between the antennas is chosen to be half the wavelength of the RF signals, i.e. $d = \lambda/2$. Throughout the simulations, the number N of antennas at the BS ranges between 16 and 64, the number L of RF chains ranges between 1 and 16, the number T of pilot symbols ranges between 1 and 4, and the number P of propagation paths from the MS to the BS ranges between 3 and 6. Smaller systems are also simulated to allow for an exhaustive search, in which the parameters are set as $N = 6$, $L = 2$, $T = 1$, and $P = 2$.

The three algorithms used in the simulations for channel estimation with hybrid analog/digital architecture are the proposed switch-based solution using the GA as introduced in Section 4.3 (from now on referred to as GA for conciseness), the SO method of [17] for

the phase-shifter solution with phase quantization as in Section 3.3, and the optimal switch-based solution obtained with exhaustive search. The SO method minimizes the cost function (2.14) over the feasible set \mathcal{F}^T , where $\mathcal{F} = \{e^{j\theta} \mid \theta \in [0, 2\pi[\}^{L \times N}$ and the analog combiner uses phase shifters, whereas the GA and the exhaustive search minimize the function (4.3) over the set $\mathcal{B}^{L \times N \times T}$, where the analog combiner uses switches. The number of quantization bits used for the SO method and the elements of \mathcal{B} used in the GA and the exhaustive search are specified in each figure. More details on the three algorithms are given in the next paragraphs.

The implementation of the GA in the simulations uses an initial population of $M_0 = 500$ individuals and a shrinking parameter $\alpha = 0.98$. The proportions of the population produced by selection, crossover, and mutation are respectively $m_S = 0.1$, $m_C = 0.5$, and $m_M = 0.4$. The maximum number of strikes is $c_{\max} = 5$ and the mutation probability is $\mu = 2.5 \cdot 10^{-3}$. The final values of these parameters were determined empirically based on a large number of prior experiments in order to find the best results while maintaining a reasonable execution time. In particular, it was observed that increasing M_0 or c_{\max} beyond these values did not significantly improve the performance, but increased the execution time.

The SO method from [17] was implemented as described in Section 3.2.3. The analog combiners computed by the algorithm were normalized as described in Section 3.3 and quantized to the required number of bits using (3.16). In every figure, the number of bits is chosen so that the number of possibilities for the phase angles matches the number of allowed values in the set \mathcal{B} for the switches.

Finally, the exhaustive search algorithm computes the MSE of all the $|\mathcal{B}|^{LNT}$ possible switching networks using the cost function (4.3), where $|\mathcal{B}|$ is the cardinality of the set of allowed values. The algorithm saves the minimum MSEs for each SNR and the combiners that yield it.

For all the algorithms under comparison, the channel covariance matrix \mathbf{R} is assumed to be known beforehand at the BS. For a discussion on how to estimate this matrix in practice, see [17]. In our simulations, \mathbf{R} is computed using the closed-form expression (2.10) using a fixed number P of paths and a fixed set $\Theta = \{\theta_1, \dots, \theta_P\}$ of P AoAs independently generated from the uniform distribution $\mathcal{U}(0, 2\pi)$. The matrix \mathbf{R} so generated is then used by each one of the methods under study to estimate the combiner matrices, i.e.:

- For the SO method, the eigenvalue decomposition (EVD) of $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ is com-

puted and used in the cost function (3.9)-(3.10) for the sequential determination of the combiner matrices $\mathbf{F}_1, \dots, \mathbf{F}_T$, which are based on phase shifters.

- For the proposed GA and the exhaustive search, \mathbf{R} is used directly in (4.3) for the determination of the combiner matrices $\mathbf{S}_1, \dots, \mathbf{S}_T$, which are based on switches.

Once the combiner matrices are determined, they can be used for the purpose of channel estimation. This amounts to applying the transformation $\hat{\mathbf{g}} = \mathbf{W}\mathbf{y}_c$ to the received data \mathbf{y}_c , where \mathbf{W} is given by (2.12) for the SO method and by (4.5) for the GA and the exhaustive search. The channel vectors used for the simulations are computed as $\mathbf{g} = \mathbf{R}^{1/2}\mathbf{h}$, where $\mathbf{h} \in \mathbb{C}^{N \times 1}$ is used to model the small-scale fading and is randomly distributed as $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. The signal \mathbf{y}_c received at the baseband processor is then computed using (2.5), where the noise \mathbf{n}_c is random and distributed as $\mathbf{n}_c \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{NT})$. Since the noise power is normalized to unity, the pilot power ρ in (2.1) actually corresponds to the desired SNR level.

The performance of the channel estimates computed by the different algorithms is evaluated in terms the normalized mean square error (NMSE) between the real and the estimated channel vectors. In the simulations, the NMSE obtained by averaging over a large number of channel and noise realizations. More precisely, for each point in the figures presented in the following sections, we proceed as follows:

- We generate $N_1 = 20$ different sets $\Theta_i = \{\theta_1^{(i)}, \dots, \theta_P^{(i)}\}$ of AoAs as explained above, and for each set, we compute the corresponding covariance matrix \mathbf{R}_i , where $i \in \{1, \dots, N_1\}$.
- For each \mathbf{R}_i , $N_2 = 10,000$ different channel realizations $\mathbf{g}_{i,j} = \mathbf{R}_i^{1/2}\mathbf{h}_j$ are generated, where $\mathbf{h}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $j \in \{1, \dots, N_2\}$.
- For each $\mathbf{g}_{i,j}$, the estimated channel is computed as $\hat{\mathbf{g}}_{i,j} = \mathbf{W}_{\text{opt}}\mathbf{y}_c^{(i,j)}$ using the optimal matrix \mathbf{W}_{opt} for the baseband processor, i.e. (2.12) for the SO method and (4.5) for the GA and exhaustive search.
- The NMSE is then computed as

$$\text{NMSE} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \frac{\|\mathbf{g}_{i,j} - \hat{\mathbf{g}}_{i,j}\|^2}{\|\mathbf{g}_{i,j}\|^2}. \quad (5.1)$$

5.2 Algorithm Convergence

We begin by investigating the convergence behaviour of the proposed GA for the solution of problem (4.2). Fig. 5.1 shows an example of the MSE performance of the population during the optimization process as the GA finds better individuals at each iteration. This figure was obtained at an SNR of 0 dB by considering the set $\mathcal{B} = \{-1, 1\}$ of allowed values for the switches along with the following choice of system parameters: $N = 64$, $L = 8$, $T = 1$, $P = 6$. In the figure, the blue dots represent the MSE values of each one of the individuals in the GA population, while the red crosses represent the population average. It can be seen that the performance improves until it stabilizes towards the end. The GA stops when the average MSE and the best MSE do not improve for c_{\max} iterations. A similar behaviour in convergence was obtained for other values of system parameters. The typical run time of the GA is five times that of the SO method. However, since it is based on switches, the method lends itself to more flexible and less costly implementation. Also, as will be shown shortly, the GA offers more accuracy than the SO method.

5.3 Comparison with Exhaustive Search

Since the feasible set for problem (4.2) is finite, it is theoretically possible to perform an exhaustive search to find the minimum. However, doing so is infeasible for practical values of the system parameters L , N , and T , due to the exponentially growing size of the search space $\mathcal{B}^{L \times N \times T}$. For example, if the set of allowed values \mathcal{B} contains two elements and $N = 32$, $L = 8$, and $T = 1$, then the feasible set contains more than 10^{77} possibilities.

However, with fewer antennas and RF chains, the number of possibilities can be reduced so that an exhaustive search is possible. To this end, we evaluate the performance of the various algorithms, including the exhaustive search, for a MIMO system with reduced dimensionality, i.e. with $N = 6$ antennas, $L = 2$ RF chains and $T = 1$ pilot. We also set the number of paths in the channel model to $P = 2$. Figures 5.2 and 5.3 compare the NMSE performance of the proposed GA and the SO method to the best achievable performance of the switch-based approach, provided by the exhaustive search. Fig. 5.2 uses the set of allowed values $\mathcal{B} = \{-1, 1\}$ for the GA and 1-bit quantization for the SO method, whereas Fig. 5.3 uses the set of values $\mathcal{B} = \{-1, 1, j, -j\}$ for the GA and 2-bit quantization for the SO method. The corresponding search spaces for the GA have cardinality $2^{12} = 4096$ and

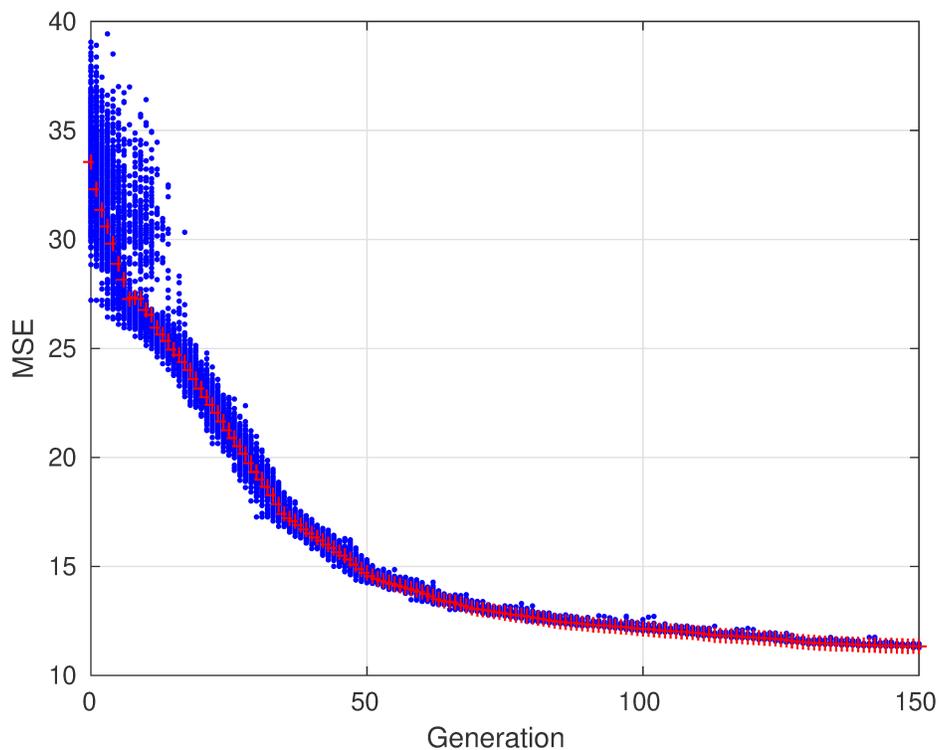


Fig. 5.1 MSE of the population versus iteration number of the GA. The blue dots are the MSEs of the individuals and the red crosses are the population averages at each iteration.

$$4^{12} \approx 1.68 \cdot 10^7.$$

Both Fig. 5.2 and Fig. 5.3 show that the performance of the GA and the SO method is close to that of the exhaustive search, although the performance of the GA exceeds that of the SO method in both cases. In the case of 1-bit quantization, shown in Fig. 5.2, the performance of the GA is within 0.05 dB of that of the exhaustive search, while in the 2-bit case, shown in Fig. 5.3, the difference between the two increases slightly to 0.07 dB. These results indicate that for search spaces with manageable dimensions, the GA can produce quite satisfactory results, which motivates its application to problems of larger dimensions.

