Technical University of Crete

Department of Electronic and Computer Engineering

Massive MIMO wireless Communication

Papakonstantinou Triantafyllos

Thesis Committee

supervisor:	Professor Athanasios Liavas
commissioner:	Professor Paterakis Michael
commissioner:	Associate Professor Karystinos George

Chania, February 2017

Acknowledgements

Abstract

In this thesis we study a novel cellular network architecture called "*Massive MIMO*". The characteristic of Massive MIMO architecture is to equip BSs with a number of antennas much larger than the number of active users. As we will see, this promising idea offers enormous enhancements in spectral and energy efficiency, simple signal processing and more.

To understand how massive MIMO works in a wireless cellular system we focus our study on the pioneering and award-winning paper [3] of Thomas L. Marzetta who introduced and popularized the concept of massive MIMO. By simulating a scenario of multicellular wireless system we see the beneficial contribution of the use of multiple antennas in signal processing (more than 100) and learn about the problem of pilot contamination which is generated during the channel estimation phase. This phenomenon limits the capacity of the system and its elimination is of primary importance for engineers. To mitigate pilot contamination we try to use a less aggressive frequency reuse strategy in the cellular network, but as we will see, this will improve the system in terms of interference but it will not produce the desired results in terms of performance.

After having realized the restrictive role of pilot contamination, we study a very interesting novel idea which is being presented in [4] and seems to address the problem (at least at theoretical level). This is about an algorithm which, in conjunction with an improved channel estimator, tackles the pilot contamination problem by taking advantage of the multiple antenna dimensions and exploits the side information lying in the second order statistics of the channel vectors.

Contents

List of Figures

Chapter 1

Introduction

It may be hard to believe but the first cell phone research began in 1843 when Michael Faraday conducted research to see if space could conduct electricity. Fast forward to 1973, and Dr. Martin Cooper is credited with inventing the first portable handset. Four years later, cell phones go public. The first commercially automated cellular network, known as first generation (1G), was launched in Japan by NTT in 1979, initially in the metropolitan area of Tokyo. Within five years, the NTT network had been expanded to cover the whole population of Japan and became the first nationwide 1G network. In 1981, the NMT system launched in Scandinavia. NMT was the first mobile phone network to feature international roaming. The first 1G network in the USA launched in 1983 using the Motorola DynaTAC mobile phone. And then several countries have followed in the early to mid-1980s.

Second Generation (2G) technology was launched in the year 1991 in Finland, and in 1992, the new standard for Pan European digital cellular telephony known as GSM saw its first operational successes. While the radio signals on 1G networks are analog, radio signals on 2G networks are digital. Three primary benefits of 2G networks over their predecessors were that phone conversations were digitally encrypted allowing for the transfer of data in such a way that only the intended receiver can receive and read it, 2G systems were significantly more efficient on the spectrum and introduced data services for mobile, starting with SMS text messages picture messages, and MMS (multimedia messages).

The first 3G network offered for commercial use was again launched in Japan by NTT in the year of 2001. Third Generation technology generally refers to the standard of accessibility and speed of mobile devices. The standards of the technology were set by the International Telecommunication Union (ITU). This technology enables use of various services like GPS (Global Positioning System), mobile television and video conferencing. It not only enables them to be used worldwide, but also provides with better bandwidth and increased speed. 4G is the fourth generation of mobile phone communications standards. It is a successor of the 3G and provides ultra-broadband internet access for mobile devices. The high data transfer rates make 4G networks suitable for use in USB wireless modems for laptops and even home internet access.

And that's how we get to this point of time, wondering what 5G will be? As the long-term evolution (LTE) system embodying 4G has now been deployed and is reaching maturity, where only incremental improvements and small amounts of new spectrum can be expected, preliminary interest and discussions about a possible 5G standard have evolved into a full- fledged conversation that has captured the attention and imagination of researchers and engineers around the world.

Thanks to the annual visual network index (VNI) reports released by Cisco, we have quantitative evidence that the wireless data explosion is

real and will continue. Driven largely by smartphones, tablets, and video streaming and many new applications beyond personal communications. VNI report [1] and forecast makes plain that an incremental approach will not come close to meeting the demands that networks will face by 2020.

So, unlike the previous four generations of cellular technology which each one have been a major paradigm shift that has broken backward compatibility, 5G will need to be a paradigm shift that includes very 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 also be highly integrative: tying any new 5G air interface and spectrum together with LTE and WiFi to provide universal high-rate coverage and a seamless user experience. To support this, the core network will also have to reach unprecedented levels of flexibility and intelligence, spectrum regulation will need to be rethought and improved, and energy and cost efficiencies will become even more critical considerations [2].

One key technology in which researchers have focus on to achieve the needs of the upcoming years is massive MIMO. MIMO communication was introduced into WiFi systems around 2006, into 3G cellular shortly thereafter and by the time LTE was developed, MIMO was a native ingredient thereof with two-to-four antennas per mobile device and as many as eight per Base station (BS) sector. In essence, MIMO embodies the spatial dimension of the communication that arises once a multiplicity of antennas are available at Base Stations and mobile devices.

Thomas L. Marzetta was instrumental in articulating a vision in which the number of antennas increased by more than an order of magnitude. The proposal is to equip BSs with a number of antennas much larger than the number of active users per time–frequency signaling resource, and, given that under reasonable time–frequency selectivities accurate channel estimation can be conducted for at most some tens of users per resource, this condition puts the number of antennas per BS into the hundreds. This bold idea, initially termed "*large-scale antenna systems*" but now more popularly known as "*massive MIMO*" constitutes the object of our study in this thesis. As we will see, this promising idea offers enormous enhancements in spectral and energy efficiency with, as is always the case, a trade-off between them. Also, smoothed out channel responses because of the vast spatial diversity, which brings about the favourable action of the law of large numbers and simple signal processing. As it is reasonable, there are also some limiting factors the tackling of which is a challenge for researchers, so that the massive MIMO system can achieve its capabilities and become reality.

The revival of [3] will help us understand how massive MIMO works in a wireless cellular system, to see the beneficial contribution of the use of multiple antennas in signal processing (more than 100) and learn about the problem of pilot contamination which is generated during the channel estimation phase. This phenomenon limits the capacity of the system and its elimination is of primary importance for engineers. One way to deal with it is by using less aggressive frequency reuse strategy in the cellular network. As we will see, this will improve the system in terms of interference but it will not produce the desired results in terms of performance.

Thus, after having realized the restrictive role of pilot contamination, we study a very interesting novel idea which is being presented in [4] and seems to address the problem (at least at theoretical level). Since the phenomenon is generated during the procedure of channel estimation, the authors create an algorithm which, in conjunction with an improved channel estimator, tackles the pilot contamination problem by taking advantage of the multiple antenna dimensions and exploits the side information lying in the second order statistics of the channel vectors.

Chapter 2

Wireless Communications

We begin with the definition and description of some fundamental principles about wireless communications, which take part on the cellular system that we will study and simulate later on. This chapter starts with Wireless Channel Models at 2.1 and continues with MIMO systems in 2.2 and Cellular Systems at 2.3.

2.1 Wireless Channel

The wireless channels operate through electromagnetic radiation from the transmitter to the receiver. Electromagnetic waves differ in energy according to their wavelength (frequency). Their ability to propagate is also different at different propagation environments. A transmitted signal undergoes changes while traveling through the propagation path to the receiver. In cellular wireless communications, obstacles, such as houses, buildings, trees and mountains cause reflection, diffraction, scattering, and shadowing of the transmitted signals and multipath propagation. Due to the multipath, the transmitted signals arrive with different phase angles, amplitude and time delay. The amplitude fluctuation of the received signal caused by the frequency selective or time varying multipath channel is called *fading*. The fading phenomena can be classified into two main groups known as *large scale fading* and *small scale fading*. The large scale fading is used to describe the signal level at the receiver after traveling over a large area (hundreds of wavelengths). Small scale fading is used to describe the signal level at the receiver after encountering obstacles near (several wavelengths to fractions of wavelengths) the receiver [5], [6].

2.1.1 Large-scale Fading

Large-scale fading is the result of signal attenuation due to signal propagation over large distances and diffraction around large objects, such as buildings and hills.

Free Space Propagation

We consider an isotropic antenna, which radiates electromagnetic energy uniformly in all directions. The transmitter Tx and receiver Rx are located in distance d as shown in Figure 2.1(a). In this case, the electromagnetic waves are characterized by their ability to propagate without obstruction and without atmospheric effects. The path loss under these conditions, where the attenuation of signal power is solely on account of expansion of the signal wavefront, is known as *free-space pathloss*. In the far field of the transmitting antenna, the free space pathloss is described by an inverse square law, which

(a) Free Space Propagation

(b) 2-ray Model
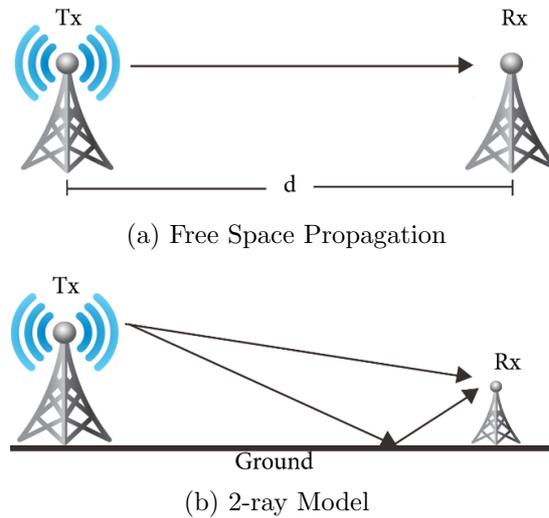
Figure 2.1: Free Space and 2-Ray propagation.

implies that the power density of the signal decreases with the square of distance from transmitting antenna.

During transmission over a ground plane, as is normally the case with cellular systems, part of the radio signal gets reflected by the earth's surface and interferes with the primary wavefront (see Figure 2.1(b)). For such a scenario, with reflection from a ground plane, the power density decreases with a pathloss exponent of 4. As we see, the received power can decrease with distance faster than d^{-2} in the presence of disturbances to free space. In practice, there are several obstacles between the transmitter and the receiver and, further, the obstacles might also absorb some power while scattering the rest. Thus, one expects the power decay to be considerably faster than d^{-2}. Indeed, empirical evidence from experimental field studies suggests that while power decay near the transmitter is like d^{-2}, at large distances the power decays exponentially with distance [5].

Shadow Fading

The density of obstacles between the transmit and receive antennas depends very much on the physical environment. For example, outdoor plains have very few of obstacles while indoor environments pose many obstacles. In a typical urban setup, there are several large structures like buildings, mountains, hills and trees that severely attenuate the radio signal and are creating shadow zones that span an area much larger than the wavelength of the signal carrier. This phenomenon is called *shadow fading* and causes large-scale attenuation for the duration that the terminal remains within the shadow of the scatterer. However, the wavefront undergoes diffraction at the edges of the scatterer, and some fraction of signal power is available to the terminal even when the primary wavefront might be completely obstructed. We also use the term *slow fading*, because the length of time it takes for a moving receiver to pass through the "shadow" of these large obstacles lasts for multiple seconds or minutes and hence occurs at a slow time-scale.

Since the location, size, and dielectric properties of the blocking objects as well as the changes in reflecting surfaces and scattering objects that cause the random attenuation are generally unknown, statistical models are widely used to characterize the attenuation of shadow fading. The most common model for this attenuation is *log-normal distribution*,

$$h_{shadow} \sim \ln \mathcal{N}(\mu, \sigma^2)$$

This model has been confirmed empirically to accurately model the variation in path loss or received power in both outdoor and indoor radio propagation environments.
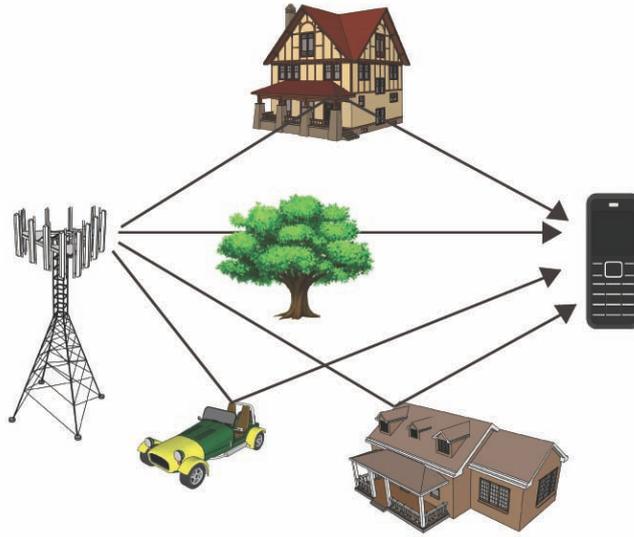
Figure 2.2: Multi-path Propagation.

2.1.2   Small-scale Fading

Small scale fading is a characteristic of radio propagation resulting from the presence of reflectors and scatterers that cause multiple versions of the transmitted signal to arrive at the receiver, each distorted in amplitude, phase, and angle of arrival.

Multi-path Propagation

The signal is reflected, refracted, and scattered from objects that are present in the path, creating multiple paths. Each one can have a different amount of attenuation, delay, and fading type. The combination of these different paths is termed *multipath fading*. At the receiver, the signals can add constructively or destructively, depending upon their path. The movement of the receiver or the transmitter, as is usual with the mobiles in cellular networks, is causing random and rapid fluctuations in the received amplitude. We use the term *fast fading* for this kind of fast changes in the channel impulse response. If there is a large number of uncorrelated scatterers and no line of sight, as is usually the case in heavily built-up urban areas, the in-phase and quadrature components of the received signal with small-scale fading can be assumed to follow independent zero-mean *Gaussian distributions*, also called *Rayleigh fading*.

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

Delay Spread

The *delay spread* is a measure of the multipath richness of a communication channel. It can be interpreted as the difference between the time of arrival of the earliest significant multipath component and the time of arrival of the latest multipath component. Delay spread can be defined as shown in the relationship below.

$$T_{max}(t) = \max_{i,j} \ |\tau_i(t) - \tau_j(t)|,$$

where τ_i and τ_j represent the time that is needed for the longest and the shortest path to reach the receiver.

The importance of delay spread is how it affects the *Inter Symbol Interference* (ISI). If the symbol duration is long enough compared to the delay spread (typically 10 times as big would be good enough), one can expect an equivalent ISI-free channel. The correspondence with the frequency domain is the notion of *coherence bandwidth*, which is the bandwidth over which the channel can be assumed flat. Coherence bandwidth is related to the inverse of the delay spread. The shorter the delay spread, the larger coherence bandwidth [6].

Flat Fading and Frequency Selective Fading Channels

As we said before, we define as *coherence bandwidth* (Bc) the bandwidth over which the channel can be assumed approximately constant. If Bc is greater than the signal bandwidth, then the measure of the frequency response of the channel is flat and at the output of the channel affects only the amplitude and the phase without introducing ISI, this type of fading is called *flat fading*. If Bc is less than the signal bandwidth, then the measure of the channel's frequency response is not flat, in general, and the channel introduces ISI, this type of fading is called *frequency selective fading*. So the statistical model that describes the *flat fading* channels is

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

In *frequency selective fading*, as we said due to large delay spread the channel introduce ISI, where in the received signal interfere signals of previous transmitted symbols. So, if the coefficients that interfere in the channel is l, the channel is modeled as the sequence of l gaussian random variables with differ in variance [6],

$$h_{m,l} \sim \mathcal{CN}(0, \sigma_l^2).$$

2.2 Multiple-input Multiple-output Antenna System

There is a number of different antenna systems configurations that can be used in wireless communication. These are termed SISO, SIMO, MISO and MIMO. These different formats offer different advantages and disadvantages, which can be balanced to provide the optimum solution for any given application. The different MIMO formats require different number of antennas as well as having different levels of complexity. Also dependent upon the format, processing may be needed at one end of the link or the other [6], [7].

2.2.1 Single-input Single-output (SISO)

The simplest form of radio link can be defined in MIMO terms as SISO - Single-Input Single-Output. This is effectively a standard radio channel in which the transmitter operates with one antenna as does the receiver. There is no diversity and no additional processing required. The advantage of a SISO system is its simplicity. It requires no processing in terms of the various forms of diversity that may be used, however in digital communications systems, such as wireless internet, it can cause a reduction in data speed and an increase in the number of errors. Interference and fading will impact the system more than a MIMO system which is using some form of diversity, and the channel bandwidth is limited by Shannon's law. The throughput depends upon the channel bandwidth and the signal to noise ratio.
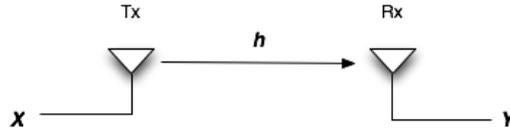
Figure 2.3: Single input Multiple output.

The input-output relationship in such a system is given by:

$$Y = hX + n,$$

where X represents the transmitted symbol, Y the received signal, h the channel between the transmitter and receiver, $n \sim \mathcal{CN}(0, \sigma_{noise}^2)$ the additive noise.

2.2.2   Single-input Multiple-output (SIMO)
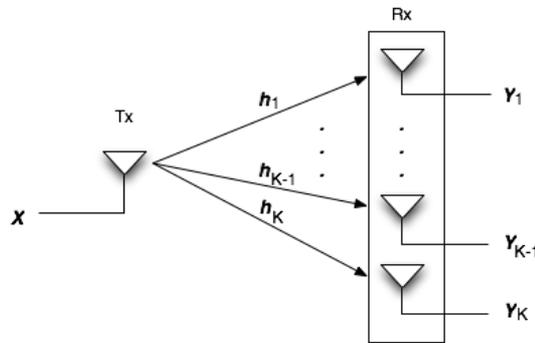
Figure 2.4: Single input Multiple output.

The SIMO occurs where the transmitter has a single antenna and the receiver has multiple antennas. This is also known as receive diversity. It is often used to enable the receiver system to combat the effects of fading. It has been used for many years with short wave listening/receiving stations to combat the effects of deep fading and interference. SIMO has the advantage that it is relatively easy to implement although it does have some disadvantages in that the processing is required in the receiver. The use of SIMO may be quite acceptable in many applications, but where the receiver is located in a mobile device such as a cellphone handset, the levels of processing may be limited by size, cost and battery drain.

The input-output relationship in a system like in Figure 2.4 is given by:

$$Y = HX + N,$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_K \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{bmatrix} X + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_K \end{bmatrix},$$

where $N \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_K)$.

