Universidade de Aveiro Departamento de Eletrónica, Telecomunicações e Informática,
2016

Daniel Filipe Pinheiro de Azevedo

Comparação do desempenho de Arquiteturas Híbridas para comunicações na banda das ondas milimétricas

Performance comparison of Hybrid Architectures for millimeter wave communications

Daniel Filipe Pinheiro de Azevedo

Comparação do desempenho de Arquiteturas Híbridas para comunicações na banda das ondas milimétricas

Performance comparison of Hybrid Architectures for millimeter wave communications

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requesitos necessários à obtenção do grau de Mestre em Engenharia Eletrónica e Telecomunicações, realizada sob a orientação científica do Professor Doutor Adão Silva (orientador), Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e do Doutor Daniel Castanheira (co-orientador), investigador no Instituto de Telecomunicações de Aveiro.

o júri / the jury

presidente / president

Professor Doutor José Carlos Esteves Duarte Pedro
Professor Catedrático da Universidade de Aveiro

vogais / examiners committee

Professor Doutor Paulo Jorge Coelho Marques
Professor Adjunto no Instituto Politécnico de Castelo Branco

Professor Doutor Adão Paulo Soares da Silva
Professor Auxiliar na Universidade de Aveiro

agradecimentos

Queria deixar um agradecimento a todas as pessoas sem as quais não teria sido possível concluir este trajecto.

E principalmente,

À minha família, aos meus pais, avós, irmã e madrinha, por todo o apoio e a ajuda ao longo destes anos, sem o qual este objetivo não teria sido concluído.

Aos professores, Adão Silva e Daniel Castanheira por toda a ajuda, orientação, e disponibilidade demonstrada ao longo deste ano.

Aos meus amigos, por toda amizade, paciência e tolerância demonstrada ao longo destes anos, principalmente nos meus períodos de ausência.

A todos um enorme Muito Obrigado.

Palavras-chave

5G, mmWave communications, massive MIMO, hybrid analog/digital architectures

Resumo

A proliferação massiva das comunicações sem fios faz prever que o número de utilizadores aumente exponencialmente até 2020, o que tornará necessário um suporte de tráfego milhares de vezes superior e com ligações na ordem dos Gigabit por segundo. Este incremento exigirá um aumento significativo da eficiência espectral e energética. Impõe-se portanto, uma mudança de paradigma dos sistemas de comunicação sem fios convencionais, imposta pela introdução da 5^a geração.

Para o efeito, é necessário desenvolver novas e promissoras técnicas de transmissão, nomeadamente a utilização de ondas milimétricas em sistemas com um número massivo de antenas. No entanto, consideráveis desafios emergem ao adotar estas técnicas. Por um lado, este tipo de ondas sofre grandes dificuldades em termos de propagação. Por outro lado, a adoção de arquiteturas convencionais para sistemas com um número massivo de antenas é absolutamente inviável, devido ao custo e ao nível de complexidade inerentes. Isto acontece porque o processamento de sinal ao nível da camada física é maioritariamente feito em banda base, ou seja, no domínio digital requerendo uma cadeia RF por cada antena.

Neste contexto as arquiteturas híbridas são uma proposta relativamente recente que visa simplificar a utilização de um grande número de antenas, dividindo o processamento entre os domínios analógico e digital. Para além disso, o número de cadeias RF necessárias é bastante inferior ao número total de antenas do sistema, contribuindo para óbvias melhorias em termos de complexidade, custo e energia consumida.

Nesta dissertação é implementada uma arquitetura híbrida para ondas milimétricas, onde cada cadeia RF está apenas conectada a um pequeno conjunto de antenas. É considerado um sistema contendo um transmissor e um recetor ambos equipados com um grande número de antenas e onde, o número de cadeias RF é bastante inferior ao número total de antenas. Pré-codificadores híbridos analógico/digital, recentemente propostos na literatura são utilizados e novos equalizadores híbridos analógico/digital são projetados. É feita uma avaliação de performance à arquitetura implementada e posteriormente comparada com uma outra arquitetura, onde todas as antenas estão conectadas a todas as cadeias RF.

Keywords 5G, mmWave communications, massive MIMO, hybrid analog/digital architectures

Abstract

The expected massive proliferation of wireless systems points out an exponential increase in the number of users until 2020, which is needed to support up to one thousand times more traffic and connections in order of Gigabit per second. However, these goals require a significantly improvement in the spectral and energy efficiency. As a result, it is essential to make a paradigm shift in conventional wireless systems, imposed by the introduction of fifth generation (5G).

For this purpose, new and promising transmission techniques will be needed, namely the use of millimeter Waves (mmWave) in systems with a massive number of antenna elements. Nevertheless, considerable challenges emerge in the adoption of these techniques. On one hand, mmWave suffer great difficulties in terms of propagation. On the other hand, the using of conventional architectures for systems with a large number of antennas is absolutely impracticable because of the costs and the level of complexity. This happens because the signal processing in physical layer is mostly done in baseband, which means, that one RF chain for each antenna is required.

In this context the hybrid architectures are a relatively recent proposal where the aim is to simplify the use of a large number of antenna elements, dividing the processing between the analog and digital domains. Moreover, the number of RF chains needed are much lower than the total number of antenna elements of the system, which contribute to obvious improvements in terms of complexity, costs and energy consumption.

In this Dissertation a hybrid mmWave based architecture, where each RF chain is only connected to a small set of antennas, is implemented. It is considered a system comprising a transmitter and a receiver both equipped with a massive number of antennas and where the number of RF chains is much lower than the number of antennas. Hybrid analog/digital precoders recently proposed in the literature are used and a new hybrid analog/digital equalizer is designed. The implemented architecture is then evaluated and compared with other architecture, where all the antennas are connected to all RF chains.

Contents

List of Figures

List of Tables

List of Acronyms

1G 1st Generation

2D Two Dimensional

2G 2nd Generation

3D Three Dimensional

3G 3rd Generation

3GPP Third-Generation Partnership Project

4G 4th Generation

5G 5th Generation

AA Antenna Array

ADC Analogue to Digital Converter

AoA Angle of Arrival

AoD Angle of Departure

AWGN Added White Gaussian Noise

BER Bit Error Ratio

BS Base Station

CBSM Correlation-Based Stochastic Model

CoMP Coordinated Multipoint

CSI Channel State Information

DDCM Double Directional Channel Model

DoF Degrees of Freedom

EDGE Enhanced Data Rate for Global Evolution

EE Energy Efficiency

EGC Equal Gain Combining

eNB evolved Node B

FCC Federal Communications Commission

FDD Frequency Division Duplex

FDD-TDD CA Frequency Division Duplex - Time Division Duplex Carrier Aggregation

FDMA Frequency Division Multiple Access

GBSM Geometry-Based Stochastic Model

GPRS General Packet Radio Service

GSM Global System for Mobile Communications

HA-1 Hybrid Architecture 1

HA-2 Hybrid Architecture 2

HetNets Heterogeneous Networks

HSDCSD High-Speed Circuit Switched Data

HSDPA High Speed Downlink Packet Access

HSPA High Speed Packet Access

HSUPA High Speed Uplink Packet Access

i.i.d. independent identically distributed

IMT-Advanced International Mobile Telecommunications Advanced

ITU-R International Telecommunications Union Radiocommunication Sector

IEEE Institute of Electrical and Electronics Engineers

LNA Low Noise Amplifiers

LoS Line-of-Sight

LTE Long Term Evolution

MBMS Multimedia Broadcast/Multimedia Services

MIMO Multiple-Input Multiple-Output

MISO Multiple-Input Single-Output

MIT Massachusetts Institute of Technology

mMIMO massive Multiple-Input Multiple-Output

MMIC Monolithic Microwave Integrated Circuits

MMS Multimedia Message Services

MMSE Minimum Mean Square Error

mmWave Millimeter Wave

MRC Maximal Ratio Combining

MTC Massive machine-type communication

MU-MIMO Multi-user MIMO

NLoS Non-Line-of-Sight

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

PA Power Amplifiers

PIC Parallel Interference Cancellation

PMI Precoding Matrix Indicator

ProSe Proximity Services

PSM Parametric Stochastic Model

QAM Quadrature Amplitude Modulation

QoS Quality of Service

RF Radio Frequency

SC Selection Combining

SC-FDMA Single Carrier Frequency Division Multiple Access

SCN Small Cell Network

SE spectral efficiency

SFBC Space-frequency Block Code

SIC Successive Interference Cancellation

SIMO Single-Input Multiple-Output

SISO Single-Input Single-Output

SNR Signal-to-Noise Ratio

STBC Space-time Block Code

STBC - TSTD STBC - Time Shift Transmit Diversity

SU-MIMO Single-User MIMO

SVD Singular Value Decomposition

TDD Time Division Duplex

TDMA Time Division Multiple Access

UE User Equipment

ULA Uniform Linear Array

UMTS Universal Mobile Telecommunication System

UPA Uniform Planar Array

UWB Ultra-Wideband

VRM Virtual Ray Model

WCDMA Wideband Code-Division Multiple Access

WiMAX Worldwide Interoperability for Microwave Access

ZF Zero-Forcing

Chapter 1

Introduction

This dissertation begins with an introduction to the technologies used in the developments of the implemented work. To easily understand the technologies, this chapter starts with an evolution of the mobile communications system since their beginning until the used systems nowadays. The future deployments, are described in the next sections. Then the motivations, objectives and the main contributions are presented. Finally, a brief summary of the following chapters is made.

1.1 Mobile Communications Evolution

The technology or concept of wireless technology is not recent. In fact, the story of wireless phones is dated way back to the middle of the 20th century [1]. The world first mobile phone call was made on April 3, 1973, when Martin Cooper, a senior engineer at Motorola, called a rival telecommunications company and informed them that he was speaking via a mobile phone [2]. Since those days a lot innovations have been made and the perception of mobile phones changed.