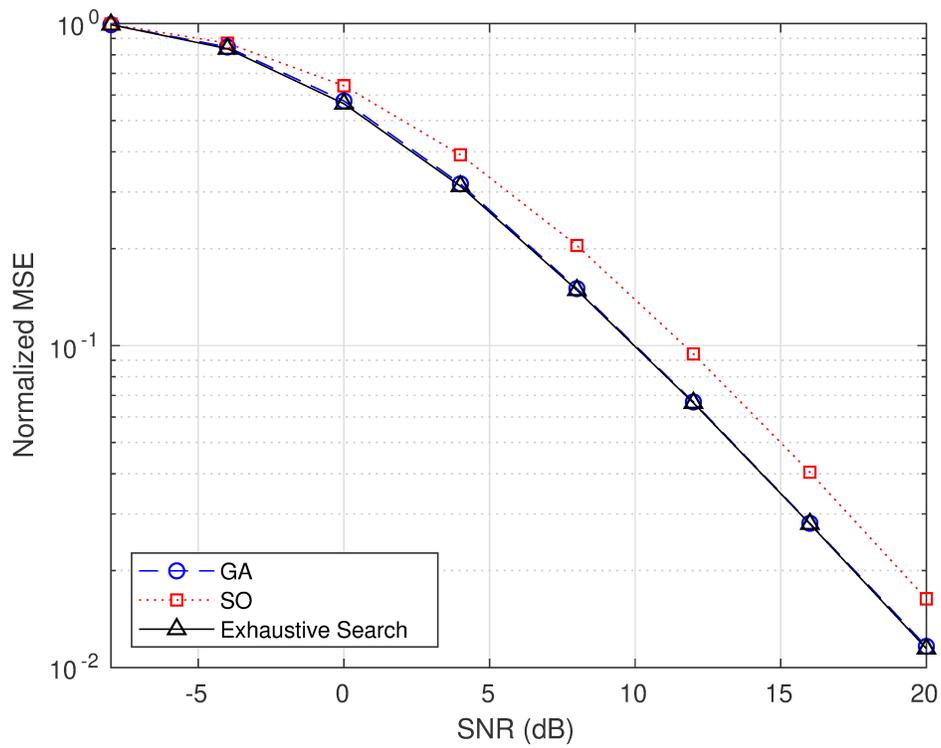


Fig. 5.2 NMSE of channel estimate versus SNR for the GA, the SO method, and the exhaustive search. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $N = 6$, $L = 2$, $T = 1$, $P = 2$ are used.

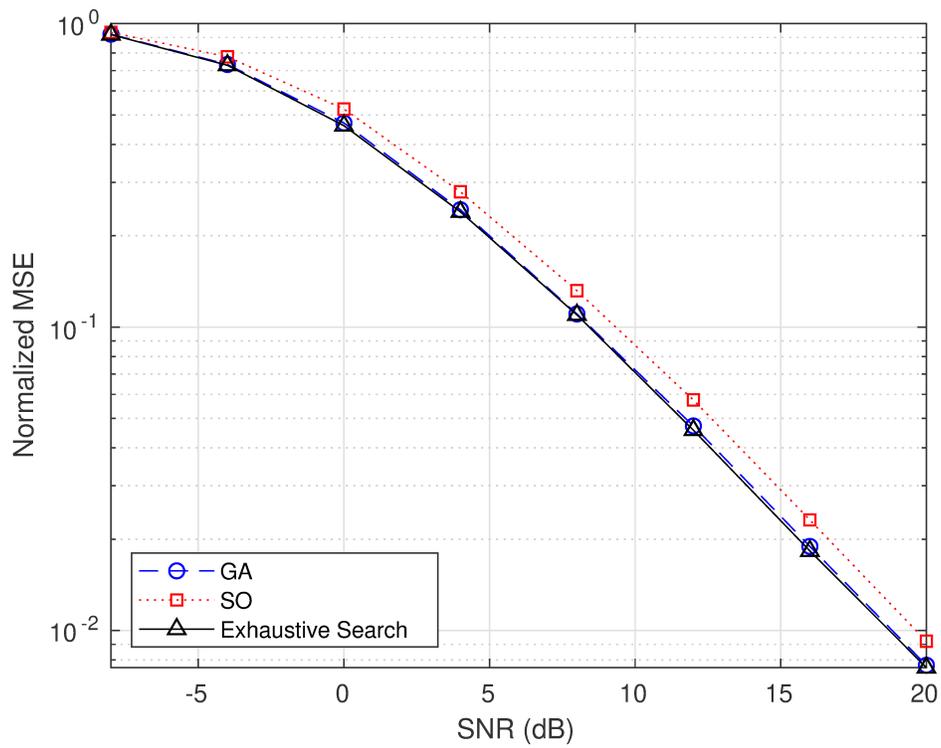


Fig. 5.3 NMSE of channel estimate versus SNR for the GA, the SO method, and the exhaustive search. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, j, -j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $N = 6$, $L = 2$, $T = 1$, $P = 2$ are used.

5.4 Results for Massive MIMO Systems

This section presents simulation results for massive MIMO systems with larger values of system parameters: number of antennas $16 \leq N \leq 64$, number of RF chains $1 \leq L \leq 16$, and number of pilots $1 \leq T \leq 16$. For these values of system parameters, performing exhaustive searches is not feasible and thus only the SO method is used as a benchmark to assess the effectiveness of the GA. For the GA method, unless otherwise indicated, we focus on two possible sets of allowed values for the switches, i.e. $\mathcal{B} = \{-1, 1\}$ and $\mathcal{B} = \{-1, 1, -j, j\}$, corresponding to the configurations shown in Fig. 4.1b and Fig. 4.2a. For the sake of fairness, the performance of the GA with these sets is compared with the performance of the SO method with the number of quantization bits respectively set to 1 and 2. Hence, for both cases, the corresponding values of the quantized phase shifts in the SO method, as given by (3.16), are identical to the allowed values in the set \mathcal{B} .

In Figures 5.4 and 5.5, the NMSE versus SNR performance of the two methods is compared for different numbers of antennas. In general, using more antennas when the other parameters are fixed results in better estimation accuracy, as would be expected. In Fig. 5.4, channels are estimated with 1-bit quantization for the RF components in the analog combiner. It can be seen that the GA outperforms the SO method by about 2 dB over the considered range of SNR for all antenna array sizes and that the GA needs half as many antennas as the SO method to exhibit a similar performance. In Fig. 5.5, the same comparison is performed with 2-bit quantization for the RF components. The GA still outperforms the SO method for all sizes of antenna arrays, although the performance gain in this case is reduced to about 1 dB.

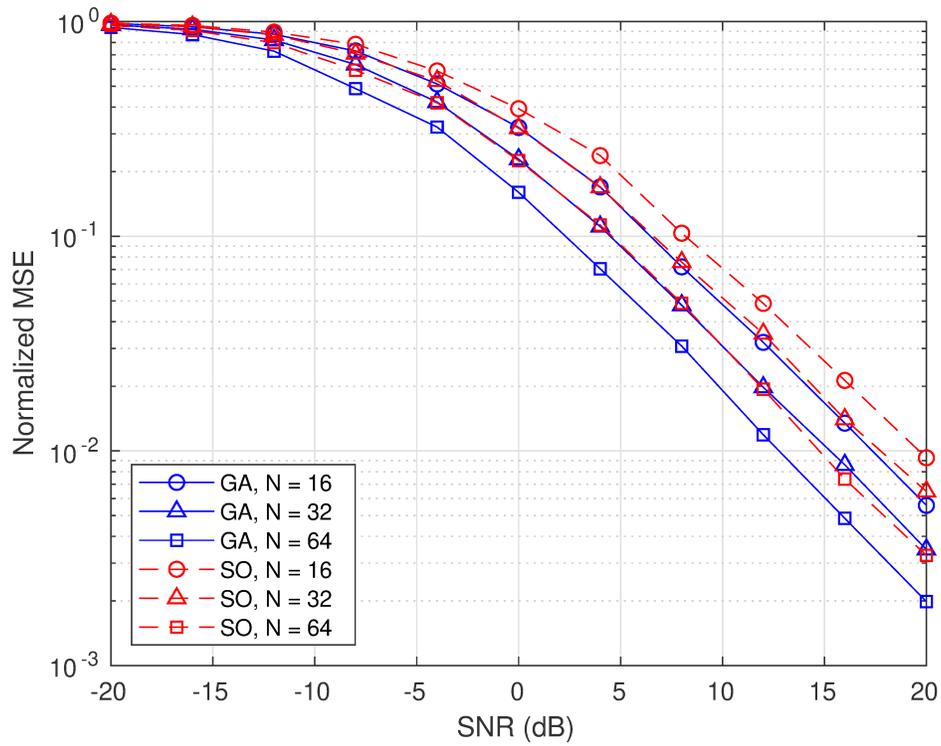


Fig. 5.4 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers N of antennas. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $L = 8$, $T = 1$, $P = 6$ are used.

