



Technical University of Crete
School of Electrical and Computer Engineering

Estimation of Sparse Channels in 5G Wireless Systems

Paktitis Spyros

Diploma Thesis

Supervisor: *Prof.* Athanasios Liavas

Committee: *Prof.* Georgios Karystinos

Prof. Konstantinos Berberidis

February 7, 2022

Acknowledgements

As this long journey comes to an end, and a new page of my life is about to start, its only natural to look back and better comprehend how everything turned out the way they did.

The moments to be grateful for are beyond counting, but the people who made them treasurable are not. I am talking about the family I made at Chania. Together we went through good and bad, we grew as personalities and most importantly we concluded our studies for the engineering diploma. I am talking for none other than Stefanos, two Georges, Ilias, Iwanna, Marilena, Sofia, and Persefwnh.

Furthermore, there are no words to describe my gratitude towards my real family back at Corfu, my father Yiannis, my mother Helen, my brother Theodore, my cousins Stamatis and Costas and the rest of our big family who cared for me and helped become the person I am today.

But also, the family I chose, and it chose me throughout my school years, my childhood friends, Yiannis, Micheal, Alex and Dani. The moments we shared and the feelings I get as we all grow and carry on with our lives having each other on our side, are what make me get up from the bed everyday, fill me with drive to pursuit my dreams and become the best version of myself.

But more importantly, I would like to thank and express my gratitude towards my supervisor, Professor Athanasios Liavas, for his valuable guidance. I consider myself very lucky to have had a supervisor who would help me every time i needed it, support me and have the patience to see this trough with me.

Abstract

In this Diploma Thesis, we study a novel Channel State Information on the Transmitter side (CSIT) algorithm for the estimation of the CSIT in a multiuser, Frequency Division Multiplexing (FDD), massive MIMO wireless system. The main characteristic of a massive MIMO system is the large number of antennas at the Base Station (BS). This fact puts significant difficulties at the channel estimation process but offers significant benefits regarding spectral and energy efficiency, reliability, and capacity.

First, we present the concept of Compressive Sensing and the important research results of J.A. Tropp and A.C. Gilbert, including an algorithm for sparse signal recovery from random measurements via Orthogonal Matching Pursuit [29]. In their work, Tropp and Gilbert propose a remarkably simple algorithm for tackling signal estimation when dealing with sparse channel matrices.

Then, we familiarize ourselves with the Angular Domain representation of signals, mainly based on the book by D. Tse and P. Viswanath [10].

Last, we present an algorithm for efficient channel estimation in Massive MIMO FDD systems proposed in a paper by X. Rao and V. K. N. Lau [23]. We also studied the work of J. C. Shen, J. Zhang, K. C. Chen and K. B. Lataief [26] (see also the work of M. Massod, L.H. Afify and T.Y. Al-Naffouri [20]). We apply and test the algorithm in various scenarios of interest, revealing that efficient massive MIMO channel estimation is possible under the hypothesis of sparsity in the angular domain.

Contents

1	Introduction	8
2	Wireless Communications	11
2.1	Wireless Channels	12
2.2	Free space propagation	12
2.3	2-ray Model	14
2.4	Multi-path propagation	14
2.5	Delay Spread and Coherence Bandwidth	16
2.6	Flat fading and Frequency selective fading	17
3	Massive MIMO	18
3.1	Time division duplexing - TDD	18
3.2	Frequency division duplexing - FDD	21
3.3	Compressive Sensing - CS	22
4	Orthogonal Matching Pursuit	24
4.1	The concept	24
4.2	Signal Recovery from Random Measurements via Orthogonal Matching Pursuit	25
4.3	Experiments	27
5	Spatial modeling of MIMO channels	31
5.1	Line-of-sight channels	31
5.2	Line-of-Sight SIMO channel	31
5.3	Line-of-Sight MISO channel	34
5.4	Antenna arrays with only a line-of-sight path	35

5.5	Geographically separated transmit antennas	36
5.6	Resolvability in the angular domain	37
5.7	Pictorial Representation of Angular Domain Resolvability	40
5.8	Geographically separated receive antennas	43
5.9	Line-of-sight plus one reflected path	44
6	Modeling of MIMO Fading Channels	48
6.1	Basic Approach	48
6.2	MIMO Multipath Channel	50
6.3	Angular Domain Representation of Signals	51
6.4	Angular Domain Representation of MIMO Channels	52
7	Joint Channel Estimation	54
7.1	Joint Orthogonal Matching Pursuit	54
7.2	The system	55
7.3	Joint Sparsity Channel Model	56
7.4	Distributed Compressive CSIT Estimation and Feedback	59
7.5	Joint CSIT Recovery Algorithm Design	60
7.6	Design of Pilot training Matrix	64
8	Conclusion	74

List of Figures

2.1	Isotropic transmission	13
2.2	2-ray propagation model	14
2.3	multi-path propagation	15
3.1	Pilot reuse in multiple cells. (a) FDD downlink training. (b) TDD uplink training.	19
3.2	frame duration	22
4.1	OMP performance with Gaussian distribution	28
4.2	OMP performance from another point of view	29
4.3	OMP performance with Gaussian distribution	30
5.1	Line-of-sight MISO and SIMO transmission	33
5.2	Geographically separated transmit antennas	37
5.4	Behavior of the $ f(\Omega_r) $ function	39
5.6	Receive Beamforming patterns for fixed antenna array size	41
5.7	Beamforming patterns for different antenna array size	42
5.8	Geographically separated receive antennas	43
5.9	Channel with a direct and a reflected transmission path	44
5.10	Channel is viewed as a concatenation of two channels \mathbf{H}' and \mathbf{H}'' with intermediate (virtual) relays A and B.	45
5.11	Signal's angular separation	46
6.1	A representation of the MIMO channel in the angular domain.	50
7.1	Scattering environment	58
7.2	Training process	60

7.3	JOMP processing flow	63
7.4	NMSE of CSIT versus T	67
7.5	NMSE of CSIT versus the SNR	68
7.6	NMSE of CSIT versus s_c	69
7.7	NMSE of CSIT versus s	70
7.8	NMSE of CSIT versus N	71
7.9	NMSE of CSIT versus M	71
7.10	NMSE of CSIT versus K	73

Chapter 1

Introduction

South Korea became the first country to offer 5G (the fifth generation mobile wireless standard) on December 1, 2018 [24], and it is fair to say that the mobile industry has made astounding advances since the first mobile phone call was made in 1973. Mobile devices have changed our world in ways we could never have predicted. Many countries intend to implement 5G as soon as possible, which will help drive the Internet of Things (IoT) and Big Data. Every generation of wireless standard —abbreviated to ‘G’— has brought dizzying advances in data-carrying capacity and latency reduction, and 5G will be no exception. 5G is predicted to be at least three times quicker than current 4G standards, despite the fact that actual 5G standards have yet to be established. It’s helpful to look back at the unstoppable rise of wireless standards from the first generation (1G) to where we are now, on the verge of a global 5G rollout, to understand how we got here.

Nippon Telegraph and Telephone (NTT) launched the first generation of mobile networks in Tokyo in 1979. By 1984, NTT had rolled out 1G to fill Japan’s gap. In 1983, 1G technology arrived in the United States, and Motorola’s DynaTAC was one of the first ‘mobile’ phones to see widespread use. A few years later, other countries, such as Canada and the United Kingdom, launched their own 1G networks. However, the 1G was far from perfect, as it had poor communication quality and only covered a small area. There was no roaming support between different operators. Since different systems operated at different frequency ranges, there was no system compatibility. Worse, calls were not encrypted, so anyone with a radio scanner could listen in.

The GSM standard was utilized in Finland to launch the second generation of mobile networks, sometimes known as 2G. For the first time, calls could be encrypted, and digital

voice communications were noticeably clearer, with less static and crackling in the background. But 2G was about so much more than just telecoms; it lay the foundation for a cultural revolution. For the first time, people could send text messages (SMS), photo messages, and multimedia communications (MMS) on their phones. The digital future of 2G has replaced the analog past of 1G. As a result, consumers and businesses alike have embraced the technology in unprecedented numbers. Operators hastened to invest in new infrastructure such as mobile cell towers, despite the fact that 2G transfer speeds were initially only approximately 9.6 kbit/s. Speeds of 40 kbit/s were achievable by the end of the era, and EDGE (Enhanced Data for Global Evolution) connections offered speeds of up to 500 kbit/s. Despite relatively slow speeds, 2G revolutionized the business landscape and forever changed the world.

NTT DoCoMo launched 3G in 2001 with the goal of standardizing the network protocol used by vendors. Users could access data from anywhere in the world since the ‘data packets’ that fuel web connectivity were standardized. International roaming services became a reality for the first time. New services such as video conferencing, video streaming, and voice over IP have evolved as a result of 3G’s improved data transfer capabilities (4 times faster than 2G) (such as Skype).

In Stockholm, Sweden, and Oslo, Norway, the Long Term Evolution (LTE) 4G technology was initially introduced in 2009. It was then introduced all across the world, allowing millions of people to enjoy high-quality video streaming. 4G enables gaming, HD videos, and high-definition video conferencing by providing fast mobile web access (up to 1 gigabit per second for stationary users). The hitch was that, unlike upgrading from 2G to 3G, mobile devices for 4G had to be created particularly for it. [24].

Unlike the previous four generations of cellular technology, which each required a major paradigm shift with backward compatibility, 5G will necessitate a paradigm shift that includes extremely high carrier frequencies with massive bandwidths, extreme base station and device densities, and unprecedented numbers of antennas. However, unlike the previous four generations, it will be extensively integrated, combining any new 5G air interface and spectrum with LTE and WiFi to give global high-rate coverage and a consistent user experience. To support this, the core network must achieve unprecedented levels of flexibility and intelligence, spectrum regulation must be rethought and enhanced, and energy and cost efficiencies will become even more important considerations. [22].

Massive MIMO is a crucial technology that researchers have been focusing on. MIMO

communication was first implemented in WiFi systems about 2006, then in 3G cellular shortly after, and by the time LTE was developed, MIMO was a standard component, with two to four antennas per mobile device and up to eight each Base station (BS) sector. MIMO, in essence, embodies the spatial dimension of communication that emerges when a variety of antennas are accessible at Base Stations and mobile devices.

Massive MIMO, finds the base station equipped with a large number of antennas, usually more than 100. This has offered the researchers significant new potential to exploit but has also introduced quite a few challenges to face. One of them being the large dimensions of the channel matrices, which make the complexity of channel estimation significantly larger. This is exactly the problem the authors of [23] are trying to deal with. Proposing a novel algorithm that can exploit hidden attributes the propagation environment creates, along with useful theories drawn from the area of *Compressive Sensing*, successful channel estimation is achieved with a surprisingly small complexity. This thesis is dedicated at understanding the theory behind the above concept and successfully recreating the algorithm and the accompanied figures.

Chapter 2

Wireless Communications

Nowadays, when discussing about communication, most people think of their personal smart phones or, in other words, wireless communication. Despite being the most common type of communication the last decade, wireless communication has been a topic of scientific research since the 60's. The reason why this type of communication is the dominating one is not difficult to comprehend. To begin with, there has been an increasing demand for wire-free connectivity, mostly due to cellular telephony, with data applications gradually embracing the same requirement. VLSI technology also had a role in this by allowing for the implementation of complex signal processing algorithms and coding techniques in a small-area and in a low-power environment. However, the extraordinary success of the second generation (2G) digital wireless standards, notably the IS-95 Code Division Multiple Access, has had a considerable impact on cementing this status quo (CDMA). Further research done into wireless communications the recent years has illuminated new paths and approaches to reliable communication with many more-approaches- expected to come in the near future.

This section has drawn so much scientific attention for several reasons and most notably for the following. Firstly, the phenomenon of *fading*, which is the time-variation of the channel strengths due to small-scale effects, like multipath fading, and large-scale effects, such as path loss via distance attenuation. Furthermore, the medium on which wireless communications are based is the air, where different users trying to communicate have to deal with significant *interference*. This interference can be between transmitters communicating with a common receiver (e.g. uplink of cellular system), between signals from a single transmitter to multiple receivers (e.g. downlink of a cellular system), or

between different transmitter-receiver pairs (e.g. interference between users in different cells).

So far, when developing a wireless network, the engineer have mostly focused on mitigating the two major issues, *fading* and *interference*, in order to establish a reliable communication. Despite the fact that there are new ways focusing on boosting the *spectral* efficiency of the overall connection, a novel point of view argues that *fading* offers an opportunity to be capitalized on rather than a problem to be fixed.

2.1 Wireless Channels

A decisive feature of the wireless channel of mobile communication equipment is its quality fluctuation dependency on time and frequency. Generally, these fluctuations can be grouped in two categories.

Large scale fading, because of path loss with increasing distance and shadowing by huge objects such as buildings and hills This happens when the mobile travels a distance analogous to the cell size, and it is usually frequency independent.

Small-scale fading, because of constructive and destructive interference between the transmitter and receiver's various signal routes. This happens on a spatial scale on the order of the carrier wavelength and is frequency dependant.

Large-scale fading is more important in topics like cell-site design. Small-scale multipath fading is more important in the design of trustworthy and efficient communication systems.

It is critical to be able to figure out and solve equations of electromagnetic wave propagation in order to analyze wireless communication systems, a process that necessitates extensive understanding of the propagation environment. This is not a viable option. The theoretical technique of *channel modeling* aids us in overcoming this challenge. The wireless channel can be viewed as a (slowly) time-varying system, allowing us to continue researching this topic.

2.2 Free space propagation

Let the transmitter Tx and the receiver Rx be placed in distance d from each other in free space -there are no obstacles between them. The propagation is isotropic, which means

that the transmitted electromagnetic energy is uniformly distributed on concentric spheres with the center being the starting point of the transmission [18]. If the transmitting power is P_{T_x} , the power of the signal reaching the receiver is

$$P_{R_x}(d) = P_{T_x} \frac{1}{4\pi d^2} A_{R_x}, \quad (2.1)$$

with A_{R_x} being the effective area of the receiver antenna. This equation follows the law of *conservation of energy*.

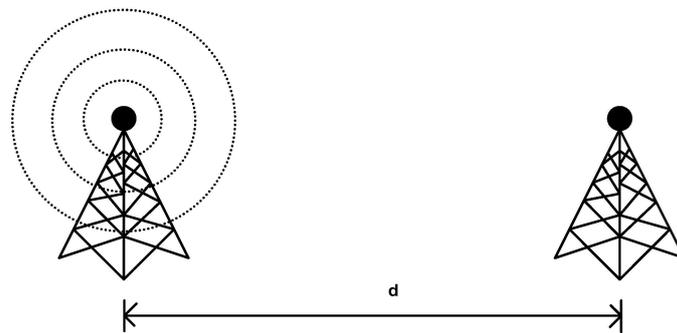


Figure 2.1: Isotropic transmission in a free space environment

In case the transmission is not isotropic, there is another quantity entering this equation, and this is the *power gain* (G_{T_x}) of the transmitting antenna in the direction of the receiving antenna. With this change, the power of the received signal is given by

$$P_{R_x}(d) = P_{T_x} G_{T_x} \frac{1}{4\pi d^2} A_{R_x}. \quad (2.2)$$

At this point, it is important to mention that, when we speak about transmitting power gain along a specific direction, we have in mind the amount described from the following equation

$$G_{T_x} = \frac{\text{Power Density at a distance } d \text{ along this direction}}{P_{T_x}/4\pi d^2}. \quad (2.3)$$

In case of isotropic transmission, $G_{T_x} = 1$.

Generally, the above mentioned equations are in fact an approximation since the power gains depend on the wavelength (λ) and, thus, also from the modulating frequency.

To summarize this section, it can be demonstrated that when transmitting in open space, the received signal power is proportional to the square of the distance between transmitter and receiver. Power gain is also affected by the frequency of the transmitted signal, but this relation is complicated and also depends on the antenna's characteristics. Overall, one could argue that if the distance between the two antennas remains constant, increasing the frequency of the transmitted signal results in reduced received signal power.

2.3 2-ray Model

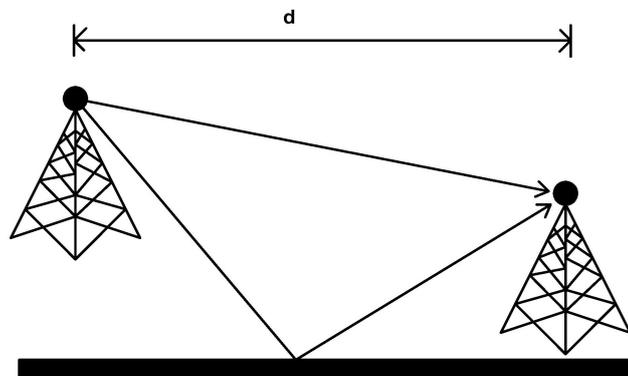


Figure 2.2: 2-ray Propagation Model

If we want to include obstacles and barriers to the propagation process of the electromagnetic waves, that exist in a real life environment, then things get tricky. An ideal model of propagation is presented in Figure 2.2, where the signal arrives at the receiver from 2 different paths, directly and scattered from the ground. This is called the 2-ray model, and under these circumstances, the signal loss is much larger than in the free-space scenario. This time the signal power is decreased proportionally to the fourth power of the distance d .

2.4 Multi-path propagation

The phrase *multi-path* propagation refers to a condition in which the signal arrives at the receiver after several reflections, significant scattering, and nearly no line-of-sight. In this

situation, each path to the communication's final point has a different attenuation, delay, and fading type. What also changes is the phase of the different electromagnetic waves upon arrival at the receiver. The two waves add *constructively* when the phase difference is an integer multiple of 2π , and the received signal is strong. When the phase difference is an odd integer multiple of π , on the other hand, the two waves add *destructively*, weakening the received signal. This phenomenon creates a spatial pattern of constructive and destructive wave interference during communication and specifically, in this case, from a spatial point of view one can talk about a *coherence distance*. Although the same pattern occurs when the modulating frequency takes certain values, putting also the term of *coherence bandwidth* in the literature. Of course, when we talk about frequency there is always its counterpart, the symbol period T . From that point of view, we can talk about *delay spread*, which has similar meaning with coherence frequency and distance but from the time perspective.

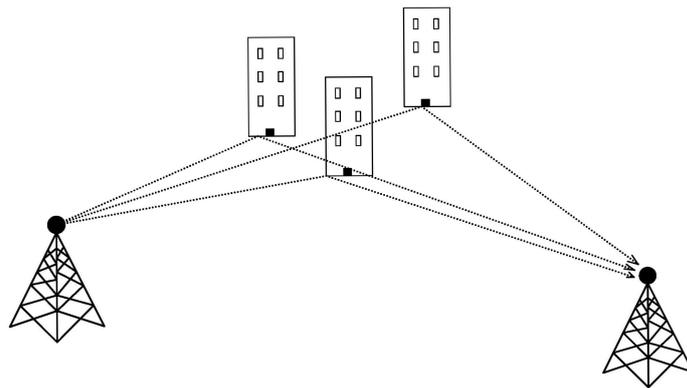


Figure 2.3: Multi-path Propagation

As it was mentioned before, for a wireless communication to be successful, one should take into consideration all the details of the surrounding physical environment in order to calculate the precise propagation equations for this system. Obviously, this is a rather tricky goal to accomplish, and is far from an ideal solution that could realistically be used in every-day communication. Luckily, this is where *Channel Modeling* techniques come to untie our hands. In an urban environment, with much scattering and reflections (a multipath propagation environment) the channel coefficients tend to take values quite similar to a *Gaussian distribution* with zero mean. This kind of channel is called *Rayleigh*

fading, and is defined as

$$h_{\text{Rayleigh}} \sim \mathcal{CN}(0, \sigma^2).$$

2.5 Delay Spread and Coherence Bandwidth

An important parameter of wireless systems is the multipath delay spread T_d , defined as the time difference between the longest and shortest path -taking into consideration only the paths with significant energy. The following equation presents how *delay spread* is defined.

$$T_d := \max_{i,j} |\tau_i(t) - \tau_j(t)|. \quad (2.4)$$

To get a better grasp of this, let us think of a cell or a LAN. If one of these two (cell or LAN) has a linear extend of a few kilometers, it is quite rare for 2 random paths to differ more than 300 to 600 meters. This distance corresponds to path delays of no more than $2 \mu\text{s}$. Growing cellular use decreases the size of the cell, which makes T_d to shrink. The delay spread T_d is substantially lower than the coherence time T_c in most wireless channels, indicating that they are underspread. The *frequency coherence* of the channel is likewise determined by the delay spread. Wireless channels alter in frequency and time. The time coherence of a channel determines how rapidly it changes in time, while the frequency coherence determines how quickly it changes in frequency.