Two kinds of processing at the receiver are:

- **Switched diversity SIMO:** Looks for the best signal among all the signals received from different antennas at the receiving end and switches to that antenna. This option is clearly not the optimal solution as $(K - 1)$ elements of the array are ignored.

- **Maximum ratio combining SIMO:** Combining all the signals in a co-phased manner, with such weights for each signal level as to have the highest achievable SNR at the receiver at all times. MRC is optimal in terms of SNR and it is commonly used because of its simplicity and effectiveness. This is achieved by filtering the received signals (vector $Y = [Y_1 \ Y_2 \ \cdots \ Y_M]^T$) with the *matched filter* : $\frac{\hat{h}}{||\hat{h}||}$, where $\hat{h} = [\hat{h}_1 \ \hat{h}_2 \ \cdots \ \hat{h}_M]^T$ is the channel estimation vector. Thus, we have the following input-output relationship:

$$\hat{X} = \frac{\hat{h}}{||\hat{h}||} Y.$$
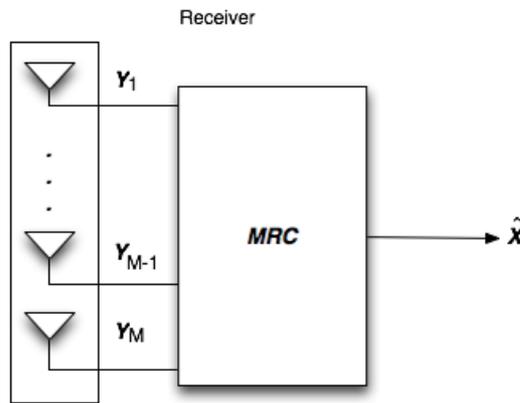
Figure 2.5: Matched Filtering.

2.2.3 Multiple-input Single-output (MISO)

MISO is another antenna technology for wireless communications in which multiple antennas are used at the source (transmitter) and only one antenna is used at the destination (receiver). The advantage of using MISO is that the multiple antennas and the redundancy coding/processing is moved from the receiver to the transmitter. In instances such as cellphones, this can be a significant advantage in terms of space for the antennas and reducing the level of processing required in the receiver for the redundancy coding. This has a positive impact on size, cost and battery life as the lower level of processing requires less battery consumption.

The received signal raised by the relation:

$$Y = HX + n$$
$$Y = h_1 X_1 + h_2 X_2 + \ldots + h_M X_M + n$$

where $H = [h_1 \ h_2 \ \cdots \ h_M]$, $X = [X_1 \ X_2 \ \cdots \ X_M]^T$ and $n \sim \mathcal{CN}(0, \sigma_{noise}^2)$.

An effective technique used by MISO systems is the *transmit beamforming*, where the transmitter adds information in transmitted signal in order to avoid the processing in the receiver. This requires the knowledge of the channel, so the transmitter instead of X, will send $\mathbf{X} := \frac{\hat{h}}{||\hat{h}||} X$.
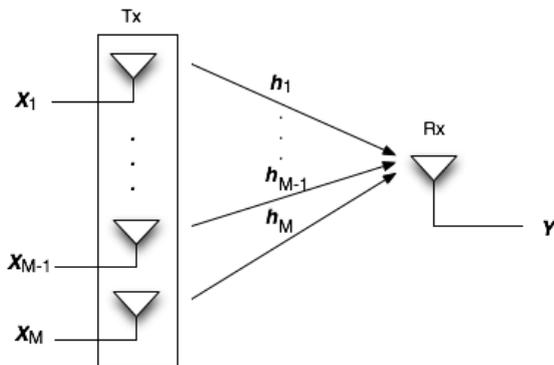
Figure 2.6: Multiple input Single output.

2.2.4   Multiple-input Multiple-output (MIMO)

When there are more than one antenna at both ends of the radio link, this is termed MIMO - Multiple-Input Multiple-Output. The use of two or more antennas, along with the transmission of multiple signals at the source and the destination, eliminates the trouble caused by multipath wave propagation, and can even take advantage of this effect. MIMO can be used to provide improvements in both channel robustness as well as channel throughput. In order to be able to fully benefit from MIMO, it is necessary to be able to utilise coding on the channels to separate the data from the different paths. This requires processing, but provides additional channel robustness/data throughput.
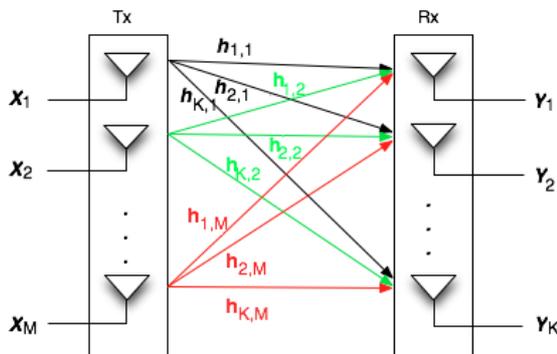
Figure 2.7: Multiple input Multiple output.

The system is described by the input-output relationship:

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_K \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,M} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ h_{K,1} & h_{K,2} & \cdots & h_{K,M} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_K \end{bmatrix}.
$$

2.3   Cellular Systems

Cellular systems constitute networks which serve a large number of mobile users interested in communicating with a common wireline network infrastructure. This form of wireless communication is different from radio or TV in two important respects: first, users are interested in messages specific to them as opposed to the common message that is broadcast in radio and TV. Second, there is two-way communication between the users and the network. In particular, this allows feedback from the receiver to the transmitter, which

(a) half duplex
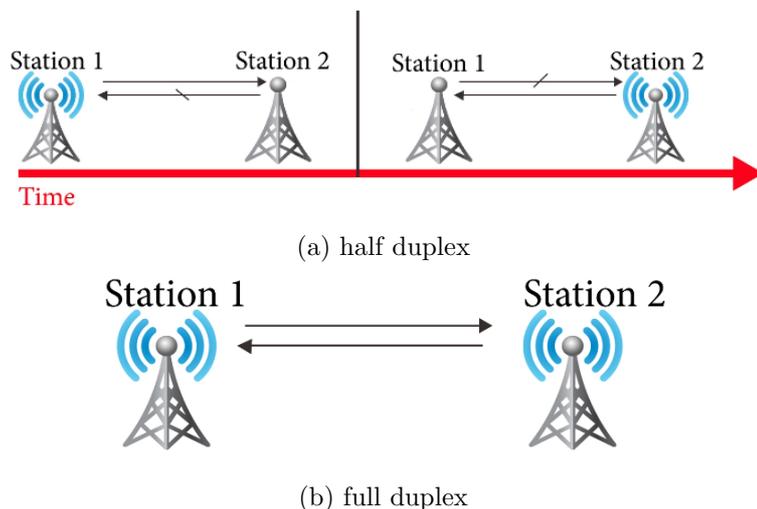
(b) full duplex

Figure 2.8: Half and full duplex.

is missing in radio and TV. This form of communication is also different from the all-wireless walkie-talkie communication since an access to a wireline network infrastructure is demanded.

A cellular network consists of a number of fixed base stations, one for each cell. The total coverage area is divided into cells and a mobile communicates with the base station close to it (or in some systems with the base station which provides "stronger" channel). At the physical and medium access layers, there are two main issues in cellular communication: multiple access and interference management. The first issue addresses how the overall resources (time, frequency and space) of the system are shared by the users in the same cell (intra-cell) and the second issue addresses the interference caused by simultaneous signal transmissions in different cells (inter-cell). At the network layer, an important issue is that of seamless connectivity to the mobile as it moves from one cell to the other (and thus switching communication from one base station to the other) [5].

In addition to resource sharing between different users, there is also an issue of how the resources are allocated between the uplink (the communication from the mobile users to the base station, also called the reverse link) and the downlink (the communication from the base station to the mobile users, also called the forward link). There are two natural strategies for separating resources between the uplink and the downlink: time division duplex (TDD) separates the transmissions in time and frequency division duplex (FDD) achieves the separation in frequency [5].

Duplexing

Duplexing is a process of achieving two-way communication over a communication channel. It takes two forms: half duplex and full duplex.

In half duplex, the two communicating parties take turns transmitting over a shared channel. Two-way radios work this way. As one party talks, the other listens (Figure 2.8(a)). Speaking parties often say "Over" to indicate that they are finished and it is time for the other party to speak. In networking, a single cable is shared as the two computers communicating take turns sending and receiving data.

Full duplex refers to simultaneous two-way communication. The two communicating stations can send and receive at the same time (Figure 2.8(b)). Landline telephones and cell phones work this way. Some forms of networking permit simultaneous transmit and receive operations to occur. This is the more desirable form of duplexing, but it is more complex and

expensive than half duplexing. There are two basic forms of full duplexing:
frequency division duplex (FDD) and *time division duplex* (TDD) [8].

Frequency division duplex

FDD in wireless systems needs two separate frequency bands or channels
(Figure 2.9). A sufficient amount of guard band separates the two bands so
the transmitter and receiver do not interfere with one another. Good filtering
or duplexers and possibly shielding are a must to ensure the transmitter does
not desensitize the adjacent receiver. The greater the spectrum separation,
the more effective the filters.

Although the FDD is very widely used in cellular telephone systems, such
as the widely used GSM system, it has some serious disadvantages. First
of all, FDD uses lots of frequency spectrum, at least twice the spectrum
needed by TDD. In addition, there must be adequate spectrum separation
between the transmit and receive channels. These so-called guard bands
are not useable, so they are wasteful. Given the scarcity and expense of
spectrum, that is a big negative factor of FDD.

Another disadvantage with FDD is the difficulty of using special antenna
techniques like MIMO and beamforming, and as we will see later these
technologies are a core part of the next generation cellular systems to increase
data rates. It is difficult to make antenna bandwidths broad enough to cover
both sets of spectrum. More complex dynamic tuning circuitry is required.
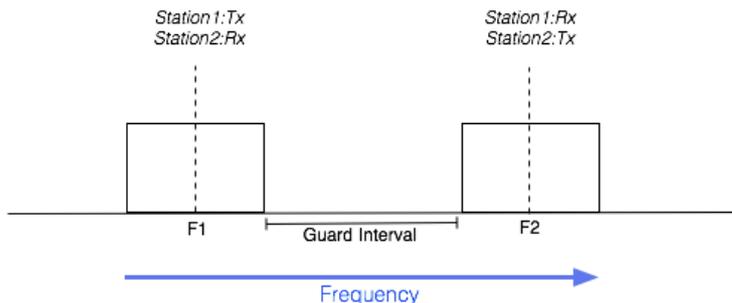
Figure 2.9: The required spectrum of FDD for the uplink and downlink
channels.

Time Division Duplex

TDD uses a single frequency band for both transmit and receive. Then it
shares that band by assigning alternating time slots to transmit and receive
operations (Figure 2.10). The information to be transmitted—whether it is
voice, video, or computer data—is in serial binary format. Each time slot
may be 1 byte long or could be a frame of multiple bytes.

Because of the high-speed nature of the data, the communicating parties
cannot tell that the transmissions are intermittent. The transmissions are
concurrent rather than simultaneous.

In some TDD systems, the alternating time slots are of the same duration
or have equal DL and UL times. However, the system does not have to be
50/50 symmetric. The system can be asymmetric as required. For instance,
in Internet access, download times are usually much longer than upload times
so more or fewer frame time slots are assigned as needed. Some TDD formats
offer dynamic bandwidth allocation where time-slot numbers or durations
are changed on the fly as required.

The real advantage of TDD is that it only needs a single channel of fre-
quency spectrum. Furthermore, no spectrum-wasteful guard bands or channel
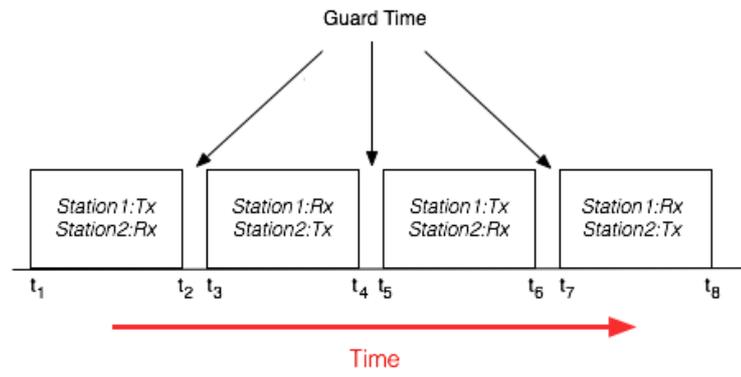
Figure 2.10: Time slots of TDD.

separations are needed. The downside is that successful implementation of TDD needs a very precise timing and synchronization system at both the transmitter and receiver to make sure time slots do not overlap or otherwise interfere with one another.

Chapter 3

Massive MIMO for Next Generation Wireless Systems

Multiple-input multiple-output (MIMO) technology has been widely studied during the last two decades and has been applied to many wireless standards, since it can significantly improve the capacity and reliability of wireless systems, but it has yet to be adopted on a scale commensurate with its true potential. The reason behind this is that telecommunication companies had preferred cheaper alternatives to increase throughput, but the needs of the time, which require rapid rise in data transfer, have necessarily led to more expensive and technologically sophisticated solutions. In this chapter, we discuss the Massive MIMO system which is a game-changing technology with regard to theory, systems, and implementation. The potential of this technology is enormous and is promising to be the answer to the todays needs in wireless communication.

3.1 MU-MIMO

Conventional MIMO systems, known as *point-to-point MIMO*, require both the transmitter and the receiver of a communication link to be equipped with multiple antennas. There are various reasons to why this technology does not really catch on the todays needs of wireless communication.

According to Thomas Marzetta, the problem of point-to-point MIMO system is that it is not really scalable, because, as the number of antennas increases, the sum throughput does not really increase significantly and the actual MIMO performance falls further behind the theoretical gains. In practice, as the number of antennas increases, the system is not really doing multiplexing and just does not get the maximum number of transmit receiver multiplexing gains. Marzetta also says that in cellular systems, it is critically important to provide uniformly good service throughout cells and the point-to-point system does not meet this, because multiplexing gains may disappear near the edges of the cell, where signal levels are low relative to interference or in a propagation environment which is dominated by scattering.

Furthermore, while MIMO has become an essential element of wireless communication, many wireless devices cannot support multiple antennas due to size, cost, and/or hardware limitations. More importantly, the separation between antennas on a mobile device and even on fixed radio platforms is often insufficient to allow meaningful performance gains.

So, in recent years, the scientific community has turned its attention to the so called *Multi-User MIMO* (MU-MIMO) system. As we see in Figure 3.1, in MU-MIMO we split one user with multiple antennas into K users each with a single antenna. So, typically, in MU-MIMO a base station (BS) with multiple antennas simultaneously serves a set of single-antenna users
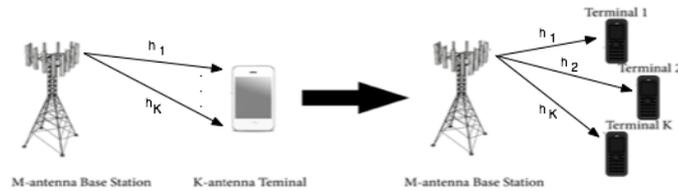
Figure 3.1: From point to point to MU-MIMO.

and the multiplexing gain can be shared by all users. In this way, expensive equipment is only needed on the BS end of the link, and the user terminals can be relatively cheap single-antenna devices.

Thomas Marzetta in his presentation of *Massive MIMO and Beyond*, which presented in Munich's workshop on massive MIMO in 2015, argues that the most remarkable things of MU-MIMO is that by the splitting up the multi-antenna user into autonomous single-antenna users does not decrease the sum-throughput. He also said that, due to multi-user diversity, the performance of MU-MIMO systems is generally less sensitive to the propagation environment than in the point-to-point MIMO case. For example, under line-of-sight propagation conditions, multiplexing gains can disappear for a point-to-point system, but are retained in the multi-user system, provided the angular separation of the terminals exceeds the Rayleigh resolution of the array.

However, even if the theory of multi-user MIMO was promising by a number of researchers, as originally envisioned with roughly equal numbers of service-antennas and terminals and frequency division duplex operation, is not a scalable technology. Especially because there is no way to get economically the requirement of the channel state information (CSI) at both ends, as the number of antennas grows. In contrast with the point-to-point MIMO, which only needs CSI at the receiver, in MU-MIMO apart from the need of CSI at the receiver we also need CSI at the transmitter.

3.2 Massive MIMO
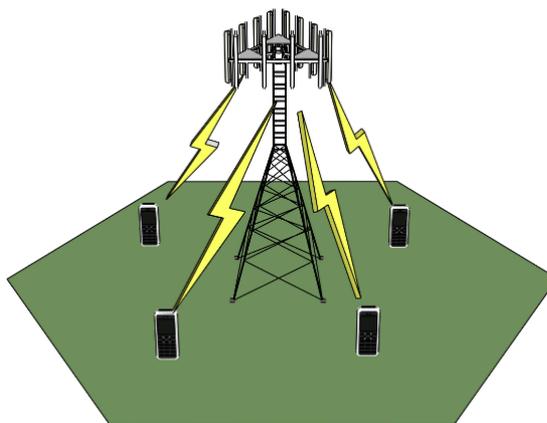
Figure 3.2: Single-Cell Massive MIMO.

For most MIMO implementations, the BS typically employs only a few (i.e., fewer than 10) antennas, and the corresponding improvement in spectral efficiency, while important, is still relatively modest. In an effort to achieve more dramatic gains and to reap all the benefits of conventional MIMO on a much greater scale as well as to simplify the required signal processing,

"*Massive MIMO*" systems are currently investigated as a novel cellular network architecture. Where in Massive MIMO operation each BS is equipped with a large number of antennas, 100 or more, which simultaneously serve many tens of terminals in the same time-frequency slot. There are many different configurations and deployment scenarios for the placing of the antenna arrays used by a massive MIMO system, such as linear arrays, cylindrical arrays, distributed antennas in different buildings and more [9]. Each antenna unit would be small and active, with antenna spacing greater than the half of the signal wavelength so the signals to be uncorrelated. The principal findings of recent researches have shown that the addition of more base station antennas is always beneficial and overall massive MIMO is an enabler for the development of future broadband networks which will be energy-efficient, secure, robust, and will use the spectrum efficiently.

3.2.1   Massive MIMO benefits

Here, we mention some of the key advantages of massive MIMO systems.

Energy Efficiency

An important advantage of massive MIMO lies in its potential energy efficiency. The fundamental principle that makes the dramatic increase in energy efficiency possible is that with large number of antennas, energy can be focused with extreme sharpness into small regions in space [9]. The underlying physics is coherent superposition of wavefronts. By appropriately shaping the signals sent out by the antennas, the base station can make sure that all wavefronts collectively emitted by all antennas add up constructively at the locations of the intended terminals, but destructively (randomly) almost everywhere else.