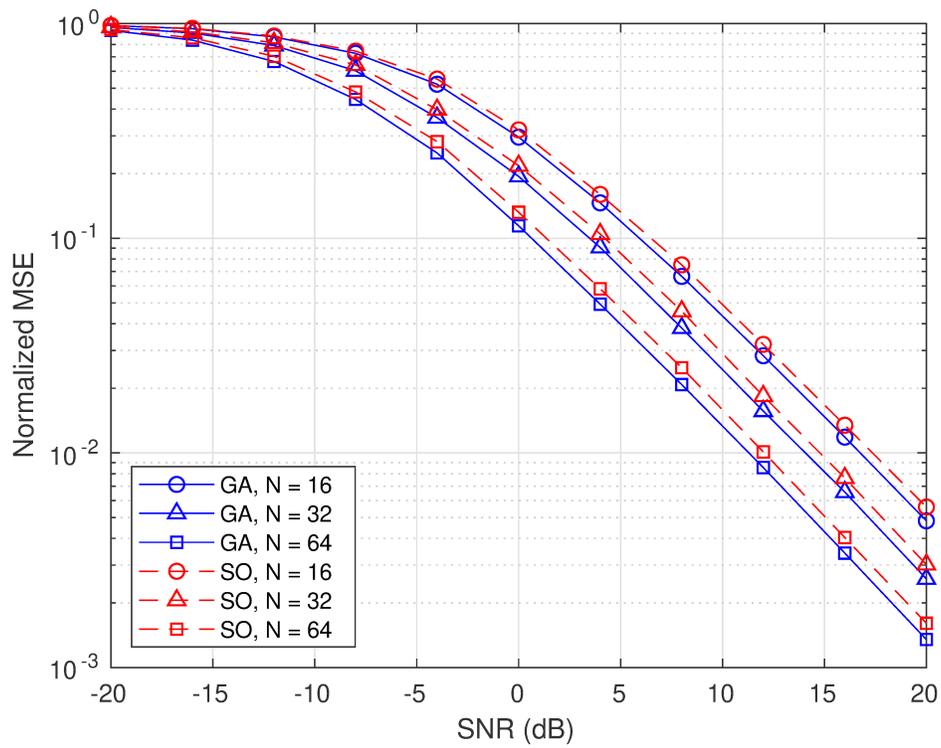


Fig. 5.5 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers N of antennas. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $L = 8$, $T = 1$, $P = 6$ are used.

In Figures 5.6 and 5.7, the effect of the number P of propagation paths between the MS and the BS is studied. In Fig. 5.6, the GA uses the values $\mathcal{B} = \{-1, 1\}$, which is equivalent to the 1-bit quantization of the SO method. A significant gap between the GA's curves and the SO method's curves can be observed. In Fig. 5.7, the GA and the SO method both use the values $\{-1, 1, -j, j\}$. The GA performs better than the SO method, but the difference between the two is smaller than in the 1-bit scenario. We note that in both Figures 5.6 and 5.7, the estimation accuracy as given by the MSE decreases with an increase in the number of propagation paths.

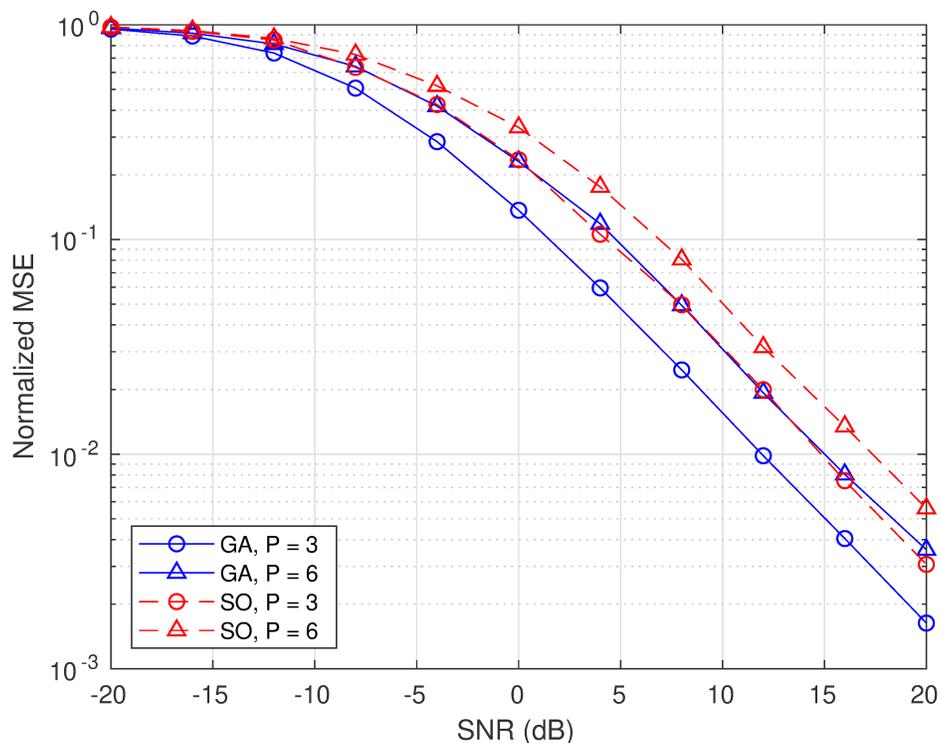


Fig. 5.6 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers P of propagation paths. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $L = 8$, $N = 32$, $T = 1$ are used.

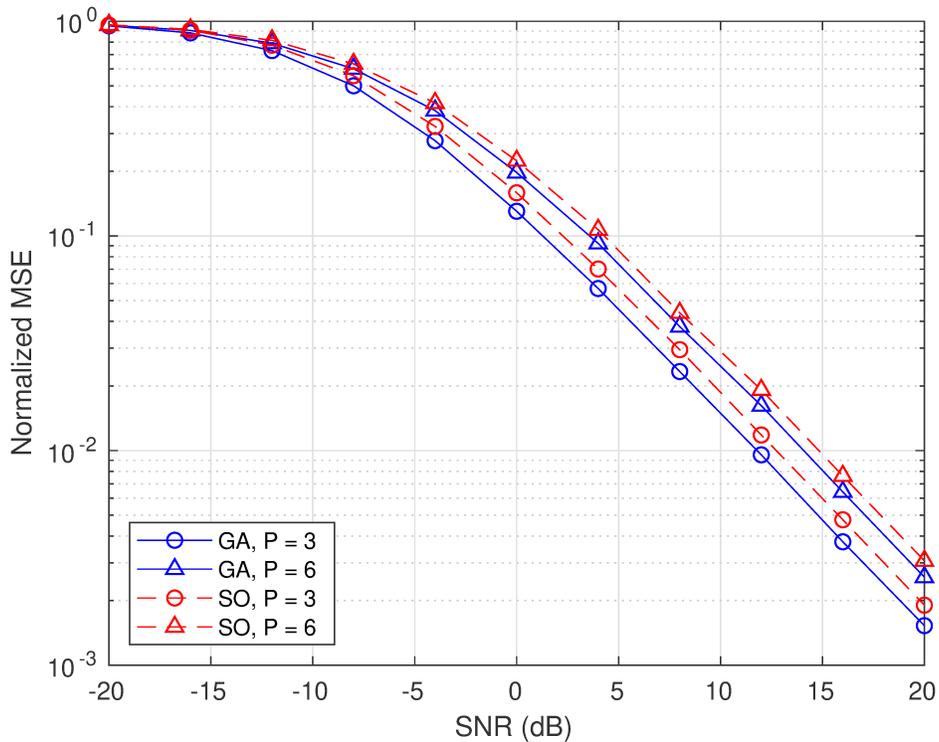


Fig. 5.7 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers P of propagation paths. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $L = 8$, $N = 32$, $T = 1$ are used.

In Figures 5.8 and 5.9, the effect of the number L of RF chains at the BS is studied. When there are more RF chains, more information about the channel can be gathered at once and thus more accurate estimates can be obtained. On the opposite, a system with only a few RF chains cannot adequately capture the channel information and would need a much longer pilot sequence to provide a decent estimation, as discussed later. In Fig. 5.8, the GA uses the values $\mathcal{B} = \{-1, 1\}$, equivalent to 1-bit quantization for the SO method. When $L \leq 4$ RF chains are available, both the GA and the SO method perform poorly. With $L \geq 6$ RF chains, the performance of the two methods is substantially better. We also note that in this case, the MSE performance of the GA over the SO method exhibits an SNR gain of more than 2 dB at high SNR. In Fig. 5.9, as in the 1-bit case, the two algorithms perform very poorly when the number of RF chains is insufficient to adequately perform a channel estimation. For larger numbers of RF chains, the GA performs better than the SO method, even if the difference between the two is smaller than in the 1-bit scenario.