The true significance of delay spread is the fact that it is a measure of Inter Symbol Interference (ISI) in wireless channels. What this actually means is that, in case the duration of the symbol outlasts the delay spread by a certain amount of time, the channel is considered ISI-free. This time difference typically finds symbol duration to be 10 times longer than the delay spread. Sifting into the frequency domain, the correspondence of delay spread is the *coherence bandwidth*, which is the bandwidth over which the channel's frequency response is flat. Obviously, the coherence bandwidth is the inverse of the delay spread, and this is why these two amounts grow disproportionately. The smaller the delay spread the larger the coherence bandwidth [18].

2.6 Flat fading and Frequency selective fading

Similarly to the time domain, channels are characterised as flat fading or frequency selective fading according to their frequency response. In this case, the key to making this separation is the *coherence bandwidth* (B_c).

If B_c is quite bigger than the signal bandwidth, then the symbol transmission is free of interference and the only things changing are the amplitude and phase of the received signal. Hence, this fading type is called *flat fading*. This kind of channels are statistically described by

$$h_m \sim \mathcal{CN}(0, \sigma^2). \quad (2.5)$$

On the other hand, when B_c is smaller than the signal's bandwidth, then the channel creates ISI and its response is not flat. This type of fading is called *frequency selective fading* and every new transmission contains parts of the next or the previous transmitted symbol. If we assume that there is a finite number of coefficients interfering in the channel, and that amount is L , then this kind of channel can be modelled as

$$h_{m,l} \sim \mathcal{CN}(0, \sigma_l^2), \quad l = 0, \dots, L - 1. \quad (2.6)$$

Chapter 3

Massive MIMO

As the dawn of fifth-generation (5G) wireless networks becomes brighter by the day, Massive multiple-input-multiple-output (MIMO) systems dominate the wireless technology scene with their spectral and energy efficiency. Having available high-dimensional channel state information (CSI) can lead to significant performance gains, though the overhead needed to acquire this CSI may turn out to be devastating for the radio resources.

In order to exploit the spatial multiplexing and array gains of this system in full, it is crucial for the CSI to be retrieved at the base station (BS), a task that turns out to be extremely challenging [3]. The barriers that come along with very high-dimension MIMO channels are confronted both at time-division duplexing (TDD) and frequency-division duplexing (FDD) systems. Specifically, techniques, like CSI acquisition based on feedback overhead and pilot-aided training, grow similarly to the BS antenna size, in FDD systems. Moreover, the channel's coherence period sets some serious restrictions on radio resources available for CSI acquisition, and this situation is only getting worse in an environment with high user equipment (UE) mobility.

3.1 Time division duplexing - TDD

Due to these circumstances, research has focused on TDD massive MIMO systems, exploiting channel reciprocity. The notion of *reciprocity* is one of the best advantages TDD architecture has to offer, since the reverse channel is used as an estimate of the forward one. This way, the need of feedback is eliminated, once the desired CSI can be estimated by the uplink training together with the reciprocity of the wireless medium [19]. However,

this estimation is not free of issues that need to be dealt with before this approach becomes useful.

At TDD systems, training overhead grows proportionally to the number of UEs [26]. Usually, in order to acquire precise CSI, each user has to transmit an orthogonal pilot sequence to its corresponding BS.

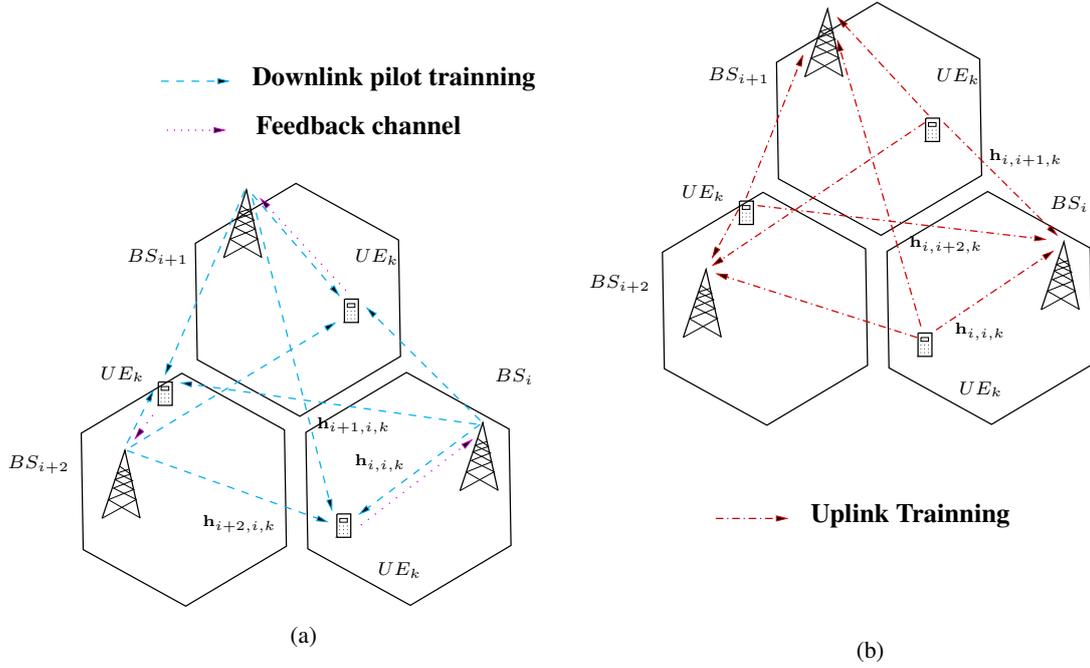


Figure 3.1: Pilot reuse in multiple cells. (a) FDD downlink training. (b) TDD uplink training.

The availability of orthogonal pilot sequences is restricted by the ratio of the channel coherence interval to the channel's delay spread [17], which is usually small due to the movement of users or the unfavorable physical environment. As the number of UEs becomes large, it gets even harder to ensure the orthogonality of the pilot sequences, making inevitable the reuse of correlated pilot sequences in different cells, in a multi-cell scenario. Proceeding the communication process with the acquired spoiled data, intercell interference occurs immediately and the performance instantly drops. *Pilot contamination*, as this phenomenon is known, is a problem that cannot be dealt with simply adding more antennas at the BS.

Consider a massive MIMO system consisting of L hexagonal cells. Each cell has a BS equipped with linear arrays of M elements -the size of BS antenna is much larger than the

number of served UEs [26]. This system serves K users, each one carrying a single-antenna UE. The $M \times 1$ vector $\mathbf{h}_{i,j,k}$ represents the channel between BS i and UE k in cell j . During the uplink training, the received signal at BS i is

$$\mathbf{Y}_i^{UL} = \sum_{l=1}^L \mathbf{S}_i^{UL} \mathbf{H}_{i,l} + \mathbf{Z}_i^{UL}, \quad (3.1)$$

where $\mathbf{H}_{i,l} = [\mathbf{h}_{i,l,1}, \dots, \mathbf{h}_{i,l,k}, \dots, \mathbf{h}_{i,l,K}]'$ consists of the channel vectors from K users in the l_{th} cell, communicating with the i_{th} BS; \mathbf{S}_i^{UL} is a set of $\tau \times 1$ pilot sequences $\{\mathbf{s}_{l,k}\}_{k=1}^K$, and \mathbf{Z}_i^{UL} , is the additive noise matrix. For better understanding of the intercell interference scenario, let us assume that the same set of orthogonal pilot sequences are reused in each cell, i.e. $\mathbf{S}_1^{UL} = \mathbf{S}_l^{UL} = \dots = \mathbf{S}_L^{UL}$ and $\mathbf{s}'_{l,k_1} \mathbf{s}_{l,k_2} = 0$ for $k_1 \neq k_2$, as shown in Figure 3.1(b). The estimated channel, using the LS method, is

$$\begin{aligned} \hat{\mathbf{H}}_{i,i} &= [(\mathbf{S}_i^{UL})^H \mathbf{S}_i^{UL}]^{-1} (\mathbf{S}_i^{UL})^H \mathbf{Y}_i^{UL} \\ &= \mathbf{H}_{i,i} + \sum_{l \neq i} \mathbf{H}_{i,l} + [(\mathbf{S}_i^{UL})^H \mathbf{S}_i^{UL}]^{-1} (\mathbf{S}_i^{UL})^H \mathbf{Z}_i^{UL}. \end{aligned} \quad (3.2)$$

When ignoring the noise, the rows of $\hat{\mathbf{H}}_{i,i}$ consist of $\hat{\mathbf{h}}_{i,i,k} = \sum_{l=1}^L \mathbf{h}_{i,l,k}$.

During downlink transmission, the uplink channel estimates $\hat{\mathbf{h}}_{i,i,k}$ are used to form the transmitted signal $\mathbf{x}_i = \sum_{k=1}^K \mathbf{w}_{i,k}^{TDD} x_{i,k}$, where $\mathbf{w}_{i,k}^{TDD} = \sum_{l=1}^L (\mathbf{h}_{i,l,k}^H)'$ are the maximum ratio transmission (MRT) precoding vectors, which consequently causes interference

$$\begin{aligned} \mathbf{I}_{i,j,m} &= \mathbf{h}'_{i,j,m} \mathbf{x}_i \\ &= \|\mathbf{h}_{i,j,m}\|_2^2 x_{i,m} + \sum_{\substack{k \neq m \\ l \neq j}} \mathbf{h}_{i,l,k}^H \mathbf{h}_{i,j,m} x_{i,k} \end{aligned} \quad (3.3)$$

to UE m in cell j . Despite the fact that increasing the number of antennas at the BS diminishes the second term of (3.3), the first term is here to stay, setting, thus, a limit to the received signal-to-interference-plus-noise ratio (SINR) at UE m in cell j .

Considerable research has been made to tackle pilot contamination in high-dimensional TDD CSI acquisition. Current investigation focuses on the impact of this phenomenon on the received SINR or the sum rate when linear precoders/detectors are applied, with very little known about the non-linear precoders/detectors scenario. Some other work in [9] proposes a open-/closed-loop training which utilizes temporal and spatial channel statistics in order to reduce the amount of downlink training overhead. Also, in [13], optimal design of precoding matrices aimed at minimizing the squared errors caused by pilot reuse turned out to be superior over linear precoding.

3.2 Frequency division duplexing - FDD

On the other hand, real-time data and applications tend to be the new status quo nowadays, consequently systems with symmetric traffic using delay-sensitive applications are on demand. A TDD system cannot cope with these requirements, making research focus on investigating the capabilities of FDD massive MIMO systems. In this case, CSI acquisition on the BS follows a different procedure than the one at TDD.

Specifically, first the BS transmits pilot symbols; each user receives the signal and estimates the downlink CSI locally. This acquired information is then transmitted back to the BS via uplink channels [27], so that the benefits of massive MIMO systems can be fully exploited.

Consider the massive MIMO system described in the previous section with L hexagonal cells, each with a base station. During downlink training, the channel of UE k , in cell i receives

$$\mathbf{y}_{i,k}^{DL} = \underbrace{\mathbf{S}_i^{DL} \mathbf{h}_{i,i,k}}_A + \underbrace{\sum_{l \neq i} \mathbf{S}_l^{DL} \mathbf{h}_{l,i,k}}_B + \mathbf{z}_{i,k}^{DL}, \quad (3.4)$$

where \mathbf{S}_l^{DL} stands for the $N \times M$ pilot training matrix used in cell l , $\mathbf{z}_{i,k}^{DL}$ is the additive noise, amount A is the desired CSI and B comes from intercell interference.

Repeatedly using the same training matrix in multiple cells, i.e. $\mathbf{S}_1^{DL} = \dots = \mathbf{S}_L^{DL}$, can be considered as pilot contamination in FDD systems. This means that BS i will collect the aggregated channel $\sum_{l=1}^L \mathbf{h}_{l,i,k}$, instead of the wanted one, $\mathbf{h}_{i,i,k}$, considering a noiseless environment and an error free feedback channel. Despite that, there is a possibility to tackle intercell interference, once the composite CSI is utilized to form a precoding vector. That is, in a maximum ratio transmission (MRT) scenario, where precoding is employed, the transmitted signal from BS i can be expressed as $\mathbf{x}_i = \sum_{k=1}^K \mathbf{w}_{i,k}^{FDD} x_{i,k}$, where $x_{i,k}$ is the signal destined for UE k within the cell, and $\mathbf{w}_{i,k}^{FDD} = \sum_{l=1}^L (\mathbf{h}_{l,i,k}^H)'$ stands for the MRT precoding vector. From the UE m 's point of view, during the downlink transmission, the received interference in cell j due to BS i is

$$I_{i,j,m} = \mathbf{h}'_{i,j,m} \mathbf{x}_i = \sum_{k=1}^K \sum_{l=1}^L \mathbf{h}_{l,i,k}^H \mathbf{h}_{i,j,m} x_{i,k}.$$

As the number of BS antennas rapidly increases, channel vectors become asymptotically orthogonal. When this occurs, the products $\mathbf{h}_{l,i,k}^H \mathbf{h}_{i,j,m}$ approach zero, and so does the in-

interference $I_{i,j,m}$. In other words, intercell interference caused by pilot contamination diminishes asymptotically when increasing the number of BS antennas. Thus, in the asymptotic scenario, intercell interference vanishes together with the need to create distinct training matrices. This is why pilot contamination of FDD systems is an issue rarely mentioned in the existing literature.

Some implementations of the above process use least squares (LS) [5] or minimum mean square error (MMSE) [31] in order to estimate the downlink CSI at the user side. However, the increasing number of antennas at the BS makes these two approaches somewhat inefficient, since the numbered pilot symbols necessary for CSI acquisition grow proportionally. Consequently, the complexity of the estimation problems to be solved becomes analogous to the ensemble of BS antennas (i.e. $\mathcal{O}(M)$). Since massive MIMO systems is the theme of this discussion, one could easily comprehend the prohibitively large overhead created not only for the pilot training (downlink), but also for the CSI feedback (uplink). To make matters worse, in such large channel dimensions, channel's coherent time and bandwidth set an extra limitation to the number of independent pilot symbols available for CSI acquisition [17]. Figure 3.2 helps illustrate this.

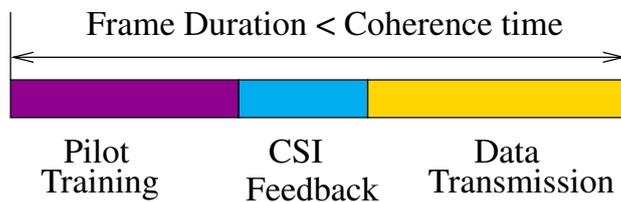


Figure 3.2: Frame structure with pilot training to obtain CSIT in massive MIMO FDD system.

It is now becoming clear that, as the number of BS antennas grow, FDD estimation of CSI does not stand up to the expectations, due to very large training overhead and limited number of available pilot symbols, forcing research to look for a new, more promising direction for high-dimensional CSIT estimation.

3.3 Compressive Sensing - CS

A game changing observation was made, at various experiments in massive MIMO system tests, which helps to tackle the above mentioned obstacles. The observation was the

following, as M increases, the user channel matrices tend to be sparse, due to limited local scattering on the BS side -a signal is considered “sparse” when most of its coefficients are zero. Also, m -sparse signal is one that has m non-zero entries regardless its dimension. Under these circumstances, estimating the entire channel matrix, which is accompanied by long pilot training symbols becomes a naive solution. Instead, exploiting the newly discovered hidden sparsity caused by the physical environment is a much more promising approach to this task.

At this point, compressive sensing (CS) enters the telecommunications world with interesting potential. Compressive sensing is a relatively new concept which made its first appearance a few years ago, and ever since gained much attention, especially at the applied mathematics and signal processing fields. Imaging, radar, speech recognition and data acquisition are a few of the areas where CS is applied. In communications, many scientific papers utilize its potential to estimate sparse channels. To clarify things, most channels in wireless massive MIMO systems are characterized by sparse multipath. This means that there are much fewer distinct arrivals than there are channel taps. What CS promises to do is to estimate the channel signal with much less training overhead (pilot symbols) or at higher accuracy with a specific number of training symbols. A CS-based low-rank approximation strategy has been presented to improve channel estimation performance for TDD systems [21], as well as a CS-based channel estimation method that tries to leverage the per-link sparse multipath channels [1]. The later finds applications in TDD, FDD as well as spatial domains, and also achieves a substantial reduction to training overhead compared to the conventional LS approach. Specifically, by exploiting spatial sparsity in a massive MIMO system, only $\mathcal{O}(s \log M)$ training overhead is needed to estimate the CSI.

Chapter 4

Orthogonal Matching Pursuit

4.1 The concept

Sparse approximation techniques have been in the spotlight of very active research since the 90s and have done nothing but evolve ever since. Their overwhelming benefit has been the compression of high-dimensional data with wavelets. However, approximating a signal or an image with a sparse linear expansion from a possibly overcomplete dictionary of basis functions (called atoms) has turned out to be an extremely useful tool to solve many other signal processing problems; blind source separation, feature extraction and classification, denoising, and detection, are only some of them.

When it comes to sparse signals, Joel A. Tropp and Anna C. Gilbert proposed a very interesting greedy algorithm at their paper [29], which deals with the channel recovery problem with outstanding simplicity and effectiveness. Orthogonal Matching Pursuit (OMP) can reliably recover a sparse signal, given some random linear measurements of that signal.

The proposition goes as follows: Let \mathbf{s} be a m -sparse, d -dimensional signal (m -sparse translates as m non-zero components in a vector signal). Given a sequence of vector measurements $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \mathbb{R}^d$, one could create N linear measurements of the signal as follows

$$\langle \mathbf{s}, \mathbf{x}_1 \rangle, \langle \mathbf{s}, \mathbf{x}_2 \rangle, \dots, \langle \mathbf{s}, \mathbf{x}_N \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the inner product. However, concerns were raised over two issues. The first one being how many measurements were enough to reconstruct the signal and the second one what algorithm could succeed in reconstructing the signal given these measurements.

Both of these questions are answered by the OMP algorithm and the experimental results which follow.