On the uplink, reducing the transmit power of the terminals will drain their batteries slower and it is shown in [10] that each single-antenna user can scale down its transmit power proportional to the number of antennas at the BS with perfect channel state information (CSI) or to the square root of the number of BS antennas with imperfect CSI (in realistic scenarios where CSI is estimated from uplink pilots), to get the same performance as a corresponding single-input single-output system.

On the downlink, the drastically improved energy efficiency enables the massive MIMO systems to operate with a total output radio frequency power two orders of magnitude less than with current technology. Much of the electrical power consumed by a BS is spent by power amplifiers and associated circuits and cooling systems. Hence reducing the emitted radio frequency power would help in cutting the electricity consumption of the BS. This matters, because the energy consumption of cellular base stations is a growing concern worldwide. In addition, base stations that consume many orders of magnitude less power could be powered by wind or solar, and hence easily deployed where no electricity grid is available. As a bonus, the total emitted power can be dramatically cut and therefore the base station will generate substantially less electromagnetic interference. This is important owing to the increased concerns of electromagnetic exposure [9].

Spectral Efficiency

The energy-efficiency of a system is defined as the spectral-efficiency divided by the transmit power expended. Typically, increasing the spectral efficiency is associated with increasing the power and, hence, with decreasing the energy-efficiency. Therefore, there is a fundamental tradeoff between the energy efficiency and the spectral efficiency. However, in [9], [10], the study of this

tradeoff in massive MIMO multi-cellular system shows that for imperfect CSI in the low transmit power regime can simultaneously increase the spectral-efficiency and energy-efficiency. The reason that the overall spectral efficiency is higher than in conventional MIMO is that many tens of terminals are served simultaneously, in the same time-frequency slot. In the first three generations of cellular technology, the BS served multiple terminals by separating them in time, frequency or code. Each terminal was assigned a unique fraction of spectrum resources for communication over the forward and reverse-links, to minimize intra-cell interference. A massive system opens up the spatial dimension that allows it to discriminate the signals to/from each terminal based on its location. This enables each terminal to use all available spectrum resources, improving the throughput without the need for additional (expensive) resources.

Simple Signal Processing

By increasing the number of BS antennas relative to the number of users (relative to MU-MIMO) one could think that the signal processing problem could be even more difficult but in fact the opposite happens. In massive MIMO, users do not need CSI and no duty of dirty paper coding. In fact, the terminals do not have to do any signal processing at all, signal processing only needed at the base stations. This in itself is very important as it considerably reduces the construction requirements of phones. But there are more benefits about signal processing in massive MIMO.

In [9], it is shown that, in large-scale MIMO, very high spectral efficiency can be obtained even with simple maximum-ratio combining processing at the same time as the transmit power can be cut back by orders of magnitude and this holds true even with imperfect CSI. The attractiveness of MRC compared with ZF is not only its computational simplicity-multiplication of the received signals by the conjugate channel responses, but also that it can be performed in a distributed fashion, independently at each antenna unit. While ZF also works fairly well for a conventional or moderately-sized MIMO system, MRC generally does not. The reason for why MRC works so well for massive MIMO is that the channel responses associated with different terminals tend to be nearly orthogonal when the number of base station antennas is large, and in the limit of an infinite number of antennas (scenario we adopt later in our system for the simulations), the effects of fast fading and uncorrelated noise vanish [3].

Massive MIMO also enables a significant reduction of latency on the air interface. It relies on the law of large numbers and beamforming in order to avoid fading dips, which are due to multi-path fading and limit the performance of wireless communications systems makes it hard to build low-latency wireless links [9].

Low cost components in Base Stations

Another advantage of massive MIMO is that BS can be built with inexpensive, low-power components. Unlike the conventional systems which use ultra-linear amplifiers, massive MIMO BS can be equipped with hundreds of low-cost amplifiers with output power in the milli-Watt range. The contrast to classical array designs, which use few antennas fed from high-power amplifiers, is significant. Also, several expensive and bulky items, such as large coaxial cables, can be eliminated altogether. Massive MIMO reduces the constraints on accuracy and linearity of each individual amplifier and radio frequency chain. All what matters is their combined action. In a way, massive MIMO relies on the law of large numbers to make sure that noise,

fading and hardware imperfections average out when signals from a large number of antennas are combined in the air together. The same property that makes massive MIMO resilient against fading also makes the technology extremely robust to failure of one or a few of the antenna units [9].

3.2.2   Limiting Factors of Massive MIMO

In massive MIMO operation there are some factors that are limiting the performance, which need to be addressed in order to achieve system's true potentials.

Channel Reciprocity

For reasons we will explain below (section 4.1.3), massive MIMO is more efficient with TDD operation. The TDD operation relies on reciprocity of the channel in the uplink and downlink and there appears to be a reasonable consensus that the propagation channel itself is essentially reciprocal, unless the propagation is affected by materials with strange magnetic properties. However, the hardware chains in the base station and terminal transceivers may not be reciprocal between the uplink and the downlink. Calibration of the hardware chains does not seem to constitute a serious problem and there are calibration-based solutions that have already been tested to some extent in practice. Specifically, [11] treats reciprocity calibration for a 64-antenna system in some detail and claims a successful experimental implementation.

Pilot Contamination

Ideally, every terminal in a Massive MIMO system is assigned an orthogonal pilot sequence. However, the maximum number of orthogonal pilot sequences that can exist is upper-bounded by the duration of the coherence time. *Coherence time* is the time interval in which the channel can be assumed constant and depends on the mobility of the terminals. This time-interval is divided between reverse-link pilots and transmitting data, either on the forward link or the reverse link or both. So, if the coherence time suffices for sending about τ symbols, the number of terminals that can be served by a base station is proportional to the symbols that BS will use for pilots. Therefore, the number of terminals that can be served in each cell is limited by the coherence time of the channel and independent of the size of the cell and the number of BS antennas.

It is critical for the system performance to eliminate interference from terminals within the same cell. Thus, in order to serve the maximum number of terminals in each cell and to eliminate intra-cell interference, we exhaust all the available pilot sequences in all of the users in every single cell.

In a multi-cellular scenario the same band of frequencies and the same orthogonal pilot sequences are re-used among the cells. In the course of learning the channels of its own terminals, a base station inadvertently learns the channel of terminals in other cells who share the same pilot sequences, or whose pilot sequences are merely correlated with the pilot sequences of its own terminals. Thus, while BS is transmitting data to its own terminals, it also selectively transmits data to terminals in other cells. Similarly, when the base station combines its reverse-link signals to receive the individual data transmissions of its terminals, it is also coherently combining signals from terminals in other cells. The resulting inter-cellular interference persists even with an infinite number of antennas. This effect is known as *pilot contamination* and it is a fundamental problem which adversely affects the performance of the system.

Pilot contamination as a basic phenomenon is not really specific to massive MIMO, but its effect on massive MIMO appears to be much more profound than in classical MIMO. As we will see in our analysis below, in which we study a scenario where a noncooperative cellular wireless network with unlimited numbers of Base station antennas with receivers that rely on pilot-based channel estimation, pilot contamination constitutes an ultimate limit on performance. Due to the significance of the pilot contamination problem in massive MIMO, a large part of researchers interested in wireless communications have turned their attention to finding ways to combat it. We will analyze two promising ideas below and we will see the results of their application in the simulations below.

Chapter 4

Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas

In this chapter, we present the system and the analysis of [3], which constitutes the state of the art in which references are made by the largest amount of researches relevant to massive MIMO. We study an integrated massive MIMO multicellular system to see how it operates and we perform simulations to find the limits of its capacity and how this is affected by pilot contamination. Then, we present the idea of frequency re-use in the cellular systems and we simulate to see how good it works against pilot contamination and how much improves the capacity of the system.
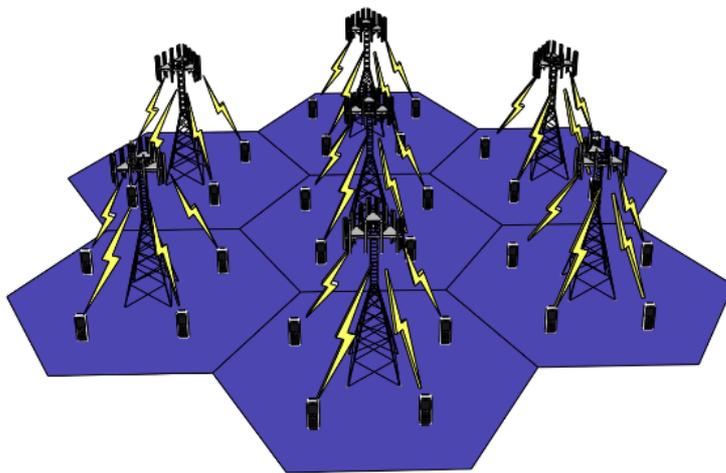
Figure 4.1: Cellular Systems.

4.1   Scenario

We consider a cellular system consisting of noncooperative hexagonal cells with frequency re-use of one, three, or seven, TDD operation, Orthogonal Frequency Division Multiplexing (OFDM), and base station arrays comprising M antennas where $M \to \infty$, where each base station serves K single-antenna terminals. We will describe all these analytically.

(a) Isotropic radiation   (b) Omnidiractional radiation
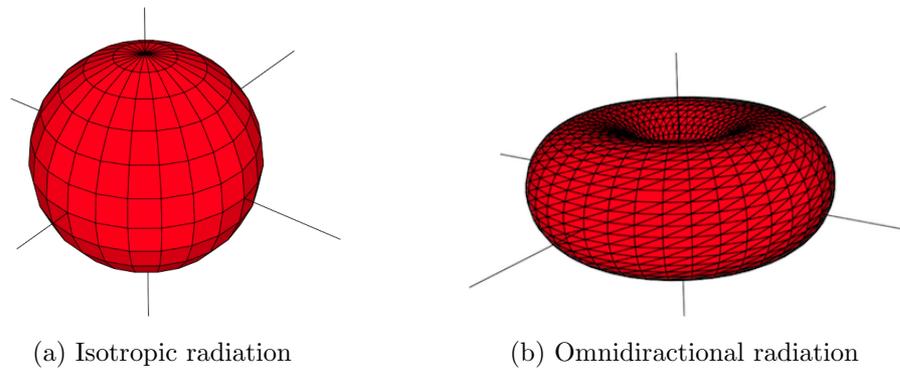
Figure 4.2: Radiation of Isotropic and Omnidiractional antennas.

4.1.1   Hexagonal cells

We consider a circular area filled of cells, in which each cell is hexagonal with a radius from center to vertex of r_c. The propagation models represent a cell as a circular area but approximate cell coverage with a hexagon allows easier analysis because hexagons can fill an area without gaps and overlaps. In the center of the cells are positioned the base stations which consist of an array of M omnidirectional antennas, where in the subsequent analysis M grows without limit. Omnidirectional antennas oriented vertically are widely used for nondirectional antennas on the surface of the Earth because they radiate equally in all horizontal directions, while the power radiated drops off with elevation angle so little radio energy is aimed into the sky or down toward the earth and wasted. As we see in Figure 4.2, the omnidirectional radiation pattern is different from isotropic antennas, as an isotropic one radiates equal power in all directions and has a "spherical" radiation pattern. Within each cell there are K single-antenna terminals, uniformly distributed over the cell with the exclusion of a central disk of radius r_h where supposedly is placed the base station.

4.1.2   OFDM

We assume that OFDM is utilized. OFDM has become the dominant signaling format for high-speed wireless communication, forming the basis of all current WiFi standards and of LTE, and due to its qualities OFDM constitutes a promising method for next generation cellular wireless systems.

Some of the advantages of OFDM are that first it constitutes a natural way to cope with frequency selectivity. Second it has a computationally efficient implementation due to low complexity of FFT/IFFT transforms which are needed at the transmitting and receiving process. Also, the OFDM constitutes an excellent pair with MIMO, since OFDM allows for the spatial interference from multi-antenna transmission to be dealt with at a subcarrier level, without the added complication of inter-symbol interference. From a multiple access point, OFDM invites dynamic fine-grained resource allocation schemes in the digital domain. OFDMA (Orthogonal Frequency-Division Multiple Access) is a multi-user version of OFDM, which offers orthogonal multiple access by assigning subsets of subcarriers to individual users. Being able to do frequency and time slot allocation digitally also enables more adaptive and sophisticated interference management techniques such as fractional frequency reuse. Finally, given its near-universal adoption, industry has by now a great deal of experience with its implementation, and tricky aspects of OFDM, such as frequency offset correction and synchronization have been essentially conquered [2].
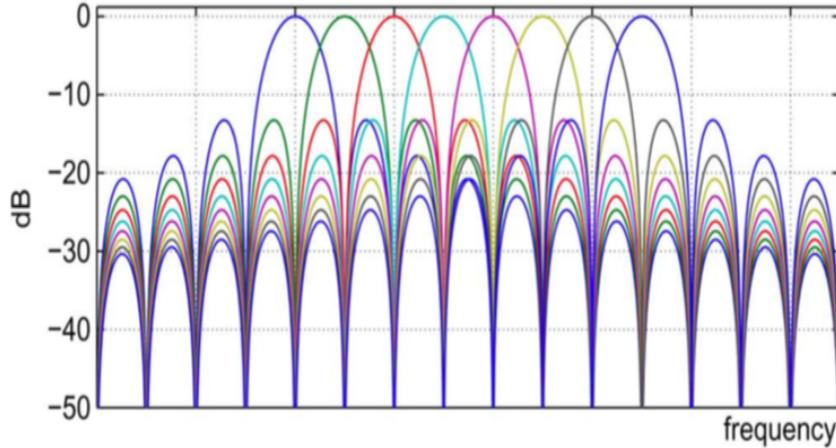
Figure 4.3: Subcarriers of OFDM.

For our analysis, we denote the OFDM symbol interval by T_s, the subcarrier spacing by Δ_f, the useful symbol duration by $T_u = \frac{1}{\Delta_f}$, and the guard interval (duration of the cyclic prefix) by $T_g = T_s - T_u$. We call the reciprocal of the guard interval, when measured in subcarrier spacings, the "frequency smoothness interval",

$$N_{Smooth} = \frac{1}{T_g \Delta_f}. \tag{4.1}$$

4.1.3   Duplexing

For regular MIMO systems, multi-user precoding in the downlink and detection in the uplink require Channel State Information at the BS. The resources, time or frequency, required for channel estimation in a MIMO system are proportional to the number of the transmit antennas and independent of the number of the receive antennas.

If FDD is used, uplink and downlink use different frequency bands, the CSI corresponding to the uplink and downlink is different. Channel estimation for the uplink is done at the BS by letting all users send different pilot sequences. The time required for uplink pilot transmission is independent of the number of antennas at the BS. However, to get CSI for the downlink channel in FDD systems, a two-stage procedure is required. The BS first transmits pilot symbols to all users, and then all users feed back estimated CSI for the downlink channels to the BS. The time required to transmit the downlink pilot symbols is proportional to the number of antennas at the BS. As the number of BS antennas grows large, the traditional downlink channel estimation strategy for FDD systems becomes infeasible. For example, consider a 1 ms x 100 kHz channel coherence interval, which can support transmission of 100 complex symbols. When there are 100 antennas at the BS, the whole coherence interval will be used for downlink training if orthogonal pilot waveforms are used for channels to each antenna, while there is no symbol left for data transmission [12].

Figure 4.4: A typical protocol of Time Division Dupplex time slots.

Fortunately, the channel estimation strategy in TDD systems can be utilized to solve the problem, because, based on the assumption of channel reciprocity, only CSI for the uplink needs to be estimated in BSs, and the total training time is proportional only to the number of users and independent of the number of antennas. A typical TDD protocol [13] is shown in Figure 4.4. The part of "pilots" is the most important where the users transmit uplink orthogonal pilots and the BS learns the uplink channels and by reciprocity the downlink channels. Each BS uses this information to demultiplex the uplink data and multiplex the downlink data. According to this protocol, all the users in all the cells first synchronously send uplink data signals to their base station. Next, the users send pilot sequences, BSs use these pilot sequences to estimate CSI to the users located in their cells. Then, BSs use the estimated CSI to detect the uplink data and to generate beamforming vectors for downlink data transmission. Each part in Fig. 4.4 has different duration, depending on the needs of the system.

Thus, is easy after the above analysis for someone to understand why the TDD dominates at massive MIMO systems.
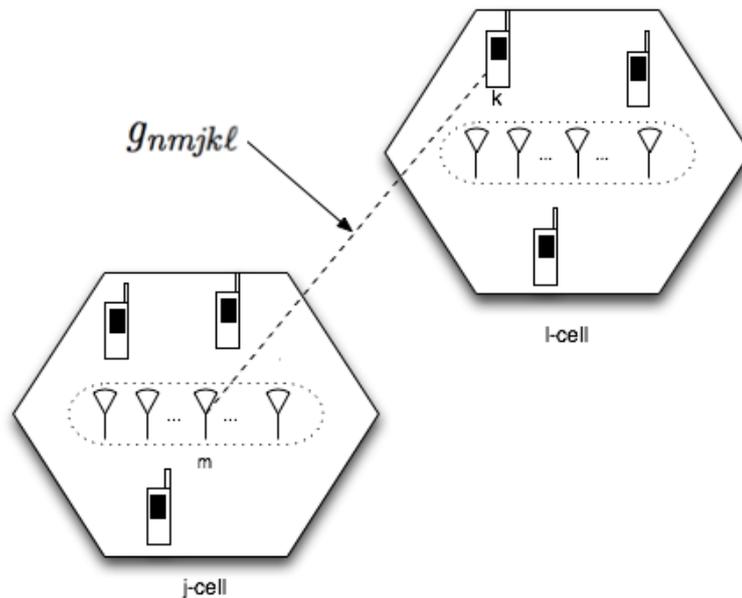
Figure 4.5: The propagation coefficient between k-th terminal in the ℓ-th cell, and the m-th base station antenna of the j-th cell, in the n-th subcarrier, is denoted by $g_{nmjk\ell}$.

4.1.4 Propagation

Because of pilot contamination, there is a complex channel coefficient between each BS antenna in each cell and each terminal in every cell which belongs in the group of same frequency reuse. Using TDD operation, the propagation is the same for either a downlink or an uplink transmission.