At the beginning, they were considered just luxury items although nowadays they are an important instrument in our lives. The newest, sleek phones support mobile customers by providing applications in communication, navigation, multimedia, entertainment among others. Once perceived as luxury items, cell phones have become a necessity for many [3]. A recent Massachusetts Institute of Technology (MIT) survey, listed cell phones as the most hated invention that Americans cannot live without, even more so than alarm clocks, televisions, razors, microwave ovens, computers, and answering machines [4]. In figure 1.1, we can see the four generations and some technologies that represent them.

1st Generation (1G) mobile systems used analog transmissions for voice services, and the first company that implemented the first cellular system in the world was Nippon Telephone and Telegraph from Tokyo, in Japan. However, the two most popular analog systems were Nordic Mobile Telephones (NMT) and TotalAccess Communication Systems (TACS). The system was allocated a 40-MHz bandwidth within the 800 to 900MHz frequency range [5]. This system was not so efficient in the use of the spectrum and was used in large cells with a smallest reuse factor.

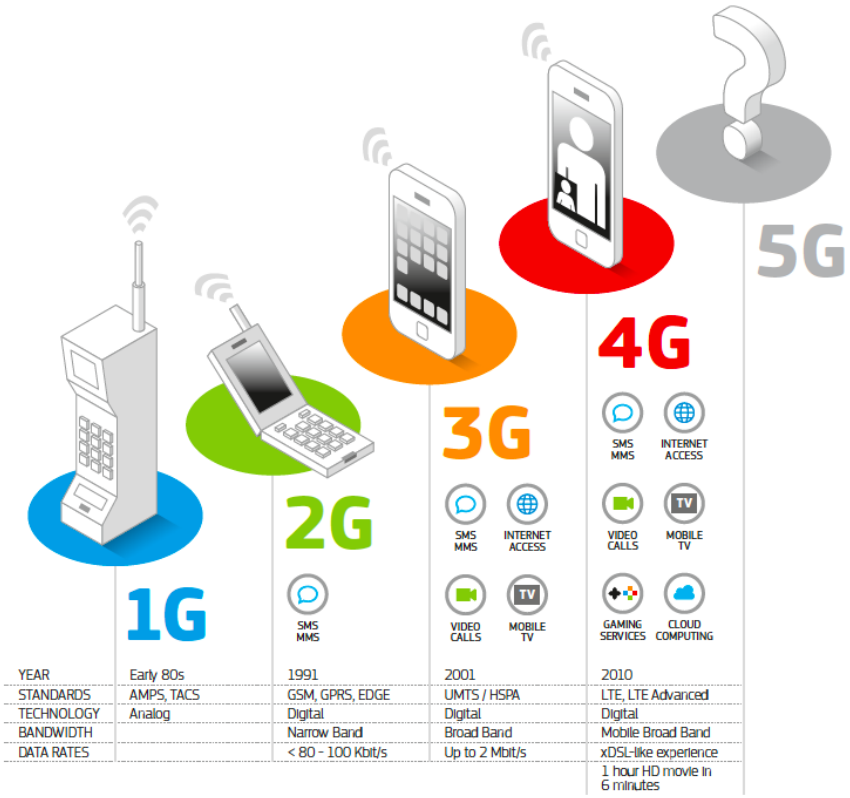


Figure 1.1: Cellular Technology Evolution [6]

In 1990s, the 1G was gradually replaced by the 2nd Generation (2G). This technology was also mostly based on circuit-switched that allowed a more efficient use of the radio spectrum and an introduction to smaller and cheaper devices. 2G wireless technologies could handle some data capabilities such as fax and short message service at the data rate up to 9.6 kbps, but it is not suitable for web browsing and multimedia applications. In order to support multiple users, Global System for Mobile Communications (GSM) can use Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA). FDMA, which is a standard that lets multiple users access a group of radio frequency bands and eliminates interference of message traffic, used to split the available 25MHz of bandwidth into 124 carrier frequencies of 200 kHz each. Each frequency is then divided using a TDMA scheme into eight time-slots and allows eight simultaneous calls on the same frequency [7]. The success of 2G communication systems came at the same time as the early growth of the Internet. It was natural for network operators to bring the two concepts together, by allowing users to download data onto mobile devices. To do this, so-called 2.5G systems built on the original ideas from 2G, by introducing the core network packet switched domain and by modifying the air interface so that it could handle data as well as voice. The General Packet Radio Service (GPRS) incorporated these techniques into GSM [8]. The GSM system has repercussions to the present day despite the number of users decreasing, this still remains the most widely used system, with a percentage of 49% worldwide at the end of 2015, as it can be seen in figure 1.2.

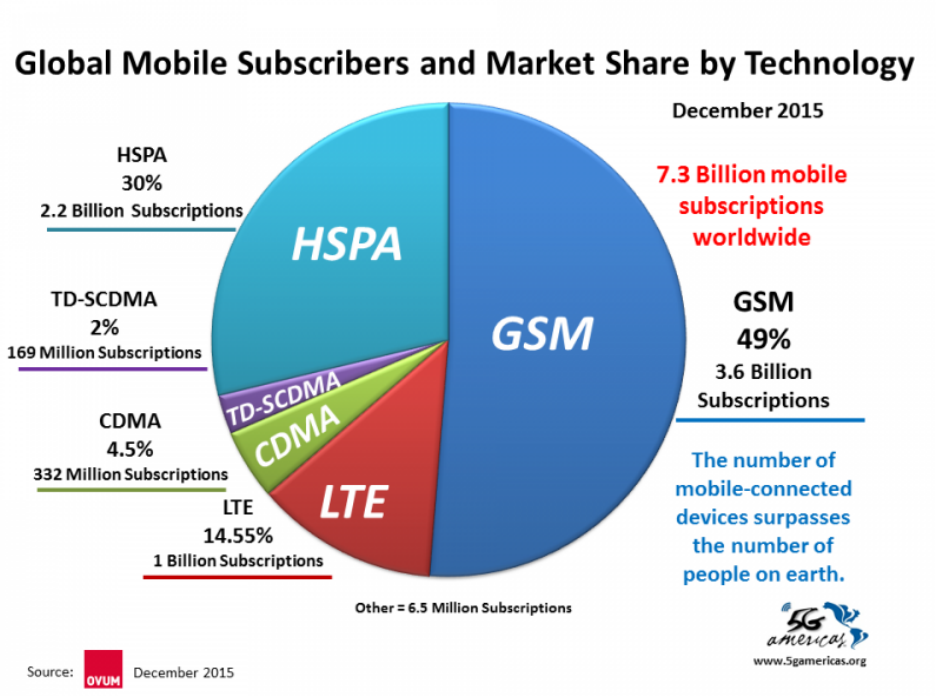


Figure 1.2: Global Mobile Subscribers and Market Shares [9]

The continuous need for higher data rates made another evolution to the 2G system, happen the so-called Enhanced Data Rate for Global Evolution (EDGE), through the use of 8-PSK modulation, using the same TDMA frame structure which allowed a three times higher data rate [10]. EDGE also introduced High-Speed Circuit Switched Data (HSDCSD) service which allowed transferring more data in each time-slot. The cost of increased capacity is that the signal quality therefore must be better.

This improvement in data speed continued and as faster and higher quality networks started supporting better services like video calling, video streaming, mobile gaming and fast Internet browsing, resulting in the introduction of the 3rd Generation (3G) mobile telecommunication standard, Universal Mobile Telecommunication System (UMTS) [11]. UMTS uses Wideband Code-Division Multiple Access (WCDMA) as a stand technique and a pair of 5 MHz channels. Today's 3G specifications call for 144 Kb/s while the user is on the move in an automobile or train, 384 Kb/s for pedestrians, and ups to 2 Mb/s for stationary users. That is a big step up from 2G bandwidth using 8 to 13 Kb/s per channel to transport speech signals [7].

At first, its appearance was quite unsuccessful due to high performance expectations and only after 3.5G introduction its performance was sufficed. The system was later enhanced for data applications, by introducing the 3.5G technologies of High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA), which are collectively known as High Speed Packet Access (HSPA).

In the 3G has appeared another technology that marked the migration for 4th Generation (4G), so-called Worldwide Interoperability for Microwave Access (WiMAX). This was developed by the Institute of Electrical and Electronics Engineers (IEEE) under IEEE standard 802.16 and has a very different history from other 3G systems. WiMAX supported point-to-multipoint communications between an omni-directional base station and a number of fixed devices further WiMAX allowing the devices to move and to hand over their communications from one base station to another [8].

With the increasing of mobile devices smartphones being more attractive and user-friendly than their predecessors such as Apple Iphones and Google Android, combining multimedia capabilities with applications that require higher mobile broadband data speeds. Between January 2007 to July 2011, the amount of data traffic increased by a factor of over 100 [8]. In response to the new necessities, and the fully exploited 3G systems, Third-Generation Partnership Project (3GPP) wrote the specifications for 4G like as it had done for 2G and 3G. The most notable requirements were on high data rate at the cell edge and the importance of low delay, in addition to the normal capacity and peak data rate requirements. Furthermore, spectrum flexibility and maximum commonality between Frequency Division Duplex (FDD) and Time Division Duplex (TDD) solutions are pronounced [12].