4.2 Signal Recovery from Random Measurements via Orthogonal Matching Pursuit

As we mentioned above, let \mathbf{s} be an m -sparse signal and $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a family of N measurement vectors, where $\mathbf{x}_i \in \mathbb{R}^d$. In order to collect N measurements of the signal, we form a matrix Φ of size $N \times d$ and fill its N rows with the \mathbf{x}_i vectors. At this point, we calculate the quantity $\mathbf{u} = \Phi\mathbf{s}$, creating thus an N -dimensional data vector \mathbf{u} . Since the original signal \mathbf{s} is m -sparse, the data vector \mathbf{u} is a linear combination of m columns of Φ . So, signal recovery turns out to be the much simpler problem of identifying the m columns of Φ participating in the measurement vector \mathbf{u} .

THE ALGORITHM:

INPUT:

- A $N \times d$ measurement matrix Φ
- A N -dimensional data vector \mathbf{u}
- The sparsity level m of the original signal

OUTPUT:

- An estimate $\hat{\mathbf{s}} \in \mathbb{R}^d$ of the original signal
- An index set Λ_m with the positions of the non-zero entries
- An N -dimensional approximation \mathbf{a}_m of the data vector \mathbf{u}
- An N -dimensional residual $\mathbf{r}_m = \mathbf{u} - \mathbf{a}_m$

Algorithm 1: OMP for signal recovery

Initialization:Residual $\mathbf{r}_0 = \mathbf{u}$ Index set $\Lambda_0 = \emptyset$ $\Phi_0 = \emptyset$ Iteration counter $t = 1$ **while** $t < m$ **do**(1) Find the index λ_t such that:

$$\lambda_t = \arg \max_{j=1, \dots, d} | \langle \mathbf{r}_{t-1}, \phi_j \rangle | .$$

In the case of a double max, break the tie deterministically.

(2) Update the index set $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$ and the matrix of chosen atoms

$$\Phi_t = [\Phi_{t-1} \quad \phi_{\lambda_t}] .$$

(3) Solve a least-squares problem to obtain a new signal estimate:

$$\mathbf{x}_t = \arg \min_x \|\Phi_t \mathbf{x} - \mathbf{u}\|_2 .$$

(4) Calculate the new approximation of the data and the new residual:

$$\mathbf{a}_t = \Phi_t \mathbf{x}_t, \quad \mathbf{r}_t = \mathbf{u} - \mathbf{a}_t .$$

end

(5) The estimate $\hat{\mathbf{s}}$ of the original signal has non-zero entries at the indexes listed in Λ_m . The value of the entry with index λ_j has a value that equals the j th component of \mathbf{x}_t

As one can deduce, the idea is to pick columns in a greedy fashion. With each iteration, the column of matrix Φ which has the strongest correlation with the remaining part of \mathbf{u} is selected, then its contribution is subtracted from \mathbf{u} and the iteration continues using the new residual. The goal is to identify the correct set of columns after m iterations. What makes this task easier is the fact that the residual \mathbf{r}_t is always orthogonal to the columns of Φ_t , and therefore the algorithm always selects a new column at every iteration.

As for the complexity of the above procedure, step 2 is the dominating step with a total cost of $\mathcal{O}(mNd)$. In order to face this problem, the authors proposed to deal with the least-squares problem by maintaining a **QR** factorization of Φ_t achieving, thus, a complexity of $\mathcal{O}(tN)$. Often, the measurement matrix is unstructured and dense, therefore, the modified Gram-Schmidt (MGS) algorithm was used, with alternative approaches to the issue already existing like the paper [6] published in 1996. In the case of a structured measurement matrix, more efficient implementations are proposed, like [15].

Despite the fact that algorithms dealing with unstructured and dense measurement matrices already exist, the authors focused their work in the case where d is much larger than m or N , so there would be a substantial difference at the complexity cost from the rest of the research. As for the construction of the measurement matrix Φ , the two distributions that can be used are (a) Gaussian and (b) Bernoulli, both of them normalized for mathematical convenience.

4.3 Experiments

Let us now take a look at the experimental results and see for ourselves the power of the algorithm.

The main task is to find a pattern to correctly recover a signal with high probability. For this to succeed, we need to determine the number of measurements N that are necessary. To begin with, we create an m -sparse signal by selecting m non-zero components (out of d) and give them the value 1. Just like it was described before, a $N \times d$ measurement matrix Φ is created, with entries drawn from a Gaussian distribution. By multiplying this matrix with the sparse signal \mathbf{s} the data vector \mathbf{u} is acquired ($\mathbf{u} = \Phi\mathbf{s}$) and the OMP algorithm can now begin. For each combination of the 3 variables, m , N and d the procedure is repeated 1000 times and at the end the percentage of the correctly recovered signals is calculated.

For the first experiment, the dimension d was set to 256. The curves show the per-

centage of correctly recovered signals as a function of the number of measurements N . As anticipated, there is a linear relationship between the sparsity number m and the number of measurements N , meaning that the bigger m gets the more measurements are needed to recover the signal.

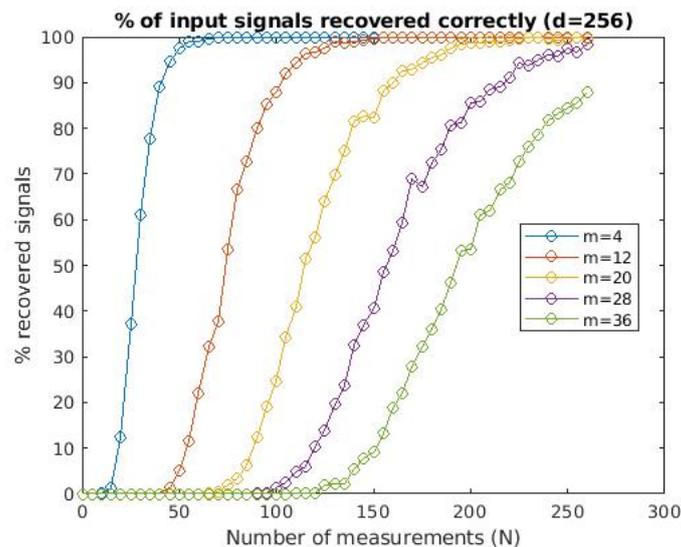


Figure 4.1: The percentages of 1000 input signals correctly recovered as a function of the number N of measurements for different sparsity levels m in dimension $d=256$.

A different point of view of the same data is drawn in Figure 4.2, where, for various number of measurements N , one can see the value of m for which a 100% recovery of signal is achieved. This representation might also be very useful from an application point of reference. Suppose there are limited resources available for a system/application; Figure 4.2 shows the expected performance when there is only enough space for, let us say, $N=100$ measurements with dimension $d=256$. Alternatively, under the same features, one can hope for a 90% recovery of the signal with 16 non-zero entries and 50% with 20 entries.

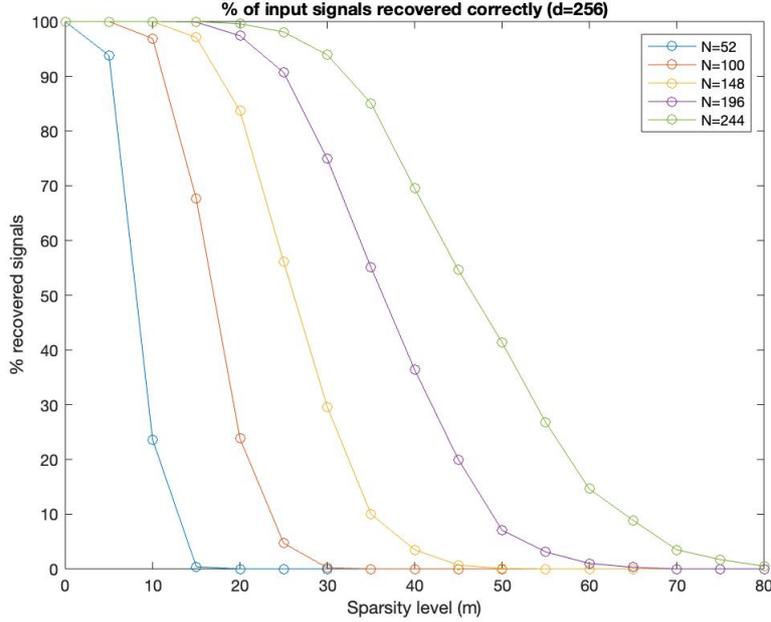


Figure 4.2: The percentages of 1000 input signals correctly recovered as a function of the sparsity level m for different number N of measurements in dimension $d=256$.

To follow the above concept, it is interesting to find out the number of measurements needed to achieve a certain success rate. The following table displays the relationship between N and m necessary to recover a sparse signal which belongs in \mathcal{R}^d , when $d = 256$, 1024, with 95% probability. It also presents the value $N/(m \ln d)$ which helps visualize a pattern of values that N should take in order to have the specific recovery ratio. When $d = 256$, $N \approx 1.67m \ln 256$, and when $d = 1024$ $N \approx 1.62m \ln 1024$. On the other hand, theoretical results were much more pessimistic on the above mentioned value of N .

d = 256			d = 1024		
m	N	$N/(m \ln d)$	m	N	$N/(m \ln d)$
4	43	1.94	5	60	1.73
8	74	1.67	10	112	1.61
12	108	1.62	15	159	1.52
16	142	1.6			
20	174	1.56			

At this point, it should be noted that the above 2 experiments were repeated using the entries with equal probability from the Bernoulli distribution in order to create the

measurements. The results were pretty much identical and for the sake of brevity only the first figure is included in this document with different distribution this time.

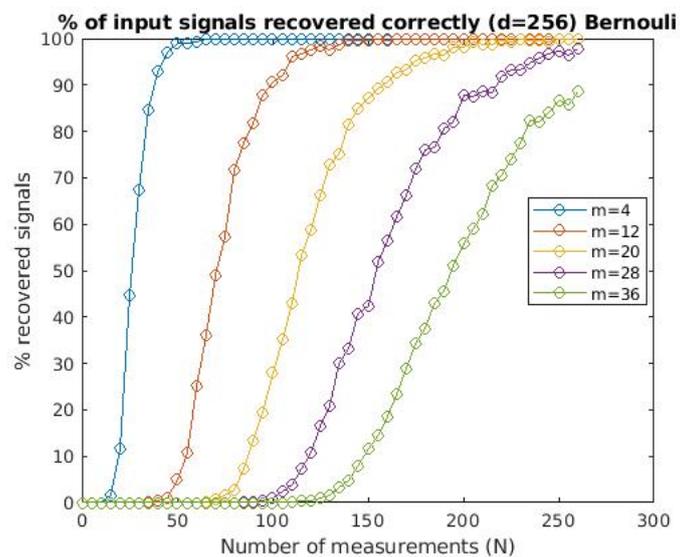


Figure 4.3: The percentage of 1000 input signals correctly recovered as a function of the number N of Bernoulli measurements for different sparsity levels m in dimension $d = 256$.

Chapter 5

Spatial modeling of MIMO channels

In this section, we focus on how the spatial multiplexing capability of MIMO channels depends on the physical environment. This study has a big range, from specifics on the condition and rank of the channel matrix, to the layout of the antennas at the transmitter and the receiver respectively (i.e. rectangular antenna array). We will be focusing on specific examples of idealized models and get a first taste on physical modeling so that later we shall look into a natural approach to statistical modeling of MIMO channels. To be concrete, we restrict ourselves to *uniform linear arrays* where the antennas are evenly spaced on a straight line.

5.1 Line-of-sight channels

The simplest wireless channel is the line-of-sight channel, where there are no reflectors or scatterers or any physical obstacles between an antenna pair. In this kind of wireless communication we can distinguish two types of channel models, MISO (Multiple Input-Single Output) and SIMO (Single Input-Multiple Output). It is important to point out that in order to proceed with the modeling of wireless channels, the distance between the transmit and receive antennas must be much greater than the size of the antenna arrays.

5.2 Line-of-Sight SIMO channel

The antenna separation in a line-of-sight SIMO channel is $\Delta_r \lambda_c$, where λ_c is the carrier wavelength and Δ_r is the normalized receive antenna separation, normalized to the carrier

wavelength's unit.

Between the transmit antenna and the i_{th} receive antenna, the continuous-time impulse response $h_i(\tau)$ is given by:

$$h_i(\tau) = \alpha \delta(\tau - d_i/c), \quad i = 1, \dots, n_r,$$

which is easy to understand if we think that α is the attenuation of the path for which we speak and d_i/c expresses the time delay of this path. The letter c stands for the speed of light. Assuming $d_i/c \ll 1/W$, where W is the transmission bandwidth, the baseband channel gain is given by:

$$h_i = \alpha \exp\left(\frac{-j2\pi f_c d_i}{c}\right) = \alpha \exp\left(\frac{-j2\pi d_i}{\lambda_c}\right),$$

where f_c is the carrier frequency. The SIMO channel can be written as

$$\mathbf{y} = \mathbf{h}x + \mathbf{w},$$

where x is the transmitted symbol, $\mathbf{w} \sim \mathcal{CN}(0, N_0\mathbf{I})$ is the noise and \mathbf{y} is the received vector. From a spatial point of view, we usually name the channel gain vector $\mathbf{h} = [h_1, \dots, h_{n_r}]^T$ signal direction or the spatial signature induced on the receive antenna array by the transmitted signal.

On the system model we are currently studying, the arrays at the transmitter and the receiver are spaced at a distance far larger than their size. So, its safe to assume that the paths from the transmit antenna to each receive antennas are, to a first order, parallel, so

$$d_i \approx d + (i - 1)\Delta_r \lambda_c \cos \phi, \quad i = 1, \dots, n_r,$$

where d is the distance from the transmit antenna to the first receive antenna and ϕ is the angle of incidence of the line of sight onto the receive antenna array.

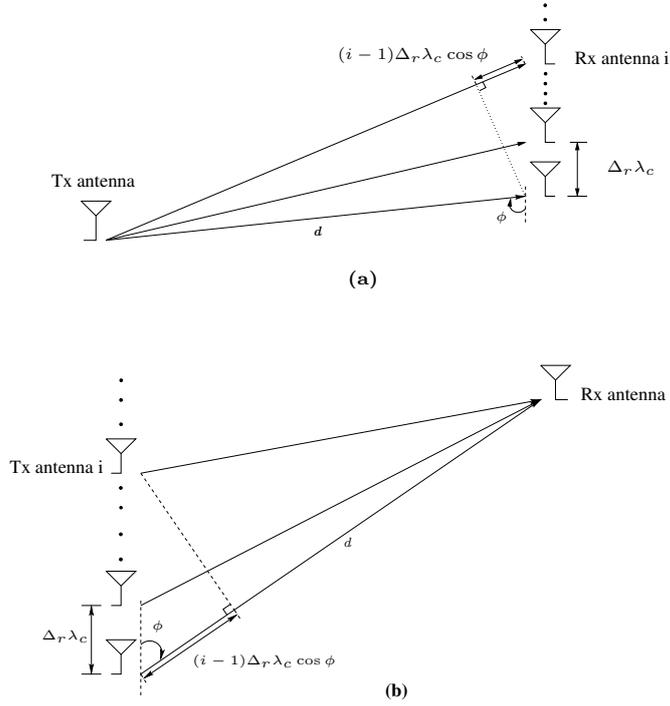


Figure 5.1: Line-of-sight channel with single transmit antenna and multiple receive antennas. The signals from the transmitting antenna arrive almost in parallel at the receiving antennas. (b). Line-of-sight channel with multiple transmit antennas and single receive antenna.

The quantity $(i-1)\Delta_r\lambda_c \cos \phi$ is the displacement of receive antenna i from receive antenna 1 in the direction of the line of sight. The quantity

$$\Omega \equiv \cos \phi$$

is often called the directional cosine with respect to the receive antenna array. The spatial signature $\mathbf{h} = [h_1, \dots, h_{n_r}]^T$ is therefore given by

$$\mathbf{h} = \alpha \exp\left(-\frac{j2\pi d}{c}\right) \begin{bmatrix} 1 \\ \exp(-j2\pi\Delta_r\Omega) \\ \exp(-j2\pi 2\Delta_r\Omega) \\ \vdots \\ \exp(-j2\pi(n_r-1)\Delta_r\Omega) \end{bmatrix}. \quad (5.1)$$

i.e., the signals received at consecutive antennas differ in phase by $2\pi\Delta_r\Omega$ due to the relative delay. For notational convenience, we define

$$\mathbf{e}_r(\Omega) := \frac{1}{\sqrt{n_r}} \begin{bmatrix} 1 \\ \exp(-j2\pi\Delta_r\Omega) \\ \exp(-j2\pi2\Delta_r\Omega) \\ \vdots \\ \exp(-j2\pi(n_r - 1)\Delta_r\Omega) \end{bmatrix}, \quad (5.2)$$

as the unit spatial signature in the directional cosine Ω .

The optimal receiver simply projects the noisy received signal onto the signal direction, i.e., maximal ratio combining or receive beamforming. It adjusts for the different delays so that the received signals at the antennas can be combined constructively, yielding a n_r -fold power gain. The SIMO channel thus provides a power gain but no degree-of-freedom gain.

5.3 Line-of-Sight MISO channel

It is easy to realise that the MISO line-of-sight model with proper spacing of the antennas at the transmitter array can actually yield to same results as the SIMO model we considered above. Specifically, when the transmit antennas are separated by $\Delta_t\lambda_c$ and there is a line of sight with angle of departure of ϕ (directional cosine $\Omega \equiv \cos\phi$), the MISO channel is given by

$$y = \mathbf{h}^* \mathbf{x} + w$$

where

$$\mathbf{h} = \alpha \exp\left(\frac{-j2\pi d}{\lambda_c}\right) \begin{bmatrix} 1 \\ \exp(-j2\pi\Delta_t\Omega) \\ \exp(-j2\pi2\Delta_t\Omega) \\ \vdots \\ \exp(-j2\pi(n_t - 1)\Delta_t\Omega) \end{bmatrix}. \quad (5.3)$$

The optimal transmission (transmit beamforming) is performed along the direction $\mathbf{e}_t(\Omega)$ of \mathbf{h} , where

$$\mathbf{e}_t(\Omega) := \frac{1}{\sqrt{n_t}} \begin{bmatrix} 1 \\ \exp(-j2\pi\Delta_t\Omega) \\ \exp(-j2\pi2\Delta_t\Omega) \\ \vdots \\ \exp(-j2\pi(n_t-1)\Delta_t\Omega) \end{bmatrix}, \quad (5.4)$$

is the unit spatial signature in the transmit direction of Ω . The phase of the signal from each of the transmit antennas is adjusted so that they add constructively at the receiver, yielding a n_t -fold power gain. Again, there is no degree-of-freedom gain.

5.4 Antenna arrays with only a line-of-sight path

Now we'll look at a MIMO channel with just direct line-of-sight routes between the antennas. Uniform linear arrays are used in both the transmitter and the receiver. Δt and Δr are the normalized transmit and receive antenna separations, respectively. Between the k_{th} transmit antenna and the i_{th} receive antenna, the channel gain is

$$h_{ik} = \alpha \exp(-j2\pi d_{ik}/\lambda_c), \quad (5.5)$$

where a is the attenuation along the line-of-sight path and d_{ik} is the distance between the antennas (assumed to be the same for all antenna pairs). Assuming that the antenna array sizes are significantly lower than the distance between the transmitter and receiver:

$$d_{ik} = d + (i-1)\Delta_r\lambda_c \cos\phi_r - (k-1)\Delta_t\lambda_c \cos\phi_t, \quad (5.6)$$

where d is the distance between the first transmit and the first receive antenna, and ϕ_t, ϕ_r are the angles of incidence of the line-of-sight path on the transmit and receive antenna arrays, respectively. Defining $\Omega_t \equiv \cos\phi_t$ and $\Omega_r \equiv \cos\phi_r$ and substituting (5.6) into (5.5), we get

$$h_{ik} = \alpha \exp\left(-\frac{j2\pi d}{\lambda_c}\right) \cdot \exp(j2\pi(k-1)\Delta_t\Omega_t) \cdot \exp(-j2\pi(i-1)\Delta_r\Omega_r) \quad (5.7)$$

and we can write the channel matrix as

$$\mathbf{H} = \alpha\sqrt{n_r n_t} \exp\left(-\frac{j2\pi d}{\lambda_c}\right) \mathbf{e}_r(\Omega_r) \mathbf{e}_t(\Omega_t)^*, \quad (5.8)$$

where $\mathbf{e}_r(\cdot)$ and $\mathbf{e}_t(\cdot)$ have been defined previously. Thus, \mathbf{H} is a rank-one matrix with a unique non-zero singular value $\lambda_1 = \alpha\sqrt{n_t n_r}$.