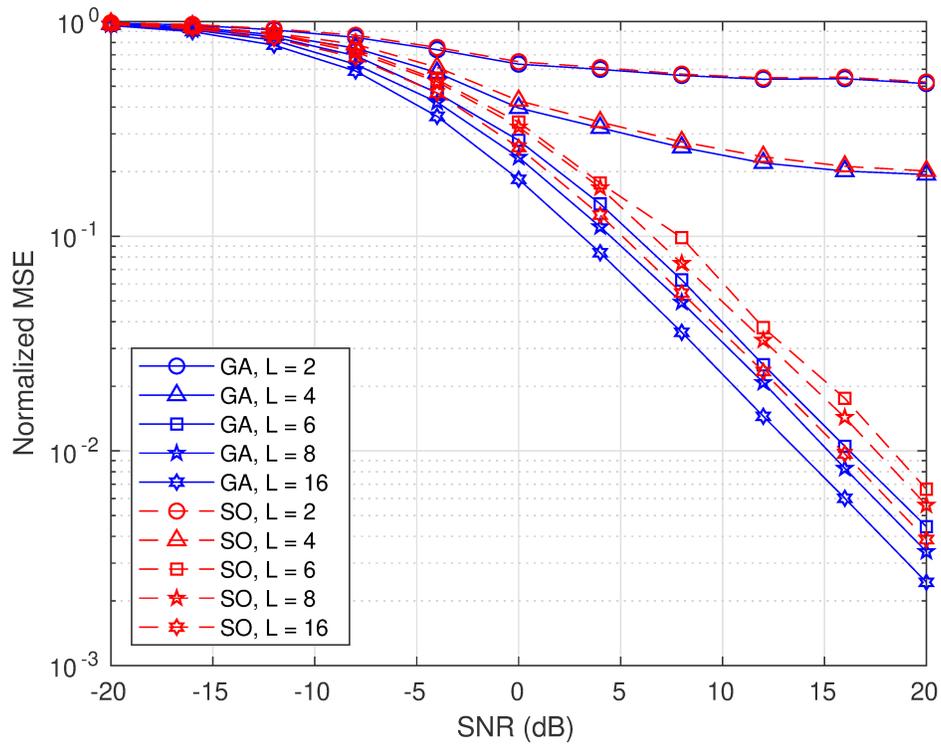


Fig. 5.8 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers L of RF chains at the BS. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $N = 32$, $T = 1$, $P = 6$ are used.

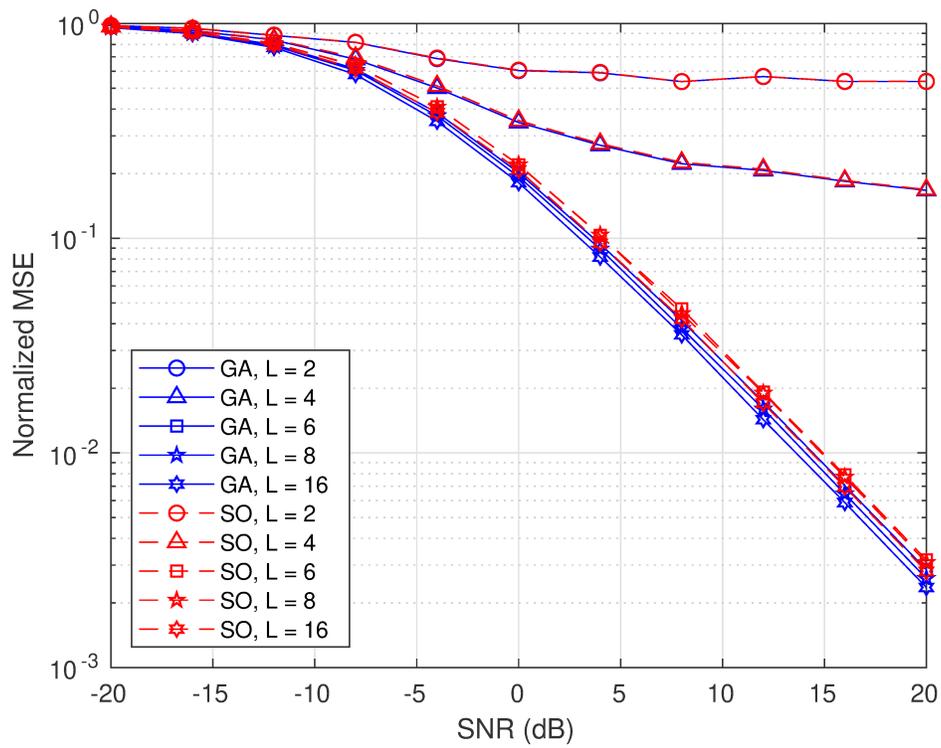


Fig. 5.9 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers L of RF chains at the BS. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $N = 32$, $T = 1$, $P = 6$ are used.

Figures 5.10 and 5.11 show the results of simulations for different numbers $T \in \{1, 2, 4\}$ of pilot symbols; the other key parameters are set to $N = 32$, $L = 8$, $P = 6$. In Fig. 5.10, the comparison is carried out for the set $\mathcal{B} = \{-1, 1\}$ in the GA method with corresponding 1-bit quantization for the SO method. The curves show that regardless of the number of pilot symbols used, the estimation made by the GA is significantly more accurate than that of the SO method, allowing an SNR gain of nearly 3 dB. Fig. 5.11 shows the same comparison, but with the set $\mathcal{B} = \{-1, 1, -j, j\}$ for the GA and a 2-bit quantization for the SO method. The results show again that for any number of pilot symbols used, the estimation made by the GA is more accurate than that of the SO method, although here the SNR gain is only about 1 dB.

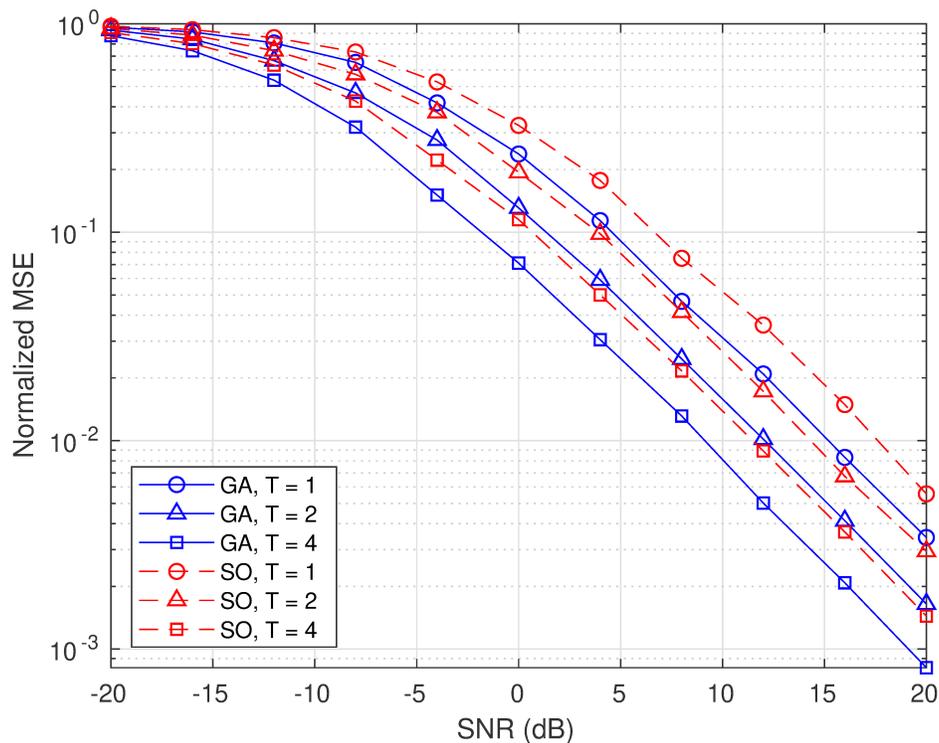


Fig. 5.10 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers T of pilot symbols. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1\}$. For the SO method, an equivalent 1-bit quantization is applied. The parameters $L = 8$, $N = 32$, $P = 6$ are used.

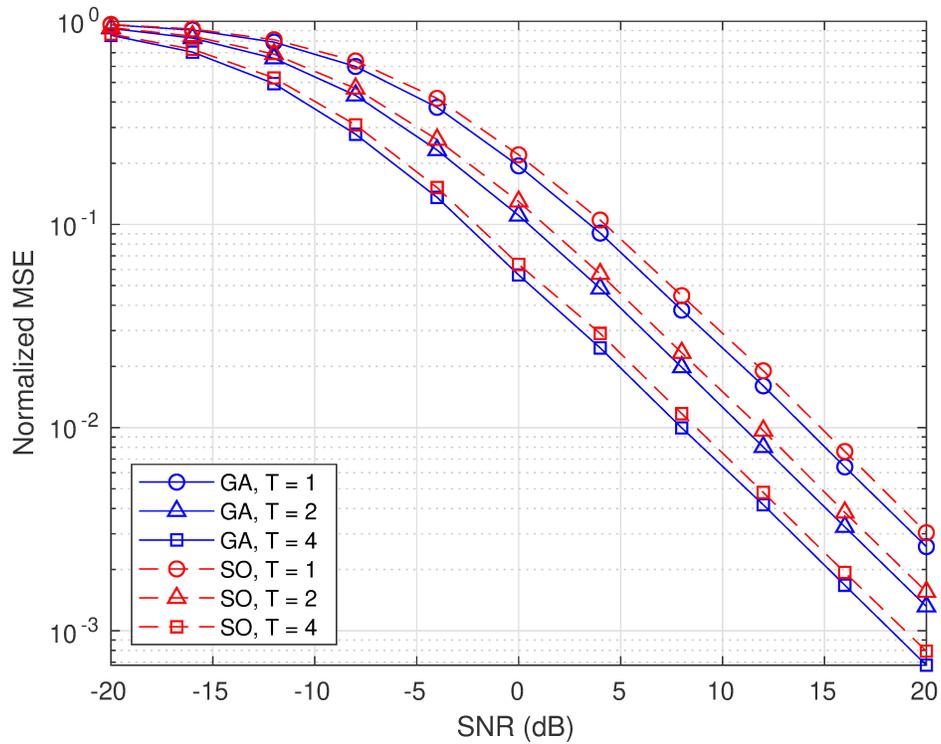


Fig. 5.11 NMSE of channel estimate versus SNR for the GA and the SO method with different numbers T of pilot symbols. For the proposed GA, the set of allowed switch values is $\mathcal{B} = \{-1, 1, -j, j\}$. For the SO method, an equivalent 2-bit quantization is applied. The parameters $L = 8$, $N = 32$, $P = 6$ are used.

Finally, we compare the performance of the proposed switch-based channel estimator for the five different types of switches introduced in Section 4.1. Fig. 5.12 shows the corresponding NMSE curves versus SNR for the system parameters $L = 8$, $N = 64$, $T = 1$, $P = 6$. The results show that the new switches introduced in Fig. 4.2 perform better than the simpler types in Fig. 4.1. An improvement of 1 dB distinguishes the switches implementing j and $-j$ from the simpler ones. A general trend seen in this and other figures is that using more allowed values for the entries of the analog combiner provides more freedom and thus allows the channels to be estimated more accurately.