The multiple access scheme in Long Term Evolution (LTE) downlink uses Orthogonal Frequency Division Multiple Access (OFDMA) and uplink uses Single Carrier Frequency Division Multiple Access (SC-FDMA). These multiple access solutions provide orthogonality between the users, reducing the interference and improving the network capacity [13]. LTE uses higher order modulations, large bandwidths up to 20 MHz and Multiple-Input Multiple-Output (MIMO) spatial multiplexing, that allowed peak rates of 300 Mbps in the downlink and 75 Mbps in the uplink, in release 8 [14]. In release 9 more capacities, were added like Multimedia Broadcast/Multimedia Services (MBMS) or dual beamforming layers, but the peak data rates were maintained [10].

The International Telecommunications Union Radiocommunication Sector (ITU-R) introduced the International Mobile Telecommunications Advanced (IMT-Advanced) list. This list was intended to recognize new cellular system and begin a brand new generation. Since 2009, 3GPP has tried to perform enhancements in LTE technology in order to meet the IMT-Advanced requirements. Later that year, a new ITU system, was proposed represented by LTE release 10 or LTE-Advanced. Besides using LTEs spectrum bands, it also uses some bands predicted by IMT-Advanced [15]. On this development the aimed peak data rates are up to 3 Gbps in the downlink and up to 1.5 Gbps in the uplink [16] [17]. This can be achieved with some enhancements, like carrier aggregation, that increases the bandwidth to 100 MHz, and a higher number of antennas in the MIMO schemes (8x8 in the downlink and 4x4 in the uplink [14]).

The network planning is essential to cope with the increasing number of mobile broadband data subscribers and bandwidth intensive services competing for limited radio resources. In earlier releases of LTE has introduced the Heterogeneous Networks (HetNets) which are composed by different cell sizes has resulted on a network with large macro cells in combination with small cells providing higher capacities and performances in the network and the cover-

age area has also improved [18]. The introduction of this networks allow the decreasing of implementation costs for the operators, because the costs associated to small cells are lower than the ones of macro-cells.

The next step to the LTE-Advanced is fixed in the release 11, with enhancements of the previous implementations. However, the main improvement in this release are the Coordinated Multipoint (CoMP) transmissions. This technique allows the User Equipment (UE) in the cell edge to receive and transmit data from and to multiple cells, which consequently improve its performance.

The last frozen release, 12, brings more enhancements to the system, which is to be implemented during the current year of 2016. In [19] various enhancements are described. Among them are the Frequency Division Duplex - Time Division Duplex Carrier Aggregation (FDD-TDD CA), that allows a UE to operate in TDD and FDD spectrum jointly, HetNets Mobility, to better handle the handover between cells and to allow handover between cells of different sizes, and Proximity Services (ProSe) to allow near UE to communicate between them without the use of cellular network and the small cells. Moreover, release 12 also brings an optimal power ON/OFF to interference mitigation and an increased modulation of 256 Quadrature Amplitude Modulation (QAM) to improve spectral efficiency.

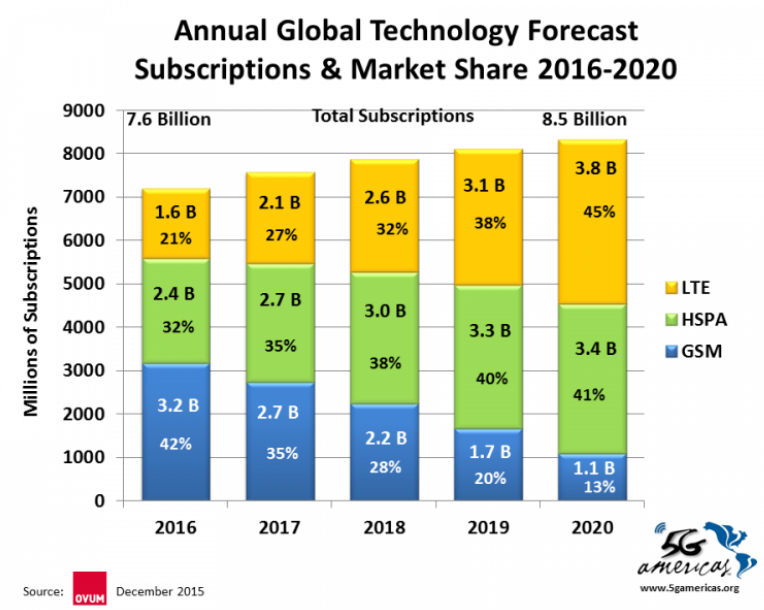


Figure 1.3: Annual Global Technology Forecast Subscriptions & Market Share 2016-2020 [9]

In the future two more releases will be implemented, release 13 and 14. They can be seen as an approach to the 5th Generation (5G). Release 13, will bring enhancements in multi-user transmission techniques, where the goal is using superposition coding to increase spectral efficiency. Another focus of release 13 is the aggregation of a primary cell, operating in licensed spectrum to deliver critical information and guaranteed Quality of Service (QoS), with a secondary cell, operating in unlicensed spectrum to opportunistically boost data rate.

The currently systems support MIMO systems with up to 8 antennas, release 13 will look into high-order MIMO systems with up to 64 antenna at the evolved Node B (eNB), to become more relevant with the use of higher frequencies and also exploiting the vertical dimension in beamforming [20]. Release 14 will mark the start of 5G, and the last step in 4G. Unlicensed spectrum has received a lot of attention in release 13 and will continue to be in focus. Reducing latency is another improvement to fully exploit the high data rates provided by LTE, it can also provide better support for new use cases, for example critical machine-type communication. Massive machine-type communication (MTC) will be a vital part of the overall vision of a networked society. LTE has already been enhanced in previous releases and is well positioned to combine low device cost with long battery lifetime.

LTE is currently on the market and the number of LTE subscribers will increase exponentially in the next years, with the result that in 2020 this will be the most used technology. As shown in figure 1.3, LTE is the fastest growing in percentage while the old GSM systems will be almost extinct with only a percentage of 13% of subscribers in 2020. In fact, GSM is currently being the most used technology as we can see in figure 1.2, although the paradigm will change over the next few years.

1.2 Towards 5G systems

During the second half of 2014, the noise around what will 5G mean and who will dominate, started rising beyond the research community. In mature LTE markets like the US, Korea, and Japan, the talk has shifted to the next generation technology evolution but in Europe which still has a long way to go before their 4G is built out have set their sights on 5G to recapture the mantle and the pride of the GSM days. The history said that Europe has not done much innovative work on a broad scale in mobiles since the glory days of GSM. On the other hand Japan is anxious to regain old glory and the Japanese government has set the ambitious goal of having 5G by the Tokyo Olympics in 2020. China which has had some excursions to set its own standards has warmed up to the global standards and is very keen on leaving its own stamp on 5G . Then of course, the US market is clearly the current leader in almost every dimension. US does have an extreme importance in the development of 5G since some of the most important companies are theirs. Qualcomm sets the baseline for growth in mobile and the mobile growth is led by the software such as Google, Facebook, Amazon, Microsoft, and Apple [21].

To be successfully implemented 5G systems will need to cope with three fundamental requirements for building wireless networks. One of them is the capability for supporting massive capacity and massive connectivity. The predictions show an increase up to one hundred times in the number of connections between different devices [22]. The second requirement relates to the necessity of supporting an increasingly diverse set of services, application and users all with extremely diverging requirements for work and life. Virtual Reality, Immerse or Tactile Internet, Autonomous driving, Connected cars, Wireless cloud-based office, Multi-people videoconferencing and Machine-to-machine connectivity are some examples of different types of services, as it is shown in figure 1.4.

The last requirement is to find a balance between flexibility and efficiency in use with all available non-contiguous spectrum for wildly different network deployment scenarios [23]. As a result of the developments in these three areas 5G will be able to support 1-10 Gbps connections to end points in the field with hundreds or thousands of users per cell, 1 millisecond end-to-end round trip delay (latency) is equal expected [23].

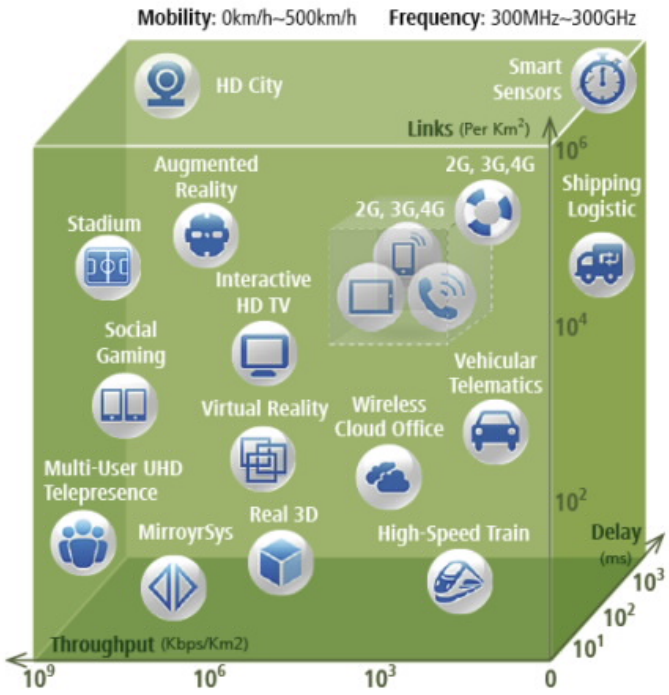


Figure 1.4: 5G service and scenario requirements [24]

For this purpose it is necessary to develop promising new transmission techniques, particularly the use of Millimeter Wave (mmWave) communications alongside with Multi-user MIMO (MU-MIMO) and small-cell (pico and femto) that are expected to be a crucial part of 5G systems, and there are already some standards like IEEE 802.11ad (WiGig), and IEEE 802.15.3c [25].