Despite the fact that we are discussing a MIMO system, all transmitted signals are projected onto a single-dimensional space (one non-zero eigenmode), implying that there is only one spatial degree of freedom available. All transmit antenna spatial signatures (i.e. the columns of \mathbf{H}) are received in the same direction $\mathbf{e}_r(\cdot)$, preserving the spatial degree of freedom at 1.

The factor $n_t n_r$ determines the MIMO channel's power gain. If $n_t = 1$, the power gain is equal to the number of receive antennas, and the power gain is produced at the receiver using maximum ratio combining (receive beamforming). The power gain is equal to the number of transmit antennas and is produced via transmit beamforming in the case of $n_r = 1$. We profit from both transmit and receive beamforming in the general situation, since the transmitted signals are constructively added in phase at each receive antenna, and the signals at each receive antenna are further constructively combined. [10].

In summary, in an line-of-sight only environment, a MIMO channel provides a power gain but no degree-of-freedom gain.

Geographically separated antennas

5.5 Geographically separated transmit antennas

Let us now investigate how to achieve degree-of-freedom gain. We focus on the transmitting side, putting the 2 antennas far apart, with the distance being equal to the distance between transmitter-receiver.

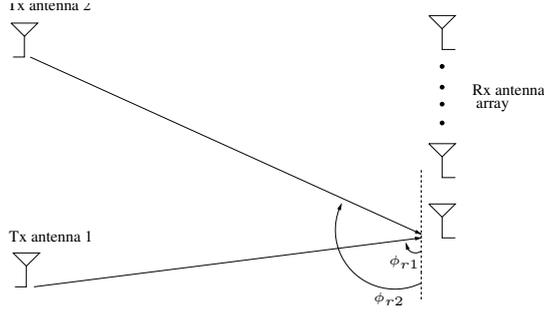


Figure 5.2: Two geographically separated transmit antennas with line-of-sight paths to the receiver antenna array.

Each path at the above figure has its own attenuation, α_1 , α_2 , and ϕ_{r1} , ϕ_{r2} are the angles of incidence, respectively. Under the assumption that the delay spread of the signals from the receiver is much smaller than $1/W$, we can write the spatial signature (channel vector) that transmit antenna k impinges on the receive array as

$$\mathbf{h}_k = \alpha_k \sqrt{n_r} \exp\left(-\frac{j2\pi d_{1k}}{\lambda_c}\right) \mathbf{e}_r(\Omega_{rk}), \quad k = 1, 2, \quad (5.9)$$

where d_{1k} is the distance between transmit antenna k and receive antenna 1, $\Omega_{rk} \equiv \cos \phi_{rk}$ and $\mathbf{e}_r(\cdot)$ is defined in (5.2). The channel matrix $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2]$ has linearly independent columns under the condition

$$\Omega_r := \Omega_{r2} - \Omega_{r1} \neq 0 \bmod \frac{1}{\Delta_r}, \quad (5.10)$$

where Ω_{ri} are each channel vector's directional cosines. In this case, we achieve two degrees of freedom, since the matrix \mathbf{H} has two non-zero singular values λ_1^2 and λ_2^2 . The transmitted signal can now be received from two different directions that can successfully be resolved at the receiver array. Following that thought, we will proceed with studying the resolvability of signals in the angular domain.

Before that, it is important to note that Ω_{r1}, Ω_{r2} , being directional cosines, have a range of $[-1, 1]$ and therefore, cannot differ by more than 2. This simplifies the condition in (5.10) to $\Omega_{r1} \neq \Omega_{r2}$, whenever the antenna spacing is $\Delta_r \leq 1/2$.

5.6 Resolvability in the angular domain

As we saw before, if $\Omega_r \neq 0 \bmod 1/\Delta_r$, then the channel matrix \mathbf{H} has full column rank. However, this is not enough to efficiently yield two or more spatial degrees of freedom. It

is important that the matrix \mathbf{H} is well conditioned, which means the angular separation of signals must be up to a specific point.

Specifically, the less aligned the spatial signatures are, the smaller the condition number of \mathbf{H} and vice versa. The angle θ between the two spatial signatures is:

$$|\cos \theta| := |\mathbf{e}_r(\Omega_{r1})^* \mathbf{e}_r(\Omega_{r2})|. \quad (5.11)$$

Since the second part of the above equation depends only on $\Omega_r := \Omega_{r2} - \Omega_{r1}$, we define:

$$f_r(\Omega_{r2} - \Omega_{r1}) := \mathbf{e}_r(\Omega_{r1})^* \mathbf{e}_r(\Omega_{r2}). \quad (5.12)$$

By direct computation,

$$f_r(\Omega_r) = \frac{1}{n_r} \exp(j\pi \Delta_r \Omega_r (n_r - 1)) \frac{\sin(\pi L_r \Omega_r)}{\sin(\pi L_r \Omega_r / n_r)}, \quad (5.13)$$

where $L_r := n_r \Delta_r$ is the normalized length of the receive antenna array. So,

$$|\cos \theta| = \left| \frac{\sin(\pi L_r \Omega_r)}{n_r \sin(\pi L_r \Omega_r / n_r)} \right|. \quad (5.14)$$

It is safe to say that the condition of the matrix \mathbf{H} depends directly on the above parameter (5.14). Considering $a_1 = a_2 = a$, for simplicity, the squared singular values of \mathbf{H} are:

$$\lambda_1^2 = \alpha^2 n_r (1 + |\cos \theta|), \quad \lambda_2^2 = \alpha^2 n_r (1 - |\cos \theta|) \quad (5.15)$$

and the condition number is

$$\frac{\lambda_1}{\lambda_2} = \sqrt{\frac{1 + |\cos \theta|}{1 - |\cos \theta|}}. \quad (5.16)$$

The matrix is ill-conditioned whenever $|\cos \theta| \approx 1$ and well conditioned otherwise. In Figure 5.4, we plot this quantity $|\cos \theta| = |f_r(\Omega_r)|$ as a function of Ω_r for a fixed array size and different values of n_r .

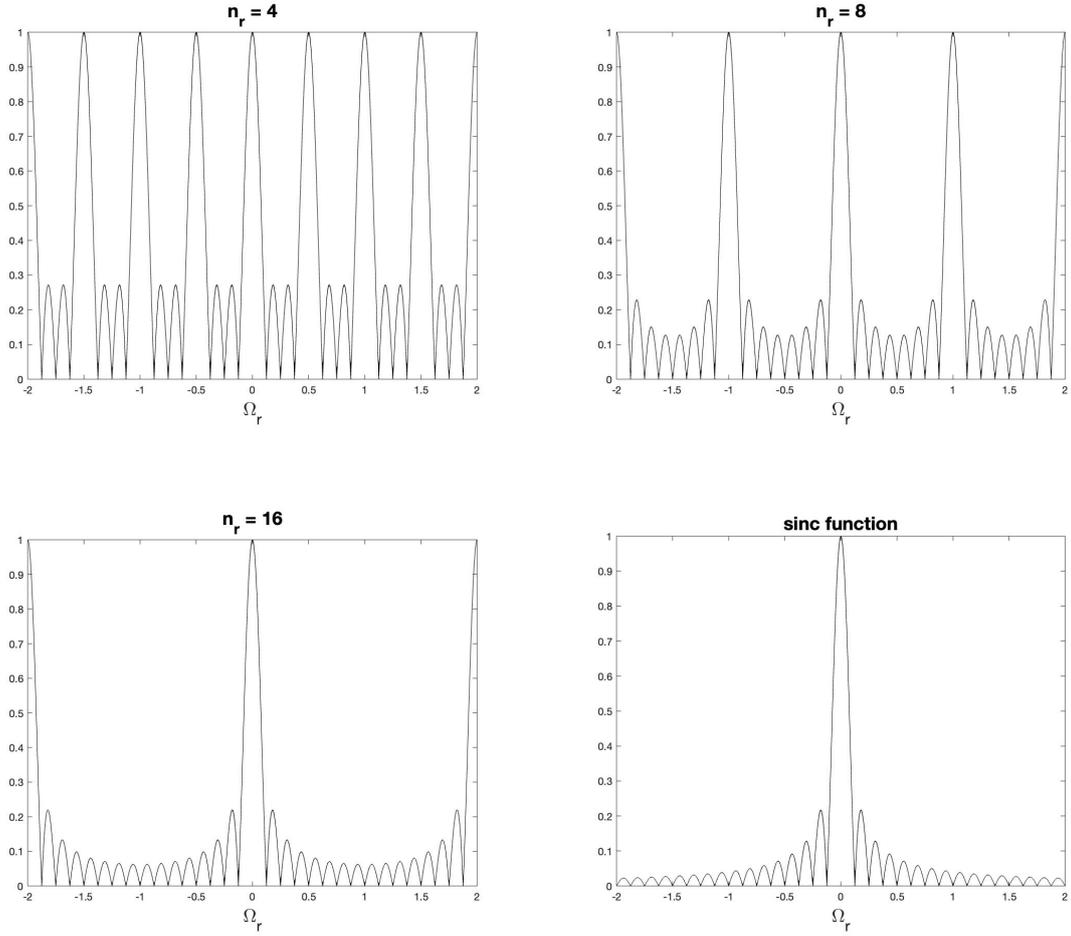


Figure 5.4: The function $|f(\Omega_r)|$ plotted as a function of Ω_r for fixed $L_r = 8$ and different number of receiving antennas n_r .

As we can see, the function $f_r(\cdot)$ has the following properties:

- $f_r(\Omega_r)$ is periodic with period $n_r/L_r = 1/\Delta_r$.
- $f_r(\Omega_r)$ peaks at $\Omega_r = 0$. $f(0) = 1$.
- $f_r(\Omega_r) = 0$ at $\Omega_r = k/L_r$, for $k = 1, \dots, n_r - 1$.

$f_r(\cdot)$ has a periodicity that is derived from the periodicity of $\mathbf{e}_r(\cdot)$. It has a $2/L_r$ main lobe that is focused on integer multiples of $1/\Delta_r$. The peaks of the other lobes are substantially

lower. If the signatures are near to being aligned and the channel matrix is badly conditioned, this indicates the signatures are close to being aligned and the channel matrix is ill conditioned whenever

$$\left| \Omega_r - \frac{m}{\Delta_r} \right| \ll \frac{1}{L_r} \quad (5.17)$$

for some $m \in \mathcal{Z}$. Now, if the antenna separation is $\Delta_r \leq \frac{1}{2}$ and since Ω_r ranges in $[-2, 2]$, (5.17) leads to

$$|\Omega_r| \ll \frac{1}{L_r}. \quad (5.18)$$

One may think, that increasing the number of antennas for a fixed antenna length L_r will increase the above qualitative results. However, that is not the case. In fact, when $n_r \rightarrow \infty$ and $\Delta_r \rightarrow 0$,

$$f_r(\Omega_r) \rightarrow e^{j\pi L_r \Omega_r} (L_r \Omega_r), \quad (5.19)$$

and the dependency of $f_r(\cdot)$ on n_r does not exist.

If $\Omega_r \ll \frac{1}{L_r}$, the signals from two different transmit antennas cannot be successfully resolved at the receive antenna array, consequently acquiring only one spatial degree of freedom. Packing more and more antennas in a given amount of space does not increase the angular resolvability of the receive antenna array. So, according to what has been argued above, one may say that the parameter $\frac{1}{L_r}$ is a *measure of resolvability* in the angular domain.

5.7 Pictorial Representation of Angular Domain Resolvability

Let us try to see how an antenna array's angular resolvability is represented. We use the receive beamforming pattern to do this. That is, if the signal comes from a single direction ϕ_0 , the best receiver will project the signal onto the vector $\mathbf{e}_r(\cos \phi_0)$, which is known as the (receiving) beamforming vector. A signal coming from any other direction is attenuated by a factor of ϕ .

$$|\mathbf{e}_r(\cos \phi_0)^* \mathbf{e}_r(\cos \phi)| = |f_r(\cos \phi - \cos \phi_0)|. \quad (5.20)$$

The beamforming pattern associated with the vector $\mathbf{e}_r(\cos \phi)$ is the polar plot

$$(\phi, |f_r(\cos \phi - \cos \phi_0)|) \quad (5.21)$$

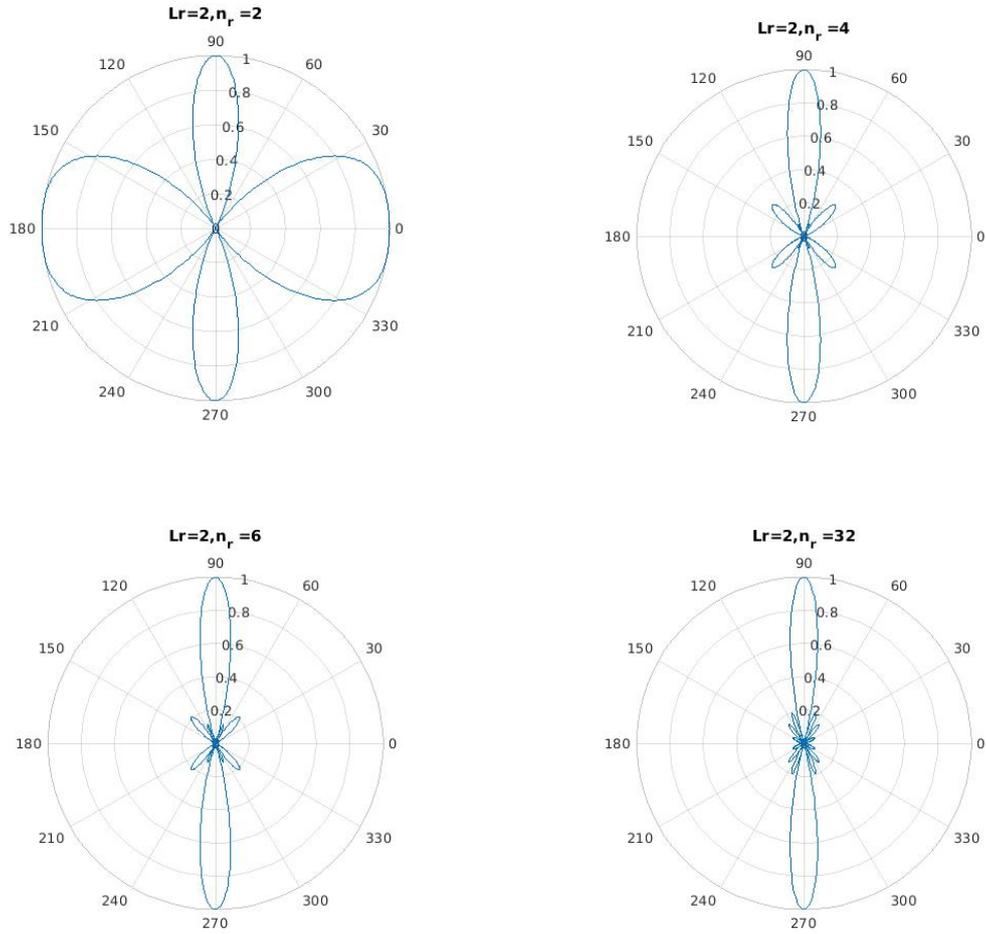


Figure 5.6: Receive beamforming patterns aimed at 90° , with antenna array length $L_r = 2$ and different numbers of receive antennas n_r . Note that the beamforming pattern is always symmetrical about the $0^\circ - 180^\circ$ axis, so lobes always appear in pairs. For $n_r = 4, 6, 32$, the antenna separation $\Delta_r \leq 1/2$, and there is a single main lobe around 90° (together with its mirror image). For $n_r = 2$, $\Delta_r = 1 > 1/2$ and there is an additional pair of main lobes.

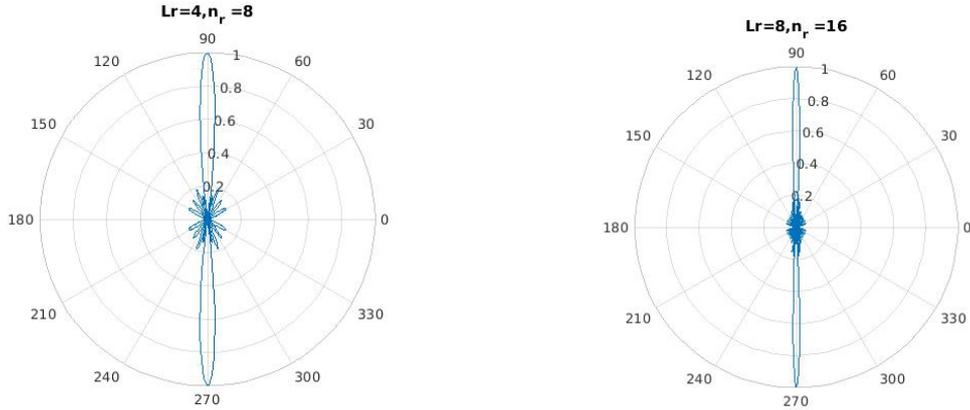


Figure 5.7: Beamforming patterns for different antenna array lengths. (Left) $L_r = 4$ and (Right) $L_r=8$. Antenna separation is fixed at half the carrier wavelength. The larger the length of the array, the narrower the beam.

Two important notes yielded from the above figures about the beamforming pattern are as follow:

- It has main lobes around ϕ_0 and any angle ϕ for which

$$\cos \phi = \cos \phi_0 \pmod{\frac{1}{\Delta_r}}, \quad (5.22)$$

which follows the periodicity of $f_r(\cdot)$. If the antenna separation is $\Delta_r < 1/2$, there is only one main lobe at ϕ and if $\Delta_r > 1/2$, there can be several main lobes. At both these situations, there also exist a mirror image at $-\phi$ (Figure 5.6).

- The main lobe has a beam width (directional cosine width) of $2/L_r$. The larger the array length L_r , the narrower the beam and the higher the angular resolution (Figure 5.7).

In a wireless channel, the antenna array size L_r and bandwidth W are clearly analogous. The parameter $1/W$ evaluates the time domain resolvability of signals: multipaths arriving at time separations smaller than $1/W$ cannot be resolved by the receiver. The parameter $1/L_r$ assesses the resolvability of angular domain signals: signals arriving at an angle less than $1/L_r$ cannot be resolved by the receiver. Adding additional antenna components cannot raise angular-domain resolvability beyond $1/L_r$, just as oversampling cannot increase time-domain resolvability beyond $1/W$. This comparison will be used in

the statistical modeling of MIMO fading channels and will be explored in further detail in the next section.

5.8 Geographically separated receive antennas

Similarly to the case of the transmit separated antennas, we can achieve spatial diversity by separating the receive antennas and having an antenna array at the transmitter. See Figure 5.8.