It is interesting to note that the proposed GA always stays within the constraint imposed by problem (4.2), which is not the case with the SO method from [17] and the phase-only constraint of problem (2.11). Specifically, during the application of the sequential optimization steps in (3.9) and (3.10), the constant modulus constraint on the entries of

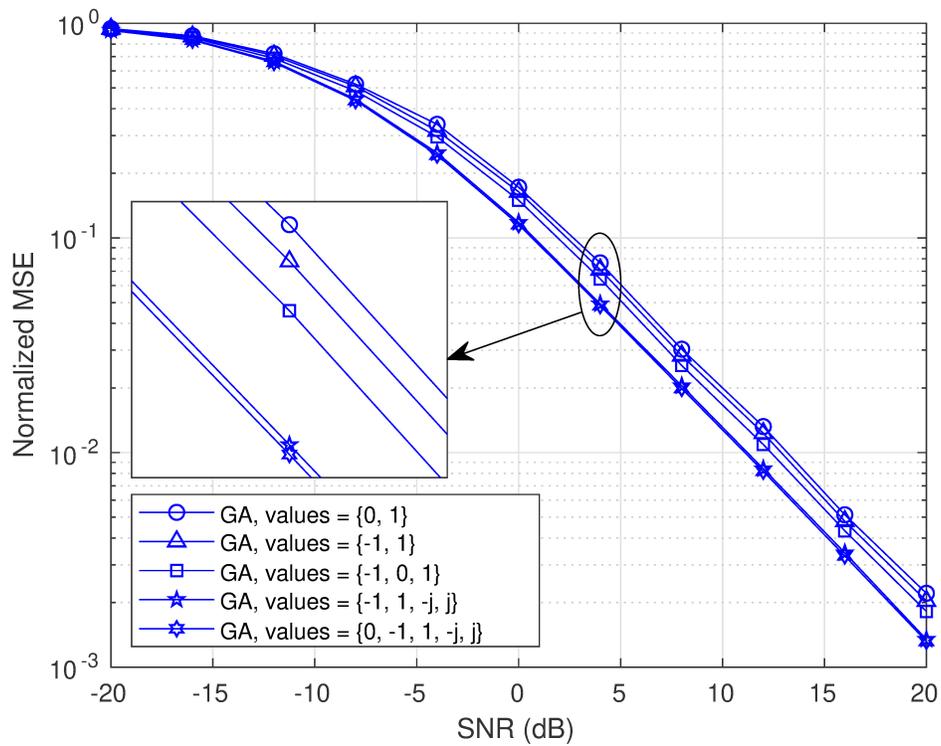


Fig. 5.12 NMSE of channel estimate versus SNR for the GA with different sets of allowed values. The parameters $L = 8$, $N = 64$, $T = 1$, $P = 6$ are used.

the matrices \mathbf{F}_t is relaxed; it is only reintroduced at the end using the projection operation (3.14). This might be one of the reasons why the proposed GA performs better than the SO method despite the simplicity of the hardware it uses.

Chapter 6

Conclusion and Future Prospects

This final chapter concludes the thesis by restating its objectives and contributions, and by summarizing its chapters. Then follows a discussion of possible directions for future research projects building on the work of this thesis.

6.1 Thesis Summary

The goal of this thesis was to address the problem of channel estimation for hybrid mmWave massive MIMO telecommunication systems. Massive MIMO and mmWave-enabled systems are among the main technologies that will allow the requirements of the 5G standard to be met. However, the hybrid analog/digital architecture is necessary for massive MIMO transceivers to be viable because it allows the number of RF chains, which are expensive and power-hungry, to be reduced. But this architecture introduces a problem because its analog component acts as an interface that limits the processing functionality of the baseband processor over the antenna signals. This makes it harder for the MIMO system to estimate the wireless channels, which is a crucial step for any wireless communication to take place between the BS and an MS.

In Chapter 1, a detailed overview of the problem and its context was given. This was followed by a survey of the technical literature on the problem of channel estimation for mmWave massive MIMO systems, including references to works focusing on different approaches, like neural networks and Bayesian learning, as well as different but related problems, like pilot contamination and realistic limitations sometimes neglected in some models.

In Chapter 2, the specific channel model used in this thesis was defined. A massive MU

MIMO channel was considered, but the use of time-orthogonal pilots allowed the problem to be reduced to a one-user channel estimation problem. A short description of the properties of ULAs and the spatial channel covariance matrix \mathbf{R} was included. The formulation of the analog/digital hybrid-architecture channel estimation problem as an MMSE optimization problem followed.

In Chapter 3, four already-existing methods for estimating channels in the above context were presented. The first one works only for a single pilot symbol, whereas the other three work for a pilot sequence of arbitrary length. All four methods were designed for hybrid-architecture systems with phase shifters. They more or less follow the same basic idea, where the cost function is rewritten in such a way that the problem can be reduced into computing block generalized Rayleigh quotients. Due to their differences, these methods offer a trade-off between simplicity and performance. Furthermore, it is assumed for each of these methods that the channel covariance matrix \mathbf{R} is full-rank, which is not a valid assumption for mmWave channels. The method with the best performance, called the SO method, was used as a benchmark in this thesis.

In Chapter 4, new types of switches were proposed and compared to the existing phase shifters from a practicality and an efficiency point of view. These new switches can implement more values for the matrix representation of the analog processing unit in the transceiver than traditional switches. The channel estimation problem was then reformulated in terms of these switches. Finally, a GA was proposed to solve the combinatorial problem of switch-based channel estimation. A review of the important properties of GAs and examples of where they were applied was provided.

In Chapter 5, the proposed solution was evaluated and compared to the SO method in various conditions. The methodology explained the simulation process and enumerated the values used for the different parameters. The GA and the SO method were first compared to an exhaustive search for a system of small dimensions. Then, the comparison between the two was carried out for massive MIMO systems with different numbers of allowed switch values, and different choices of system parameters, i.e. number of antennas, RF chains, pilot symbols, and propagation paths. Finally, the performance of the different switches was compared.

Throughout the thesis, the advantages of the very low hardware complexity of the switches were highlighted. It was shown that in spite of this complexity, they can still outperform phase shifters. Moreover, since they are also faster, less expensive, and less

energy-consuming, the approach presented here proves that they are solid candidates for real-life implementation.

6.2 Future Prospects

This last section lists possible research directions for future work in the area.

- The proposed GA yields effective estimates but is more complex than the SO method. In this regard, it would be of interest to further investigate if and how the complexity of its implementation could be reduced by further optimizing its internal parameters and operations.
- As explained in Section 4.1, RF switches do not have the same drawbacks as phase shifters, since switches work for any frequency, not only for a limited frequency range. They are thus well adapted to work with broadband frequency-selective channels. The solution proposed in this thesis could be adapted to this type of channel and provide a realistic hybrid transceiver architecture.
- The proposed GA could be applied on more general switching networks, such as ones with non-homogeneous sets of switches. The analog component of a transceiver could for example be equipped with switches implementing $\{0, -1, 1, -j, j\}$ and simple on/off switches. This would provide a compromise between the performance of the more complex switches and the simplicity of the more basic switches.
- The simulations show that it is better to use the GA and optimize directly over a discrete space than to use the SO method and optimize over a continuous space, project the optimized results on the unit circle, and then quantize their phases. Since the GA introduced in this thesis is very general, it could be applied directly to a phase shifter network with 3 or more quantization bits. More work could be done to find parameter values that provide good results with this architecture.
- A GA could be used to estimate channels on even more general systems, such as ones with switches and phase shifters. Because of its high versatility, the GA could actually be used to optimize the sizes and the configurations of each of the components. The cost function used for this search could be modified to put more emphasis on reducing

the size of the system, reducing the number of RF chains, or simply obtaining the lowest estimation error.

- The method presented in this thesis assumes that the channel covariance matrix \mathbf{R} is known beforehand with negligible estimation error. It would be interesting to compare how the proposed GA and the SO method behave under different levels of imperfection in the estimate of \mathbf{R} and see how the quality of their results are affected.

Appendix A

Derivation of the Channel Covariance Matrix and its Properties

A.1 Derivation for an Arbitrary Antenna Array

Combining the definitions

$$\mathbf{g} = \frac{1}{\sqrt{\sigma_\alpha^2 P}} \sum_{p=1}^P \alpha_p \mathbf{a}(\theta_p), \quad (\text{A.1})$$

for the multipath vector channel and $\mathbf{R} = \mathbb{E}[\mathbf{g}\mathbf{g}^H]$ for the spatial covariance matrix yields

$$\mathbf{R} = \mathbb{E}[\mathbf{g}\mathbf{g}^H] \quad (\text{A.2})$$

$$= \mathbb{E} \left[\frac{1}{\sqrt{\sigma_\alpha^2 P}} \sum_{p=1}^P \alpha_p \mathbf{a}(\theta_p) \left(\frac{1}{\sqrt{\sigma_\alpha^2 P}} \sum_{q=1}^P \alpha_q \mathbf{a}(\theta_q) \right)^H \right] \quad (\text{A.3})$$

$$= \frac{1}{\sigma_\alpha^2 P} \sum_{p,q} \mathbb{E}[\alpha_p \mathbf{a}(\theta_p) \alpha_q^* \mathbf{a}(\theta_q)^H] \quad (\text{A.4})$$

$$= \frac{1}{\sigma_\alpha^2 P} \left(\sum_p \mathbb{E}[|\alpha_p|^2] \mathbf{a}(\theta_p) \mathbf{a}(\theta_p)^H + \sum_{p \neq q} \mathbb{E}[\alpha_p \alpha_q^*] \mathbf{a}(\theta_p) \mathbf{a}(\theta_q)^H \right). \quad (\text{A.5})$$

The second term in the above sum vanishes because α_p and α_q are independent and identically distributed (i.i.d.) with mean zero, which means that

$$\mathbb{E}[\alpha_p \alpha_q^*] = \mathbb{E}[\alpha_p] \mathbb{E}[\alpha_q^*] = 0. \quad (\text{A.6})$$