1.3 Motivation and Objectives

The number of subscribers that use mobile communications increases everyday. Today the number of terminals that have connections to the mobile network is higher than the number of people in the planet and continues increasing. The evolution of technology made possible for people that can not only make calls, but also have access to email, news, multimedia and navigation services almost in every terrestrial place of the planet. With this came the larger need for more data, for more capacity and for better performances that were demanded by the increasing number of users. The actual cellular architectures are obsolete and unable to cope with the increased demand for wireless services.

The emergence of mmWave with massive Multiple-Input Multiple-Output (mMIMO) is regarded as a promising technique for future 5G systems, where it could be possible to achieve multi Gb/s in wireless systems. Nowadays, mmWave communications and mMIMO have been considered as two of the key enabling technologies needed to meet the QoS requirements for future wireless communication [25]. In order to solve the global bandwidth shortage new mmWave spectrum frequencies were exploited for future broadband cellular communication networks [26]. mmWave based technologies have been standardized for short-range services, IEEE 802.11ad, and deployed for niche applications such as small cell backhaul, although the technology is not yet fully mature.

Together, more bandwidth and a large number of antennas, have also been considered an enabling technology for meeting the ever increasing demand of higher data rates in future wireless networks [27]. The use of mmWave with mMIMO is very attractive, since it allows packing more antennas in the same volume due to the smaller wavelength compared to current microwave communication systems and also terminals can be equipped with large number of antennas. Moreover, large antenna arrays in mMIMO can provide sufficient antenna gains to compensate the serious signal attenuation introduced at high frequencies [28]. MmWave mMIMO may exploit new and efficient spatial processing techniques such as beamforming/precoding and spatial multiplexing at the transmitter and/or receiver sides [28]. However, the precoding schemes for mMIMO are mainly performed in the baseband domain, where a fully digital precoder is utilized to eliminate the interferences by controlling both the amplitude and phase of transmitted signals and can achieve the optimal performance. Unfortunately, requires a higher number of RF chains, one for each antenna. As a result, the total hardware complexity and energy consumption of digital precoding becomes a serious problem due to the increasing number of antennas, specially for mmWave communications where the number of antennas at the Base Station (BS) can be huge. Another important issue is that the mmWave propagation characteristics are quite different from the ones at lower frequencies since the channels do not have such rich multipath propagation effects [29] [30], and this should be taken into account in the design of beamforming techniques.

To solve these problems a simple and immediate approximate was proposed to overcome the limitation on the number of RF chains, that is to make beamforming only in the analog domain by using phase shifters [31]. Nevertheless, the performance of the pure analog signal processing approach is limited by the availability of only quantized phase shifters and the constraints on the amplitudes of these phase shifters, and thus analog beamforming is usually limited to single-stream transmission. To overcome these limitations, hybrid analog/digital architectures, where some signal processing is done at the digital level and some left to the analog domain, have been discussed in [32] [33]. Recently, some beamforming and/or combining/equalization schemes have been proposed for hybrid architectures [34] [35] [36].

The aim of this dissertation is to implement a hybrid mmWave based architecture, where each RF chain is only connected to a small set of antennas, referred in Chapter 3 as *sub-array architecture*. The performance of this architecture is compared with the *full-array* based counterpart, where all the antennas are connected to all RF chains. The full-array architecture achieves close to optimal performance, but involves very high computational complexity. Furthermore, as each RF chain connects to all the antennas its hardware implementation is

complex [34]. On the other hand, the first one is a more realistic hybrid precoding structure where each RF chain is connected to only a small set of transmit antennas. For this architecture, the precoding scheme proposed in [36] will be implemented, where the analog and digital precoders are selected from a predefined candidate set by searching the optimal pair of analog and digital precoders. Finally, a hybrid analog/digital equalizer will be designed to efficiently separate the spatial streams. The implemented schemes will be evaluated under mmWave realistic channel models.

1.4 Contributions

The main contributions of this dissertation includes,

- Design of a new hybrid analog/digital equalizer for the implemented mmWave sub-array based architecture.

- Performance comparison of two architectures: the implemented sub-array based one and the full-array architecture proposed in [34].

1.5 Outline

This dissertation started with the introduction chapter, where a brief introduction about the evolution of mobile communications until today, the main LTE releases and the evolution of mobile systems towards to 5G systems is presented. In this first chapter the motivations and objectives are also described and presented. The rest of the dissertation is organized as follows,

In Chapter 2 the main characteristics of multiple antenna systems is presented. Namely, we briefly describe the diversity and spatial data multiplexing advantages brought by multiple antenna systems and corresponding enabling techniques.

Following, in Chapter 3, the mmWave propagation conditions are presented. After that, we discuss mMIMO as a promising technique to increase the capacity and energy efficiency. As a result of the higher frequencies the dimension of antennas is reduced and we show some ways to group them to efficiently form massive beamforming that makes the existence of networks composed by small cells possible. Finally some considerations regarding mmWave channel modeling are done.

In Chapter 4, a detailed description and comparison of two hybrid architectures are done. We start by presenting the characterization of the system, then the transmitter and receiver model as well as the precoding and combining schemes. Finally, the performance of mMIMO mmWave based architectures are compared in terms the Bit Error Ratio (BER).

The last Chapter 5, where the work conclusion and possible ways for future research and enhancements are presented, which finalizes this dissertation.

1.6 Notation

In this dissertation the following notation will be used: Lowercase letters, boldface lowercase letters and boldface uppercase letters are used for scalars, vectors and matrices, respectively. $(.)^{\mathrm{T}}$, $(.)^{\mathrm{H}}$, $(.)^{*}$ and $tr(.)$ represent the transpose, the Hermitian transpose, the conjugate and the trace of a matrix, E[.] represents the expectation operator. Consider a matrix $\mathbf{A}$, diag($\mathbf{A}$) which correspond to a diagonal matrix with entries equal to the diagonal entries of matrix $\mathbf{A}$. $\mathbf{A}$(i,l) denotes the element at row i and column l of a matrix $\mathbf{A}$. $\mathbf{I}_{\mathrm{N}}$ is the identity matrix of size $N \times N$.

Chapter 2

MIMO Systems

The field of wireless communication systems and network have experienced explosive growth and the number of connections among devices increased in the last years, leading to more interference. As an example, customers are using mobile phone applications like Multimedia Message Services (MMS), an extension of text messaging (SMS), that adds pictures, sound and video elements. Further, it is expected that the number of connections continues to increase as increasing the number of antennas per device. Along with this rapid growth comes the costumer demand for more and better applications, improved performance, and increased data rates with better QoS. The use of multiple-antenna techniques have an important role in the evolution of the wireless communications, since they made the increase of rates, coverage and capacity of the wireless systems possible. Additionally to the time and frequency techniques the MIMO systems brought the opportunity to exploiting the spatial dimension by creating different paths between the transmitter and receiver [37].

So, the aim of this chapter is to discuss some signal processing issues related with specific MIMO schemes. It starts by presenting several multiple-antenna schemes and the different ways to achieve diversity and follows with a detailed description of spatial multiplexing, detection and beamforming techniques, which are the essence of the MIMO transmissions.

2.1 Multiple Antenna Schemes

There are four multi-antenna formats that can be used, which differ from each other in the number of transmitter and/or receiver antennas and in level of complexity. These are termed Single-Input Single-Output (SISO), Single-Input Multiple-Output (SIMO), Multiple-Input Single-Output (MISO) and MIMO. These different MIMO formats offer different advantages and disadvantages, which can be balanced to provide the optimum solution for any given application.

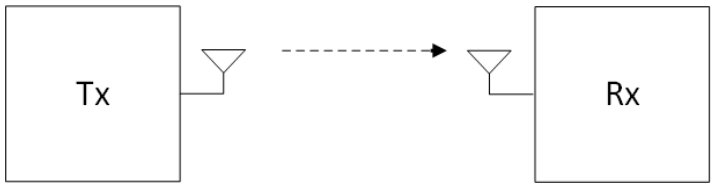


Figure 2.1: SISO configuration

The first configuration is a SISO which is the simplest form of radio link. This transmitter operates with one antenna as does the receiver and there is no diversity and no additional processing required. Obviously the advantage of SISO is the simplicity despite offering no protection against interference and fading and these events will have impact in the system. An example is shown in figure 2.1.

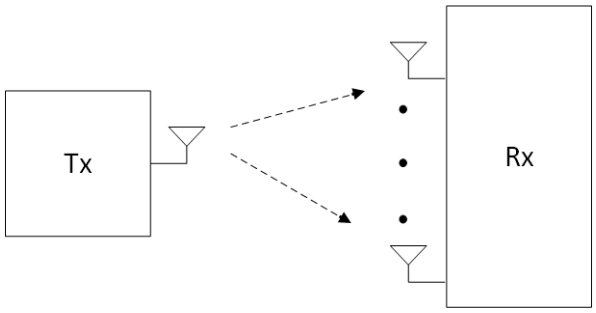


Figure 2.2: SIMO configuration

In the second configuration, the receiver has multiple antennas (M_R) while the transmitter remains with only one. This scheme is also known as receive diversity since the received signals can arrive from several independent sources, what has as consequence a higher processing requirement on the receiver that can be a limitation because of size, cost and battery drain in the case of a mobile device. An example is shown in figure 2.2.