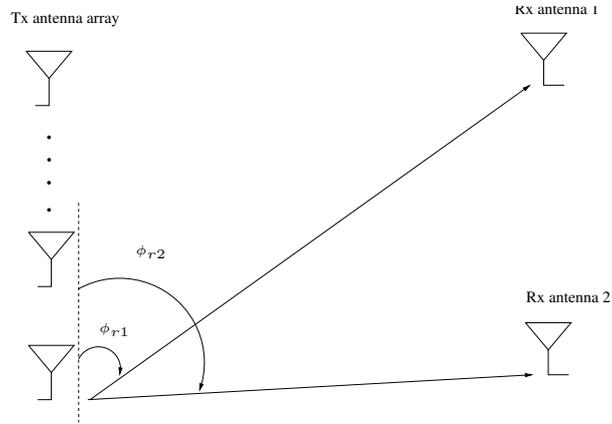


Figure 5.8: Two geographically separated receive antennas each with line of sight from a transmit antenna array.

In this case, the channel matrix is

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^* \\ \mathbf{h}_2^* \end{bmatrix}, \quad (5.23)$$

where

$$\mathbf{h}_i = \alpha_i \exp\left(-\frac{j2\pi d_{i1}}{\lambda_c}\right) \mathbf{e}_r(\Omega_{ti}), \quad (5.24)$$

and, just like before, Ω_{ti} is the directional cosine of the signal from transmit antenna array to the receive antenna i and d_{i1} is the distance between the first transmit antenna and the i th receive. In order for \mathbf{H} to have two linearly independent rows, the angle separation of the transmit directional cosines must be

$$\Omega_t := \Omega_{t2} - \Omega_{t1} \neq 0 \bmod \frac{1}{\Delta_t}. \quad (5.25)$$

As for the channel's matrix condition, only when the angular separation of the two receive antennas Ω_t is of the order or larger than $1/L_t$, where $L_t := n_t \Delta_t$, can we successfully utilize the two degrees of freedom since the matrix is well conditioned.

Some transmitters, in order to measure the amount of energy dissipated in directions other than the one transmitting, use the *transmit beamforming* pattern. Analogous to the receive beamforming technique, the beam has width of $2/L_t$, and the longer the antenna array, the sharper the transmitter can focus the energy along the desired direction and the better it can spatially multiplex information to multiple receive antennas.

5.9 Line-of-sight plus one reflected path

Let us now study whether we can achieve the same result as before (spatial diversity) working with antenna arrays instead of putting two single antennas at the transmitter or the receiver, far apart from each other. We consider the scenario where we have a line-of-sight and a reflected path from transmitter to receiver array just like Figure 5.9.

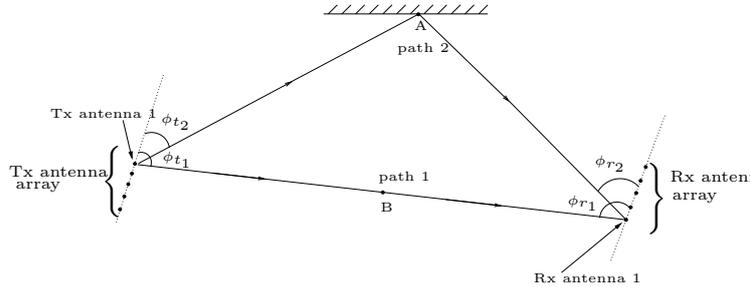


Figure 5.9: A MIMO channel with a direct path and a reflected path.

Name the direct path, path 1, and the reflected one, path 2. Path i has an attenuation of a_i , an angle of ϕ_{ti} ($\Omega_{ti} := \cos \phi_{ti}$) and ϕ_{ri} ($\Omega_{ri} := \cos \phi_{ri}$) with the transmit and the receive antenna array respectively. Using the principle of superposition, we can write the channel matrix \mathbf{H} as:

$$\mathbf{H} = a_1^b \mathbf{e}_r(\Omega_{r1}) \mathbf{e}_t(\Omega_{t1})^* + a_2^b \mathbf{e}_r(\Omega_{r2}) \mathbf{e}_t(\Omega_{t2})^* \quad (5.26)$$

where, for $i = 1, 2$,

$$a_i^b := a_i \sqrt{n_t n_r} \exp\left(-\frac{j2\pi d^{(i)}}{\lambda_c}\right), \quad (5.27)$$

and $d^{(i)}$ is the distance between the first transmit and first receive antenna along the path i . In order for \mathbf{H} to have rank 2, it must hold true that

$$\Omega_{t1} \neq \Omega_{t2} \bmod \frac{1}{\Delta_t}, \quad (5.28)$$

and

$$\Omega_{r1} \neq \Omega_{r2} \bmod \frac{1}{\Delta_r}. \quad (5.29)$$

As for the the condition number of matrix \mathbf{H} , we need the angle separation between the signals, not only at the transmitter but also at the receiver, to fulfill the criteria mentioned in previous sections. Specifically, angular separation $|\Omega_t|$, of the transmitter needs to be of the same order or larger than $1/L_t$, with the same going for the receiver about the amounts $|\Omega_r|$ and $1/L_r$, where

$$\Omega_t = \cos \phi_{t2} - \cos \phi_{t1}, \quad L_t := n_t \Delta_t \quad (5.30)$$

and

$$\Omega_r = \cos \phi_{r2} - \cos \phi_{r1}, \quad L_r := n_r \Delta_r. \quad (5.31)$$

Let us now consider an approach which will make clear how the multipath effect can help significantly in the rank and condition number of channel matrix \mathbf{H} . Suppose we rewrite \mathbf{H} as $\mathbf{H}''\mathbf{H}'$, where

$$\mathbf{H}'' = \begin{bmatrix} a_1^b \mathbf{e}_r(\Omega_{r1}), & a_2^b \mathbf{e}_r(\Omega_{r2}) \end{bmatrix}, \quad \mathbf{H}' = \begin{bmatrix} \mathbf{e}_t^*(\Omega_{t1}) \\ \mathbf{e}_t^*(\Omega_{t2}) \end{bmatrix}. \quad (5.32)$$

\mathbf{H}' is a 2 by n_t matrix and \mathbf{H}'' is a n_r by 2. Suppose we have a scenario where there are two imaginary receivers at point A and B just like in Figure 5.10.

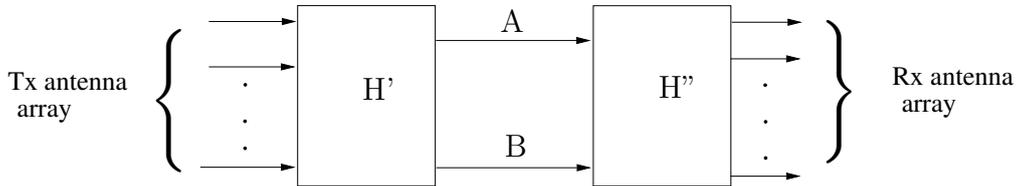


Figure 5.10: Channel is viewed as a concatenation of two channels \mathbf{H}' and \mathbf{H}'' with intermediate (virtual) relays A and B.

Point A is the point where the first signal is reflected on the wall and point B is located along the line-of-sight path. It is safe to assume that the two matrices above have rank 2,

since points A and B are geographically far apart. As for their condition number, matrix \mathbf{H}' depends on $L_t\Omega_t$ and matrix \mathbf{H}'' on $L_r\Omega_r$. If both matrices are well-conditioned, then the overall channel matrix \mathbf{H} is well-conditioned as well.

As someone could easily derive, the MIMO channel with two multipaths is actually a concatenation of the n_t by 2 channel in Figure 5.9 and the 2 by n_r channel in Figure 5.10. Despite the position of the antenna arrays, a multipath environment provides virtual “relays” which are geographically far apart, which leads to multiple degrees of spatial diversity, making spatial multiplexing possible. In this context, multipath fading provides an advantage that can be exploited.

Last but not least, it is crucial to point out that, in this example, we had a significant angular separation both at the receiver and the transmitter, which led to the well-conditioning of matrix \mathbf{H} . This is not the case though, at every environment. In a case where the reflector is close to the receiver, and the transmitter-receiver distance is considerably large, the angular separation of the transmitter Ω_t is small. Respectively, in the case where the reflector is close to the transmitter and far from the receiver, Ω_r is small. Figure 6.1 helps us visualize these scenarios.

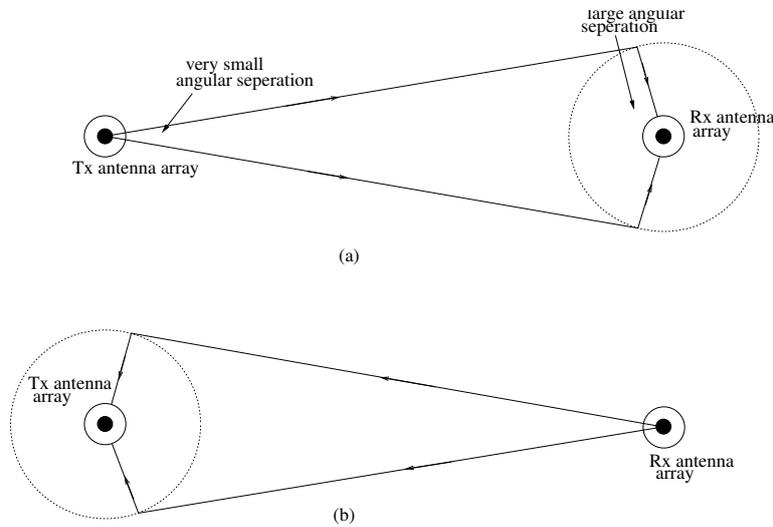


Figure 5.11: (a) The reflectors and scatterers are in a ring around the receiver; the angular separation of the transmitter is small. (b) The reflectors and scatterers are in a ring around the transmitter; the receiver angular separation is small.

At both the above cases, the channel matrix \mathbf{H} is not very well-conditioned. What these all mean, in real life applications is that if the base station is located high, on top of

a tower with hardly any scatterers around it, and, in contrast, a mobile user has plenty of reflectors at his area, then in order for the BS to exploit the spatial multiplexing effect, its size has to be many wavelengths long.

Chapter 6

Modeling of MIMO Fading Channels

In the previous section, we dealt with deterministic channels. Based on the insights we obtained, we can move towards statistical MIMO models and get a better understanding of spatial multiplexing.

6.1 Basic Approach

It was made clear that, sometimes, despite the rank of \mathbf{H} being 2, spatial multiplexing was not possible if the matrix is ill-conditioned. To deal with this, we looked into the angular separation between two spatial signatures Ω_i (with $i = r, t$ for the receiver and the transmitter, respectively) and compared it to the angular resolvability of the transmitter and receiver array, $1/L_t$ and $1/L_r$. If Ω_t was greater or equal to $1/L_t$, we had a successful angular resolvability at the transmitter array, with the same applying for the receiver array if $\Omega_r \geq 1/L_r$.

At this point, we come up with an insightful idea. Individual paths between the antenna arrays can be multiple, but, yet, not all of them are resolvable. Therefore, in order to design and analyse a communication system, it might be more productive to approach statistical modeling of physical MIMO channels based only on spatially resolvable paths rather than individual paths.

This method is not, in fact, novel. The statistical modeling of frequency selective fading channels is based on the gains of the discrete-time sampled channel taps rather than the gains of the individual physical routes. In this case, each tap may be viewed as a (time-)resolvable path made up of a collection of physical pathways. These physical paths can

be neatly or coarsely divided into resolvable pathways based on the system's bandwidth. Communication is influenced by the behavior of these (resolvable) pathways.

Our objective from the start was to transfer this reasoning from the time-resolution of a bandlimited system to the finite angular-resolution of an array-size-limited system. The transmit and receive antenna array lengths, L_t and L_r , determine the degree of angular resolvability: pathways with transmit directional cosines that vary by less than $1/L_t$ and receive directional cosines that differ by less than $1/L_r$ are not resolvable by the arrays. This implies that we need "sample" the angular domain at fixed angular spacings of $1/L_t$ at the transmitter and $1/L_r$ at the receiver, and describe the channel using these new input and output coordinates. In these angular coordinates, the $(k, l)_{th}$ channel gain is basically the sum of all pathways whose transmit directional cosine is within an angular window of width $1/L_t$ around l/L_t and whose receive directional cosine is within an angular window of width $1/L_r$ around k/L_r . We shall now develop our strategies specifically for uniform linear arrays (ULA).

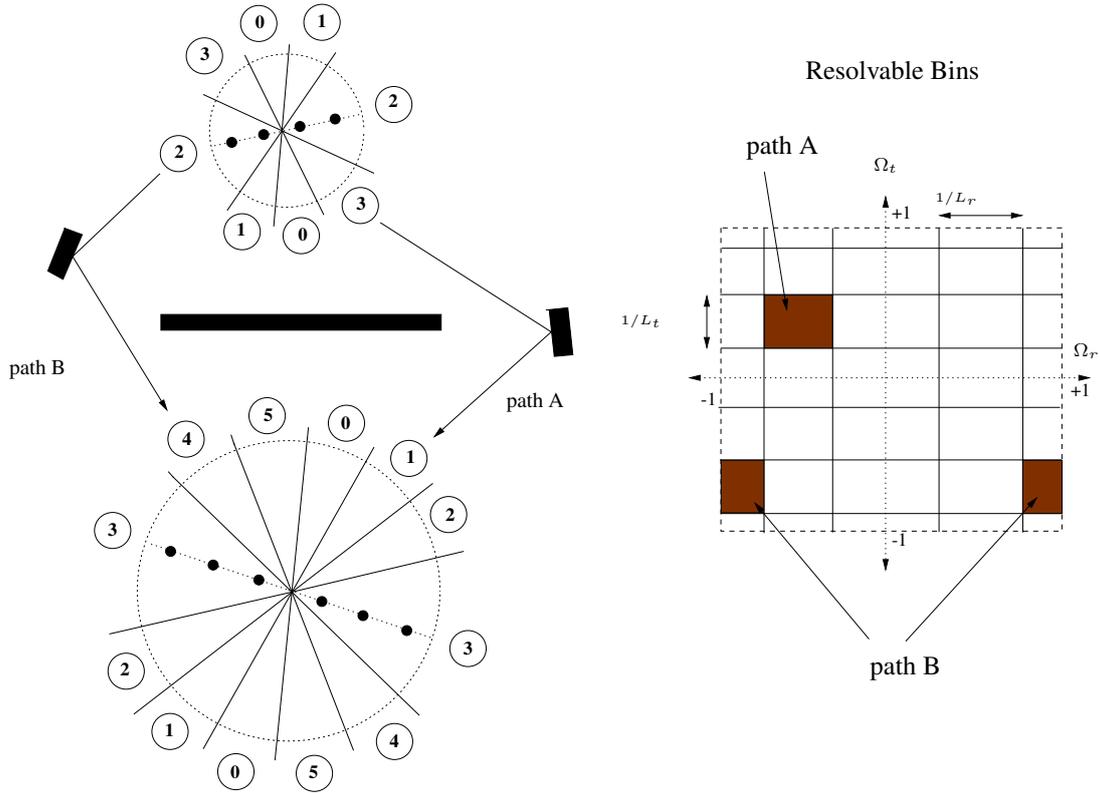


Figure 6.1: A representation of the MIMO channel in the angular domain. Due to the limited resolvability of the antenna arrays, the physical paths are partitioned into resolvable bins of angular widths $1/L_r$ by $1/L_t$. Here are the 4 receive antennas ($L_r=2$) and 6 transmit antennas ($L_t=3$).

6.2 MIMO Multipath Channel

Let us consider the narrowband MIMO channel:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}. \quad (6.1)$$

At the two points of communication we place uniform linear arrays (ULA) with normalized lengths L_t, L_r and n_t, n_r number of antennas, respectively, for the transmitter and the receiver. We use for normalization the wavelength of the passband transmitted signal λ_c ; Δ_t and Δ_r denote the normalized spacings of the ULA's antennas with $\Delta_t = L_t/n_t$ and $\Delta_r = L_r/n_r$ regarding the transmitter and the receiver. For simplicity, let the channel \mathbf{H} be fixed. Time-variation is easy to be dealt later on.

We consider the a scenario where we have an arbitrary number of physical paths between the transmitter and the receiver. As we mentioned before, a_i is the attenuation of the i_{th} path, which makes an angle ϕ_{ti} ($\Omega_{ti} := \cos \phi_{ti}$) and angle ϕ_{ri} ($\Omega_{ri} := \cos \phi_{ri}$) with the transmitter and the receiver array respectively. Then, \mathbf{H} is given by:

$$\mathbf{H} = \sum_i a_i^b \mathbf{e}_r(\Omega_{ri}) \mathbf{e}_t(\Omega_{ti})^*, \quad (6.2)$$

where, as it was mentioned earlier,

$$a_i^b := a_i \sqrt{n_t n_r} \exp\left(-\frac{j2\pi d^{(i)}}{\lambda_c}\right), \quad (6.3)$$

$$\mathbf{e}_r(\Omega) := \frac{1}{\sqrt{n_r}} \begin{bmatrix} 1 \\ \exp(-j2\pi\Delta_r\Omega) \\ \exp(-j2\pi2\Delta_r\Omega) \\ \vdots \\ \exp(-j2\pi(n_r - 1)\Delta_r\Omega) \end{bmatrix} \quad (6.4)$$

$$\mathbf{e}_t(\Omega) := \frac{1}{\sqrt{n_t}} \begin{bmatrix} 1 \\ \exp(-j2\pi\Delta_t\Omega) \\ \exp(-j2\pi2\Delta_t\Omega) \\ \vdots \\ \exp(-j2\pi(n_t - 1)\Delta_t\Omega) \end{bmatrix} \quad (6.5)$$

Vectors $\mathbf{e}_r(\Omega)$ and $\mathbf{e}_t(\Omega)$ denote the transmitted and received unit spatial signatures along the direction Ω , and $d^{(i)}$ is the distance between first transmit and first receive antenna of path i .

6.3 Angular Domain Representation of Signals

When a signal arrives onto the receiver array with a directional cosine Ω it is along the unit spatial signature $\mathbf{e}_r(\Omega)$, given by (6.4). It is important to remember that

$$f_r(\Omega) := \mathbf{e}_r(0)^* \mathbf{e}_r(\Omega) = \frac{1}{n_r} \exp(j\pi\Delta_r\Omega(n_r - 1)) \frac{\sin(\pi L_r \Omega)}{\sin(\pi L_r \Omega / n_r)}. \quad (6.6)$$

Specifically (see Figure 5.4),

$$f_r\left(\frac{k}{L_r}\right) = 0 \text{ and } f_r\left(\frac{-k}{L_r}\right) = f_r\left(\frac{n_r - k}{L_r}\right), k = 1, 2, \dots, n_r - 1. \quad (6.7)$$

Consequently, the n_r original vectors

$$\mathcal{S}_r := \left\{ \mathbf{e}_r(0), \mathbf{e}_r\left(\frac{1}{L_r}\right), \dots, \mathbf{e}_r\left(\frac{n_r-1}{L_r}\right) \right\} \quad (6.8)$$

form an orthogonal basis for the received signal space \mathcal{C}^{n_r} . This basis help us represent the received signals in the angular domain. Similarly, from the transmitter perspective, the n_t original vectors are:

$$\mathcal{S}_t := \left\{ \mathbf{e}_t(0), \mathbf{e}_t\left(\frac{1}{L_t}\right), \dots, \mathbf{e}_t\left(\frac{n_t-1}{L_t}\right) \right\}. \quad (6.9)$$

Let us recall Figs 5.6 and 5.7, where we exploited the beamforming pattern of $\mathbf{e}_r(\Omega)$. From what we saw, it is conceivable that a received signal along any physical direction has almost all of its energy along one $\mathbf{e}_r(k/L_r)$ vector/angle/direction. Since different $\mathbf{e}_r(k/L_r)$ s have different main lobes, it is fairly reasonable to argue that the orthonormal basis formed by the vectors in \mathcal{S}_t and \mathcal{S}_r provide a simple (but approximate) decomposition of the received signal into multipaths received along the different physical directions, with a maximum resolution of $1/L_t$ or $1/L_r$ respectively.