As we assume that OFDM is used, we consider a flat-fading channel model for each OFDM subcarrier. As shown in Figure 4.5, we denote the complex propagation coefficient between the m-th base station antenna in the j-th cell, and the k-th terminal in the ℓ-th cell in the n-th subcarrier by g_{nmjkl} which, in turn, is equal to a complex fast fading factor times an amplitude factor that accounts for geometric attenuation and shadow fading,

$$g_{nmjk\ell} = h_{nmjk\ell} \cdot \beta_{jk\ell}^{\frac{1}{2}}, \qquad\qquad (4.2)$$

$$\text{with} \quad n = 1, \ldots, N_{FFT},$$
$$m = 1, \ldots, M,$$
$$j = 1, \ldots, L,$$
$$k = 1, \ldots, K,$$
$$\ell = 1, \ldots, L,$$

where N_{FFT} is the number of subcarriers, M is the number of base station antennas in each cell, L is the number of active cells (i.e., re-using the same band of frequencies), and K is the number of terminals in each cell. The fast fading coefficients, $h_{nmjk\ell} \sim \mathcal{CN}(0,1)$, are assumed to be a complex normal random variables with zero-mean and unit-variance. With respect to the frequency index, n, the fast fading is assumed to be piecewise-constant over N_{Smooth} successive subcarriers, where N_{Smooth} is the frequency smoothness interval (4.1). Only one pilot symbol per smoothness interval is required. The second factor in (4.2) is assumed constant with respect to both frequency and with respect to the index of the base station antenna since the geometric and shadow fading change slowly over space, and it factors as follows:

$$\beta_{jk\ell} = \frac{z_{jk\ell}}{r_{jk\ell}^{\gamma}}. \qquad\qquad (4.3)$$

Here, $r_{jk\ell}$ is the distance between the k-th terminal in the ℓ-th cell and the base station in the j-th cell, γ is the decay exponent, and $z_{jk\ell}$ is a log-normal random variable, i.e., the quantity $10\log_{10}(z_{jk\ell})$ is distributed zero-mean Gaussian with standard deviation σ_{shad}. The shadow fading $z_{jk\ell}$ is statistically independent over all three indices. The ranges $r_{jk\ell}$ are statistically independent over k and ℓ, but statistically dependent over j, because the only randomness that affects $r_{j1k\ell}$ and $r_{j2k\ell}$ is the position of the k-th terminal in the ℓ-th cell.

Throughout we assume that both the terminals and the base station are ignorant of the propagation coefficients.

4.2   Reverse-Link Pilots

As we mentioned before, CSI plays a key role in the above MIMO system, as the knowledge of the channel is required in the base station either for the forward-link data transmission or the reverse-link data transmission. In channel estimation phase, terminals transmit a known sequence of symbols (pilots) and the base station evaluates the effect of the channel on these symbols. Our analysis assumes that exactly the same set of pilot sequences is used in all active cells, and that each cell serves the maximum possible number of terminals. We assume that a total of τ OFDM symbols are pilots and the remainder of the coherence interval is used for transmitting data.

4.2.1   Maximum number of terminals

We note again that the number of terminals that can be served in each cell is limited by the coherence time of the channel and independent of the size of the cell and the number of BS antennas. So if the channel response changed arbitrarily fast with frequency, then, over τ OFDM symbols, the base station could only learn the channel for τ terminals in its cell. In general, the channel response is constant over N_{smooth} consecutive subcarriers and the base station can learn the channel for a total of $K_{max} = \tau N_{Smooth}$ terminals. This number has a simple interpretation in the time domain. The guard

interval T_g is chosen to be greater than the largest possible delay-spread, T_d (we assume that $T_d = T_g$). Then, according to (4.1), the maximum number of terminals is

$$K_{MAX} = \tau N_{Smooth}$$
$$= \frac{\tau}{T_d \Delta_f} = \frac{(\tau T_s) T_u}{T_d T_s} = \left(\frac{T_{pilot}}{T_d}\right)\left(\frac{T_u}{T_s}\right), \qquad (4.4)$$

where $T_{pilot} = \tau T_s$ is the time spent on sending reverse pilots. Training could be accomplished directly in the time-domain (i.e., without OFDM) by transmitting impulses from a succession of different terminals spaced by the delay-spread. The factor $\frac{T_u}{T_s}$ reflects the inefficiency of OFDM due to the cyclic prefix. Another interpretation for the frequency smoothness interval is that the quantity $N_{Smooth}\Delta_f$ is the Nyquist sampling interval in frequency for the time-limited channel impulse response.

The simplest way to send reverse-link pilots would be to assign each terminal one unique time-frequency index for its pilot (e.g., one subcarrier within each smoothness interval and within one OFDM symbol).

4.2.2 Pilot Contamination

We assume that the number of cells that are using the same band of frequencies is L and, as we said, the total number of terminals is the same for each cell. Furthermore, we assume synchronized transmissions and reception (synchronized transmission constitutes a worse-case scenario from the standpoint of pilot contamination). So, when the ℓ-th cell estimates the channel of its k-th terminal by its pilot sequence, it also decodes the signals of L terminals from the other cells which are using the same pilot sequence. This results in the base station to obtain a channel estimate for its k-th terminal that is contaminated by a linear combination of channels to the other terminals (in other cells) that share the same pilot sequence.

Let $\hat{G}_{jj}$ denote the estimate for the $M \times K$ propagation matrix between the M base station antennas of the j-th cell, and the K terminals in the j-th cell; for notational simplicity we suppress the dependence of $\hat{G}_{jj}$ in the sub-carrier index:

$$\hat{G}_{jj} = \sqrt{\rho_p} \sum_{n=1}^{L} G_{j\ell} + V_j, \qquad (4.5)$$

where $G_{j\ell}$ is the $M \times K$ propagation matrix between the K terminals in the ℓ-th cell and the M base station antennas in the j-th cell,

$$[G_{j\ell}] = g_{nmjk\ell}, m = 1, \dots, M, k = 1, \dots, K. \qquad (4.6)$$

V_j is $M \times K$ matrix of receiver noise whose components are zero-mean, mutually uncorrelated, and uncorrelated with the propagation matrices, and ρ_p is a measure of pilot signal- to-noise ratio. We need not quantify ρ_p because, as M grows without limit, the effects of the noise vanish.

4.3 Frequency-reuse Factor

In an effort to mitigate interference, it was proposed the separation of the available spectrum in different parts of frequencies, which each part being used by different clusters of cells. This method is known as frequency reuse. Thus the number of cells that interfere is reduced to L/a_{factor} but the same applies to the available spectrum of each terminal, B/a_{factor}. By using a frequency reuse factor of $a_{factor} = 3$ and $a_{factor} = 7$, clustering of cells are

(a) $a_{factor} = 3$ (b) $a_{factor} = 7$
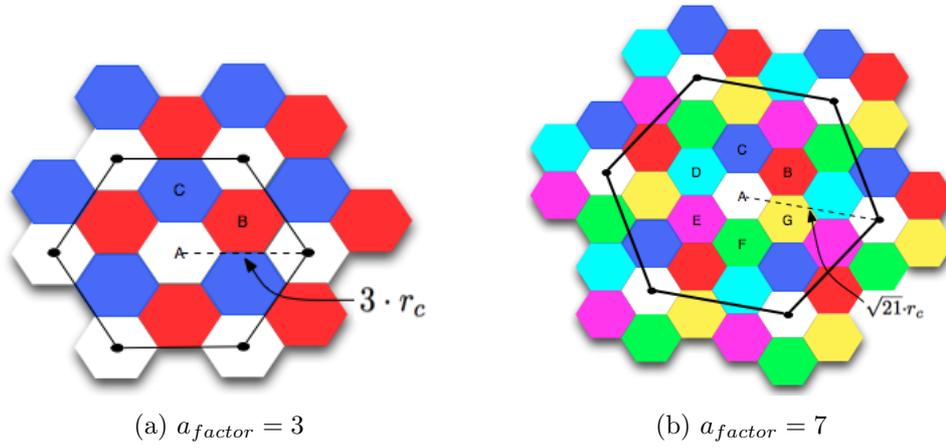
Figure 4.6: Frequency reuse models.

formed like in Figure 4.6. The advantage of these formations is that the cells of a cluster are not adjacent, the distances between them are relatively large and, therefore, in accordance with (4.3), the power of interfering channel diminishes.

4.4   Reverse-Link Data Transmission

In uplink operation, the K terminals in each cell independently transmit data streams to their respective base station. The base station uses its channel estimate to perform maximum-ratio combining and receives the message from its own K terminals, as shown Figure 4.8. The problem is that, due to pilot contamination, by using the estimated channel (4.5) each base station also receives the signals from the terminals in the other cells (Figure 4.7). So, as analysed in the next subsection, the receiving signal suffers of interference from the terminals in the $L-1$ cells which share the same frequency band.
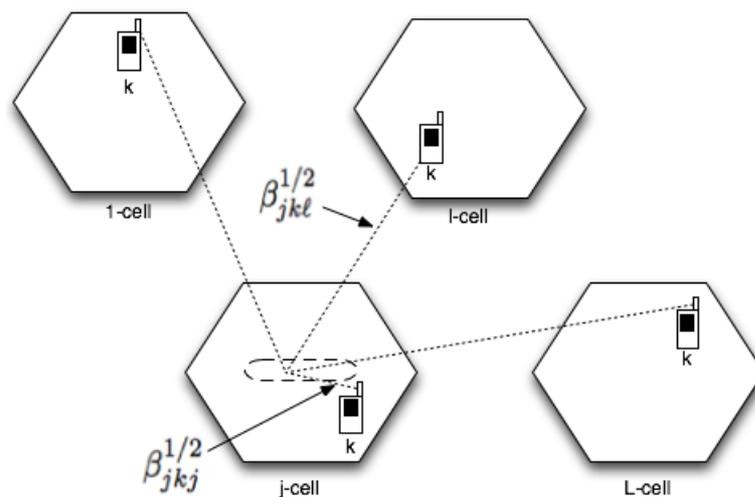
Figure 4.7: Reverse-Link interference due to pilot contamination for an unlimited number of base station antennas: transmissions from terminals in other cells who use the same pilot sequence interfere with the transmission from k-th terminal in the j-th cell to his own base station.

4.4.1 Signal Model

The j-th base station receives, within each sub-carrier, and within each OFDM symbol, a $M \times 1$ vector comprising transmissions from all of the terminals in the L cells. Again, we suppress the dependence on the sub-carrier index,

$$\overline{x}_j = \sqrt{\rho_r} \sum_{\ell=1}^{L} G_{j\ell} \overline{a}_\ell + \overline{w}_j, \tag{4.7}$$

where $\overline{a}_\ell$ is the $M \times 1$ vector of message-bearing symbols from the terminals of the ℓ-th cell, $\overline{w}_j$ is a vector of receiver noise whose components are zero-mean, mutually uncorrelated, and uncorrelated with the propagation matrices, and ρ_r is a measure of signal-to-noise ratio. In the subsequent analysis, we assume that the message-bearing signals which are transmitted by the terminals are independent and distributed as zero-mean, unit-variance, complex Gaussian.
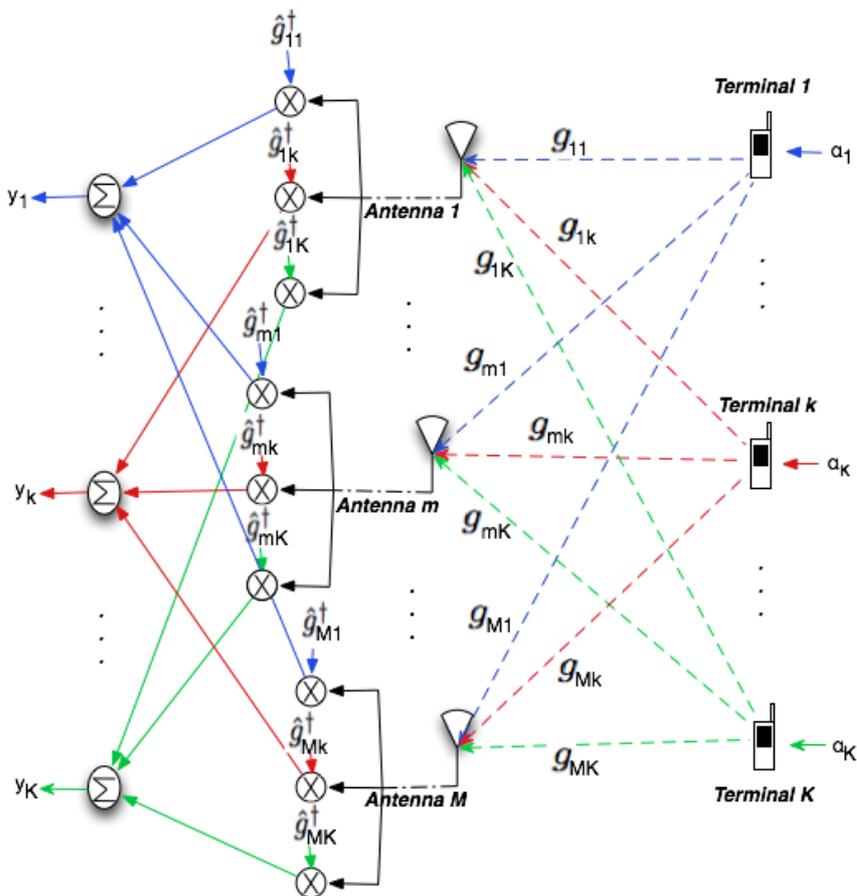
Figure 4.8: Reverse-Link operation.

4.4.2 Maximum-ratio combining

The base station processes its received signal by multiplying it by the conjugate-transpose of the channel estimate which, according to (4.5) and (4.7), yields

$$\overline{y}_j = \hat{G}_{jj}^\dagger \overline{x}_j$$
$$= \left[\sqrt{\rho_p} \sum_{\ell_1=1}^{L} G_{j\ell_1} + V_j \right]^\dagger \left[\sqrt{\rho_r} \sum_{\ell_2=1}^{L} G_{j\ell_2} \overline{a}_{\ell_2} + \overline{w}_j \right], \tag{4.8}$$

where the superscript "$\dagger$" denotes "conjugate transpose". The components of $\overline{y}_j$ comprise sums of inner products between M-component random vectors.

(a) M=100

(b) M=500
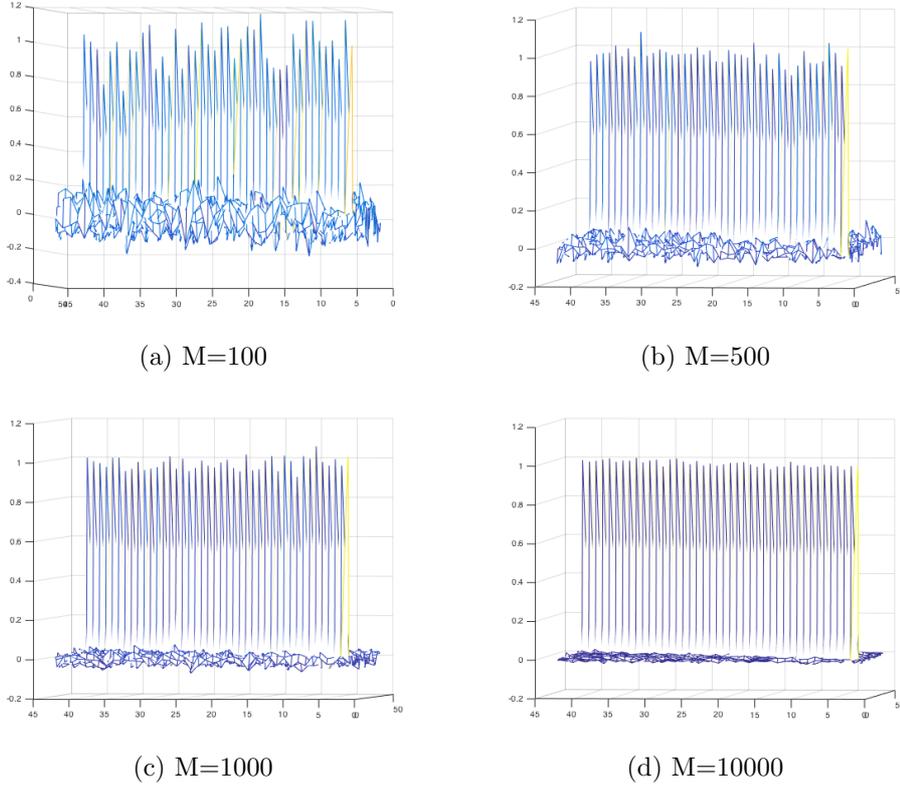
(c) M=1000

(d) M=10000

Figure 4.9: Simulations for relation 4.10 as antennas in each base station grows from 100 to 10000. As we see, as the number of antenna grows the elements except the diagonal are getting closer to zero.

As M grows without limit, the L_2-norms of these vectors grow proportional to M, while the inner products of uncorrelated vectors, by assumption, grow at a smaller rate. For large M, only the products of identical quantities remain significant, i.e., the propagation matrices which appear in both of the bracketed expressions. According to (4.2) and (4.6),

$$\frac{1}{M}G_{j\ell_1}^\dagger G_{j\ell_2} = D_{\overline{\beta}_{j\ell_1}}^{\frac{1}{2}} \left(\frac{H_{j\ell_1}^\dagger H_{j\ell_2}}{M}\right) D_{\overline{\beta}_{j\ell_2}}^{\frac{1}{2}}, \tag{4.9}$$

where $H_{j\ell}$ is the $M \times K$ matrix of fast fading coefficients between the K terminals of the ℓ-th cell, and the M antennas of the j-th base station, $[H_{j\ell}]_{mk} = h_{nmjkl}$, and $D_{\overline{\beta}_{j\ell}}$ is a $K \times K$ diagonal matrix whose diagonal elements comprise the vector $[\overline{\beta}_{j\ell}]_k = \beta_{jk\ell}, k = 1, \ldots, K$. As M grows without bound we have

$$\frac{1}{M}H_{j\ell_1}^\dagger H_{j\ell_2} \to I_K \delta_{\ell_1\ell_2}, \tag{4.10}$$

where I_K is the $K \times K$ identity matrix. The substitution of (4.10) and (4.9) into (4.8) yields

$$\frac{1}{M\sqrt{\rho_p\rho_r}}\overline{y}_j \to \sum_{\ell=1}^{L} D_{\overline{\beta}_{j\ell}}\overline{a}_\ell . \tag{4.11}$$

The k-th component of the processed signal becomes

$$\frac{1}{M\sqrt{\rho_p\rho_r}}y_{kj} \to \beta_{jkj}a_{kj} + \sum_{\ell\neq j}\beta_{jk\ell}a_{k\ell}. \tag{4.12}$$

The salutary effect of using an unlimited number of base station antennas is that the effects of uncorrelated receiver noise and fast fading are eliminated

completely, and transmissions from terminals within one's own cell do not interfere. However transmission from terminals in other cells that use the same pilot sequence constitute a residual interference. The effective signal-to-interference ratio (SIR), which is identical for all sub-carriers but which depends on the indices of the cell and the terminal, is

$$SIR_{rk} = \frac{\beta_{jkj}^2}{\sum_{\ell \neq j} \beta_{jk\ell}^2}. \tag{4.13}$$

The effective signal-to-interference ratio is a random quantity which depends exclusively on the random positions of the terminals. Because according to (4.3) the β's are only depended of distance and shadow fading coefficients.