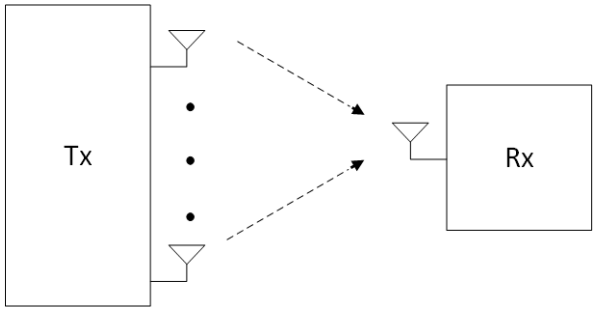


Figure 2.3: MISO configuration

Similarly, MISO is a transmission diversity technique which has multiple antennas on the transmitter side (M_T) and just one in the receiver, also known as a transmit diversity scheme because the data is transmitted redundantly by different antennas. The main advantage in this case is the change of the complexity from the receiver to the transmitter, compared with the previous scheme. The receiver is able to reconstruct the original data from the best signal received. This can be a great advantage because of the reduction of processing in the mobile equipment that consequently is better in terms of battery consumption [38]. An example is shown in figure 2.3.

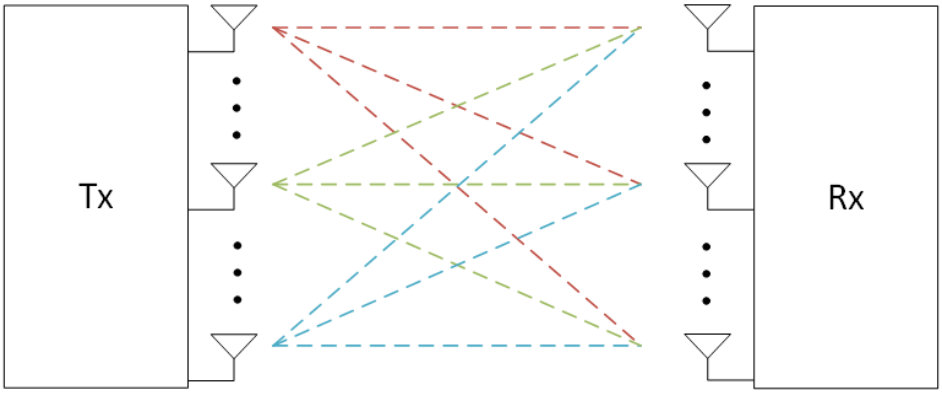


Figure 2.4: MIMO configuration

MIMO has multiple antennas both the transmitter (M_T) and the receiver (M_R), as shown in figure 2.4, that provides channel robustness and a higher throughput in data capacity. This scheme has more Degrees of Freedom (DoF) than the previous ones to improve the system performance and it is expected that it may be the most widely used. Unfortunately MIMO has some disadvantages, in terms of hardware; the complexity increase because each antenna needs a Radio Frequency (RF) unit. The software complexity also increases and most of signal processing algorithms are computationally intensive and obviously the power consumption grows and reduces the battery lifetime of mobile devices as a consequence of thermal problems. Moreover, the high number of antennas generate RF interference and antenna correlation [39]. To solve these setbacks various solutions have already been proposed, making MIMO a very promising system.

2.2 Diversity

If a fading radio signal is received through only one channel, then in a deep fade, the signal could be lost, and there is nothing that can be done. The key is to create multiple channels or branches that have uncorrelated fading. Therefore, increasing the number of independent paths across which the information symbol is repeated increases the received Signal-to-Noise Ratio (SNR), and consequently decreases the error probability, resulting in BER curves that tend to Added White Gaussian Noise (AWGN) BER, which is characterized by just affect the link with noise. This principle is called diversity and it is a way of protection against deep fades. Diversity can be applied in three main different ways which will be presented. These

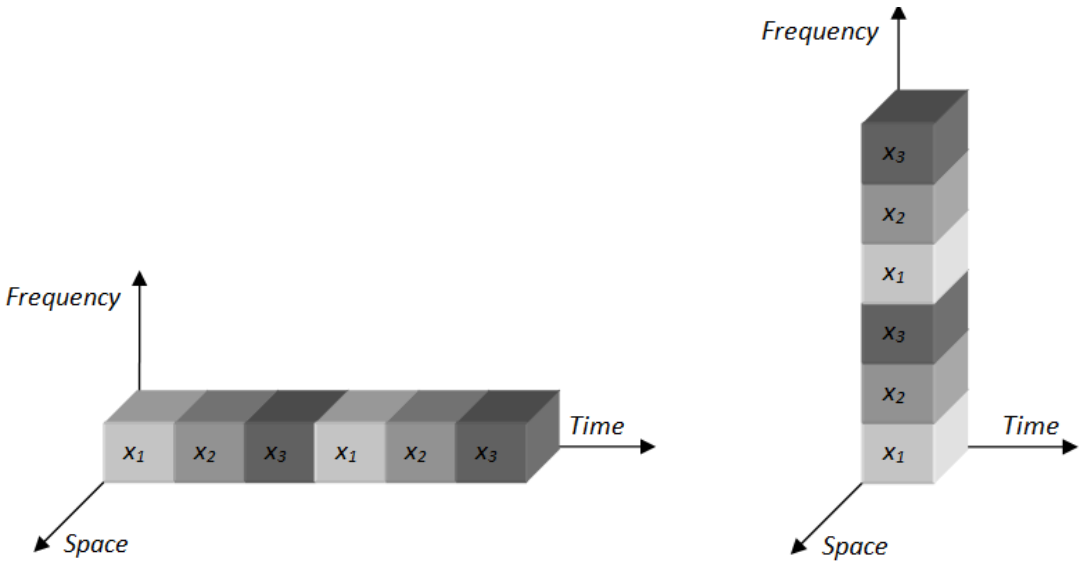
methods, therefore, reduce the probability that all the replicas are simultaneously affected by severe attenuation and noise [40] [41].

Commonly used diversity techniques are:

- **Temporal diversity**, can be achieved by sending the same information at different times if these are separated by much more than the coherence time. In practice time diversity is usually associated with interleaving and coding over symbols across different coherent time periods. An example is shown in figure 2.5(a).

- **Frequency diversity**, is achieved by transmitting the same narrowband signal at different carriers, where the carriers are separated by the coherence bandwidth of the channel. An example is shown in figure 2.5(b).

- **Antenna or Spatial diversity**, the same information-bearing signal is transmitted or received via different antennas where the maximum gain can be achieved when the fading occurring in the channel is independent (or low correlated). In the receiver, diversity gain is achieved by combining the redundant signals arriving via independent channels, without any loss in bandwidth efficiency [42].



(a) Time diversity

(b) Frequency diversity

Figure 2.5: Time and Frequency Diversity [43]

Unfortunately Temporal diversity and Frequency diversity has some drawbacks. Temporal diversity decrease the data rate by a factor of L (where L is the number of independent paths where the information flows) with repetition code, whereas the Frequency diversity usually requires more bandwidth (due to coding repetition), which is a problem because

this is a limited resource. However, the Spatial diversity is achieved without increasing the bandwidth and the transmitted power, therefore there is a great interest for cellular communications [44] [45].

2.2.1 Receive Diversity

It is in the uplink transmissions that the receive diversity is mostly used. We can use not only time and frequency diversity, but also spatial diversity at the reception. With multiple antennas at the receiver (M_R), we can decrease the influence of the multipath channel effect using spatial diversity. However, the space between antennas must be wide enough to prevent the mutual correlation between channels. In figure 2.6 we can observe that the fading occurs at different times because antennas process or filter the multipath in different ways allowing deep fade reducing by the combination of two signals received by two antennas [45].

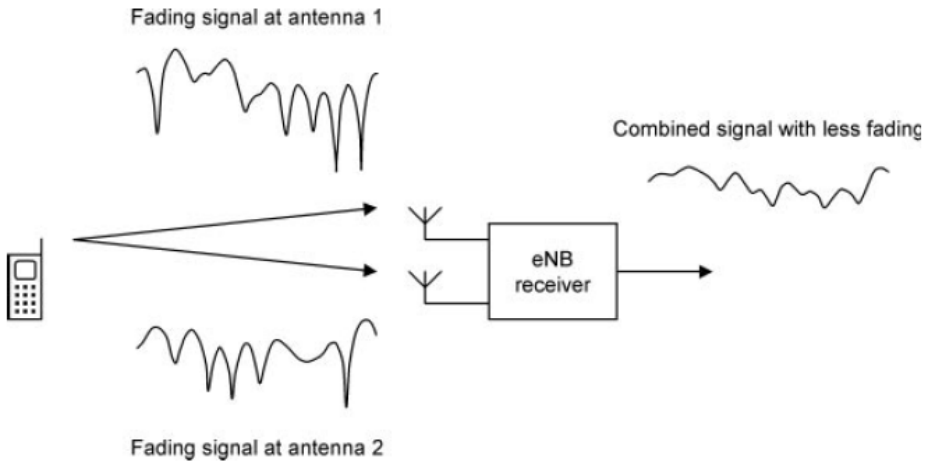


Figure 2.6: Reducing fading by using a combination of two uncorrelated signals [8]

In general, the performance of communication systems depends on how multiple signal replicas are combined at the receiver to increase the overall received SNR. This scheme can achieve diversity gain and antenna gain. The diversity gain is related to the fact that the channels are independent and the antenna gain is related to the fact that the noise terms added at each receiver are independent [46]. The use of M_R antennas at the receiver will allow the reception of the symbols across h_{M_R} channels. We can use different processing combination techniques to improve the SNR. The aim of these type of combination schemes is to use the M_R copies of the signal that arrive at each antenna of the receiver to combine them like it has shown in figure 2.7.