6.4 Angular Domain Representation of MIMO Channels

Let \mathbf{U}_t be the unitary matrix $n_t \times n_t$, the columns of which are the basis vectors in \mathcal{S}_t , with the same standing for \mathbf{U}_r and \mathbf{S}_r . Using \mathbf{U}_t and \mathbf{U}_t^* , the following transformation is defined

$$\mathbf{x} = \mathbf{U}_t \mathbf{x}^a, \quad \mathbf{x}^a = \mathbf{U}_t^* \mathbf{x}, \quad (6.10)$$

with \mathbf{x} being the original n_t -dimensional signal transmitted from the antenna array and \mathbf{x}^a its angular domain representation. In this case, the $(k, l)^{th}$ element of \mathbf{U}_t is

$$\frac{1}{\sqrt{n_t}} \exp\left(-\frac{j2\pi kl}{n_t}\right) \quad k, l = 0, \dots, n_t - 1. \quad (6.11)$$

Hence, \mathbf{x}^a is nothing more than the Inverse Discrete Fourier Transform of \mathbf{x} . It is important, however, to note that it is due to the uniform linear arrays that the specific transformation is only a DFT.

Similar to vector-signal transformation, we can transform the MIMO channel to the angular domain. \mathbf{U}_t and \mathbf{U}_r are the $n_t \times n_t$ and $n_r \times n_r$ unitary matrices which will help

us achieve the following representation:

$$\mathbf{y}^{\mathbf{a}} = \mathbf{U}_r^* \mathbf{H} \mathbf{U}_t \mathbf{x}^{\mathbf{a}} + \mathbf{U}_r^* \mathbf{w} = \mathbf{H}^{\mathbf{a}} \mathbf{x}^{\mathbf{a}} + \mathbf{w}^{\mathbf{a}}, \quad (6.12)$$

where

$$\mathbf{H}^{\mathbf{a}} := \mathbf{U}_r^* \mathbf{H} \mathbf{U}_t \quad (6.13)$$

is the angular domain transformation of the channel matrix and

$$\mathbf{w}^{\mathbf{a}} := \mathbf{U}_r^* \mathbf{w} \sim CN(0, N_0 \mathbf{I}_{n_r}). \quad (6.14)$$

Chapter 7

Joint Channel Estimation

Let us now enter the channel estimation sector. So far, the most common technique for estimating the channel is the Least-Squares method, which works quite well for relatively simple channel models and propagation scenarios. However, this is not the case here. As we discussed in previous chapters, massive MIMO systems consist of multiple antennas both at the BS and the receiver, making thus, the communication model quite complex. Furthermore, this kind of system is mostly used for wireless communication in an urban environment. This adds several challenges to the signal propagation process, but also offers opportunities to exploit. The channel matrices become quite large and phenomena like multipath propagation make the channel estimation complexity grow significantly. For this reason, researchers had to find new, more efficient ways of estimating the channel.

7.1 Joint Orthogonal Matching Pursuit

Following this concept, we focus on a very specific scenario, upon which this thesis is based. This new proposition for CSIT estimation and feedback takes advantage of the common local scatterers surrounding the users, thus, achieving a substantial reduction of CSIT estimation and feedback overhead, since it is able to exploit the *hidden joint channel sparsity* among different users.

First, we need to introduce a joint sparsity model able to comply with the channel matrix features in a multi-user massive MIMO FDD system. Based on this model, a communication procedure is established so that users can obtain the compressed channel observations and transmit them right back to the BS. At this point, CSIT is reconstructed

using a joint recovery algorithm which is based on these compressed measurements [23].

7.2 The system

Let us think of a flat block-fading, multi-user massive MIMO system that operates in FDD mode. This system consists of one BS with M antennas and K users. Each user carries a mobile equipment with N antennas. The transmission starts when the BS uses one downlink pilot channel to broadcast a sequence of T training symbols with its M antennas as Figure 7.1 illustrates. At the j th time slot, the transmitted signal from the BS is $\mathbf{x}_j \in \mathbb{C}^{M \times 1}$, $j = 1, \dots, T$. The user i , receives a vector signal at the same time slot $\mathbf{y}_{ij} \in \mathbb{C}^{N \times 1}$ that can be expressed as

$$\mathbf{y}_{ij} = \mathbf{H}_i \mathbf{x}_j + \mathbf{n}_{ij}, \quad j = 1, \dots, T. \quad (7.1)$$

The quantity $\mathbf{H}_i \in \mathbb{C}^{N \times M}$ is the channel matrix of user i ; $\mathbf{n}_{ij} \in \mathbb{C}^{N \times 1}$ is the complex Gaussian noise with zero mean and unit variance.

Of course, we can address the same system from a matrix point of view. Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T] \in \mathbb{C}^{M \times T}$ denote the concatenated pilot symbols and $\mathbf{Y}_i = [\mathbf{y}_{i1}, \dots, \mathbf{y}_{iT}] \in \mathbb{C}^{N \times T}$ the received signal. Last, let $\mathbf{N}_i = [\mathbf{n}_{i1}, \dots, \mathbf{n}_{iT}] \in \mathbb{C}^{N \times T}$ are the concatenated noise vectors. Then, (7.1) can be written as

$$\mathbf{Y}_i = \mathbf{H}_i \mathbf{X} + \mathbf{N}_i, \quad (7.2)$$

where $\text{tr}(\mathbf{X}\mathbf{X}^H) = PT$ is the sum transmit SNR in the T training time slots and P is the transmit SNR per time slot at the BS.

At this point, it is important to mention that, in order to fully exploit the array gains and spatial degrees of freedom of a massive MIMO system, it is crucial for the CSIT to be available at the BS. For this to happen, in an FDD system, a two-step procedure needs to be followed. First, each user needs to perform the CSI estimation of \mathbf{H}_i locally, and then, feed the estimate $\hat{\mathbf{H}}_i$ back to the BS. Each user uses a Least-Squares based estimation technique ([5], [25]), getting thus the LS channel estimate

$$\hat{\mathbf{H}}_i = \mathbf{Y}_i \mathbf{X}^\dagger. \quad (7.3)$$

The quantity $\mathbf{X}^\dagger = \mathbf{X}^H (\mathbf{X}\mathbf{X}^H)^{-1}$ is the Moore-Penrose pseudoinverse and \mathbf{Y}_i stands for the noisy pilot symbols received at the user side. Now, it is time for the estimated CSI

to be fed back to BS. However, this LS technique requires $T \geq M$, which translates to a prohibited large training overhead for the estimation and feedback of CSIT, especially when we talk about massive MIMO multi-user systems where M is large. Therefore, it is easy to comprehend that the LS approach is not suitable for this kind of systems and there is need for a more efficient approach to the CSIT estimation challenge, one that can exploit the hidden channel sparsity of the channel matrices in the network.

7.3 Joint Sparsity Channel Model

At this point, we are going to deal with signal propagation from a spatial point of reference and use the angular domain representation we presented in a previous section.

To begin with, we assume that we will always deal with uniform linear arrays (ULA) at the base station and the users. This way, the virtual angular representation of the channel matrix of users is given by

$$\mathbf{H}_i = \mathbf{A}_R \mathbf{H}_i^w \mathbf{A}_T^H, \quad (7.4)$$

with $\mathbf{A}_R \in \mathbb{C}^{N \times N}$ and $\mathbf{A}_T \in \mathbb{C}^{M \times M}$ being the unitary matrices for the angular domain transformation for the i_{th} user and the BS side, respectively. $\mathbf{H}_i^w \in \mathbb{C}^{N \times M}$ is the angular domain channel matrix. When its (k, l) entry is non-zero, this means that there is a spatial path from the l_{th} transmit direction to the k_{th} receive direction for user i . It is widely accepted that, in a multi-user massive MIMO system, angular domain channel matrices are sparse due to limited local scattering at the BS side.

Let us now take a closer look at the channel's matrix composition. Let the j_{th} row of \mathbf{H}_i^w be \mathbf{h}_{ij} . A much used concept is the one of the vector-signal's support. The support $\text{-supp}(\mathbf{h})$ - denotes the positions of the non-zero entries of a vector or a matrix, i.e. $\text{supp}(\mathbf{h}) = \{i : \mathbf{h}(i) \neq 0\}$. After a extended research on massive MIMO channel matrices form, several observations have been made regarding their support.

Specifically, one of these observations (Observation I), addresses the sparsity support within an individual channel matrix. From [16], [14] one can see that the channels in a massive MIMO system are usually correlated at the BS side but not at the user's side. Using the spatial point of view, we can justify this phenomenon thinking of the rich scattering surrounding a user, that makes the sparsity support take various forms. In contrast, a BS is usually placed somewhere high with little to no scattering objects around. Hence,

there are only a few active transmit directions reaching it. This is the reason why the row vectors in \mathbf{H}_i^w tend to have the same sparsity support

$$\text{supp}(\mathbf{h}_{i1}) = \text{supp}(\mathbf{h}_{i2}) = \cdots = \text{supp}(\mathbf{h}_{iN}).$$

Another important observation (Observation II) was made regarding the *interchannel correlation* or, in other words, a partially shared support between different channel matrices. It is widely accepted [14] that users who are physically close to each other, i.e. two cellphone users at a city center, usually have correlated channel matrices. This happens because these users share some common local scatterers and consequently their channel matrices $\{\mathbf{H}_i^w\}$ at the BS side usually share a partially common support. In mathematical terms, there exists a non-zero index set Ω_c of the common support, such that $\Omega_c \subseteq \Omega_i, \forall i$.

Of course, in the case where users have a single antenna ($N = 1$), the channel matrix becomes a vector \mathbf{h}_i^w , so there is no point discussing about joint sparsity.

Channel Model

Based on what was mentioned above, we assume that the channel matrices $\{\mathbf{H}_i^w\}$ we are going to deal with have certain properties.

First, there is the *Individual sparsity due to local scattering at the BS*. Let $\mathbf{h}_{i,j}$ be the j th row of \mathbf{H}_i^w ; then, every row-vector of this matrix has the same sparsity structure. In other words, there exists an index set $\Omega_i, 0 < |\Omega_i| \ll M, \forall i$, such that

$$\text{supp}(\mathbf{h}_{i1}) = \text{supp}(\mathbf{h}_{i2}) = \dots = \text{supp}(\mathbf{h}_{iN}) \triangleq \Omega_i. \quad (7.5)$$

Another property is the *Distributed joint sparsity due to common scattering at the BS*. This means that different channel matrices \mathbf{H}_i^w share a common sparsity support Ω_c . This index set Ω_c can be expressed as

$$\bigcap_{i=1}^K \Omega_i = \Omega_c. \quad (7.6)$$

It is important to mention that the entries of $(\mathbf{H}_i^w)_{\Omega_i}$ follow an i.i.d. complex Gaussian distribution with zero mean and unit variance. The notation $(\mathbf{H}_i^w)_{\Omega_i}$ is frequently used and refers to the sub matrix formed by collecting the column vectors of \mathbf{H}_i^w whose indices belong to Ω_i . Figure 7.1 helps us understand the sparsity concept just described.

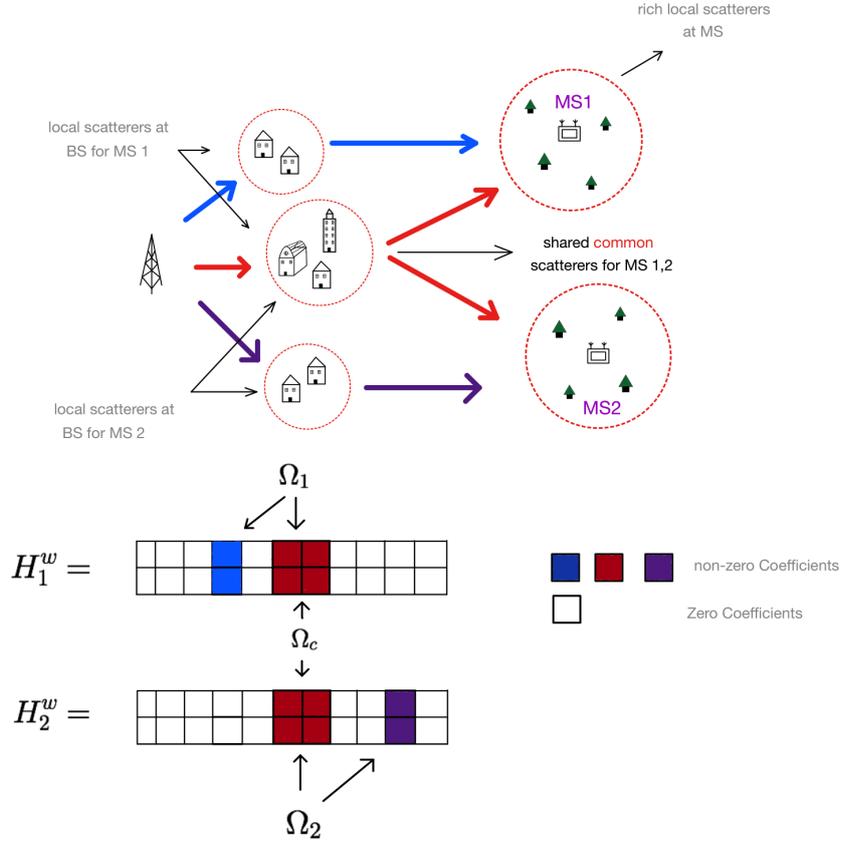


Figure 7.1: Due to the restricted and shared local scattering impact on the BS side, the joint channel sparsity structure is shown. The support for the *common* scatterers shared by all users is Ω_c , but the support for the individual scatterers for the i_{th} user is Ω_i .

Massive MIMO channel sparsity support can be parameterized by $P = \{\Omega_c, \{\Omega_i : \forall i\}\}$, where, as mentioned before, Ω_c and Ω_i denote the *common* and *individual* sparsity support, respectively. Some special cases occur when $\Omega_c = \Omega_i, \forall i$. In this case, one can understand that all the users share the same local scatterers at the BS side. Furthermore, if $\Omega_c = \emptyset$, then there are no common scatterers shared among the users. At this point, we can make an assumption and argue that there is a statistical bound on the channel sparsity levels ($|\Omega_c|, |\Omega_i|$). Specifically, let Λ be an event such that

$$\Lambda : |\Omega_c| \geq s_c, |\Omega_i| \leq s_i, \forall i. \quad (7.7)$$

The likelihood of this event is $\Pr(\Lambda) > 1 - \epsilon$ for a small ϵ ; $\mathbb{S} = \{s_c, \{s_i : \forall i\}\}$ refers to

the statistical sparsity bound. This way we ensure that, at each user's channel matrix, there will be at least s_c and at most s_i common and individual sparsity indices, respectively, but both of them will be much less than the number of transmit antennas of the BS ($s_c, s_i \ll M$). However, neither the BS nor the MS have prior knowledge of these sparsity bounds. In practice, information about \mathbb{S} can be obtained when the features of the propagation environment are already known to the system. For example, in [2], offline channel propagation measurements at the BS are used to obtain the desired sparsity bound knowledge, and [7] utilizes a long term stochastic learning and estimation technique for the same purpose. As one can see, \mathbb{S} depends on the large scale properties of the scattering environment and changes slowly. However, in order to proceed with the development of the CSIT estimation procedure, we consider that the statistical sparsity bound \mathbb{S} is available to the BS.

7.4 Distributed Compressive CSIT Estimation and Feedback

It now becomes clear that, in order to avoid the prohibitively large pilot training and feedback overhead in a massive MIMO system, this joint sparsity structure described above must be used. For this to be feasible, the CSIT estimation needs to be done at the BS side. This leaves the users with a simple purpose, to distributively observe the compressed measurements $\{\mathbf{Y}_i\}$ and feed them back to the BS, where $\{\mathbf{H}_i\}$ is recovered jointly. Let us now take a look at an algorithm trying to estimate the CSIT utilizing the above technique, and a figure depicting its functionality.

Algorithm for Distributed Compressive CSIT Estimation and Feedback

- **Step 1** (Pilot Training): The BS sends the compressive training symbols $\mathbf{X} \in \mathbb{C}^{M \times T}$, with $T \ll M$.
- **Step 2** (Compressive Measurements and Feedback): The i_{th} mobile user observes the compressed measurements \mathbf{Y}_i from the pilot symbols given in (7.2) and feeds back to the BS side.
- **Step 3** (Joint CSIT Recovery at BS): The BS recovers the CSIT $\{\mathbf{H}_1^e, \dots, \mathbf{H}_K^e\}$ jointly based on the compressed feedback $\{\mathbf{Y}_1, \dots, \mathbf{Y}_K\}$.

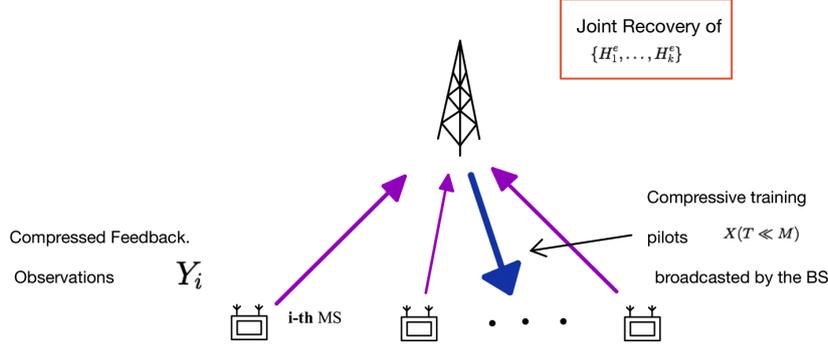


Figure 7.2: To all K users, the BS broadcasts a compressive training pilot $\mathbf{X} \in \mathbb{C}^{M \times T}$ (recall that $T \ll M$). The compressive measurement \mathbf{Y}_i is obtained locally by each user and fed back to the BS. The BS jointly recovers the CSIT $\{\mathbf{H}_1^e, \dots, \mathbf{H}_K^e\}$ based on the obtained compressive measurements $\{\mathbf{Y}_1, \dots, \mathbf{Y}_K\}$.

The letter T refers to the pilot training and feedback overhead in the above algorithm. Although our attention is focused elsewhere, and specifically at step 3. That is because this is where the hidden joint sparsity model can be exploited and drastically reduce the training overhead T needed to estimate the CSIT in a massive MIMO multi-user system. This problem can be formulated as follows:

Joint CSIT Recovery at BS:

$$\min_{\{\mathbf{H}_i, \forall i\}} \sum_{i=1}^K \|\mathbf{Y}_i - \mathbf{H}_i \mathbf{X}\|_F^2 \quad (7.8)$$

s.t. $\{\mathbf{H}_i^w : \forall i\}$ satisfy the joint sparsity model as described in (7.5) and (7.6).

This problem, under the specific constraint, turns out to be extremely challenging. The design and implementation of an algorithm able to tackle the above problem is of utmost importance.

7.5 Joint CSIT Recovery Algorithm Design

In order to deliver a low complexity greedy algorithm capable of solving the problem in (7.8), the channel equation in (7.2) needs to be modified to meet the CS model standards.