Note that the SIR expression (4.13) is independent of the quantities ρ_p and ρ_r, and therefore it is independent of the transmitted powers. This is intuitively reasonable: we are operating in a regime where performance is limited only by inter-cell interference, so if every terminal reduces its power by the same factor then the limit SIR is unchanged. Hence we conclude that for an arbitrarily small transmitted energy-per-bit, the SIR (4.13) can be approached arbitrarily closely by employing a sufficient number of antennas.

A curious thing about the effective SIR is its dependence on the *squares* of the β's. This occurs because the system is operating in a purely interference-limited rather than a noise-limited regime and because of the particular processing which is employed. Prior to maximum ratio combining, the desired signal and the inter-cellular interference are both proportional to the square-roots of their respective β's, while the receiver noise has unit-variance. After maximum-ratio combining, the desired signal and the interference are both proportional to their respective β's, while the noise has standard deviation proportional to the sum of the square-roots of the β's. If the noise were the dominant impairment then the SNR would be the ratio of the β^2 of the desired signal to β of the desired signal, or $SNR \propto \beta$. But interference is the dominant impairment, so the SIR is proportional to a ratio of squares of β's.

The SIR (4.13) is constant with respect to frequency because the slow-fading coefficients are independent of frequency. The SIR is constant with respect to the absolute size of the cell, for the following reason. Each of the β-terms is inversely proportional to a range that is raised to the decay exponent, $\beta \propto \frac{1}{r^\gamma}$. The replacement of the range by the nondimensional quantity, $r \to \frac{r}{r_c}$, does not alter the value of the SIR because the terms r_c^γ appear in both the numerator and denominator and therefore cancel. Consequently the throughput per terminal and the number of terminals which the base station can handle is independent of the cell-size.

4.4.3 Reverse-Link Capacity

Subject to the assumption that the terminals transmit Gaussian message-bearing symbols, the instantaneous capacity of the terminal within each subcarrier is equal to the logarithm of one plus the signal-to-interference ratio. The net throughput per terminal, in units of bits/sec/terminal, accounts for the total bandwidth and frequency re-use, the pilot overhead (the ratio of the time spent sending data to the total slot-length), and the overhead of the cyclic prefix:

$$C_{r_k} = \left(\frac{B}{\alpha}\right) \left(\frac{T_{slot} - T_{pilot}}{T_{slot}}\right) \left(\frac{T_u}{T_s}\right) \log_2 (1 + SIR_{r_k}), \tag{4.14}$$

where B is the total bandwidth in Hz, α is the frequency re-use factor (equal to either one, three, or seven in our subsequent analysis), T_{slot} is the slot

length, T_{pilot} is the time spent transmitting reverse-link pilots, T_u is the useful symbol duration, and T_s is the OFDM symbol interval, where the times are measured in seconds.

The net sum throughput per cell, measured in bits/sec/cell, is equal to the sum of the net throughputs per terminal.

$$C_{r_{sum}} = \sum_{k=1}^{K} C_{r_k}. \tag{4.15}$$

Since the number of terminals that can be served is proportional to the time spent sending pilots, while the instantaneous sum-throughput is proportional to the number of terminals served, it follows that net sum-throughput is maximized by spending approximately half of the slot on sending pilots, and half sending data.

4.5   Forward-Link Data Transmission
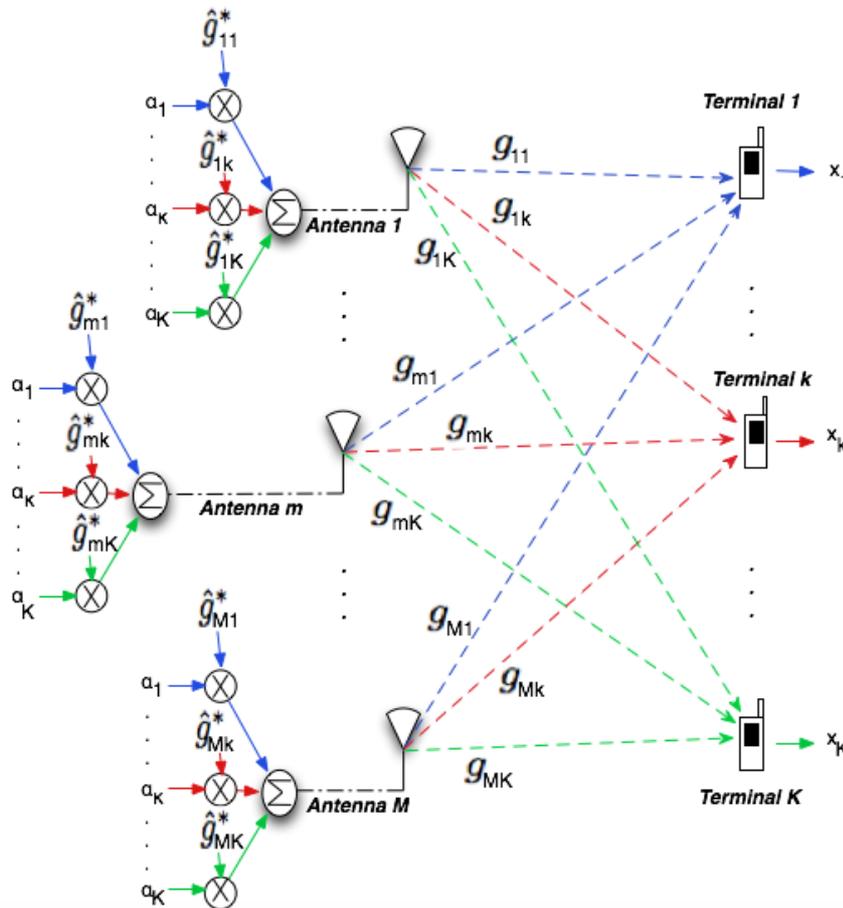
Figure 4.10: Forward-Link operation.

The simplest sort of massive MIMO downlink entails conjugate beamforming which works very well and can be very effective. In each cell, the base station transmits data in its own terminals. It provides each antenna with QAM symbols which intended for each of the users. So, let a_1 be the QAM symbol which are intended for terminal T_1, $antenna_1$ multiplies the symbol a_1 by the conjugate of its estimated channel to user one, each of the other antennas does the same thing. This happens for each terminal, so as shown in Figure 4.10 the final transmitted signal of each antenna is the sum of the above process. Thus, we achieve that all of the signals entailing a_1 will arrive in phase to user one and the amplitude beamforming gain grows with M

(number of the antennas), and they tend to arrive out of phase at the other users. Of course, this applies for each user.

The problem is that, as shown in Figure 4.11, the transmission from the base station in the ℓ-th cell to its k-th terminal suffers interference from transmissions from the base stations in other cells to their own k-th terminals due to pilot contamination. So the k-th terminal in the ℓ-th cell will also receive the signals from all the other $L-1$ cells which are intended for their own terminals with the same pilot sequence.
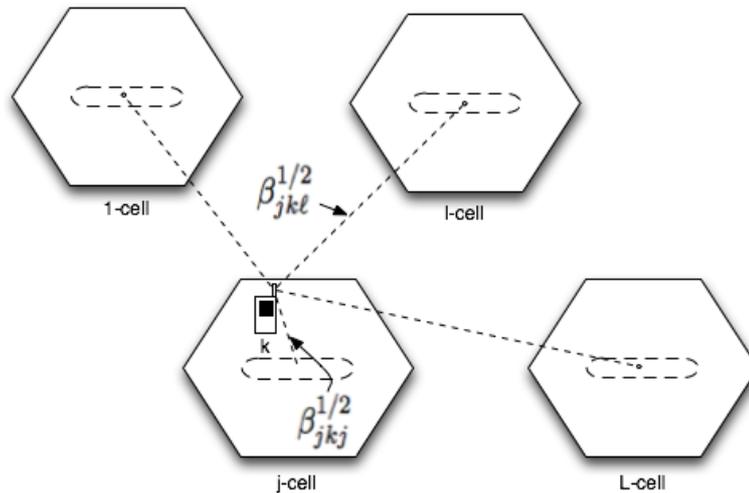
Figure 4.11: Forward-Link interference due to pilot contamination for an unlimited number of base station antennas: transmissions from base stations in other cells intended for their own k-th terminal interfere with the transmission from the base station in the j-th cell to his k-th terminal.

4.5.1 Pre-coding Matrix

The j-th base station transmits a $M \times 1$ vector, $\hat{G}_{jj}^* \bar{a}_j$, where the superscript "*" denotes "complex conjugate", and $\bar{a}_j$ is the vector of message-bearing signals which is intended for the K terminals of the j-th cell. In practice a normalizing factor would be included in order to conform to power constraints. We merely assume that this normalizing factor is the same for all base stations. As M grows without limit, the exact value of the normalizing factor is unimportant.

4.5.2 Signal Model

The K terminals in the ℓ-th cell receive their respective components of a $K \times 1$ vector comprising transmissions from all L base stations.

$$\bar{x}_\ell = \sqrt{\rho_f} \sum_{n=1}^{L} G_{j\ell}^T \hat{G}_{jj}^* \bar{a}_j + \bar{w}_\ell \tag{4.16}$$

$$= \sqrt{\rho_f} \sum_{n=1}^{L} G_{j\ell}^T \left[\sqrt{\rho_p} \sum_{n=1}^{L} G_{j\ell} + V_j \right]^* \bar{a}_j + \bar{w}_\ell, \tag{4.17}$$

where $\bar{w}_\ell$ is uncorrelated noise, $\sqrt{\rho_f}$ is a measure of the forward signal-to-noise ratio, the superscript "T" denotes "unconjugated transpose", and we have utilized (4.5).

We now let the number of base station antennas increase without limit, and again we utilize (4.9) and (4.10) to conclude that

$$\frac{1}{M\sqrt{\rho_p\rho_f}}\overline{x_\ell} \to \sum_{j=1}^{L} D_{\beta_{j\ell}}\overline{a}_j. \tag{4.18}$$

The k-th terminal in the ℓ-th cell receives the following:

$$\frac{1}{M\sqrt{\rho_p\rho_f}}x_{k\ell} \to \beta_{\ell k\ell}a_{k\ell} + \sum_{j\neq\ell}\beta_{jk\ell}a_{jk}. \tag{4.19}$$

The effective signal-to-interference ratio is

$$SIR_{fk} = \frac{\beta_{\ell k\ell}^2}{\sum_{j\neq\ell}\beta_{jk\ell}^2}. \tag{4.20}$$

While the forward and the reverse SIRs, (4.20) and (4.13), are described by similar-looking expressions, they in fact have somewhat different statistical characteristics. The numerators have identical statistics. The denominator for the reverse-link SIR (4.13) is a sum of squares of $L-1$ slow fading coefficients from different terminals to the same base station. These coefficients are statistically independent. The denominator for the forward-link SIR (4.20) is a sum of squares of $L-1$ slow fading coefficients from different base stations to the same terminal. These coefficients are correlated because motion of the one terminal affects all of the geometric decay factors.

4.5.3   Forward-link capacity

As in 3.3.3 we translate the forward SIR into the net capacity per terminal (bits/sec/terminal):

$$C_{f_k} = \left(\frac{B}{\alpha}\right)\left(\frac{T_{slot} - T_{pilot}}{T_{slot}}\right)\left(\frac{T_u}{T_s}\right)\log_2\left(1 + SIR_{f_k}\right), \tag{4.21}$$

and the net capacity per cell (bits/sec/cell):

$$C_{f_{sum}} = \sum_{k=1}^{K} C_{f_k}. \tag{4.22}$$

4.6   Simulations

As we saw, SIR constitutes the most important factor in the wireless cellular network that we study, since it determines the throughput of the system. What interests us most is to see how the performance is influenced by the different frequency reuse strategy, considering also the statistical analysis of the SIR. It is easy to assume that less aggressive frequency reuse improves the SIR since, as we have said, reduces the number of cells that interfere and grows the distances between them, but we do not know how exactly this affects the net capacity per terminal, since the available bandwidth subdivided by the respective α factor. Our analysis is based on the calculation of cumulative distribution function (cdf) which helps us to discover the probability to achieve high values of SIR (greater than a specific value). We do the same for the net capacity per terminal. To have a more clear view of the network throughput, we also calculate the mean net capacity per terminal, and the mean net capacity per cell for frequency-reuses of one, three and seven.

Our simulation comprises the evaluation of the signal-to-interference ratios (4.13) and (4.20) for 10^5 independent trials, which translate directly

into distributions for SIRs and capacities. We determine the set of cells that interfere with a particular cell by finding all cells which a) reuse the same frequency band, and b) are within eight cell-diameters of that cell. We assume a coherence time of 500 microseconds (which could accommodate TGV - Train à Grande Vitesse - speeds), of which three symbols are spent sending reverse pilots, and three symbols are spent sending data, either reverse or forward. The remaining one symbol is considered to be additional overhead. The OFDM parameters are identical to LTE (Long- Term Evolution) forward-link parameters: a symbol interval of $T_s = 500/7 \approx 71.4$ microseconds, a subcarrier spacing of $\Delta_f = 15kHz$, a useful symbol duration $T_u = 1/\Delta_f \approx 66.7$ microseconds, and a guard interval $T_g = T_s - T_u \approx 4.76$ microseconds. The basic parameters are given in table below.

Basic Parameters		
Bandwidth	B	20MHz
Cell Radius	r_c	1600
Cell-hole Radius	r_h	100
Frequency Reuse Factor	α	1, 3, 7
Number of Cells	L	343, 109, 49
Number of Subcarriers	N_{smooth}	14
Number of Terminals	K	42
Decay Exponent	γ	3.8
Shadow-fading Standard Deviation	σ_{shadow}	8.0dB

4.6.1   Performance

Figures 4.12 and 4.13 show the cumulative distribution of SIR for the reverse and uplink for frequency reuse factors of one, three, and seven. From the observation of these two figures, we conclude that the statistics of SIR are the same for both links. As we see, the probability to have high SIR is greater as the frequency reuse factor grows, which as we said is reasonable. Based on the above figures it could be argued that the use of greater frequency reuse factor will be beneficial for the cellular systems of massive MIMO operation, but this is not right because, as we will see from the statistical analysis of net capacity per terminal, the choice of the frequency reuse factor depends on the standards that the system wants to meet.
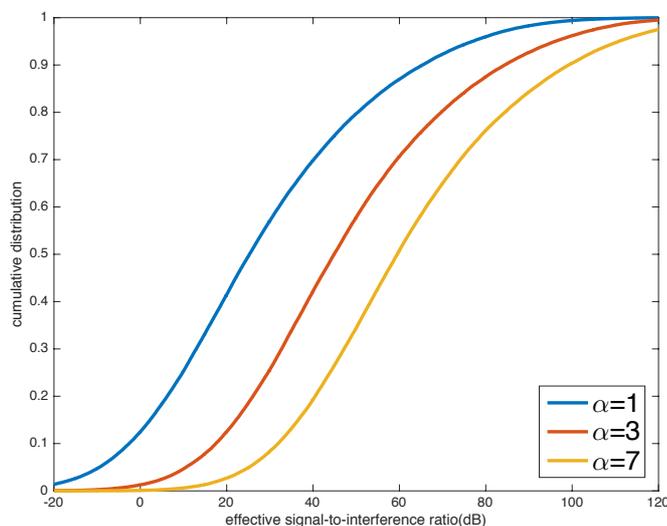

Figure 4.12: Cumulative distribution of the reverse effective SIR(dB).
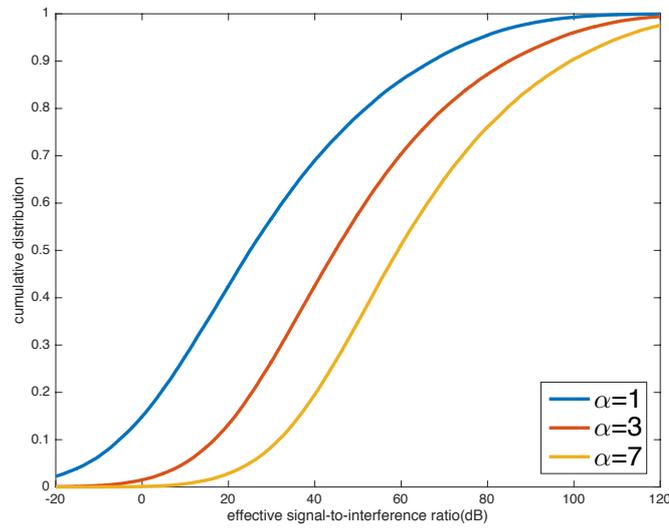
Figure 4.13: Cumulative distribution of the forward effective SIR(dB).
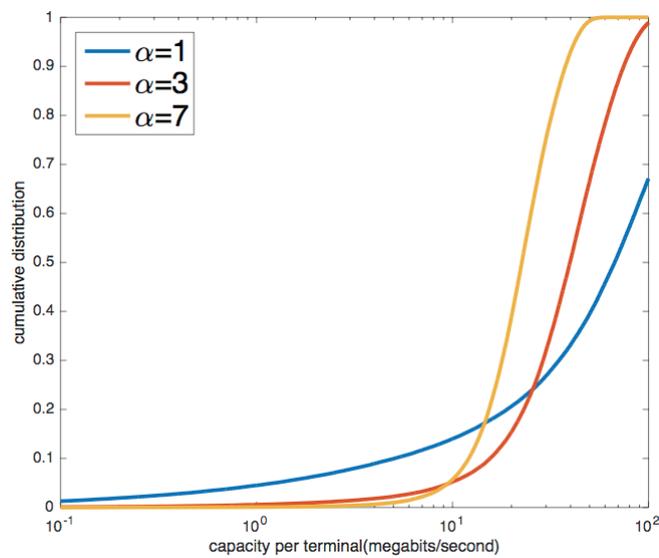
Figure 4.14: Cumulative distribution of the net reverse capacity per terminal(megabits/second).
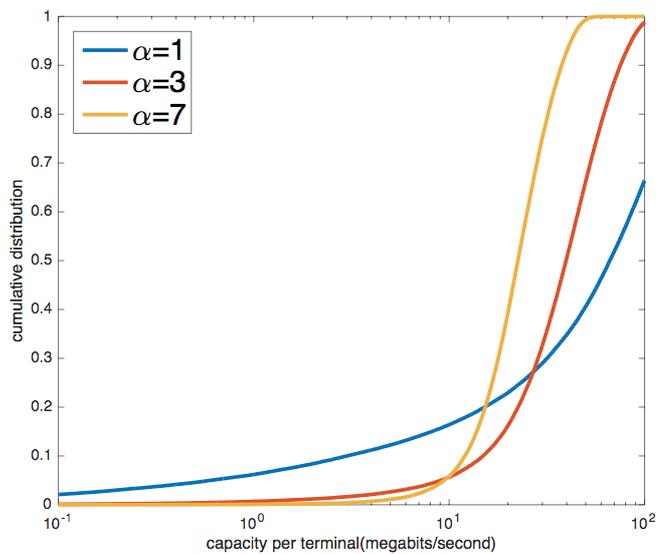

Figure 4.15: Cumulative distribution of the net forward capacity per terminal(megabits/second).