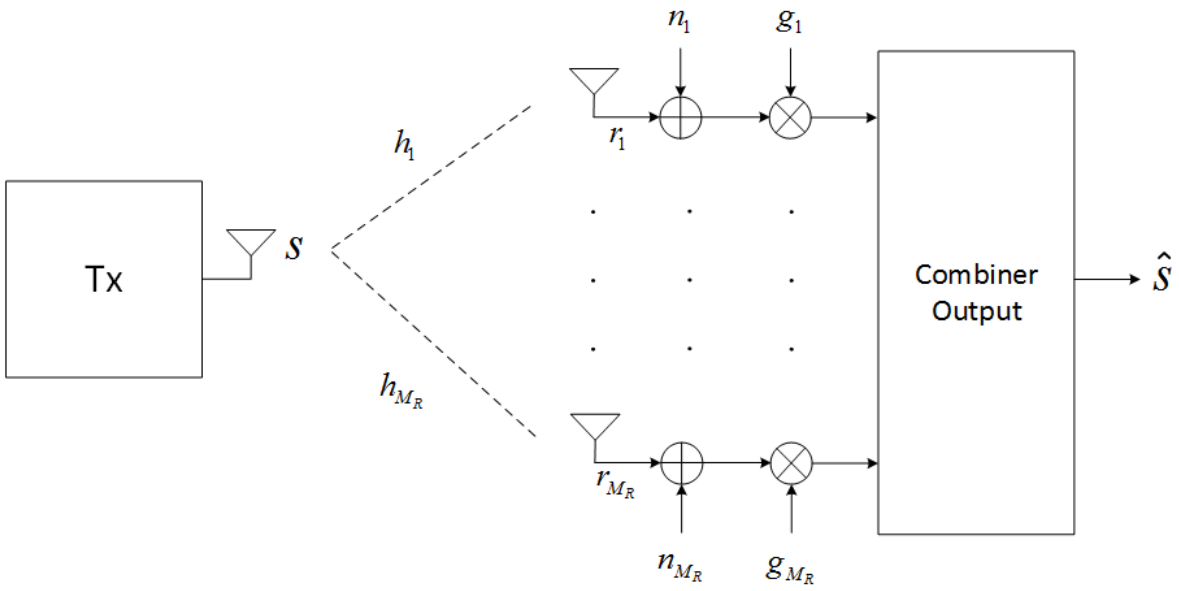


Figure 2.7: Spatial receive antenna diversity

The overall signal model for the received symbol will be the following

$$\begin{bmatrix} r_1 \\ \vdots \\ r_{M_R} \end{bmatrix} = \begin{bmatrix} h_1 \\ \vdots \\ h_{M_R} \end{bmatrix} s + \begin{bmatrix} n_1 \\ \vdots \\ n_{M_R} \end{bmatrix} \tag{2.1}$$

and the estimated symbols are given by

$$\hat{s} = \begin{bmatrix} g_1 & \dots & g_{M_R} \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_{M_R} \end{bmatrix} \tag{2.2}$$

$$\hat{s} = \begin{bmatrix} g_1 & \dots & g_{M_R} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_{M_R} \end{bmatrix} s + \begin{bmatrix} g_1 & \dots & g_{M_R} \end{bmatrix} \begin{bmatrix} n_1 \\ \vdots \\ n_{M_R} \end{bmatrix} \tag{2.3}$$

According to the implementation complexity and the level of channel state information required by the combining method, there are three main types of combining techniques, including Maximal Ratio Combining (MRC), Equal Gain Combining (EGC) and Selection Combining (SC) [44] [46] [47].

Maximal Ratio Combining

MRC is a linear combining. Various signal inputs are individually weighted and added together to get an output signal. The weighting factor of each receiver antenna can be chosen to be in proportion to its own signal voltage to noise power ratio. This scheme requires the knowledge of channel fading amplitude and signal phases. MRC is usually used when we want to maximize the SNR in order to eliminate bad noise conditions at the reception. Hence, the g_m coefficients computed, are equal to the conjugate transpose $(.)^{\mathbf{H}}$ of the channel.

The optimal coefficients/weights for maximal-ratio combining in fading is

$$g_m = h_m^* \qquad m = 1, \ldots, M_R \tag{2.4}$$

Therefore, with accurate channel knowledge $\mathbf{h}$ at the receiver and using these weights the signal at the output combiner is

$$\hat{s} = \sum_{m=1}^{M_R} \left|h_m\right|^2 s + \sum_{m=1}^{M_R} h_m^* n_m \tag{2.5}$$

In this case the antenna gain A_g will be

$$A_g = M_R \tag{2.6}$$

Note that MRC maximizes SNR aligning the phases of all M_R channel coefficients, and also giving more weight to the best channels.

Equal Gain Combining

EGC is a suboptimal but simple linear combining method. This scheme just rotates the phases of the arrived signals at each antenna and then adds them together with equal gain.

The weights are given by

$$g_m = \frac{h_m^*}{|h_m|} \qquad m = 1, \ldots, M_R \tag{2.7}$$

Using these weights the signal at the output combiner is

$$\hat{s} = \sum_{m=1}^{M_R} \left|h_m\right| s + \sum_{m=1}^{M_R} \frac{h_m^*}{|h_m|} n_m \tag{2.8}$$

In this case the antenna gain A_g will be smaller than the MRC case

$$A_g = 1 + \frac{\pi}{4}(M_R - 1) \tag{2.9}$$

If we compare 2.6 and 2.9 the performance of EGC is marginally inferior to maximum ratio combining and the implementation complexity for EGC is less than the MRC.

Selection Combining

A Selection Combining is a simple diversity combining method. This algorithm compares the instantaneous amplitude of each channel and chooses the antenna branch with the highest amplitude. Thus the received signals in the other antennas are ignored. In practice, the signal with the highest sum of signal and noise power is usually used, since it is difficult to measure the SNR. This may be implemented with a single receiver or with two receivers. A single receiver is clearly desirable from a cost, size and power consumption perspective, however some time must be made available for the antenna signals to be compared.

The selection algorithm comes to compare the instantaneous amplitude of each channel and choose the branch with the largest amplitude

$$|h|_{max} = max\Big[|h|_1, \ldots, |h|_{M_R}\Big] \tag{2.10}$$

The weight is given by

$$g_{max} = h^*_{max} \tag{2.11}$$

and the estimated symbols are

$$\hat{s} = |h_{max}|^2 x + h^*_{max} n_{max} \tag{2.12}$$

Increasing M_R yields diminishing returns in terms of antenna gain, therefore the biggest gain is obtained from $M_R = 1$ to $M_R = 2$, as shown by the following equation.

$$A_g = \sum_{m=1}^{M_R} \frac{1}{m} \tag{2.13}$$

So we can see that in SC scheme A_g keeps growing with the receiver antenna number, but in a non-linear way, and lesser than in MRC case and EGC case.

2.2.2 Transmit Diversity

Transmit diversity uses two or more antennas at the transmitter side, and consequently reduces the fading. This case, may be a quite similar to the receive diversity, however it can provide an opportunity for diversity without any need for additional receive antennas and the corresponding additional receiver chains at the terminal. Sending a signal from the transmitter antennas to the receiver without any coding/precoding, the signal just will be added together, which brings a risk of destructive interference, so it cannot achieve diversity. To obtain it, a precoder or a space-time/frequency coding can be used. There are two ways to obtain it: *closed loop transmit diversity* and *open loop transmit diversity.*

Closed loop

In the closed loop form, the transmitter sends copies of the signal over the different antennas, but before the transmission the signal is formated to improve the system performance by taking into account the Channel State Information (CSI). At the receiver, the signals will be phased that avoid destructive interference. To ensure that the signals are received in phase, the phase shift is calculated by the receiver and fed back to the transmitter as a Precoding

Matrix Indicator (PMI). A simple PMI might indicate if the signals must be transmitted without a phase shift, or instead, must be applied to a phase shift. In case of the first option it leads to destructive interference, then the second will automatically work. This implies that the transmitter has previous acknowledge of the CSI [46]. The phase shifts occur in the frequency domain and the PMI comes as a function of frequency.

In conclusion, if UEs are moving fast a delay is introduced in the feedback loop, the PMI will change and cannot correspond to the best choice when it is received because the channel state changes very quickly. This technique is suitable to relatively slowly moving mobiles. For users in fast movement the best choice could be the open loop techniques [8].

Open loop

In the open loop form, diversity is usually performed with space-time/frequency coding. Instead of a closed loop, the open loop techniques don't require CSI, which offers robustness against non-ideal conditions (like antenna correlation, channel estimation errors and Doppler effects). Typically, there are mainly two types of space-time codes: trellis and block codes. The first ones usually outperform the block codes but at cost of higher complexity, and for that reason their use in practical systems is limited. There are two types of block codes, Space-time Block Code (STBC) and Space-frequency Block Code (SFBC) [46] [48].

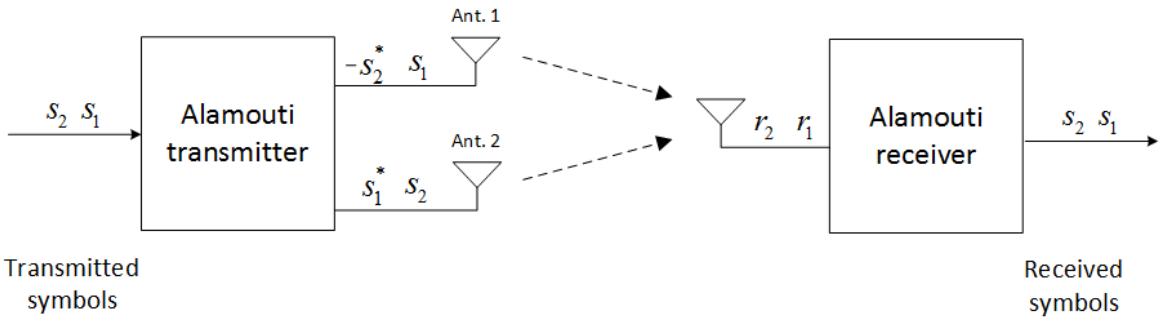


Figure 2.8: Alamouti technique for open loop transmit diversity

Figure 2.8 shows Alamouti's block diagram for the open loop transmit diversity. After modulating the information bits, using an M-ary modulation (e.g QAM), a block of two data symbols s_1 and s_2 are coded in two antennas and in two frequency instants.