Thus, using $\mathbb{E}[|\alpha_p|^2] = \sigma_\alpha^2$, \mathbf{R} is expressed as

$$\mathbf{R} = \frac{1}{\sigma_\alpha^2 P} \sum_{p=1}^P \sigma_\alpha^2 \mathbf{a}(\theta_p) \mathbf{a}(\theta_p)^H \quad (\text{A.7})$$

$$= \frac{1}{P} \sum_{p=1}^P \mathbf{a}(\theta_p) \mathbf{a}(\theta_p)^H. \quad (\text{A.8})$$

Note that because $\mathbf{a}(\theta_p)$ is an array response vector, it cannot be zero. This means that $\mathbf{a}(\theta_p) \mathbf{a}(\theta_p)^H$ has rank 1. Therefore, \mathbf{R} is a linear combination of P rank-1 matrices and its rank must be at most P . Note however that the rank of \mathbf{R} can be lower if for example $\theta_p = \theta_q$ for some $p \neq q$. Hence, in general, we conclude that

$$1 \leq \text{rank } \mathbf{R} \leq P. \quad (\text{A.9})$$

A.2 Derivation for a Uniform Linear Array

For a ULA, the array response vector can be expressed as

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 \\ e^{j\frac{2\pi d}{\lambda} \sin \theta} \\ \vdots \\ e^{(N-1)j\frac{2\pi d}{\lambda} \sin \theta} \end{bmatrix}. \quad (\text{A.10})$$

Hence, the (m, n) th entry of $\mathbf{a}(\theta_p) \mathbf{a}(\theta_p)^H$ is

$$[\mathbf{a}(\theta_p) \mathbf{a}(\theta_p)^H]_{(m,n)} = e^{(m-1)j\frac{2\pi d}{\lambda} \sin \theta_p} e^{-(n-1)j\frac{2\pi d}{\lambda} \sin \theta_p} = e^{(m-n)j\frac{2\pi d}{\lambda} \sin \theta_p}. \quad (\text{A.11})$$

From (A.8), it follows that the (m, n) th entry of \mathbf{R} can be expressed as

$$[\mathbf{R}]_{(m,n)} = \frac{1}{P} \sum_{p=1}^P e^{(m-n)j\frac{2\pi d}{\lambda} \sin \theta_p}. \quad (\text{A.12})$$

Using (A.12), it is easy to see that the diagonal entries of \mathbf{R} are

$$[\mathbf{R}]_{(n,n)} = \frac{1}{P} \sum_{p=1}^P e^{(n-n)j\frac{2\pi d}{\lambda} \sin \theta_p} = \frac{1}{P} \sum_{p=1}^P 1 = 1, \quad (\text{A.13})$$

from which it follows that

$$\text{tr}(\mathbf{R}) = \sum_{n=1}^N [\mathbf{R}]_{(n,n)} = \sum_{n=1}^N 1 = N. \quad (\text{A.14})$$

Appendix B

Derivation of the Optimal Baseband Processor Matrix

Given the expressions of the concatenated signal $\mathbf{y}_c = \sqrt{\rho}\mathbf{F}_c\mathbf{g} + \mathbf{F}_d\mathbf{n}_c$ received at the baseband processor and of the estimated channel $\hat{\mathbf{g}} = \mathbf{W}\mathbf{y}_c$, the objective function in (2.11) can be written as

$$\begin{aligned} \mathcal{M}(\mathcal{F}, \mathbf{W}) &= \mathbb{E}[\|\mathbf{g} - \hat{\mathbf{g}}\|^2] \\ &= \mathbb{E}[(\mathbf{g} - \hat{\mathbf{g}})^H (\mathbf{g} - \hat{\mathbf{g}})] \\ &= \underbrace{\mathbb{E}[\mathbf{g}^H \mathbf{g}]}_{(1)} - \underbrace{\mathbb{E}[\mathbf{y}_c^H \mathbf{W}^H \mathbf{g}]}_{(2)} - \underbrace{\mathbb{E}[\mathbf{g}^H \mathbf{W} \mathbf{y}_c]}_{(3)} + \underbrace{\mathbb{E}[\mathbf{y}_c^H \mathbf{W}^H \mathbf{W} \mathbf{y}_c]}_{(4)}. \end{aligned} \quad (\text{B.1})$$

The four terms in the previous equation are evaluated separately for convenience.

Term (1):

$$\mathbb{E}[\mathbf{g}^H \mathbf{g}] = \mathbb{E}[\text{tr}(\mathbf{g}^H \mathbf{g})] = \mathbb{E}[\text{tr}(\mathbf{g} \mathbf{g}^H)] = \text{tr}(\mathbb{E}[\mathbf{g} \mathbf{g}^H]) = \text{tr} \mathbf{R} \quad (\text{B.2})$$

Term (2): Under the assumption that the channel and noise are uncorrelated with zero

mean, i.e. $\mathbb{E}[\mathbf{g}\mathbf{n}^H] = \mathbf{0}$:

$$\begin{aligned}
 \mathbb{E}[\mathbf{y}_c^H \mathbf{W}^H \mathbf{g}] &= \mathbb{E}[\sqrt{\rho} \mathbf{g}^H \mathbf{F}_c^H \mathbf{W}^H \mathbf{g} + \mathbf{n}_c^H \mathbf{F}_d^H \mathbf{W}^H \mathbf{g}] \\
 &= \sqrt{\rho} \mathbb{E}[\mathbf{g}^H \mathbf{F}_c^H \mathbf{W}^H \mathbf{g}] + \underbrace{\mathbb{E}[\mathbf{n}_c]^H}_{=0} \mathbf{F}_d^H \mathbf{W}^H \underbrace{\mathbb{E}[\mathbf{g}]}_{=0} \\
 &= \sqrt{\rho} \mathbb{E}[\text{tr}(\mathbf{g}^H \mathbf{F}_c^H \mathbf{W}^H \mathbf{g})] \\
 &= \sqrt{\rho} \text{tr}(\mathbb{E}[\mathbf{W}^H \mathbf{g} \mathbf{g}^H \mathbf{F}_c^H]) \\
 &= \sqrt{\rho} \text{tr}(\mathbf{W}^H \mathbf{R} \mathbf{F}_c^H)
 \end{aligned} \tag{B.3}$$

Term (3): Applying the conjugate transpose operation to (B.3) yields:

$$\mathbb{E}[\mathbf{g}^H \mathbf{W} \mathbf{y}_c] = \sqrt{\rho} \text{tr}(\mathbf{F}_c \mathbf{R} \mathbf{W}) \tag{B.4}$$

Term (4):

$$\begin{aligned}
 \mathbb{E}[\mathbf{y}_c^H \mathbf{W}^H \mathbf{W} \mathbf{y}_c] &= \mathbb{E}[\rho \mathbf{g}^H \mathbf{F}_c^H \mathbf{W}^H \mathbf{W} \mathbf{F}_c \mathbf{g} + \sqrt{\rho} \mathbf{n}_c^H \mathbf{F}_d^H \mathbf{W}^H \mathbf{W} \mathbf{F}_c \mathbf{g} \\
 &\quad + \sqrt{\rho} \mathbf{g}^H \mathbf{F}_c^H \mathbf{W}^H \mathbf{W} \mathbf{F}_d \mathbf{n}_c + \mathbf{n}_c^H \mathbf{F}_d^H \mathbf{W}^H \mathbf{W} \mathbf{F}_d \mathbf{n}_c] \\
 &= \mathbb{E}[\text{tr}(\rho \mathbf{g}^H \mathbf{F}_c^H \mathbf{W}^H \mathbf{W} \mathbf{F}_c \mathbf{g})] + \sqrt{\rho} \underbrace{\mathbb{E}[\mathbf{n}_c]^H}_{=0} \mathbf{F}_d^H \mathbf{W}^H \mathbf{W} \mathbf{F}_c \underbrace{\mathbb{E}[\mathbf{g}]}_{=0} \\
 &\quad + \sqrt{\rho} \underbrace{\mathbb{E}[\mathbf{g}]^H}_{=0} \mathbf{F}_c^H \mathbf{W}^H \mathbf{W} \mathbf{F}_d \underbrace{\mathbb{E}[\mathbf{n}_c]}_{=0} + \mathbb{E}[\text{tr}(\mathbf{n}_c^H \mathbf{F}_d^H \mathbf{W}^H \mathbf{W} \mathbf{F}_d \mathbf{n}_c)] \\
 &= \rho \text{tr}(\mathbf{F}_c^H \mathbf{W}^H \mathbf{W} \mathbf{F}_c \mathbb{E}[\mathbf{g} \mathbf{g}^H]) + \text{tr}(\mathbf{F}_d^H \mathbf{W}^H \mathbf{W} \mathbf{F}_d \mathbb{E}[\mathbf{n}_c \mathbf{n}_c^H]) \\
 &= \rho \text{tr}(\mathbf{F}_c^H \mathbf{W}^H \mathbf{W} \mathbf{F}_c \mathbf{R}) + \text{tr}(\mathbf{F}_d^H \mathbf{W}^H \mathbf{W} \mathbf{F}_d \mathbf{I}_{NT}) \\
 &= \text{tr}(\mathbf{W}^H \mathbf{W} (\rho \mathbf{F}_c \mathbf{R} \mathbf{F}_c^H + \mathbf{F}_d \mathbf{F}_d^H))
 \end{aligned} \tag{B.5}$$

Combining the four terms results in the following expression.

$$\begin{aligned}
 \mathcal{M}(\mathcal{F}, \mathbf{W}) &= \text{tr} \mathbf{R} - \sqrt{\rho} \text{tr}(\mathbf{W}^H \mathbf{R} \mathbf{F}_c^H) - \sqrt{\rho} \text{tr}(\mathbf{F}_c \mathbf{R} \mathbf{W}) + \text{tr}(\mathbf{W}^H \mathbf{W} (\rho \mathbf{F}_c \mathbf{R} \mathbf{F}_c^H + \mathbf{F}_d \mathbf{F}_d^H)) \\
 &= \text{tr}(\mathbf{R} - \sqrt{\rho} \mathbf{W}^H \mathbf{R} \mathbf{F}_c^H - \sqrt{\rho} \mathbf{F}_c \mathbf{R} \mathbf{W} + \mathbf{W}^H \mathbf{W} (\rho \mathbf{F}_c \mathbf{R} \mathbf{F}_c^H + \mathbf{F}_d \mathbf{F}_d^H))
 \end{aligned} \tag{B.6}$$