Denote the following variables:

$$\bar{\mathbf{Y}}_i = \sqrt{\frac{M}{PT}} \mathbf{Y}_i^H \mathbf{A}_R \in \mathbb{C}^{T \times N}, \bar{\mathbf{X}} = \sqrt{\frac{M}{PT}} \mathbf{X}^H \mathbf{A}_T \in \mathbb{C}^{T \times M}, \quad (7.9)$$

$$\bar{\mathbf{H}}_i = (\mathbf{H}_i^w)^H \in \mathbb{C}^{M \times N}, \quad \bar{\mathbf{N}}_i = \sqrt{\frac{M}{PT}} \mathbf{N}_i^H \mathbf{A}_R \in \mathbb{C}^{T \times N}. \quad (7.10)$$

Substituting the variables into (7.2):

$$\bar{\mathbf{Y}}_i = \bar{\mathbf{X}} \bar{\mathbf{H}}_i + \bar{\mathbf{N}}_i \quad \forall i, \quad (7.11)$$

where $\bar{\mathbf{X}}$ is the measurement matrix with $\text{tr}(\bar{\mathbf{X}}^H \bar{\mathbf{X}}) = M$ and $\bar{\mathbf{H}}_i$ is the sparse channel matrix. Also:

$$\|\bar{\mathbf{Y}}_i - \bar{\mathbf{X}} \bar{\mathbf{H}}_i\|_F^2 = \frac{M}{PT} \|\mathbf{Y}_i - \mathbf{H}_i \mathbf{X}\|_F^2, \quad \forall i, \quad (7.12)$$

thus, solving the problem in (7.8), turns out to be equal to finding $\{\bar{\mathbf{H}}_i : \forall i\}$ that minimizes $(\sum_{i=1}^K \|\bar{\mathbf{Y}}_i - \bar{\mathbf{X}} \bar{\mathbf{H}}_i\|_F^2)$ while satisfying the sparsity constraint mentioned in (7.8). We develop an algorithm to solve problem (7.8) based on this analogous relationship and equation (7.11).

Algorithm 2 (Joint-OMP to solve Problem 7.8)

Input: $\{\mathbf{Y}_i : \forall i\}$, \mathbf{X} , $\mathbb{S} = \{s_c, \{s_i : \forall i\}\}$, η_1, η_2 ($\eta_1 < 1, \eta_2 > 1$).

Output: Estimated $\{\mathbf{H}_i^e\}$ for $\{\mathbf{H}_i : \forall i\}$.

- **Step 1** (Initialization): Compute $\bar{\mathbf{Y}}_i \forall i$, $\bar{\mathbf{X}}$ from $\{\mathbf{Y}_i : \forall i\}$ and \mathbf{X} , as in (7.9).
- **Step 2** (*Common Support Identification*): Initialize $\mathbf{R}_i = \bar{\mathbf{Y}}_i \forall i$, $\Omega_c^e = \emptyset$ and then repeat the following procedures s_c times.

–**A** (*Support Estimation*): Estimate the remaining index by $\Omega'_i = \arg \max_{|\Omega|=s_i-|\Omega_c^e|} \|(\bar{\mathbf{X}}_\Omega)^H \mathbf{R}_i\|_F, \forall i$.

–**B** (Support Pruning): Prune support Ω'_i to be $\Omega'_i = \{j : j \in \Omega'_i, \|(\bar{\mathbf{X}}(j))^H \mathbf{R}_i\|_F^2 \geq \eta_1 N\}, \forall i$.

–**C** (Support Update): Update the estimated common support as $\Omega_c^e = \Omega_c^e \cup \left\{ \arg \max_j \sum_{i=1}^K I_{\{j \in \Omega'_i\}} \right\}$.

–**D** (Residual Update): $\mathbf{R}_i = (\mathbf{I} - \mathbf{P}_{\Omega_c^e}) \bar{\mathbf{Y}}_i$, where $\mathbf{P}_{\Omega_c^e}$ is a projection matrix given by

$$\mathbf{P}_{\Omega_c^e} = (\bar{\mathbf{X}}_{\Omega_c^e}) (\bar{\mathbf{X}}_{\Omega_c^e})^\dagger. \quad (7.13)$$

- **Step 3** (Individual Support Identification): Set $\Omega_i^e = \Omega_c^e, \forall i$ and estimate the individual support Ω_i^e for each user i individually. Specifically, for the i_{th} user, stop if $\|\mathbf{R}_i\|_F^2 \leq \frac{\eta_2 NM}{P}$ or the following procedures have been repeated $(s_i - s_c)$ times.

–**A** (Support Update): Update the estimated individual support as

$$\Omega_i^e = \Omega_i^e \cup \left\{ \arg \max_j \|(\bar{\mathbf{X}}(j))^H \mathbf{R}_i\|_F \right\}.$$

–**B** (Residual Update): $\mathbf{R}_i = (\mathbf{I} - \mathbf{P}_{\Omega_i^e}) \bar{\mathbf{Y}}_i$.

- **Step 4** (Channel Estimation by *LS*): The estimated channel for user i is $\mathbf{H}_i^e = \mathbf{A}_R (\bar{\mathbf{H}}_i^e)^H \mathbf{A}_T^H$ where $\bar{\mathbf{H}}_i^e$ is given by $(\bar{\mathbf{H}}_i^e)^{\Omega_i^e} = (\bar{\mathbf{X}}_{\Omega_i^e})^\dagger \bar{\mathbf{Y}}_i, (\bar{\mathbf{H}}_i^e)^{[M] \setminus \Omega_i^e} = 0, \forall i$.

The above mentioned η_1, η_2 ($\eta_1 < 1, \eta_2 > 1$) are threshold parameters. The main goal is to identify the common (Ω_c) and individual (Ω_i) sparsity support, which is what step 2

and step 3, respectively, aim to do. As soon as this is done, step 4 utilizes the estimated Ω_i^e and recovers the channel matrices through LS. Figure 7.3 helps us visualize the above procedure.

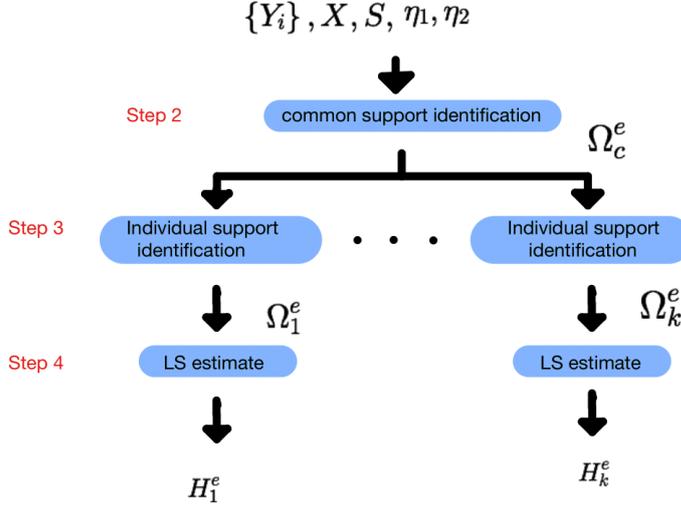


Figure 7.3: Algorithm's 2 main processing flow.

As someone can understand, identifying the common and individual sparsity support is the key factor for this technique. Towards this aim, certain strategies can be adopted; specifically:

- **Strategy to exploit the sparsity support within an individual channel matrix**

From what was previously mentioned, $\bar{\mathbf{H}}_i$, the estimation target in (7.11), can be simultaneously zero and non-zero on a single row of size N . Just like similar research [30], a row vector of $\bar{\mathbf{H}}_i$ is being identified as an *atomic unit* based on the aggregate matching effects between the residual \mathbf{R}_i and the measurement matrix $\bar{\mathbf{X}}$. For instance, the support index is selected based on the sum of the N matched terms corresponding to the N columns of the residual matrix \mathbf{R}_i ; for example, $\sqrt{\sum_{l=1}^N \|\bar{\mathbf{X}}(j)^H \mathbf{R}_i(l)\|^2} = \|\bar{\mathbf{X}}(j)^H \mathbf{R}_i\|_F$, as in Step 2.A and Step 3.A.

- **Strategy to exploit the partially shared support between different Channel Matrices.**

The different channel matrices $\{\bar{\mathbf{H}}_i\}$ for each user share a partial common support,

just like it is visualized in Figure 7.1. This means that the index set Ω_c is most likely to be estimated as part of the sparsity support by most users. For example, if a certain index is identified by most, or all of the K users, then it is very likely to be part of the common support set Ω_c . Conversely, in case an index belongs to the common sparsity support set Ω_c , it is going to be estimated as a sparsity support index by most of the users in the system. According to this thought, Step 2.B of the algorithm has been designed with the purpose of selecting the index that appears the most times in the estimated index set $\Omega'_i \forall i$, as the next entry of the Ω_c .

Assume that $s_i = s, \forall i$, for simplicity. The overall complexity of the Algorithm is $O(KsMNT)$, which is the same order as recovering each $\bar{\mathbf{H}}_i$ individually using the conventional 2-norm SOMP ([30],[8]). Furthermore, compared with some conventional CS recovery algorithms (e.g. [29]) or sparse channel estimation [2] in which knowledge of instantaneous sparsity level is needed, the proposed J-OMP requires only the *statistical channel sparsity information* \mathbb{S} , which can be estimated using slow-timescale stochastic learning [7].

7.6 Design of Pilot training Matrix

Restricted Isometry Property

The *restricted isometry property* (RIP) is commonly used in classical CS work. Its role is to evaluate the adequacy of the measurement matrix utilized in a performance analysis of a CS recovery algorithm.

Definition: A Matrix $\bar{\mathbf{X}} \in \mathbb{C}^{T \times M}$ satisfies the RIP of order k with the restricted isometry constant (RIC) δ_k , if $0 \leq \delta_k \leq 1$ and δ_k is the smallest number such that

$$(1 - \delta_k) \|\mathbf{h}\|^2 \leq \|\bar{\mathbf{X}}\mathbf{h}\|^2 \leq (1 + \delta_k) \|\mathbf{h}\|^2$$

holds for all $\mathbf{h} \in \mathbb{C}^{M \times 1}$ where $\|\mathbf{h}\|_0 \leq k$.

Let us jump onto the design of the training matrix itself. As it was mentioned above, the RIP is widely used in CS literature to characterize the quality of a measurement matrix. The recovery can be efficient and robust once the measurement matrix satisfies a proper RIC δ_s requirement. Research has shown that a matrix with entries drawn from a sub-

Gaussian distribution satisfies the RIP with overwhelming probability [8], thus it is a very popular matrix generation distribution among CS works.

According to this, and based on the signal model in (7.2), the pilot matrix $\mathbf{X} \in \mathbb{C}^{M \times T}$ can be created as $\mathbf{X} = \mathbf{A}_T \mathbf{X}_\alpha$, where $\mathbf{X}_\alpha \in \mathbb{C}^{M \times T}$ is i.i.d. drawn from $\left\{-\sqrt{\frac{P}{M}}, \sqrt{\frac{P}{M}}\right\}$, with equal probability.

Experimental Results

The authors X. Rao and VKN. Lau of [23], compared their algorithm's performance to other similar CS recovery algorithms at their experimental figures. Although, that is not the case for this thesis, since the goal was to study, understand and reproduce their work to the best of our abilities.

Without further ado, in our case, we consider a narrow band (flat fading) multi-user massive MIMO FDD system consisting of one BS and K users. The equipment used at the BS side utilizes M antennas whereas each user has N antennas on their equipment. The average transmit SNR at the BS is P and the statistical information on the channel sparsity level is denoted as $\mathbb{S} = \{s_c, \{s_i = s \ \forall i\}\}$. We assume a rich local scattering environment; the number of individual and common spatial paths from the BS broadside (corresponding to the channel sparsity levels $|\Omega_i|, |\Omega_c|$) are randomly generated as $|\Omega_i| \sim \mathcal{U}(s-2, s)$ and $|\Omega_c| \sim \mathcal{U}(s_c, s_c+2)$ respectively (with $|\Omega_c| \geq s_c$ and $|\Omega_i| \leq s$), where $\mathcal{U}(\alpha, b)$ stands for the discrete uniform distribution on the integers $\{\alpha, \alpha+1, \dots, b\}$. The spatial paths from the BS to the UE have equal path loss and the angles of departure are randomly and uniformly distributed over $[0, 2\pi]$. The threshold parameters η_1, η_2 in our algorithm are set to $\eta_1 = 0.2, \eta_2 = 2$.

At this point we need something to help us graph the performance of our algorithm, and thus we use the Normalized Mean Squared Error (NMSE) calculated as follows:

$$\mathbb{E} \left(\frac{\|\mathbf{H}_i - \mathbf{H}_i^e\|_F^2}{\|\mathbf{H}_i\|_F^2} \right) \quad (7.14)$$

where $\|\mathbf{H}_i\|_F^2$ is the Frobenius norm of the original channel matrix \mathbf{H}_i , squared. Accordingly, the amount $\|\mathbf{H}_i^e\|_F^2$ stands for the Frobenius norm of the estimated channel matrix \mathbf{H}_i^e squared. The purpose of the paper upon this thesis is based, was to compare the performance of the suggested algorithm to similar CS works of the time and present the better results that were achieved. This is why, at the original experimental figures, the perfor-

mance of the J-OMP algorithm is sketched among other algorithms such as *Conventional LS*, *OMP*, *2-norm SOMP*, *M-BP*, *SD-OMP* and *Genie-aided LS*.

However, the goal of this thesis is to comprehend, at the best possible level, the concepts this paper addresses, become familiar with the relatively new approach on the signal propagation-spatial point of reference/angular domain- comprehend how massive MIMO systems work and reproduce the J-OMP algorithm to the best of our abilities. This is why at the following figures, there will be 2 plots side by side. The right one will be the paper's results, where the J-OMP performance is among other algorithms' performances, whereas at the left figure will be the result of our reproduction of the original algorithm so we can review the similarity between them.

1. CSIT Estimation Quality Versus Overhead T

The normalized mean squared error (NMSE) of the estimated CSI is compared to the training and feedback overhead T in Figure 7.4, when the number of BS antennas is $M = 160$, number of MS antennas is $N = 2$, number of users is $K = 40$, common sparsity parameter $s_c = 9$, individual sparsity parameter $s = 17$, and transmit SNR $P = 28$

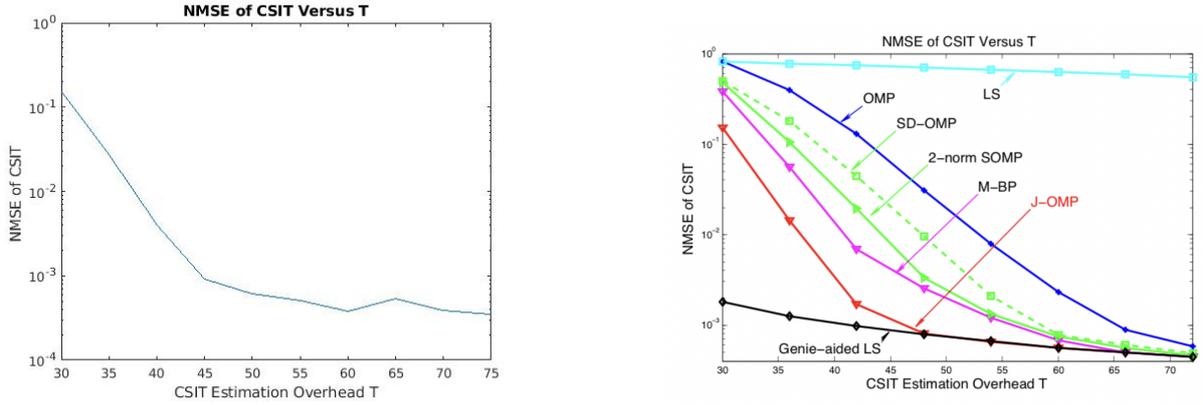


Figure 7.4: NMSE of CSIT versus the training overhead T with parameters $M=160$, $N=2$, $K=40$, $s_c=9$, $s=17$ and transmit SNR $P=28$ dB.

According to this graph, the CSIT estimate quality improves as T grows, and the J-OMP method outperforms the other algorithms significantly. This is because the proposed J-OMP takes use of the hidden joint sparsity among the user channel matrices to improve CSIT recovery. Specifically, the performance gains of the 2-norm SOMP ([11], [30]) and M-BP ([12], [28]) schemes over OMP demonstrate the benefits of exploiting the individual joint sparsity among the user channel matrices (Observation I), and the performance gain of J-OMP over the 2-norm SOMP and M-BP demonstrate the benefits of exploiting the distributed joint sparsity among the user channel matrices (Observation II). Furthermore, when T increases, the suggested J-OMP, 2-norm SOMP, M-BP, SD-OMP, and OMP all approach the genie-added LS method. This is due to the fact that the channel support recovery probability of these schemes all go to 1 as T grows. This finding also emphasizes the significance of having a better likelihood of support recovery in the CSIT reconstruction.

As for our work, the similarity between the original and the reproduced algorithm's performance is obvious. Corroborating, thus, the superiority of J-OMP over the rest, CS algorithms.

2. CSIT Estimation Quality Versus SNR P

In Figure 7.5, we compare the NMSE of the estimated CSIT versus SNR P , under $T = 45$, $M = 160$, $N = 2$, $K = 40$, $s_c = 9$ and $s = 17$.

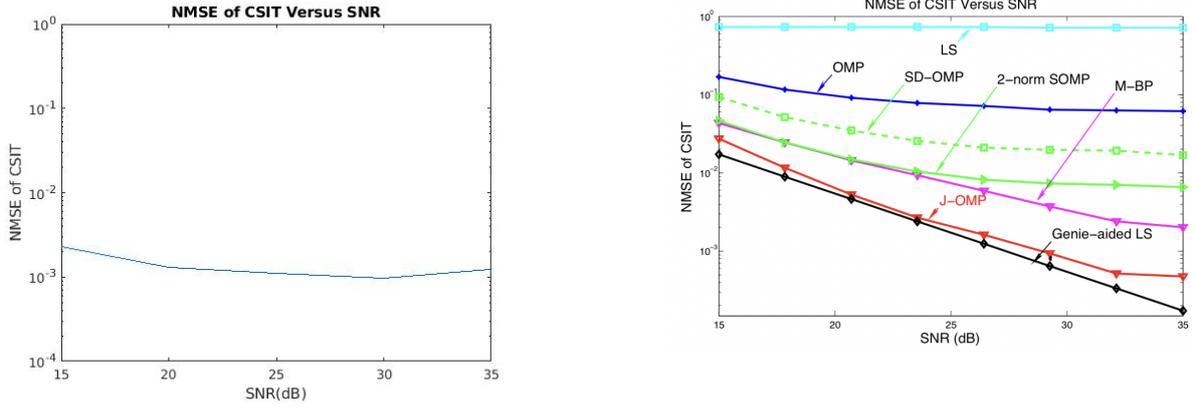


Figure 7.5: NMSE of CSIT versus the SNR P with parameters $T=45$, $M=160$, $N=2$, $K=40$, $s_c=9$, and $s=17$.

We observe that the proposed J-OMP algorithm has a substantial performance gain over the baselines and relatively larger performance gain is achieved in higher SNR regions.

In this occasion, our work slightly differs from the original algorithm, since the performance is much better from the beginning, with relatively low SNR values. As the SNR increases, the original J-OMP presents significant improvement of the CSIT recovery quality, whereas, our improvement is more smooth. At the end, both implementations achieve the same recovery quality, around 10^{-3} .