The matrix is

$$\mathbf{S} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} \tag{2.14}$$

which means that in first time slot s_1 and s_2 are transmitted by antenna 1 and 2, respectively. In the second time slot $-s_2^*$ and s_1^* are transmitted by antenna 1 and 2, respectively.

If we perform the operation

$$\mathbf{S} \times \mathbf{S}^H = \begin{bmatrix} |s_1|^2 + |s_2|^2 & 0 \\ 0 & |s_1|^2 + |s_2|^2 \end{bmatrix} \tag{2.15}$$

we can show that $\mathbf{S}$ matrix is orthogonal. s_1 and s_2 are independent thus Alamouti code is an orthogonal code.

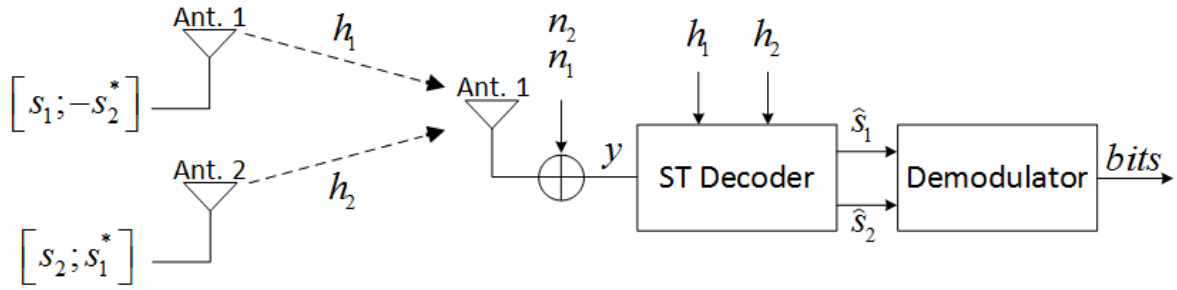


Figure 2.9: Alamouti receiver block diagram

The receiver side of Alamouti is represented in figure 2.9. Considering $\mathbf{h} = \begin{bmatrix} h_1 & h_2 \end{bmatrix}$ as the overall vector channel, h_1 and h_2 represents the channel coefficients of antenna 1 and 2, respectively, $\mathbf{n} = \begin{bmatrix} n_1 & n_2 \end{bmatrix}$ is the additive white Gaussian noise and the received signal $\mathbf{y}$ in matrix notation is given by

$$\mathbf{y} = \frac{1}{\sqrt{2}}\mathbf{hS} + \mathbf{n} \tag{2.16}$$

Each symbol is multiplied a factor $1/\sqrt{2}$ in order to normalize the power per symbol. The previous equation may be written as

$$\begin{cases} y_1 = \frac{1}{\sqrt{2}}h_1 s_1 - \frac{1}{\sqrt{2}}h_2 s_2^* + n_1 \\ y_2 = \frac{1}{\sqrt{2}}h_1 s_2 + \frac{1}{\sqrt{2}}h_2 s_1^* + n_2 \end{cases} \tag{2.17}$$

If the channels of two adjacent frequencies are highly correlated we can assume that $h_{k,n} \approx h_{k,n+1}$, the estimated symbols after dividing are given by

$$\begin{cases} \hat{s}_1 = \frac{1}{\sqrt{2}}h_1^* y_1 + \frac{1}{\sqrt{2}}h_2 y_2^* + n_1 \\ \hat{s}_2 = \frac{1}{\sqrt{2}}h_1 y_1^* + \frac{1}{\sqrt{2}}h_2 y_2^* + n_2 \end{cases} \tag{2.18}$$

With channel knowledge available at the receiver, we will decode the symbols $\hat{s}$, using the matched filter version of $\mathbf{H}_{eq}$, defined by

$$\mathbf{H}_{eq}^H = \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \tag{2.19}$$

consequently, the soft decision of data symbols are given by

$$\hat{\mathbf{s}} = \mathbf{h}_{eq}^{H}\mathbf{y} \tag{2.20}$$

$$\hat{\mathbf{s}}_n = \frac{1}{2}(h_1^* h_1 + h_2^* h_2)s_n + \frac{1}{\sqrt{2}}h_1^* n_n + \frac{1}{\sqrt{2}}h_2 n_{n+1}^* \tag{2.21}$$

From 2.21, interference caused by $n+1$ data symbol is eliminated, thus SNR is

$$SNR = \frac{1}{2}\frac{(|h_1|^2 + |h_2|^2)}{\sigma_n^2} \tag{2.22}$$

This presentation was made for a system with 2×1 antennas, where Alamouti scheme can achieve a diversity order of 2. Although it could be generalized for more than one antenna at the receiver side. i.e for $2 \times M_R$ system. In this case the diversity order that can be obtained is $2M_R$ with an antenna gain of M_R. The above Alamouti STBC scheme is only available for 2 antennas at the transmitter. One possible solution of Alamouti coding in the case of 4 antennas transmission is performing a time and space shift of Alamouti blocks using a STBC - Time Shift Transmit Diversity (STBC - TSTD) scheme as considered in the LTE, as is shown in Table 2.1.

	t_0	t_1	t_2	t_3
f_0	s_0	$-s_1^*$	0	0
f_0	s_1	s_0^*	0	0
f_0	0	0	s_2	$-s_3^*$
f_0	0	0	s_3	s_2^*

Table 2.1: STBC - TSTD mapping [8]

2.3 Spatial Multiplexing

Spatial multiplexing has a different purpose from diversity processing. SIMO and MISO systems provide diversity and antenna gains but no multiplexing gain. So if the transmitter and receiver both have multiple antennas, then we can set up multiple parallel data streams between them. We can increase the data rate achieving also maximum capacity for the same bandwidth and without additional power. Another important aspect that must be verified to achieve high spatial multiplexing gain is low antenna correlation channel conditions. Therefore a high degree of difference between the channels is needed to perform the separation of multiple layers without interference. In a system with M_T transmit and M_R receive antennas, often known as an $M_\mathrm{T} \times M_\mathrm{R}$ spatial multiplexing system, the peak data rate is proportional to $min(M_T, M_R)$ compared with the SISO system [8].

Another important thing is that we can combine diversity and multiplexing in the same multiple antenna system, but it is impossible to achieve their full benefits. If we consider a 2 x 2 system, to have full diversity each symbol must go through two independent channels so diversity is two ($d = 2$) and the multiplexing gain is one ($r = 1$). Now, to have full

multiplexing on the same system, each channel is used by only one data stream ($r = 2$) but there is no repetition of the symbols, so the diversity is one ($d = 1$). In the diversity, the data rate is constant and the BER decreases as the SNR increases and, in multiplexing, BER is constant while the data rate increases with SNR [48].

2.3.1 SU-MIMO Techniques for Spatial Multiplexing

Using optimal Singular Value Decomposition (SVD), we are able to estimate the capacity of each channel in order to select the best channel to adapt the transmission. This adaptation is done by performing a power allocation according the singular values computed using SVD. We also obtain the optimum signal precoding to perform at the transmitter, and the necessary information for ideal equalization at the receiver. Denoting that CSI must be available at both transmitter and receiver. However, sometimes just the receiver has a precise channel information and in these cases the solution to decode the signal is just based on some equalization scheme performed at the receiver [40] [12].

Channel known at the transmitter

In the diagram of figure 2.10 it is represented the way of processing two spacial streams (green and orange), with spatial multiplexing.

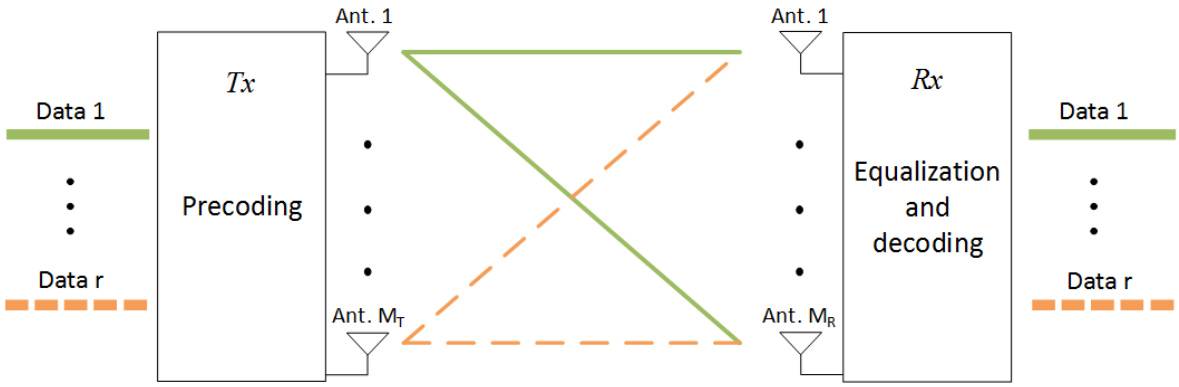


Figure 2.10: Spatial multiplexing in a MIMO System

Let's analyze in detail the whole process to achieve spatial multiplexing. Inside the transmitter data symbols in the different antennas are precoded. The precoders are computed based on the channel information that should be fed back from UE to the BS, considering FDD mode, or estimated in the uplink if the TDD mode is assumed (channel reciprocity) [15].