To simplify the subsequent developments, let $\mathbf{A} := \sqrt{\rho} \mathbf{R} \mathbf{F}_c^H$ and $\mathbf{B} := \rho \mathbf{F}_c \mathbf{R} \mathbf{F}_c^H + \mathbf{F}_d \mathbf{F}_d^H$. It is easy to see that $\mathbf{A}^H = \sqrt{\rho} \mathbf{F}_c \mathbf{R}$ and that \mathbf{B} is Hermitian. Furthermore, one can show that \mathbf{B} is positive semi-definite, which implies that \mathbf{B} has a unique Hermitian positive semi-

definite square root $\mathbf{B}^{1/2}$. If we further assume that \mathbf{B} is invertible, then so is $\mathbf{B}^{1/2}$, and its inverse can be denoted by $\mathbf{B}^{-1/2}$. Thus, the MSE in (B.6) can be rewritten as:

$$\begin{aligned}
 \mathcal{M}(\mathcal{F}, \mathbf{W}) &= \text{tr}(\mathbf{R} + \mathbf{W}^H \mathbf{W} \mathbf{B} - \mathbf{W}^H \mathbf{A} - \mathbf{A}^H \mathbf{W}) \\
 &= \text{tr}(\mathbf{R} + \mathbf{B}^{1/2} \mathbf{W}^H \mathbf{W} \mathbf{B}^{1/2} - \mathbf{B}^{1/2} \mathbf{W}^H \mathbf{A} \mathbf{B}^{-1/2} - \mathbf{B}^{-1/2} \mathbf{A}^H \mathbf{W} \mathbf{B}^{1/2}) \\
 &= \text{tr} \left(\mathbf{R} + (\mathbf{W} \mathbf{B}^{1/2})^H \mathbf{W} \mathbf{B}^{1/2} - (\mathbf{W} \mathbf{B}^{1/2})^H \mathbf{A} \mathbf{B}^{-1/2} \right. \\
 &\quad \left. - (\mathbf{A} \mathbf{B}^{-1/2})^H \mathbf{W} \mathbf{B}^{1/2} + (\mathbf{A} \mathbf{B}^{-1/2})^H \mathbf{A} \mathbf{B}^{-1/2} - (\mathbf{A} \mathbf{B}^{-1/2})^H \mathbf{A} \mathbf{B}^{-1/2} \right) \\
 &= \text{tr} \left(\mathbf{R} + (\mathbf{W} \mathbf{B}^{1/2})^H (\mathbf{W} \mathbf{B}^{1/2} - \mathbf{A} \mathbf{B}^{-1/2}) \right. \\
 &\quad \left. - (\mathbf{A} \mathbf{B}^{-1/2})^H (\mathbf{W} \mathbf{B}^{1/2} + \mathbf{A} \mathbf{B}^{-1/2}) - \mathbf{B}^{-1/2} \mathbf{A}^H \mathbf{A} \mathbf{B}^{-1/2} \right) \\
 &= \text{tr} \left(\mathbf{R} + (\mathbf{W} \mathbf{B}^{1/2} - \mathbf{A} \mathbf{B}^{-1/2})^H (\mathbf{W} \mathbf{B}^{1/2} - \mathbf{A} \mathbf{B}^{-1/2}) - \mathbf{B}^{-1/2} \mathbf{A}^H \mathbf{A} \mathbf{B}^{-1/2} \right) \quad (\text{B.7})
 \end{aligned}$$

Using the definition of the Frobenius norm, $\|\mathbf{M}\|_F := \sqrt{\text{tr}(\mathbf{M}^H \mathbf{M})}$, the MSE can be written as

$$\mathcal{M}(\mathcal{F}, \mathbf{W}) = \text{tr} \mathbf{R} + \|\mathbf{W} \mathbf{B}^{1/2} - \mathbf{A} \mathbf{B}^{-1/2}\|_F^2 - \text{tr}(\mathbf{B}^{-1/2} \mathbf{A}^H \mathbf{A} \mathbf{B}^{-1/2}). \quad (\text{B.8})$$

The second term on the right hand side in the above equation is the only dependence of $\mathcal{M}(\mathcal{F}, \mathbf{W})$ on \mathbf{W} . Since it is a norm, it is minimized if and only the term inside in its argument is zero. This means that for a given \mathcal{F} , the MSE $\mathcal{M}(\mathcal{F}, \mathbf{W})$ is minimized when

$$\mathbf{W} \mathbf{B}^{1/2} - \mathbf{A} \mathbf{B}^{-1/2} = \mathbf{0} \iff \mathbf{W} \mathbf{B}^{1/2} = \mathbf{A} \mathbf{B}^{-1/2} \iff \mathbf{W} = \mathbf{A} \mathbf{B}^{-1} \quad (\text{B.9})$$

Using the above definitions of \mathbf{A} and \mathbf{B} , the optimal value of \mathbf{W} can be expressed as

$$\mathbf{W}_{\text{opt}} = \mathbf{A} \mathbf{B}^{-1} = \sqrt{\rho} \mathbf{R} \mathbf{F}_c^H (\rho \mathbf{F}_c \mathbf{R} \mathbf{F}_c^H + \mathbf{F}_d \mathbf{F}_d^H)^{-1}, \quad (\text{B.10})$$

which proves (2.12). Finally, for this optimal choice of \mathbf{W} , the objective function (2.14) can be expressed as

$$\begin{aligned}
 \mathcal{M}(\mathcal{F}) &:= \mathcal{M}(\mathcal{F}, \mathbf{W}_{\text{opt}}) = \text{tr}(\mathbf{R} - \mathbf{A} \mathbf{B}^{-1} \mathbf{A}^H) \\
 &= \text{tr} \left(\mathbf{R} - \rho \mathbf{R} \mathbf{F}_c^H (\rho \mathbf{F}_c \mathbf{R} \mathbf{F}_c^H + \mathbf{F}_d \mathbf{F}_d^H)^{-1} \mathbf{F}_c \mathbf{R} \right). \quad (\text{B.11})
 \end{aligned}$$

Appendix C

Derivation of the Bounds on the Objective Function

The objective function $\mathcal{M} : \mathcal{F}^T \rightarrow \mathbb{R}$ in problem (2.11) is defined as $\mathcal{M}(\mathcal{F}) = \mathbb{E}[\|\mathbf{g} - \hat{\mathbf{g}}\|^2]$, where $\hat{\mathbf{g}} = \mathbf{W}\mathbf{y}_c$. By definition of a norm, $\|\mathbf{g} - \hat{\mathbf{g}}\| \geq 0$ for any $\mathbf{g}, \hat{\mathbf{g}} \in \mathbb{C}^N$. It follows that $0 \leq \mathcal{M}(\mathcal{F})$.

To prove the upper bound, the result of Appendix B is used:

$$\mathcal{M}(\mathcal{F}) = \text{tr}(\mathbf{R} - \mathbf{A}\mathbf{B}^{-1}\mathbf{A}^H), \quad (\text{C.1})$$

where $\mathbf{A} := \sqrt{\rho}\mathbf{R}\mathbf{F}_c^H$ and $\mathbf{B} := \rho\mathbf{F}_c\mathbf{R}\mathbf{F}_c^H + \mathbf{F}_d\mathbf{F}_d^H$. Since \mathbf{B} is Hermitian positive semi-definite, it has a unique Hermitian positive semi-definite square root $\mathbf{B}^{1/2}$. The objective function can thus be rewritten as

$$\begin{aligned} \mathcal{M}(\mathcal{F}) &= \text{tr}(\mathbf{R} - \mathbf{A}\mathbf{B}^{-1/2}\mathbf{B}^{-1/2}\mathbf{A}^H) \\ &= \text{tr}(\mathbf{R}) - \text{tr}\left(\left(\mathbf{B}^{-1/2}\mathbf{A}^H\right)^H \mathbf{B}^{-1/2}\mathbf{A}^H\right). \end{aligned} \quad (\text{C.2})$$

Using the definition of the Frobenius norm, $\|\mathbf{M}\|_F := \sqrt{\text{tr}(\mathbf{M}^H\mathbf{M})}$ and the normalization condition $\text{tr}(\mathbf{R}) = N$, the above expression simplifies to

$$\mathcal{M}(\mathcal{F}) = N - \|\mathbf{B}^{-1/2}\mathbf{A}^H\|_F^2. \quad (\text{C.3})$$

Since $\|\mathbf{B}^{-1/2}\mathbf{A}^{\text{H}}\|_{\text{F}}^2 \geq 0$, we finally obtain

$$\mathcal{M}(\mathcal{F}) \leq N, \tag{C.4}$$

which shows (2.15).

References

- [1] “Cisco visual networking index: Forecast and trends, 2017-2022.” <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>, Feb. 2019.
- [2] “Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022.” <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html>, Feb. 2019.
- [3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5G be?,” *IEEE J. on Selected Areas in Communications*, vol. 32, pp. 1065–1082, June 2014.
- [4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” *IEEE J. of Selected Topics in Signal Processing*, vol. 8, pp. 742–758, Oct. 2014.
- [5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Communications Magazine*, vol. 52, pp. 186–195, Feb. 2014.
- [6] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.
- [7] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Communications Magazine*, vol. 49, pp. 101–107, June 2011.
- [8] C. I. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, “Toward green and soft: a 5G perspective,” *IEEE Communications Magazine*, vol. 52, pp. 66–73, Feb. 2014.
- [9] C. Li, J. Zhang, and K. B. Letaief, “Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations,” *IEEE Trans. on Wireless Communications*, vol. 13, pp. 2505–2517, May 2014.