The cumulative distribution of net capacity per terminal for the reverse and uplink data transmission for frequency reuse factor of one, three, and seven are shown in Figures 4.14 and 4.15. At first glance, it does not seem that there is a frequency reuse factor which is clearly the best, because each factor is offering something different. We see that for $\alpha = 7$ the system ensuring a high probability for a good value of capacity but has low ceiling, compared to $\alpha = 1$, which offers the possibility of reaching really high speed (megabits/second).

Table 4.1 shows the mean net capacity per terminal, and per cell for frequency-reuses of one, three, and seven. The mean capacity is identical for reverse and forward link, which is positive, since usually there is big discrepancy between the two links in the existing systems (peaking always on theoretical level) [14].

Freq. Reuse	C_{mean} Per Terminal $(Mbits/s)$	C_{mean} Per Cell $(Mbits/s)$
1	80	3360
3	43	1806
7	24	1008

Table 4.1: Mean net capacity per terminal, and per cell for frequency-reuses of one, three, and seven.

What we conclude from the study of the above figures is that larger reuse factors are beneficial when the SIR is low, the logarithm is in its linear region, and capacity gains due to the large increase in SIR more than offset the loss due to less aggressive frequency reuse which is associated with a reduction in the actual bandwidth that each cell utilizes. When the SIR is already high, a greater frequency reuse factor causes a net decrease in throughput. If the minimum guaranteed performance per terminal is a more important consideration than the mean throughput, then a frequency reuse factor of seven should be used.

Overall, the frequency reuse factor does not seem to be the solution of the pilot contamination problem. The reason is that, by using this method to mitigate the phenomenon of pilot contamination, the system lacks of its big advantage to provide the users all the available spectrum. We said that next generation wireless systems need to achieve rapid speeds of data and, to achieve that, massive MIMO needs to take advantage of the large number of BSs antennas, which opens up the spatial dimension that allows it to discriminate the signals to/from each terminal based on its location and enables each terminal to use all available spectrum resources.

Chapter 5

A Coordinated Approach to Channel Estimation in Large-scale Multiple-antenna Systems

The study of the previous massive MIMO cellular wireless system helped us to understand the benefits of the use of large-scale multiple-antennas in BS and to discover the stumbling of the system, the phenomenon of pilot contamination, which imposes fundamental limitations in what can be achieved with a (non cooperative) massive MIMO system. An interesting idea to deal with this phenomenon is presented in [4], in which a cooperation of cellular wireless system tackles the problem by enabling a low rate coordination between cells during the channel estimation phase. The coordination makes use of the additional second-order statistical information about the user channels, which are shown to offer a powerful way of discriminating across interfering users with even strongly correlated pilot sequences.

As we will see, by using a conventional estimator, like the Least Squares (LS) in the procedure of channel estimation, the mean square error (MSE) between real and estimated channels is fairly high and it is not ameliorated by increasing the number of antennas in the BS. In a massive MIMO system, it is necessary to use estimators which take advantage of the multiple-antennas, so we can get the best results by exploiting every aspect of the system. So with that in mind, researchers have developed in [4] a Bayesian channel estimation method which exploits the dormant side-information lying in the second-order statistics of the users channels. This estimator has been shown to work much better than the LS estimator, and its performance is getting better and better by using more and more BS antennas. More importantly, in this paper the authors demonstrate a powerful result indicating that the exploitation of covariance information under certain subspace conditions on the covariance matrices can lead to a complete removal of pilot contamination effects in the large number of antennas. So, inspired by these findings, they turned their focus to create a practical algorithm to take the best out of this estimation method. The key idea behind the new algorithm is the use of a covariance-aware pilot assignment strategy within the channel estimation phase itself.

More specifically, the authors show that the channel estimation performance is a function of the degree to which dominant signal subspaces pertaining to the desired and interference channel covariance overlap with each other. Therefore, they exploit the fact that the desired user signals and interfering user signals are received at the base station with (at least approximately) finite-rank covariance matrices. This is typically the case in

realistic scenarios due to the limited angle spread followed by incoming paths originating from street-level users. Then, they suggest a coordinated pilot assignment strategy that, based on covariance matrices, assigns identical pilot sequences to appropriate selected groups of users. The gains of this coordination are shown to depend on typical angle spread measured at the base station and the number of base station antennas. In fact, in some cases, this algorithm approaches the performance of interference-free channel estimation even for moderate numbers of antennas and users.

5.1 Channel Estimation

5.1.1 Scenario

The cellular network for this scenario consists of L time-synchronized cells with full spectrum reuse. In the center of each cell there is a base station equipped with M antennas, which serves K single-antenna users. The number of terminals is the same for every cell. To avoid intra-cell interference in channel estimation phase the pilots that are used by a single-user are mutually orthogonal to the pilots of the remain $K-1$ users which belong to the same cell. However, the same pilots are reused from cell to cell, resulting in pilot contamination from $L-1$ interfering cells.

Pilot Sequence

Each terminal transmits a known pilot sequence of length τ (as we will see later, τ matches with the number of terminals in a single cell) to its own base station which in turn evaluates the effect of the channel on these symbols. This happens simultaneously for all base stations in the cellular system. The pilot sequence used in the ℓ-th cell is denoted by:

$$\mathbf{s}_l = \begin{bmatrix} s_{l1} & s_{l2} & \cdots & s_{l\tau} \end{bmatrix}^T. \tag{5.1}$$

The powers of pilot sequences are assumed equal such that $|s_{l1}|^2 + \cdots + |s_{l\tau}|^2 = \tau$, for $l = 1, 2, \ldots, L$.

Signal and Channel Models

We define the channel vector between the l-th cell user and the target base station as $\mathbf{h}_l$. To simplify the notation, we assume the 1st cell is the target cell (the one in the center), unless otherwise notified. Thus, $\mathbf{h}_1$ is the desired channel while $\mathbf{h}_l$, $l > 1$ are the interference channels. All channel vectors are assumed to be $M \times 1$ complex Gaussian, undergoing correlation due to the finite multipath angle spread at the base station side [15]:

$$\mathbf{h}_l = R_l^{\frac{1}{2}} \mathbf{h}_{Wl}, l = 1, 2 \ldots, L, \tag{5.2}$$

where $\mathbf{h}_{Wl} \sim \mathcal{CN}(0, \mathbf{I}_M)$ is the spatially white $M \times 1$ SIMO channel, and $\mathcal{CN}(0, \mathbf{I}_M)$ denotes zero-mean complex Gaussian distribution with covariance matrix $\mathbf{I}_M$. We make the assumption that covariance matrix $R_l \triangleq \mathbb{E}\{\mathbf{h}_l\mathbf{h}_l^H\}$ can be obtained separately for the desired and interference channels. This could be done in practice by exploiting resource blocks where the desired user and interference users are known to be assigned at different times. In future networks, one may imagine a specific training design for learning second-order statistics. Since covariance information varies much slower than fast fading, such training may not consume a substantial amount of resources.

So, during the pilot phase, the $M \times \tau$ signal received at the target base station is

$$\mathbf{Y} = \sum_{l=1}^{L} \mathbf{h}_l \mathbf{s}_l^T + \mathbf{N}, \tag{5.3}$$

where $\mathbf{N} \in \mathbb{C}^{M \times \tau}$ is the spatially and temporally white additive Gaussian noise (AWGN) with zero-mean and element-wise variance σ_n^2.

5.1.2   Least Squares Estimation

The conventional channel estimation method of LS relies on correlating the received signal with the known pilot sequence. Hence, using the model in (5.3), a LS estimate for the desired channel $\mathbf{h}_1$ is

$$\widehat{\mathbf{h}}_1^{LS} = \mathbf{Y}\mathbf{s}_1^*(\mathbf{s}_1^T \mathbf{s}_1^*)^{-1}. \tag{5.4}$$

As we have said, by using the same pilot sequences in the L cells, the conventional estimators, such as LS, suffer from lack of orthogonality between the desired and interfering pilots. So, due to pilot contamination the estimator can be written as

$$\widehat{\mathbf{h}}_1^{LS} = \mathbf{h}_1 + \sum_{l \neq 1}^{L} \mathbf{h}_l + \mathbf{N}\mathbf{s}^*/\tau. \tag{5.5}$$

As we see, the interfering channels leak directly into the desired channel. As a result, the estimation is then limited by SIR. As we saw in the simulations of Marzetta's scenario, these estimation methods limit the ability to design effective interference-avoiding beamforming solution. So, we will need to use something more efficient.

5.1.3   Bayesian Estimation

This is an improved channel estimation method which, by taking advantage of the multiple antenna dimensions and informations lying in the second order statistics of the channel vectors, achieves the reduction of pilot contamination effect. The role of covariance matrices is to capture structure information related to the distribution (mainly mean and spread) of the multipath angles of arrival at the base station. Due to the typically elevated position of the base station, rays impinge on the antennas with a finite angle-of-arrival (AOA) spread and a user location dependent mean angle. Assuming that the antennas of a base station are placed at the highest point in the region which they want to serve, the angle of arrival is between $[-\pi, 0]$.

Bayesian estimator is not a novel idea itself, as it has been presented before in MIMO systems, but, in the paper that we present, the researchers have turned their focus on how the Bayesian estimator works in massive MIMO, and they study the limiting behavior of covariance-based estimates in the presence of interference and large-scale antenna arrays, and how to shape covariance information for the full benefit of channel estimation quality. Subsequently, we define the Bayesian estimator for our system that we will use to implement the algorithm which tackles the pilot contamination effect.

By vectorizing the received signal and noise, our model (5.3) can be represented as

$$\mathbf{y} = \tilde{\mathbf{S}}\mathbf{h} + \mathbf{n}, \tag{5.6}$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{n} = \text{vec}(\mathbf{N})$, and $\mathbf{h} \in \mathbb{C}^{LM \times 1}$ is obtained by stacking all L channels into a vector. The pilot matrix $\tilde{\mathbf{S}}$ is defined as

$$\tilde{\mathbf{S}} \triangleq \begin{bmatrix} \mathbf{s}_1 \otimes \mathbf{I}_M & \cdots & \mathbf{s}_L \otimes \mathbf{I}_M \end{bmatrix}. \tag{5.7}$$

Applying Bayes' rule, the conditional distribution of the channels $\mathbf{h}$, given the received training signal $\mathbf{y}$, is

$$p(\mathbf{h}|\mathbf{y}) = \frac{p(\mathbf{h})p(\mathbf{y}|\mathbf{h})}{p(\mathbf{y})} \propto p(\mathbf{h})p(\mathbf{y}|\mathbf{h}). \tag{5.8}$$

We use the multivariate Gaussian probability density function (PDF) of the random vector $\mathbf{h}$ and assume its rows $\mathbf{h}_1, \ldots, \mathbf{h}_L$ are mutually independent, giving the joint PDF:

$$p(\mathbf{h}) = \frac{\exp\left(-\sum_{l=1}^{L} \mathbf{h}_l^H \mathbf{R}_l^{-1} \mathbf{h}_l\right)}{\pi^{LM}(\det\mathbf{R}_1 \ldots \det\mathbf{R}_L)^M}. \tag{5.9}$$

Note that we derive this Bayesian estimator under the standard condition of covariance matrix invertibility, although we show later this hypothesis is actually challenged by reality in the large-number-of-antennas regime. Fortunately, our final expressions for channel estimators completely skip the covariance inversion.

Using (5.6), we obtain:

$$p(\mathbf{y}|\mathbf{h}) = \frac{\exp\left(-(\mathbf{y} - \tilde{\mathbf{S}}\mathbf{h})^H(\mathbf{y} - \tilde{\mathbf{S}}\mathbf{h})/\sigma_n^2\right)}{(\pi\sigma_n^2)^{M\tau}}. \tag{5.10}$$

Combining the equations (5.9) and (5.10), (5.8) can be rewritten as

$$p(\mathbf{y}|\mathbf{h}) = \exp(-l(\mathbf{h}))AB, \tag{5.11}$$

where $A \triangleq (\pi\sigma_n^2)^{M\tau}$, $B \triangleq \pi^{LM}(\det\mathbf{R}_1 \ldots \det\mathbf{R}_L)^M = \pi^{LM}(\det\mathbf{R})^M$, and

$$l(\mathbf{h}) \triangleq \mathbf{h}^H \bar{\mathbf{R}} \mathbf{h} + (\mathbf{y} - \tilde{\mathbf{S}}\mathbf{h})^H(\mathbf{y} - \tilde{\mathbf{S}}\mathbf{h})/\sigma_n^2, \tag{5.12}$$

in which $\mathbf{R} \triangleq \mathrm{diag}(\mathbf{R}_1, \ldots, \mathbf{R}_L), \bar{\mathbf{R}} \triangleq \mathbf{R}^{-1}$.

By using the maximum a posteriori (MAP) decision rule, the Bayesian estimator yields the most probable value given the observation $\mathbf{y}$:

$$\begin{aligned}
\widehat{\mathbf{h}} &= \arg\max_{\mathbf{h} \in \mathbb{C}^{LM \times 1}} p(\mathbf{h}|\mathbf{y}) \\
&= \arg\min_{\mathbf{h} \in \mathbb{C}^{LM \times 1}} l(\mathbf{h}) \\
&= (\sigma_n^2 \mathbf{I}_{LM} + \mathbf{R}\tilde{\mathbf{S}}^H\tilde{\mathbf{S}})^{-1}\mathbf{R}\tilde{\mathbf{S}}^H\mathbf{y}.
\end{aligned} \tag{5.13}$$

Interestingly, the Bayesian estimate as shown in (5.13) coincides with the minimum mean square error (MMSE) estimate, which has the form

$$\widehat{\mathbf{h}}^{\mathrm{MMSE}} = \mathbf{R}\tilde{\mathbf{S}}^H(\tilde{\mathbf{S}}\mathbf{R}\tilde{\mathbf{S}}^H + \sigma_n^2\mathbf{I}_{\tau M})^{-1}\mathbf{y}. \tag{5.14}$$

Expressions (5.13) and (5.14) are equivalent thanks to the matrix inversion identity $(\mathbf{I} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}$.

The previous expression describes the simultaneous estimation of interfering channels and desired channel. In our analysis, by using matched filters, we require only the knowledge of the desired channel, and consider the interference channels as nuisance parametres. So, the single user channel estimation in shown below. The pilot sequence reused in all L cells is considered:

$$\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \cdots & s_\tau \end{bmatrix}^T. \tag{5.15}$$

Similar to (5.7), we define a training matrix $\bar{\mathbf{S}} \triangleq \mathbf{s} \otimes \mathbf{I}_M$. Note that $\bar{\mathbf{S}}^H\bar{\mathbf{S}} = \tau\mathbf{I}_M$. Then the vectorized received training signal at the target base station can be expressed as

$$\mathbf{y} = \bar{\mathbf{S}}\sum_{l=1}^{L}\mathbf{h}_l + \mathbf{n}. \tag{5.16}$$

Since the Bayesian and the MMSE estimators are identical, we omit the derivation and simply give the expression of this estimator for the desired channel $\mathbf{h}_1$ only:

$$\widehat{\mathbf{h}}_1 = \mathbf{R}_1 \bar{\mathbf{S}}^H \left(\bar{\mathbf{S}} \left(\sum_{l=1}^{L} \mathbf{R}_l \right) \hat{\mathbf{S}}^H + \sigma_n^2 \mathbf{I}_{\tau M} \right)^{-1} \mathbf{y} \tag{5.17}$$

$$= \mathbf{R}_1 \left(\sigma_n^2 \mathbf{I}_M + \tau \sum_{l=1}^{L} \mathbf{R}_l \right)^{-1} \bar{\mathbf{S}}^H \mathbf{y}. \tag{5.18}$$

In the following, we examine the degradation caused by the pilot contamination on the estimation performance. In particular, we point out the role played by the use of covariance matrices in dramatically reducing the pilot contamination effects under certain conditions on the rank structure.

We are interested in the mean squared error (MSE) of the proposed estimators, which can be defined as: $\mathcal{M} \triangleq \mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|_F^2\}$, or for the single user channel estimate $\mathcal{M}_1 \triangleq \mathbb{E}\{\|\hat{\mathbf{h}}_1 - \mathbf{h}_1\|_F^2\}$.

The MSE of (5.13) is

$$\mathcal{M} = \mathrm{tr} \left\{ \mathbf{R} \left(\mathbf{I}_{LM} + \frac{\tilde{\mathbf{S}}^H \tilde{\mathbf{S}}}{\sigma_n^2} \mathbf{R} \right)^{-1} \right\}. \tag{5.19}$$

Specifically, when identical pilots are used in all cells, the MSEs are

$$\mathcal{M} = \mathrm{tr} \left\{ \mathbf{R} \left(\mathbf{I}_{LM} + \frac{\tau \mathbf{J}_{LL} \otimes \mathbf{I}_M}{\sigma_n^2} \mathbf{R} \right)^{-1} \right\}, \tag{5.20}$$

$$\mathcal{M}_1 = \mathrm{tr} \left\{ \mathbf{R}_1 - \mathbf{R}_1^2 \left(\frac{\sigma_n^2}{\tau} \mathbf{I}_M + \sum_{l=1}^{L} \mathbf{R}_l \right)^{-1} \right\}, \tag{5.21}$$

where $\mathbf{J}_{LL}$ is a $L \times L$ unit matrix consisting of all 1s. Of course, it is clear from (5.20) and (5.21) that the MSE is not dependent on the specific design of the pilot sequence, but on its power. We can readily obtain the channel estimate of (5.18) in an interference-free scenario, by setting interference terms to zero:

$$\widehat{\mathbf{h}}_1^{\text{no int}} = \mathbf{R}_1 (\sigma_n^2 \mathbf{I}_M + \tau \mathbf{R}_1)^{-1} \bar{\mathbf{S}}^H (\bar{\mathbf{S}} \mathbf{h}_1 + \mathbf{n}), \tag{5.22}$$

where the superscript "no int" refers to the "no interference case", and the corresponding MSE:

$$\mathcal{M}_1^{\text{no int}} = \mathrm{tr} \left\{ \mathbf{R}_1 \left(\mathbf{I}_M + \frac{\tau}{\sigma_n^2} \mathbf{R}_1 \right)^{-1} \right\}. \tag{5.23}$$

5.2   Coordinated Pilot Assignment

Now we will try to build step-by-step the functionality on which rests the success of the algorithm.