3. CSIT Estimation Quality Versus Common Sparsity Support s_c

In Figure 7.6, we compare the NMSE of the estimated CSIT versus the common sparsity support s_c , under $T = 45$, $M = 160$, $N = 2$, $K = 40$, $s = 17$, and $P = 28$ dB.

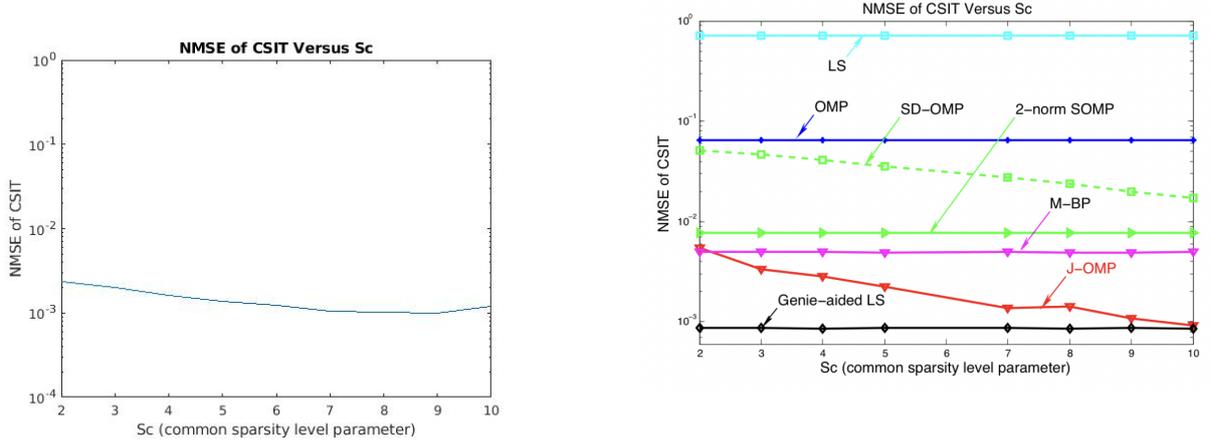


Figure 7.6: NMSE of CSIT versus the common sparsity s_c under $T=45$, $M=160$, $N=2$, $K=40$, $s=17$ and transmit SNR $P=28$ dB.

Over here, one can observe that the CSIT estimation quality of the proposed J-OMP algorithm increases as the s_c parameter becomes larger. This is because this novel algorithm is able to exploit the common sparsity shared between the K users, and this common support is more likely to be estimated by this new algorithm. Thus, as the common support increases (larger s_c), better CSIT estimation is achieved.

As for our implementation, when the users share an increasing common support, the quality of the estimation gets better, similarly to the original implementation.

4. CSIT Estimation Quality Versus Individual Sparsity Support s

In Figure 7.7, we compare the NMSE of the estimated CSIT versus the individual sparsity support s , under $T = 45$, $M = 160$, $N = 2$, $K = 40$, $s_c = 9$, and $P = 28$ dB.

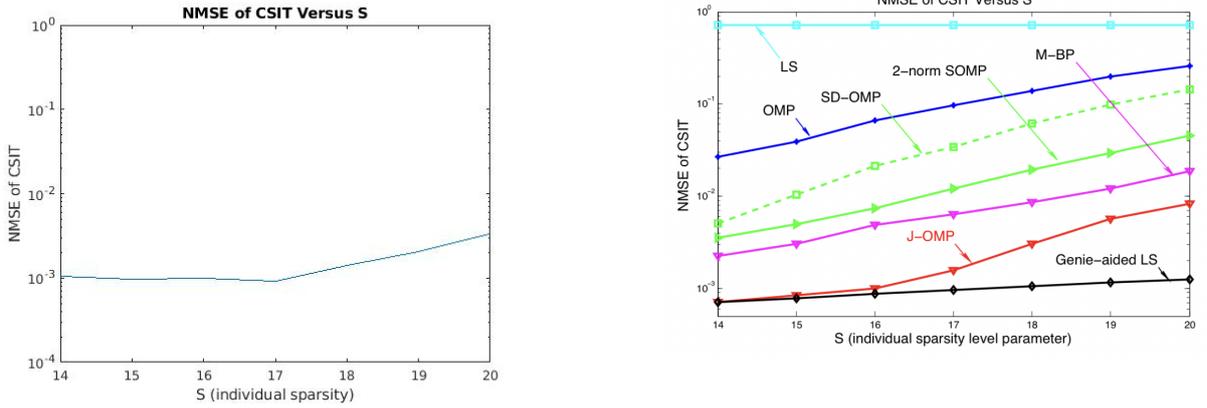


Figure 7.7: NMSE of CSIT versus the individual sparsity s under $T=45$, $M=160$, $N=2$, $K=40$, $s_c=9$ and transmit SNR $P=28$ dB.

We observe that the increase in the individual sparsity parameter s has a negative impact on the quality of the CSIT recovery. This is because as the sparsity level gets bigger, larger number of measurements are needed for a successful channel recovery, according to the classical CS theory [4]. Hence, given a specific value of overhead T , the CSIT estimation quality drops as the channel sparsity increases; a fact that is confirmed by the results of our implementation in the left figure.

5. CSIT Estimation Quality Versus MS Antennas N

In Figure 7.8, we compare the NMSE of the estimated CSIT versus MS antennas N , under $T = 45$, $M = 160$, $K = 40$, $s_c = 9$, $s = 17$ and $P = 28$ dB.

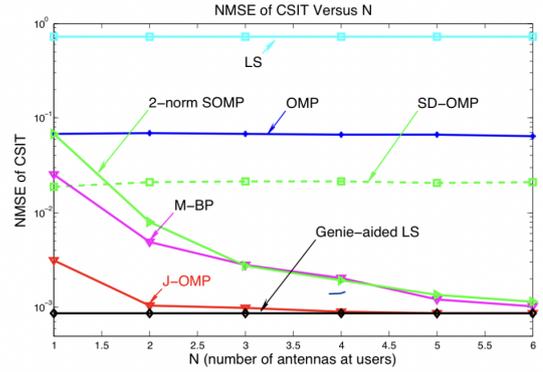
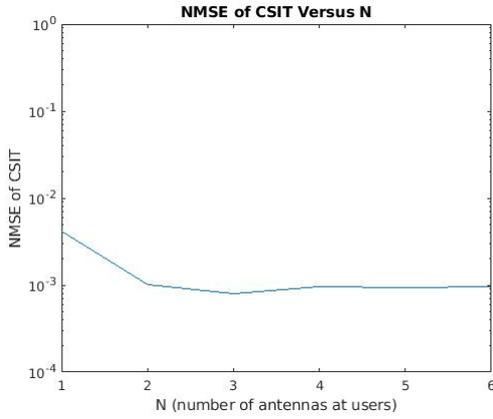


Figure 7.8: NMSE of CSIT versus the user's antennas N under $T=45$, $M=160$, $K=40$, $s=17$, $s_c=9$ and transmit SNR $P=28$ dB.

In this case, adding more antennas at the UE practically increases the rows of the channel matrix. This way, the proposed J-OMP algorithm is able to exploit the individual joint sparsity among the N row vectors of each channel matrix. Larger N means better CSIT estimation quality, exactly what our implementation results show.

6. CSIT Estimation Quality Versus BS Antennas M

In Figure (7.9), the NMSE of the estimated CSIT is sketched versus the number of BS antennas M , under $T = 45$, $N=2$, $K=40$, $s=17$, $s_c=9$ and $P=28$ dB.

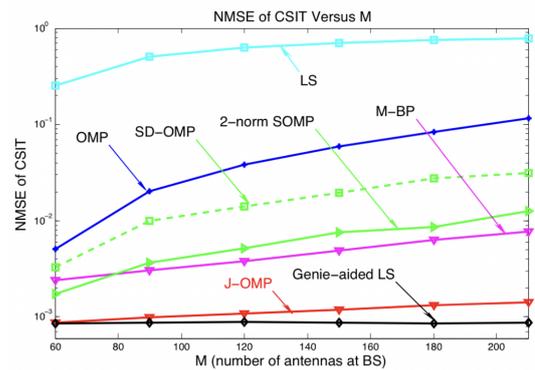
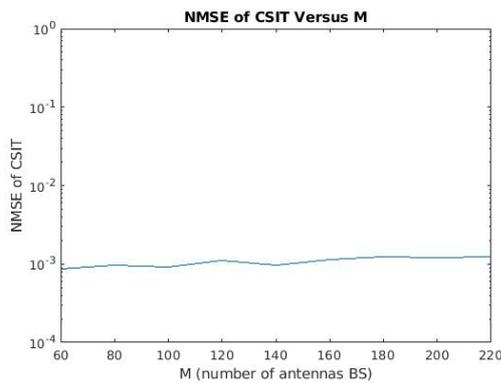


Figure 7.9: NMSE of CSIT versus the MS antennas M under $T=45$, $N=2$, $K=40$, $s=17$, $s_c=9$ and transmit SNR $P=28$ dB.

In this case, the proposed J-OMP algorithm achieves better performance than the rest. Although, the estimation quality of the CSIT decreases as M increases. This happens because, larger M leads to larger channel matrix. Consequently, more CS measurements are needed to estimate the CSIT according to classical CS theory [4]. In other words, giving a specific value to the CSIT measurements overhead T and increasing the number of BS antennas M has a significant impact on the CSIT estimation quality.

7. CSIT Estimation Quality Versus Number of MS users K

In Figure 7.10 , the NMSE of the estimated CSIT is sketched versus the number of users K , under $T = 45$, $N=2$, $K=40$, $s=17$, $s_c=9$ and $P=28$ dB.

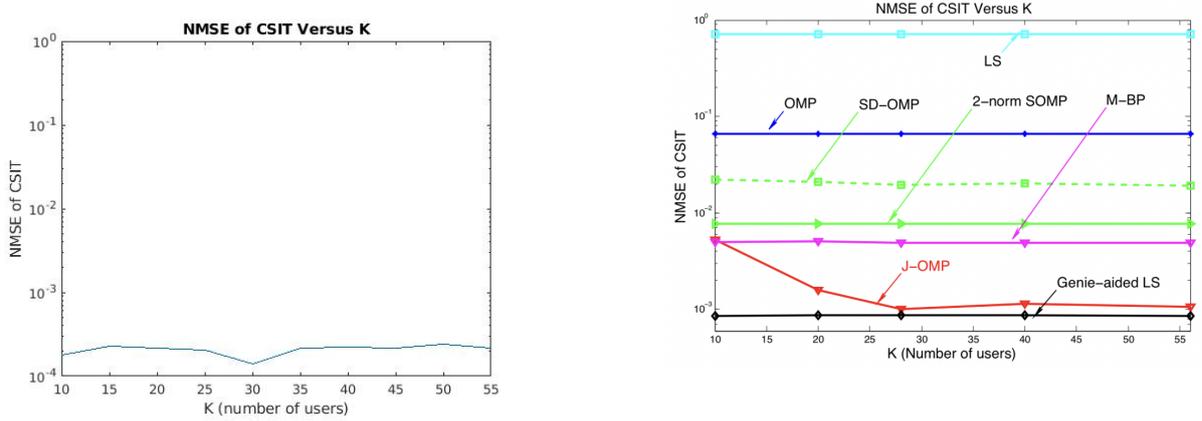


Figure 7.10: NMSE of CSIT versus the number of users K under $T=45$, $M=160$, $N=2$, $s=17$, $s_c=9$ and transmit SNR $P=28$ dB.

We observe that the novel J-OMP algorithm shows a significant increase in CSIT estimation quality as the number of users grow. Thus J-OMP exploits the distributed joint sparsity among the users' channel matrices, in order to jointly identify the common sparsity support. Hence, larger number of users (K) leads to better recovery of the common support and, thus, a better CSIT estimation quality.

Chapter 8

Conclusion

Massive multiple-input-multiple-output (MIMO) is thought by many to be the one technology describing the 5th generation of communication networks. That is only natural since it significantly improves both spectral and energy efficiency. However, this new technology does not come without challenges. The large number of antennas used, mostly on the BS, makes the complexity of computations prohibitively large. This phenomenon strongly impacts the communication since it is devastating for the radio resources and also for the real-time response of applications.

At this point, we realized that the usual ways of approaching and studying a communication system (time and frequency domain) cannot provide a viable solution to our problem. Maybe looking into signal propagation from a spatial point of view help us. This is when the *Angular Domain* representation of the signal comes to untie our hands. Studying the direction from which the signal arrived but also how the scatterers affect the communication by their place in the environment and their number, gave us new insights into how to deal with the problem of fast and reliable communication. In order to transition from time/frequency to angular domain, we use a transformation method which, under specific conditions (specific spacing of antennas on an array), turns out to be a simple IFT (Inverse Fourier transform).

Having switched to angular domain, we observe that the channel matrix is sparse due to limited scattering on the BS side. Thus, calculations with very big dimension channel matrices have a little meaning. However, *Compressive Sensing*, a relatively new concept, tackles this exact problem by greedily picking the non-zeros entries of the channel matrix and using them to recover the CSI correctly. Following this track, X.Rao and V.K.N. Lau

designed a novel algorithm for CSIT estimation in massive MIMO FDD systems. After carefully studying the propagation environment and the composition of channel matrices, they observed a pattern on the sparsity structures of the matrices. Combining this phenomenon with what Compressive Sensing theory offers, they achieved an outstanding performance of channel estimation with relatively small complexity. The experimental results, which we reproduced in this thesis, show the amazing performance of the *Joint OMP* algorithm compared with similar work trying to tackle the same problem and proves its superiority.

Overall, 5G wireless systems are the future of communication. Partially implemented by some countries, and with coverage limited to single cities, or certain blocks of a city. Its extraordinary capabilities, regarding speed and volume of data transferred, are proven in theory and there is yet to see its full potential in practice.

Bibliography

- [1] Waheed U Bajwa et al. “Compressed channel sensing: A new approach to estimating sparse multipath channels”. In: *Proceedings of the IEEE* 98.6 (2010), pp. 1058–1076.
- [2] Yann Barbotin et al. “Estimation of sparse MIMO channels with common support”. In: *IEEE Transactions on Communications* 60.12 (2012), pp. 3705–3716.
- [3] Giulio Bartoli et al. “Beamforming for small cell deployment in LTE-advanced and beyond”. In: *IEEE Wireless Communications* 21.2 (2014), pp. 50–56.
- [4] Christian R Berger et al. “Application of compressive sensing to sparse channel estimation”. In: *IEEE Communications Magazine* 48.11 (2010), pp. 164–174.
- [5] Mehrzad Biguesh and Alex B Gershman. “Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals”. In: *IEEE transactions on signal processing* 54.3 (2006), pp. 884–893.
- [6] Åke Björck. *Numerical methods for least squares problems*. SIAM, 1996.
- [7] Léon Bottou and Noboru Murata. “Stochastic approximations and efficient learning”. In: *The Handbook of Brain Theory and Neural Networks, Second edition, . The MIT Press, Cambridge, MA* (2002).
- [8] Emmanuel J Candès and Michael B Wakin. “An introduction to compressive sampling”. In: *IEEE signal processing magazine* 25.2 (2008), pp. 21–30.
- [9] Junil Choi, David J Love, and Patrick Bidigare. “Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory”. In: *IEEE Journal of Selected Topics in Signal Processing* 8.5 (2014), pp. 802–814.
- [10] Daniel J Costello. *Fundamentals of wireless communication (tse, d. and viswanath, P.)*[book review]. Vol. 55. 2. IEEE, 2009, pp. 919–920.

- [11] Marco F Duarte and Yonina C Eldar. “Structured compressed sensing: From theory to applications”. In: *IEEE Transactions on signal processing* 59.9 (2011), pp. 4053–4085.
- [12] Yonina C Eldar and Moshe Mishali. “Robust recovery of signals from a structured union of subspaces”. In: *IEEE Transactions on Information Theory* 55.11 (2009), pp. 5302–5316.
- [13] Jubin Jose et al. “Pilot contamination and precoding in multi-cell TDD systems”. In: *IEEE Transactions on Wireless Communications* 10.8 (2011), pp. 2640–2651.
- [14] Florian Kaltenberger et al. “Correlation and capacity of measured multi-user MIMO channels”. In: *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE. 2008, pp. 1–5.
- [15] Stefan Kunis and Holger Rauhut. “Random sampling of sparse trigonometric polynomials, ii. orthogonal matching pursuit versus basis pursuit”. In: *Foundations of Computational Mathematics* 8.6 (2008), pp. 737–763.
- [16] Persefoni Kyritsi et al. “Correlation analysis based on MIMO channel measurements in an indoor environment”. In: *IEEE Journal on Selected areas in communications* 21.5 (2003), pp. 713–720.
- [17] Erik G Larsson et al. “Massive MIMO for next generation wireless systems”. In: *IEEE communications magazine* 52.2 (2014), pp. 186–195.
- [18] Athanasios Liavas. *Notes for the class "Wireless Communications"*. 2018.
- [19] Thomas L Marzetta and Bertrand M Hochwald. “Fast transfer of channel state information in wireless systems”. In: *IEEE Transactions on Signal Processing* 54.4 (2006), pp. 1268–1278.
- [20] Mudassir Masood, Laila H Afify, and Tareq Y Al-Naffouri. “Efficient coordinated recovery of sparse channels in massive MIMO”. In: *IEEE Transactions on Signal Processing* 63.1 (2014), pp. 104–118.
- [21] Sinh Le Hong Nguyen and Ali Ghayeb. “Compressive sensing-based channel estimation for massive multiuser MIMO systems”. In: *2013 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2013, pp. 2890–2895.
- [22] Triantafyllos Papakonstantinou. *Massive MIMO wireless communication*. 2017.

- [23] Xiongbing Rao and Vincent KN Lau. “Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems”. In: *IEEE Transactions on Signal Processing* 62.12 (2014), pp. 3261–3271.
- [24] ReinhardtHaverans. *From 1G to 5G: A Brief History of the Evolution of Mobile Standards*. 2021. URL: <https://www.brainbridge.be/en/blog/1g-5g-brief-history-evolution-mobile-standards>.
- [25] Anna Scaglione and Azadeh Vosoughi. “Turbo estimation of channel and symbols in precoded MIMO systems”. In: *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 4. IEEE. 2004, pp. iv–iv.
- [26] Juei-Chin Shen et al. “High-dimensional CSI acquisition in massive MIMO: Sparsity-inspired approaches”. In: *IEEE Systems Journal* 11.1 (2016), pp. 32–40.
- [27] Qinfang Sun et al. “Estimation of continuous flat fading MIMO channels”. In: *2002 IEEE Wireless Communications and Networking Conference Record. WCNC 2002 (Cat. No. 02TH8609)*. Vol. 1. IEEE. 2002, pp. 189–193.
- [28] Joel A Tropp. “Algorithms for simultaneous sparse approximation. Part II: Convex relaxation”. In: *Signal Processing* 86.3 (2006), pp. 589–602.
- [29] Joel A Tropp and Anna C Gilbert. “Signal recovery from random measurements via orthogonal matching pursuit”. In: *IEEE Transactions on information theory* 53.12 (2007), pp. 4655–4666.
- [30] Joel A Tropp, Anna C Gilbert, and Martin J Strauss. “Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit”. In: *Signal processing* 86.3 (2006), pp. 572–588.
- [31] Haifan Yin et al. “A coordinated approach to channel estimation in large-scale multiple-antenna systems”. In: *IEEE Journal on selected areas in communications* 31.2 (2013), pp. 264–273.