-
- [10] A. Morsali, A. Haghighat, and B. Champagne, "Realizing fully digital precoders in hybrid A/D architecture with minimum number of RF chains," *IEEE Communications Letters*, vol. 21, pp. 2310–2313, June 2017.
- [11] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. of Selected Topics in Signal Processing*, vol. 10, pp. 501–513, Apr. 2016.
- [12] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming," *IEEE Trans. on Wireless Communications*, vol. 15, pp. 3311–3326, May 2016.
- [13] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. of Selected Topics in Signal Processing*, vol. 8, pp. 831–846, Oct. 2014.
- [14] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Communications Letters*, vol. 20, pp. 1259–1262, June 2016.
- [15] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. on Selected Areas in Communications*, vol. 35, pp. 1996–2009, Sep. 2017.
- [16] Z. Sha, Z. Wang, and S. Chen, "Harmonic retrieval based baseband channel estimation for millimeter wave OFDM systems," *IEEE Trans. on Vehicular Technology*, vol. 68, pp. 2668–2681, Mar. 2019.
- [17] L. Pan, L. Liang, W. Xu, and X. Dong, "Framework of channel estimation for hybrid analog-and-digital processing enabled massive MIMO communications," *IEEE Trans. on Communications*, vol. 66, pp. 3902–3915, Sep. 2018.
- [18] S. Buzzi and C. D'Andrea, "Subspace tracking and least squares approaches to channel estimation in millimeter wave multiuser MIMO," *IEEE Trans. on Communications*, pp. 1–1, June 2019.
- [19] F. De Figueiredo, F. Cardoso, I. Moerman, and G. Fraidenraich, "Channel estimation for massive MIMO TDD systems assuming pilot contamination and frequency selective fading," *IEEE Access*, vol. 5, pp. 17733–17741, Sep. 2017.
- [20] H. Gao, T. Zhang, C. Feng, and Y. Wang, "Clustering based pilot allocation algorithm for mitigating pilot contamination in massive MIMO systems," in *Asia-Pacific Microwave Conference (APMC)*, pp. 878–880, Nov. 2018.

-
- [21] X. Zhu, L. Dai, and Z. Wang, "Graph coloring based pilot allocation to mitigate pilot contamination for multi-cell massive MIMO systems," *IEEE Communications Letters*, vol. 19, pp. 1842–1845, Oct. 2015.
- [22] M. Feng and Y. Xu, "Low-complexity linear precoding for pilot contamination mitigation in multi-cell massive MIMO systems," in *Int. Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 807–811, Nov. 2017.
- [23] M. Boudaya, I. Kammoun, and M. Siala, "Efficient pilot design based on the divide and conquer approach for pilot contamination mitigation in massive MIMO," in *Int. Wireless Communications Mobile Computing Conference (IWCMC)*, pp. 1668–1673, June 2019.
- [24] W. Li, J. Wu, C. Hsu, and L. Wang, "Least-squares pilot sequence design for TDD massive MIMO systems under inter-cell timing misalignment," in *IEEE Annual Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–5, Sep. 2018.
- [25] T. X. Vu, T. A. Vu, and T. Q. S. Quek, "Successive pilot contamination elimination in multiantenna multicell networks," *IEEE Wireless Communications Letters*, vol. 3, pp. 617–620, Dec. 2014.
- [26] M. C. Necker and G. L. Stuber, "Totally blind channel estimation for OFDM on fast varying mobile radio channels," *IEEE Trans. on Wireless Communications*, vol. 3, pp. 1514–1525, Sep. 2004.
- [27] T. Peken, G. Vanhoy, and T. Bose, "Blind channel estimation for massive MIMO," *Analog Integrated Circuits and Signal Processing*, vol. 91, pp. 257–266, May 2017.
- [28] J.-L. Yu, B. Zhang, and P.-T. Chen, "Blind and semi-blind channel estimation with fast convergence for MIMO-OFDM systems," *Signal Processing*, vol. 95, pp. 1 – 9, Feb. 2014.
- [29] C. Y. Wu, W. J. Huang, and W. H. Chung, "Low-complexity semiblind channel estimation in massive MU-MIMO systems," *IEEE Trans. on Wireless Communications*, vol. 16, pp. 6279–6290, Sep. 2017.
- [30] P. Dong, H. Zhang, G. Y. Li, I. Gaspar, and N. Naderializadeh, "Deep CNN-based channel estimation for mmwave massive MIMO systems," *IEEE J. of Selected Topics in Signal Processing*, pp. 1–1, July 2019.
- [31] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. on Vehicular Technology*, vol. 67, pp. 8549–8560, Sep. 2018.

-
- [32] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs," *IEEE Communications Letters*, vol. 23, pp. 1989–1993, Nov. 2019.
- [33] Y. Jin, J. Zhang, B. Ai, and X. Zhang, "Channel estimation for mmWave massive MIMO with convolutional blind denoising network," *IEEE Communications Letters*, pp. 1–1, Nov. 2019.
- [34] K. Kim, J. Lee, and J. Choi, "Deep learning based pilot allocation scheme (DL-PAS) for 5G massive MIMO system," *IEEE Communications Letters*, vol. 22, pp. 828–831, Apr. 2018.
- [35] J. Dai, A. Liu, and V. K. N. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Trans. on Signal Processing*, vol. 66, pp. 2584–2599, May 2018.
- [36] W. Shao, S. Zhang, X. Zhang, J. Ma, N. Zhao, and V. C. M. Leung, "Massive MIMO channel estimation over the mmWave systems through parameters learning," *IEEE Communications Letters*, vol. 23, pp. 672–675, Apr. 2019.
- [37] Z. He, X. Yuan, and L. Chen, "Super-resolution channel estimation for massive MIMO via clustered sparse bayesian learning," *IEEE Trans. on Vehicular Technology*, vol. 68, pp. 6156–6160, June 2019.
- [38] J. Rodríguez-Fernández and N. González-Prelcic, "Channel estimation for hybrid mmWave MIMO systems with CFO uncertainties," *IEEE Trans. on Wireless Communications*, pp. 1–1, June 2019.
- [39] H. Iimori, G. Abreu, and G. C. Alexandropoulos, "MIMO beamforming schemes for hybrid SIC FD radios with imperfect hardware and CSI," *IEEE Trans. on Wireless Communications*, pp. 1–1, July 2019.
- [40] A. Morsali and B. Champagne, "Robust hybrid analog/digital beamforming for uplink massive-MIMO with imperfect CSI," in *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, Apr. 2019.
- [41] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.
- [42] F. Molina and J. Borràs, "Low-complexity switching network design for hybrid precoding in mmWave MIMO systems," in *European Signal Processing Conference (EUSIPCO)*, pp. 1–5, Sep. 2019.

- [43] R. Hu, J. Tong, J. Xi, Q. Guo, and Y. Yu, "Robust channel estimation for switch-based mmwave MIMO systems," in *Int. Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–7, Oct. 2017.
- [44] A. Alkhateeb, Y. Nam, J. Zhang, and R. W. Heath, "Massive MIMO combining with switches," *IEEE Wireless Communications Letters*, vol. 5, pp. 232–235, June 2016.
- [45] N. Wu, F. Zhu, and Q. Liang, "Evaluating spatial resolution and channel capacity of sparse cylindrical arrays for massive MIMO," *IEEE Access*, vol. 5, pp. 23994–24003, Oct. 2017.
- [46] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, ch. 13, p. 258. Society for Industrial and Applied Mathematics (Philadelphia), 2nd ed., Aug. 2002.
- [47] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, pp. 397–401, July 1988.
- [48] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. on Signal Processing*, vol. 52, pp. 546–557, Feb. 2004.
- [49] I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*, ch. 12, pp. 652–676. Wiley-Interscience, 2nd ed., Apr. 2003.
- [50] W. Banzhaf, P. Nordin, R. E. Keller, and F. D. Francone, *Genetic Programming: An Introduction On the Automatic Evolution of Computer Programs and Its Applications*, ch. 4, pp. 87–104. Morgan Kaufmann Publishers (San Francisco); Dpunkt-verlag (Heidelberg), 1st ed., Feb. 1998.
- [51] M. Patrascu, "Genetically enhanced modal controller design for seismic vibration in non-linear multi-damper configuration," *Proceedings of the Institution of Mechanical Engineers, Part I: J. of Systems and Control Engineering*, vol. 229, pp. 158–168, Sep. 2015.
- [52] J. Buurman, S. Zhang, and V. Babovic, "Reducing risk through real options in systems design: The case of architecting a maritime domain protection system," *Risk analysis : an official publication of the Society for Risk Analysis*, vol. 29, pp. 366–79, Mar. 2009.
- [53] T. Vidal, T. G. Crainic, M. Gendreau, N. Lahrichi, and W. Rei, "A hybrid genetic algorithm for multidepot and periodic vehicle routing problems," *Operations Research*, vol. 60, pp. 611–624, June 2012.
- [54] A. George, B. R. Rajakumar, and D. Binu, "Genetic algorithm based airlines booking terminal open/close decision system," in *Proceedings of the Int. Conference on Advances*

- in Computing, Communications and Informatics*, ICACCI '12, (New York, NY, USA), pp. 174–179, ACM, Aug. 2012.
- [55] K. Stanislawska, K. Krawiec, and T. Vihma, “Genetic programming for estimation of heat flux between the atmosphere and sea ice in polar regions,” in *Proceedings of the Annual Conference on Genetic and Evolutionary Computation*, GECCO '15, (New York, NY, USA), pp. 1279–1286, ACM, July 2015.
- [56] J. D. Lohn, D. S. Linden, G. S. Hornby, and W. F. Kraus, “Evolutionary design of an X-band antenna for NASA’s space technology 5 mission,” in *IEEE Antennas and Propagation Society Symposium*, vol. 3, pp. 2313–2316 Vol.3, June 2004.
- [57] J. D. Lohn, D. S. Linden, G. S. Hornby, A. Rodriguez-Arroyo, S. E. Seufert, B. Blevins, and T. Greenling, “Evolutionary design of a single-wire circularly-polarized X-band antenna for NASA’s space technology 5 mission,” in *IEEE Antennas and Propagation Society Int. Symposium*, vol. 2B, pp. 267–270 vol. 2B, July 2005.