5.2.1   Angle-of-Arrival and Channel model

We said that in every Base Station there exists a large number of antennas. These are smart antennas positioned uniformly in linear array with supercritical antenna spacing (less than or equal to half wavelength). Smart Antennas are phased array antennas with smart signal processing algorithms used to identify the angle of arrival (AOA) of the signal. AOA estimation is the process of determining the direction of an incoming signal from mobile
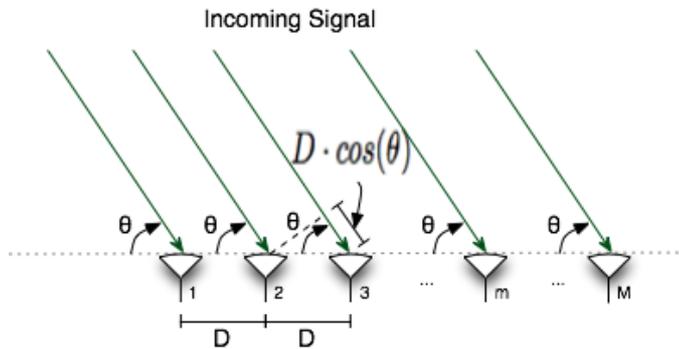
Figure 5.1: Angle of Arrival.

devices to the Base Transceiver Station. In this process, we determine the time "phase" difference of arrival (TDOA) at individual elements of the antenna array, as shown in Figure 5.1, and, from these delays, the angle (or direction) of the mobile devices can be calculated.

Hence, the multipath channel model for the large scale analysis is expressed as

$$\mathbf{h}_i = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} \mathbf{a}(\theta_{ip})\alpha_{ip}, \tag{5.24}$$

where P is the number of i.i.d. paths, $\alpha_{ip} \sim \mathcal{CN}(0, \delta_i^2)$ are independent over channel index i and path index p, where δ_i^2 is the i-th channel's average attenuation. $\mathbf{a}_\theta$ is the steering vector

$$\mathbf{a}(\theta) \triangleq \begin{bmatrix} 1 \\ e^{-j2\pi \frac{D}{\lambda} cos(\theta)} \\ \vdots \\ e^{-j2\pi \frac{(M-1)D}{\lambda} cos(\theta)} \end{bmatrix}, \tag{5.25}$$

where D is the antenna spacing at the base station and λ is the signal wavelength, such that $D \leq \lambda/2$. $\theta_{ip} \in [0, \pi]$ is a random AOA. Note that we can limit angles to $[0, \pi]$, because any $\theta \in [-\pi, 0]$ can be replaced by $-\theta$, giving the same steering vector.

5.2.2 Theorem
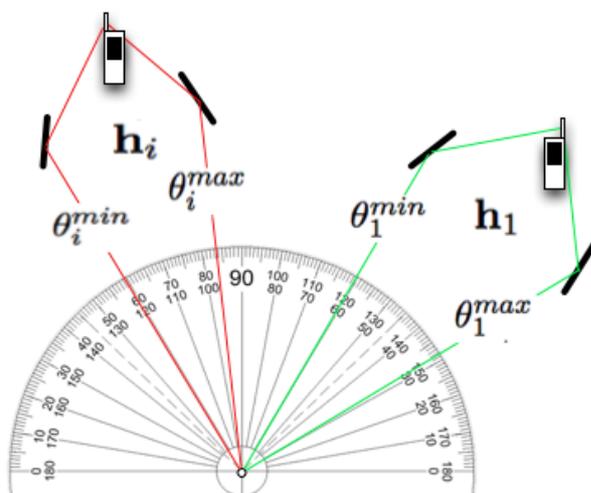
Figure 5.2: Theorem.

Assume the multipath angle of arrival θ yielding channel $\boldsymbol{h}_j, j = 1, \ldots, L$, in (5.24), is distributed according to an arbitrary density $p_j(\theta)$ with bounded support, i.e., $p_j(\theta) = 0$ for $\theta \notin [\theta_j^{min}, \theta_j^{max}]$ for some fixed $\theta_j^{min} \leq \theta_j^{max} \in [0, \pi]$. If the $L - 1$ intervals $[\theta_i^{min}, \theta_i^{max}]$, $i = 2, \ldots, L$ are strictly with non-overlapping with the desired channel's AOA interval $[\theta_1^{min}, \theta_1^{max}]$, we have

$$\lim_{M \to \infty} \widehat{\mathbf{h}}_1 = \widehat{\mathbf{h}}_1^{\text{no int}}. \tag{5.26}$$

Based on the above theorem, we conclude that the performance of the Bayesian channel estimation is sensitive to the degree with which the signals of the desired and interference channels bump in the BS antennas. The authors of [4] prove that in the ideal case in which the AOA of the desired and the interference channels do not overlap with each other, then, if the antenna-array of BSs is large, the pilot contamination effect temps to vanish. This is also confirmed by our simulations in section 5.3.

Proof:

We said before that we made the assumption that $R_i \triangleq \mathbb{E}\{\mathbf{h}_l \mathbf{h}_l^H\}$, so from the channel model (5.24), we get

$$\mathbf{R_i} = \frac{\delta_i^2}{P} \sum_{p=1}^{P} \mathbb{E}\{\mathbf{a}(\theta_{ip})\mathbf{a}(\theta_{ip})^H\} = \delta_i^2 \mathbb{E}\{\mathbf{a}(\theta_i)\mathbf{a}(\theta_i)^H\},$$

where θ_i has the PDF $p_i(\theta)$ for all $i = 1, \ldots, L$. The proof of Theorem 1 relies on three intermediate lemmas which exploit the eigenstructures of the covariance matrices. The essential ingredient is to exhibit an asymptotic (at large M) orthonormal vector basis for R_i constructed from steering vectors at regularly sampled spatial frequencies.

Lemma 1. *Define* $\alpha \triangleq \begin{bmatrix} 1 & e^{-j\pi x} & \cdots & e^{-j\pi(M-1)x} \end{bmatrix}^T$ *and* $\mathcal{A} \triangleq span\{\alpha(x), x \in [-1, 1]\}$. *Given* $b_1, b_2 \in [-1, 1]$ *and* $b_1 < b_2$, *define* $\mathcal{B} \triangleq span\{\alpha(x), x \in [b_1, b_2]\}$, *then*

- $dim\{\mathcal{A}\} = M$
- $dim\{\mathcal{B}\} \sim (b_2 - b_1)M/2$ *when M grows large.*

Lemma 1 characterizes the number of dimensions a linear space has, which is spanned by $\alpha(x)$, in which x plays the role of spatial frequency.

Lemma 2. *With a bounded support of AOAs, the rank of channel cavariance matrix R_i satisfies*

$$\frac{rank(\mathbf{R}_i)}{M} \leq d_i, \quad as \quad M \to \infty$$

where d_i is defined as

$$d_i \triangleq (cos(\theta_i^{min}) - cos(\theta_i^{max}))\frac{D}{\lambda}.$$

Lemma 2 indicates that for large M, there exists a null space $\text{null}(\mathbf{R}_i)$ of dimension $(1 - d_i)M$.

Lemma 3. *The null space $\text{null}(\mathbf{R}_i)$ includes a certain set of unit-norm vectors*:

$$null(\mathbf{R}_i) \supset span\left\{\frac{a(\Phi)}{\sqrt{M}}, \forall \Phi \notin [\theta_i^{min}, \theta_i^{max}]\right\}, \quad as \quad M \to \infty$$

This lemma indicates that multipath components with AOA outside the AOA region for a given user will tend to fall in the null space of its covariance matrix in the large-number-of-antennas case.

We now return to the proof of Theorem 1. $\mathbf{R}_i$ can be decomposed into

$$\mathbf{R}_i = \mathbf{U}_i \Sigma_i \mathbf{U}_i^H, \tag{5.27}$$

where $\mathbf{U}_i$ is the signal eigenvector matrix of size $M \times m_i$, in which $m_i \leq d_i M$. Σ_i is an eigenvector matrix of size $m_i \times m_i$. Due to Lemma 3 and the fact that densities $p_i(\theta)$ and $p_1(\theta)$ have non-overlapping supports, we have

$$\mathbf{U}_i^H \mathbf{U}_1 = 0, \forall i \neq 1, \ as \ M \to \infty. \tag{5.28}$$

Combining the channel estimate (5.18) and the channel model (5.16), we obtain

$$\widehat{\mathbf{h}}_1 = \mathbf{R}_1 \left(\sigma_n^2 \mathbf{I}_M + \tau \sum_{l=1}^{L} \mathbf{R}_l \right)^{-1} \bar{\mathbf{S}}^H \left(\bar{\mathbf{S}} \sum_{i=1}^{L} \mathbf{h}_i + \mathbf{n} \right). \tag{5.29}$$

According to (5.28), matrices $\mathbf{R}_1$ and $\sum_{l=2}^{L} \mathbf{R}_l$ span orthogonal subspaces in the large M limit. Therefore, we place ourselves in the asymptotic regime of M, when $\tau \sum_{l=2}^{L} \mathbf{R}_l$ can be eigen-decomposed according to

$$\tau \sum_{l=2}^{L} \mathbf{R}_l = \mathbf{W} \Sigma \mathbf{W}^H. \tag{5.30}$$

where $\mathbf{W}$ is the eigenvector matrix such that $\mathbf{W}^H \mathbf{W} = \mathbf{I}$ and $span\{\mathbf{W}\}$ is included in the orthogonal complement of $span\{\mathbf{U}_1\}$. Now denote $\mathbf{V}$ the unitary matrix corresponding to the orthogonal complement of both $span\{\mathbf{W}\}$ and $span\{\mathbf{U}_1\}$, so that the $M \times M$ identity matrix can now be decomposed into:

$$\mathbf{I}_M = \mathbf{U}_1^H \mathbf{U}_1 + \mathbf{W}\mathbf{W}^H + \mathbf{V}\mathbf{V}^H \tag{5.31}$$

Thus, for large M,

$$\widehat{\mathbf{h}}_1 \sim \mathbf{U}_1 \Sigma_1 \mathbf{U}_1^H (\sigma_n^2 \mathbf{U}_1 \mathbf{U}_1^H + \sigma_n^2 \mathbf{V}\mathbf{V}^H + \sigma_n^2 \mathbf{W}\mathbf{W}^H$$
$$+ \tau \mathbf{U}_1 \Sigma_1 \mathbf{U}_1^H + \mathbf{W}\Sigma\mathbf{W}^H)^{-1} \left(\tau \sum_{i=1}^{L} \mathbf{h}_i + \bar{\mathbf{S}}^H \mathbf{n} \right)$$

Due to asymptotic orthogonality between $\mathbf{U}_1$, $\mathbf{W}$, and $\mathbf{V}$,

$$\widehat{\mathbf{h}}_1 \sim \mathbf{U}_1 \Sigma_1 (\sigma^2 \mathbf{I}_{m_1} + \tau \Sigma_1)^{-1} \mathbf{U}_1^H (\tau \sum_{i=1}^{L} \mathbf{h}_i + \bar{\mathbf{S}}^H \mathbf{n})$$

$$\sim \mathbf{U}_1 \Sigma_1 (\sigma^2 \mathbf{I}_{m_1} + \tau \Sigma_1)^{-1} \tau (\mathbf{U}_1^H \mathbf{h}_1 + \sum_{i=2}^{L} \mathbf{U}_1^H \mathbf{h}_i + \frac{\bar{\mathbf{S}}^H \mathbf{n}_1}{\tau}).$$

However, since $\widehat{\mathbf{h}}_i \subset span\{\mathbf{a}(\theta)\}, \forall \theta \in [\theta_i^{min}, \theta_i^{max}]\}$, we have from Lemma 3 that $\frac{\|\mathbf{U}_1^H \mathbf{h}_i\|}{\|\mathbf{U}_1^H \mathbf{h}_1\|} \to 0$, for $i \neq 1$ when $M \to \infty$.
Therefore

$$\lim_{M \to \infty} \widehat{\mathbf{h}}_1 = \tau \mathbf{U}_1 \Sigma_1 (\sigma_n^2 \mathbf{I}_{m_1} + \tau \Sigma_1)^{-1} \left(\mathbf{U}_1^H \mathbf{h}_1 + \frac{\bar{\mathbf{S}}^H \mathbf{n}}{\tau} \right),$$

which is identical to $\widehat{\mathbf{h}}_1^{\text{no int}}$ if we apply the EVD decomposition (5.27) for $\mathbf{R}_1$ in (5.22). This proves Theorem 1. (For the proof of the lemmas see the appendices of [4])

The authors also believe that, although antenna calibration is needed as a technical assumption in the theorem, orthogonality of covariances signal subspaces will occur in non-tightly calibrated settings provided the AOA regions do not overlap.
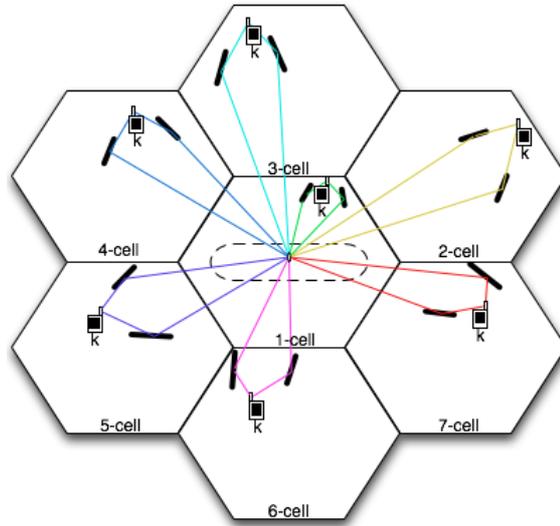
Figure 5.3: Coordinated Pilot Assignment.

5.2.3   Algorithm of Coordinated Pilot Assignment

Inspired by this theorem, the authors of [4] have invented an algorithm which improves the performance of Bayesian channel estimation in cellular systems and tackles the pilot contamination effect. This algorithm uses a coordination protocol of assigning pilot sequences to the users in the L cells, with the aim to satisfy the non-overlapping AOA constraint of the theorem, by taking advantage of the covariance informations and the property of subspace orthogonality. Recall that each individual pilot sequence corresponds to one, and only one, user of each cell. So the role of this coordination is to optimise the clustering of terminals which use the same pilot sequence, so that the degree of their signal subspaces of covariance matrices do not overlap with each other.

Let $\mathcal{G} \triangleq \{1, ..., K\}$, then $\mathcal{K}_l \in \mathcal{G}$ denotes the index of the user in the l-th cell who is assigned the pilot sequence s. The set of selected users is denoted by $\mathcal{U}$ in what follows. For a given user set $\mathcal{U}$, the authors define a network utility function

$$F(\mathcal{U}) \triangleq \sum_{j=1}^{|\mathcal{U}|} \frac{\mathcal{M}_j(\mathcal{U})}{\text{tr}\left\{\mathbf{R}_{jj}(\mathcal{U})\right\}}, \qquad (5.32)$$

where $|\mathcal{U}|$ is the cardinal number of the set $\mathcal{U}$. $\mathcal{M}_j(\mathcal{U})$ is the estimation MSE for the desired channel at the j-th base station, with a notation readily extended from $\mathcal{M}_j$ in (5.21), where this time cell j is the target cell when computing $\mathcal{M}_j$. $\mathbf{R}_{jj}(\mathcal{U})$ is the covariance matrix of the desired channel at the j-th cell.

The approach that have been followed is:

Data: KL^2 covariance matrices $\mathbf{R}$.

$\mathcal{U} = \emptyset$
for $l = 1, \ldots, L$ **do**
  $\arg\min_{\mathbf{k}\in\mathcal{G}} F(\mathcal{U} \cup \{k\})$
  $\mathcal{U} \leftarrow \mathcal{U} \cup \{\mathcal{K}_l\}$
end

Result: The set $\mathcal{U}$ with the L users, who will be assigned the pilot sequence $\mathbf{s}$.

As we see, the coordinated pilot assignment in order to satisfy the conditions of the theorem, minimises the sum of (5.32) by choosing users whose the covariance matrices show maximum signal subspace orthogonality. Getting started with the first cell, the system chooses the user with the most powerful channel (the trace of a covariance matrix equals with the power of the channel), after that the algorithm is scanning the terminals of another cell and chooses the one who shows the maximum signal subspace orthogonality with the K_1 and also has strong channel with its base station. The same happens with the remaining cells, searching in each one for the user who combined better with those who have already been selected.

As it is reasonable, the performance will improve by increasing the number of users, as it becomes more likely to find users with discriminable second-order statistics.

5.3 Simulations

Due to computational complexity, we are limited in the number of the cells that we consider. So in this scenario, the most that we consider are seven hexagonal cells with M number of base station antennas.

The basic parameters are given in Table below.

Basic Parameters	
Cell Radius	1000m
Cell Edge SNR	20dB
Distance from a user to its BS	800m
Number of Terminals	10
Number of Paths	50
Path Loss Exponent	3
Carrier Frequency	2GHz
Antenna Spacing	$\lambda/2$

For our simulations, we use the channel model of (5.24). So, the channel vector between the u-th user in the l-th cell and the target base station is

$$\mathbf{h}_{lu} = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} \mathbf{a}(\theta_{lup}) \alpha_{lup}, \tag{5.33}$$

where θ_{lup} and α_{lup} are the AOA and the attenuation of the p-th path between the u-th user in the l-th cell and the target base station, respectively. Note that the variance of α_{lup}, $\forall p$, is δ_{lu}^2, which includes the distance-based path loss β_{lu} between the user and the target base station (which can be anyone of the L cells):

$$\beta_{lu} = \frac{\alpha}{d_{lu}^{\gamma}}, \tag{5.34}$$

where d_{lu} is the geographical distance. γ is the path-loss exponent. α is a constant dependent on the prescribed average SNR at cell edge. In order to preserve fairness between users and avoid having high-SNR users being systematically assigned the considered pilot, we use the α factor to provide each user in every cell similar SNR to the others.

We consider two types of AOA distributions, a bounded one (Uniform) and a non-bounded one (Gaussian):

1) Uniform distribution: For the channel $\mathbf{h}_{lu}$, the AOAs are uniformly distributed over $[\bar{\theta}_{lu} - \theta_\Delta, \bar{\theta}_{lu} + \theta_\Delta]$, where $\bar{\theta}_{lu}$ is the mean AOA and the θ_Δ is the angle spreads. The same angle spread apply to both desired and interference channels. The characteristic of the usage uniform distribution,

is that AOAs are bounded by a standard angle spread, and thus does not consist a realistic option, since is likely the signal to hit an object and follow an undefined path. Nevertheless, this will help us to verify the theorem and take a first taste of the coordinated approach in channel estimation by making easier the finding of users with non-overlapping channels.

2) Gaussian distribution: For the channel coefficients $\mathbf{h}_{lu}$, the AOAs of all P paths are i.i.d. Gaussian random variables with mean θ_{lu} and standard deviation σ. The standard deviation of AOA is the same either for desired or interference channels. Using this distribution we take more representative results of what the coordinated pilot assignment can achieve since it covers every possible angle which a channel path can follow in real conditions.

5.3.1   Numerical Results of Normalized Channel Estimation Error

To evaluate the proposed channel estimation scheme, we first use the below normalized channel estimation error

$$\text{err} \triangleq 10 \log_{10} \left(\frac{\sum\limits_{j=1}^{L} ||\widehat{\mathbf{h}}_{jj} - \mathbf{h}_{jj}||_F^2}{\sum\limits_{j=1}^{L} ||\mathbf{h}_{jj}||_F^2} \right), \tag{5.35}$$

where $\mathbf{h}_{jj}$ and $\widehat{\mathbf{h}}_{jj}$ are the desired channel at the j-th base station and its estimate respectively. Note that after the clustering of the users, we consider the normalized channel estimation error as the sum of the estimation error for each desired channel in each cell, and not for one specific cell which it could be a chance to have advantage over others. This leads us to assess the situation in a more comprehensive manner.

First we confirm the theorem with a 2-cell network. Within each cell there is only one user, fixed in 90 degrees and 800 metres away from the BS. All P channels are uniformly distributed with $\theta_\Delta = 20$. By having fixed users' positions and bounded angle spread at 20 degrees, we secure the non-overlapping between desired and interfering multipaths. As we see in Figure 5.4 the Covariance-aided Bayesian (CB) estimation in this scenario is identical with the one in an interference-free network, and much better of LS estimation as the number of BS antennas grows.
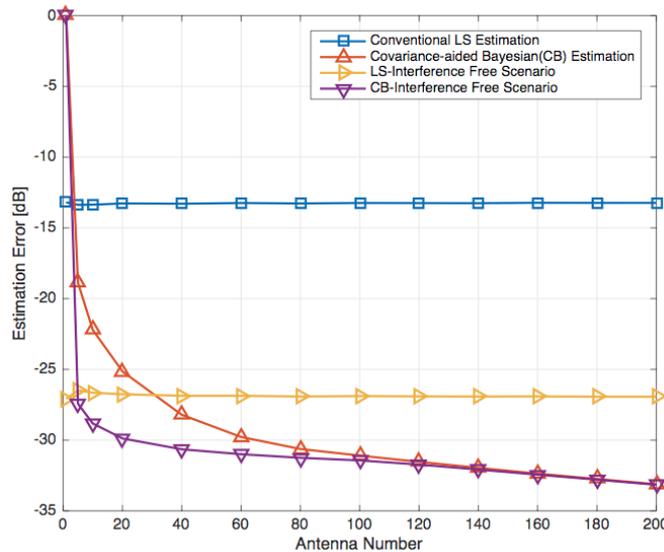

Figure 5.4: Estimation MSE vs BS antenna number in 2-cell network, fixed positions of two users, uniformly distributed AOAs with $\theta_\Delta = 20$ degrees.

We are then testing the coordinated pilot assignment (CPA) in a 2-cell network for uniform distribution AOAs with $\theta_\Delta = 10$ degrees and for a Gaussian distribution with $\sigma = 10$ degrees in Figure 5.5 (a) and (b) respectively.
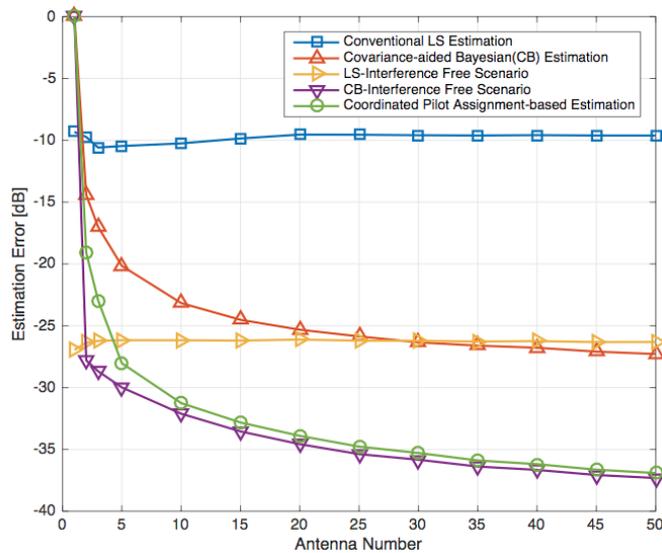
(a) Uniformly distributed AOAs with $\theta_\Delta = 10$ degrees.
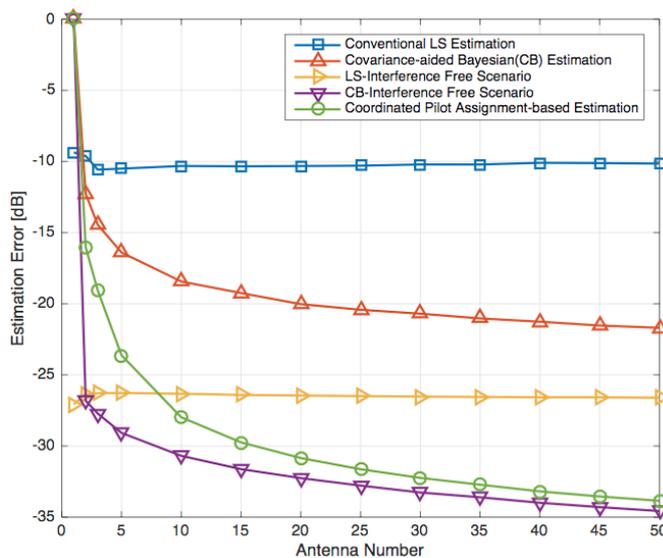
(b) Gaussian distributed AOAs with $\sigma = 10$ degrees.

Figure 5.5: Estimation MSE vs antenna number, 2-cell network.

As we see, the gains of CB and CPA compared to the classical estimator are substantial, and are getting better as the antenna number grows. Overall, the CB estimation and the coordinated one are better in Uniform distribution AOAs than in Gaussian due to the non-boundedness of the Gaussian PDF, which makes it harder to find users with absolutely non-overlapping channels. The most important finding is that the coordinated pilot assignment matches the performance of CB interference free scenario, which means that pilot contamination effect tends to vanish in the large-antenna-array case.

The difference between the two AOAs' distributions, raises the need of seeing how the standard deviation influence the performance of CB and CPA estimations. We choose the Gaussian distribution because as we mention before consists a more realistic model. We simulate a 7-cell network and a fixed number of M=10 BS antennas.

As expected, the performances of CB and CPA reduced dramatically as the standard deviation grows. As we see in Figure 5.6, the estimation error
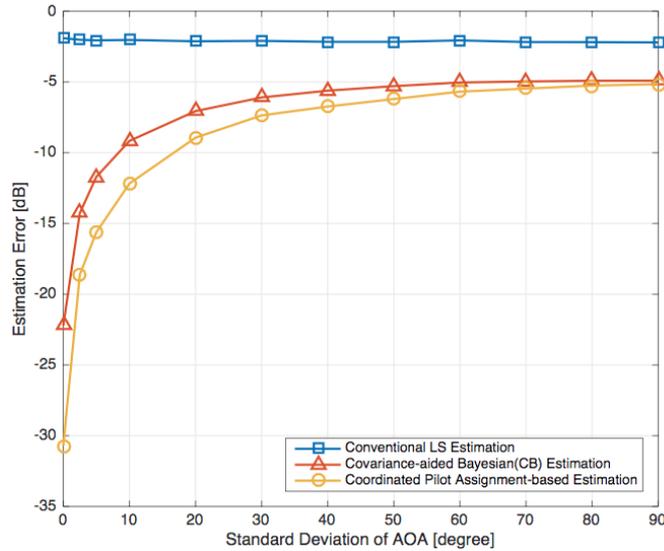
Figure 5.6: Estimation MSE vs standard deviation of Gaussian distributed AOAs with M=10, 7-cell network.

is a monotonically increasing function of σ. This makes sense since in large values of σ there will be paths of the desired channels which will overlap with paths of the interference channels. That is also why the estimation error of CPA is approaching and eventually identified with CB's for $\sigma = 90$, because it is impossible to find users whose their channels do not overlap.

5.3.2 Numerical Results of the downnlink per-cell rate

Having confirmed that in massive MIMO the CB estimation works better than conventional LS estimator in terms of estimation error, and that CPA achieves even better results in tackling the pilot contamination effect, we will now check the per-cell rate of the downlink for the three estimators (LS, CB, CPA). Assuming standard MRC beamformer based on the channel estimates we define the per-cell rate as follows :

$$\mathcal{C} \triangleq \frac{\sum_{j=1}^{L} \log_2(1 + \text{SINR}_j)}{L}$$

where SINR_j is the received signal-to-noise-plus-interference ratio by the scheduled user in the j-th cell.

As we see in Figures 5.7 and 5.8, the better functioning of CB and CPA in comparison to LS is also confirmed in per-cell rate. Specifically, in Figure 5.7 we see that as the number of antennas grows, while the performance of LS stays constant, the performances of CB and CPA are getting better. Affirming once again the CB's usefulness in massive MIMO.

In Figure 5.8, we examine the impact of standard deviation of Gaussian AOAs on the per-cell rate of the downlink. As in the corresponding diagram of estimation error (Figure 5.6), it is confirmed again the destructive effect of high standard deviation in CB and CPO estimations.

What we conclude from the study of [4] and our simulations is that Covariance-aided Bayesian estimation exploits much better the spatial dimension of massive MIMO than the conventional estimators. The theorem that the authors introduce in [4] and we present in section 5.2.2 seems very useful for exploitation in massive MIMO. The Coordinated Pilot Assignment seems to exploit greatly this Theorem and to improve the performance of CB. Nevertheless, it should be considered how this will work in a network
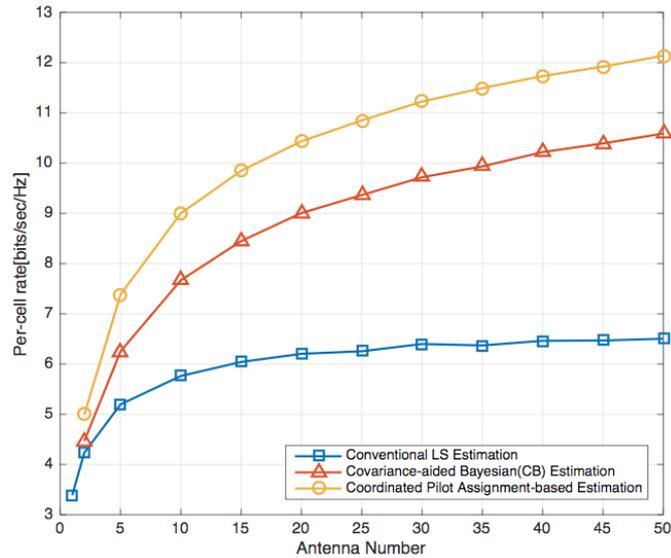
Figure 5.7: Per-cell rate vs antenna number, 2-cell network, Gaussian distributed AOAs with $\sigma = 10$ degrees.
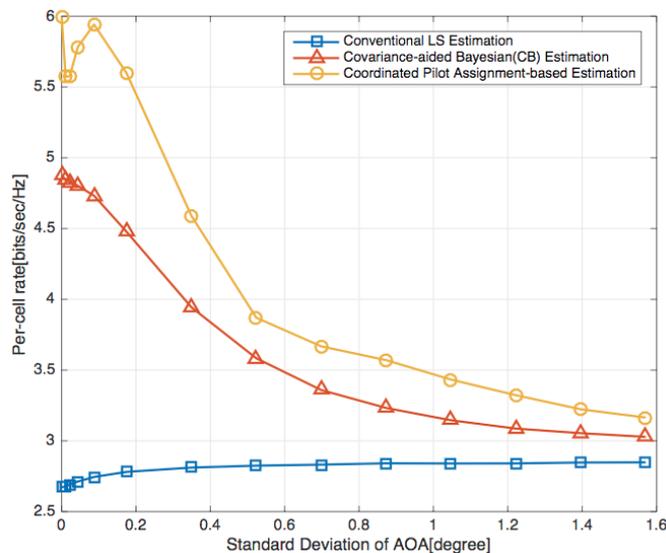

Figure 5.8: Per-cell rate vs standard deviation of Gaussian distributed AOAs with M=10, 7-cell network.

with many more cells, and if the improvement which will achieve, compared to the complexity of its functionality, worth it.

The coordination of the users requires the exchange of information between base stations, which in combination with the processing that is needed in a server in order to complete the clustering and the assignment of the pilots, takes time. If we assume that the number of the cells, and the number of the users, in fact they will be much larger than in our scenario, it is reasonable to think if the coherence time is big enough to estimate the channels, coordinate the users and send data.

The most important hurdle that requires solution, is the need of knowledge of the individual covariance matrices, which constitutes a fundamental factor of CB's operation. As we mentioned before, there is no specific training design for learning second-order statistics [4]. So if we want to apply CB and CPA in the future networks, we have to find a solution for this first.

Chapter 6

Conclusions

To discover the capabilities and needs of massive MIMO we started our study from the pioneering and award-winning paper [3] of Thomas L. Marzetta, who introduced and popularized the concept of "massive (multiuser) MIMO".

We saw that the assumption of an unlimited number of antennas at the base station greatly simplifies the analysis, and it illustrates the desirable effects of operating with a large excess of antennas compared with terminals. By using an unlimited number of antennas, we saw that the effects of uncorrelated receiver noise and fast fading are eliminated completely, so transmissions from terminals within one cell do not interfere. The acquisition of channel state information, which leads to the phenomenon known as pilot contamination, constitutes fundamental limitation on what can be achieved by a noncooperative cellular multiuser MIMO system and is not treated just from the increase of antennas. Since the transmissions are not influenced by noise and intra-cell interference, the performance is only limited by inter-cell interference expressed by the random quantity SIR.

With a view to boost the SIR and tackle the pilot contamination effect Thomas L. Marzetta proposed the method of frequency reuse. By using less aggressive frequency reuse factor, clustering the cells in three or seven different groups, we ensure greater probability for a terminal to have high SIR but this results in lesser mean net capacity per terminal. This happens because by using frequency reuse factor greater than one, the available bandwidth of each user subdivided by the same factor. So, regardless of SIR's increase, we lose the big advantage of massive MIMO system which is to enable each terminal to use all available spectrum resources, because it opens up the spatial dimension that allows it to discriminate the signals to/from each terminal based on its location.

Realising that pilot contamination effect constitute a major bottleneck for overall performance of massive MIMO systems and that frequency reuse is not a solution, we then turned our focus to study solutions which tackle the pilot contamination problem. Since the pilot contamination problem stems from procedure of channel estimation, we find a very interesting coordinated approach of channel estimation in [4]. It is about an algorithm which tackles the pilot contamination problem by enabling a low-rate coordination between cells during the channel estimation phase itself. The coordination makes use of the additional second-order statistical information about the user channels, which are shown to offer a powerful way of discriminating across interfering users with even strongly correlated pilot sequences. The simulations showed that this coordinated approach in the channel estimation phase achieves amazing results even with small antenna array in BSs which become even better as the number of antennas grows large, approaching the interference-free results and in some special cases become identical to them.

Overall, massive MIMO seems to be an importand part of the future network, since we have seen that offers many advantages and increases dramatically the cell capacity. So as we approach the time that 5G will be in use, researchers of all over the world are trying to find solution to tackle the pilot contamination effect, and generally to make massive MIMO functional.

Bibliography

[1] Cisco Visual Networking Index: *Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper*, Feb 01, 2016. Available at: cisco.com

[2] J. G. Andrews, et al., *What Will 5G Be?*. IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065–1082, June 2014

[3] T. Marzetta, *Noncooperative cellular wireless with unlimited numbers of base station antennas*. IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Sep. 2010.

[4] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, *A coordinated approach to channel estimation in large-scale multiple-antenna systems*. IEEE J. Sel. Areas Commun., vol. 31, no. 2, pp. 264–273, Feb. 2013.

[5] D. Tse and P. Viswanath, *Fundamentals of wireless communications.*. Cambridge University Press, 2005

[6] Athanasios P. Liavas, *Lectures of Wireless Communications.*. Technical University of Crete , 2014

[7] Radio electronics: Resources and analysis for electronics engineers, `http://www.radio-electronics.com/info/antennas/mimo/`

[8] Electronic Design:What's the difference between FDD and TDD `http://electronicdesign.com/communications/`

[9] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, *Massive MIMO for Next Generation Wireless Systems*. IEEE Communications Magazine, vol. 52, no. 2, pp. 186-195, Feb. 2014

[10] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, *Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems*. Transactions on Communications, vol. 61, no. 4, pp. 1436-1449, April 2013.

[11] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, *Argos: Practical many-antenna base stations*. in ACM Int. Conf. Mobile Computing and Network- ing (MobiCom), Istanbul, Turkey, Aug. 2012.

[12] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, *An Overview of Massive MIMO: Benefits and Challenges*. IEEE Journal of Selected Topics in Signal Processing, Vol. 8, No. 5, October 2014.

[13] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, *Inter-cell interference in noncooperative TDD large scale antenna systems*. IEEE Journal on Selected Areas in Communications, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[14] Ian F. Akyildiz, David M. Gutierrez-Estevez ,Elias Chavarria Reyes *The evolution to 4G cellular systems: LTE-Advanced*. Journal Physical Communication, vol. 3, no. 4, pp. 217-244, December, 2010

[15] A. J. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications.* Cambridge University Press, 2003.