Assuming that $\mathbf{H} = \begin{bmatrix} h_{11} & \dots & h_{1,j} \\ \vdots & h_{i,j} & \vdots \\ h_{i,1} & \dots & h_{i,j} \end{bmatrix}$ i = 1,..., M_R; j = 1,...,M_T is the channel matrix, $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_{M_T} \end{bmatrix}^T$ is the vector of the transmitted signal, $\mathbf{n} = \begin{bmatrix} n_1 & \dots & n_{M_T} \end{bmatrix}^T$

is the noise vector and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_{M_R} \end{bmatrix}^T$ defined by

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} \tag{2.23}$$

The transmitted signal $\mathbf{x}$ is defined as

$$\mathbf{x} = \mathbf{Ws} \tag{2.24}$$

where $\mathbf{W}$ is the precoder matrix and $\mathbf{s} = [s_1, \dots, s_r]^T$ is the data stream transmitted over the M_T antennas and $r = min(M_R, M_T)$.

Another way to describe the channel matrix is using the SVD. Channel $\mathbf{H}$ is now written as

$$\mathbf{H}=\mathbf{UDV^T} \tag{2.25}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary with the size $M_\mathrm{R} \times r$ and $M_\mathrm{T} \times r$ respectively, $r = rank(\boldsymbol{H}) \leq min(M_R, M_T)$. $\mathbf{D}$ is the diagonal matrix composed by non-negative real numbers, defined by

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_r \end{bmatrix}. \tag{2.26}$$

The precoding matrix $\mathbf{W}$ is defined by

$$\mathbf{W}=\mathbf{VP}^{\frac{1}{2}} \tag{2.27}$$

where the size is $M_\mathrm{T} \times r$, and $\mathbf{P}$ is a square diagonal power allocation matrix of size $r \times r$ defined by

$$\mathbf{p} = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_r \end{bmatrix}. \tag{2.28}$$

Setting the equalizer matrix as

$$\mathbf{G}=\mathbf{U}^H \tag{2.29}$$

Replacing in 2.23, the equations 2.24, 2.25 and 2.27, then

$$\mathbf{y}=\mathbf{UDV}^T\mathbf{VP}^{\frac{1}{2}}\mathbf{s} + \mathbf{n} \tag{2.30}$$

The estimated data symbols in the receiver are obtained and given by

$$\hat{s} = \mathbf{Gy} = \mathbf{U^H UDV^T VP}^{\frac{1}{2}}\mathbf{s} + \mathbf{n} \tag{2.31}$$

$$\hat{s} = \mathbf{DP}^{\frac{1}{2}}\mathbf{s} + \widetilde{n} \tag{2.32}$$

The soft estimate of the rth data symbol is defined as

$$\hat{s}_i = \lambda_i \sqrt{p_i} s_i + \widetilde{n}_i \qquad i = 1, \ldots, r \tag{2.33}$$

The channel capacity via SVD decomposition is the following

$$C = \sum_{i=1}^{r} \log_2 \left(1 + \frac{\lambda_i^2 p_i}{\sigma^2}\right) \qquad bits/s/Hz \tag{2.34}$$

The selection of power that we will allocate in each channel is done in order to maximize the system capacity. The amount of power put in each channel is done using the *Water Filling* power algorithm. Firstly, we select a limit called water level and then the amount of power is distributed to the different channels. If the SNR $\left(\frac{\lambda_i^2}{\sigma^2}\right)$ of a channel is below the water level the power allocated in this channel will be zero [49]. In figure 2.11 a specific realization of the *Water Filling* algorithm is presented.

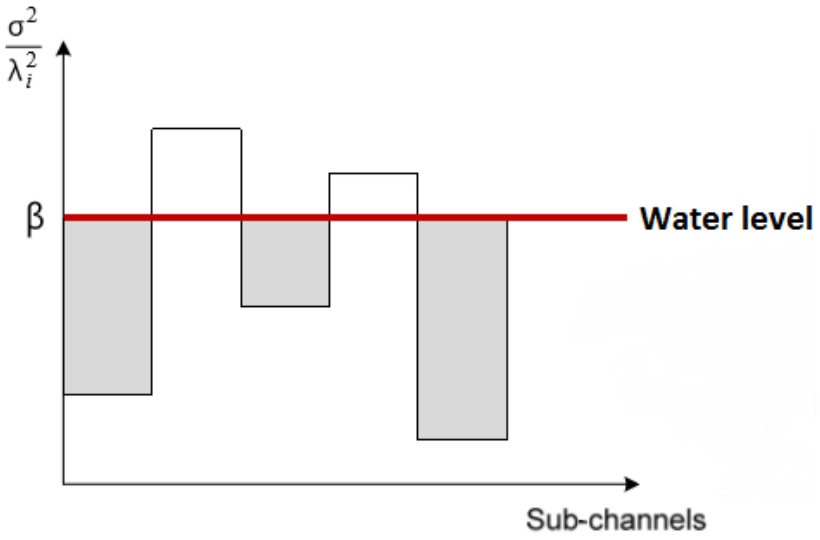


Figure 2.11: Water Filling power scheme

Like we have seen the channels have different levels and the power must be distributed according to the SNR of each channel. Obviously, more power is allocated to better channels and a reduced amount of power in bad channels. One way to find the best power allocation to optimize the system is using the Lagrangian method.

The optimal power allocation is then given by,

$$p_i = \left(\beta - \frac{1}{SNR_i}\right) = \left(\beta - \frac{\sigma^2}{\lambda_i^2}\right) \tag{2.35}$$

where i, β and σ^2 are respectively, the channel index, the water level value and the noise power.

Channel known only at the receiver

Like it was said before sometimes just the receiver has a precise channel information and we are not able to use the SVD channel decomposition. In this case any precoding is performed at the transmitter side and the solution is just based in an equalization scheme performed at the receiver. We can use linear equalizers, like Zero-Forcing (ZF) and Minimum Mean Square Error (MMSE), or non-linear equalizers like Successive Interference Cancellation (SIC) or Parallel Interference Cancellation (PIC) applied together with linear equalizers [50] [51] [52]. SIC and PIC establish a good compromise between performance and complexity improving the performance of linear equalizers ZF and MMSE.

Note, the signal model that will be considered is the same that was used in the previously subsection.

- **ZF Equalizer**

  The aim of this technique is to design an equalizer vector for each data symbol to remove the interference using at least the same number of antennas at the receiver and transmitter ($M_R \geq M_T$).

  The coefficients of this equalizer are given by

$$\mathbf{G}_{ZF} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H \tag{2.36}$$

  Despite of full symbol separation ZF amplifies the noise term which can be a problem in channels in deep fading.

- **MMSE Equalizer**

  Instead of the use of ZF, we can improve the SNR using MMSE equalization. This latter one can enhance the noise when the channels are in deep fading or the channels coefficients are correlated. This equalizer can be considered a trade-off between noise enhancement and interference removal. It does not enable to achieve full interference removal but prevents noise amplification.

  The coefficients are obtained by minimizing the mean square error between the transmitted signal and the equalized output signal. These coefficients are given by

$$\mathbf{G}_{MMSE} = \left(\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}_{M_t}\right)^{-1}\mathbf{H}^H \tag{2.37}$$

In figure 2.12 four combinations to perform equalization are compared. Two conclusions can be drawn, firstly the MMSE equalization is always better than ZF. Secondly, the SIC always improves the performance of the elementary linear equalizers, as mentioned above.

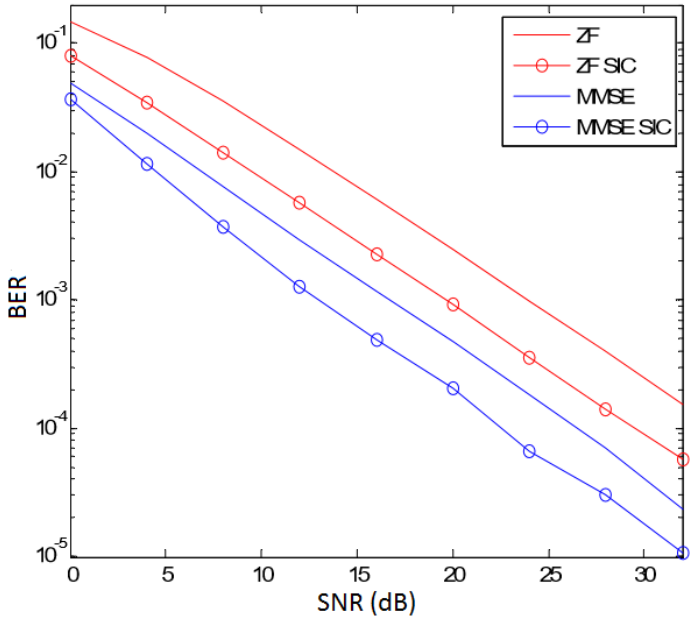


Figure 2.12: Performance comparison between the different equalizers [46]

2.3.2 MU-MIMO Techniques for Spatial Multiplexing

MU-MIMO is typically more used in the uplink case in cellular communications, where various mobiles can transmit at the same time and frequency to the same BS. The principles used in Single-User MIMO (SU-MIMO) are similar to the used in MU-MIMO, however in SU-MIMO the ZF and MMSE equalizers are performed to separate the layers at the receiver while in MU-MIMO the separation is performed at the BS. Moreover, in SU-MIMO we perform the processing in the signal already affected by the channel and noise. By contrasting in MU-MIMO we anticipate channel effect in the signal, and according to that, we adapt the transmitted signal before the channel affects it. The use of ZF and MMSE in the transmitter can only happen because the mobiles are probably sufficiently far from each other, which means that their signal paths are different. To make the reception of signals possible from different users the number of antennas on the receiver should be greater than the number of users sending information. The main challenges in this implementation are the need of CSI at the transmitter and also a scheduling procedure to make the selection of users to serve simultaneously [53]. Uplink MU-MIMO does not increase peak data rate of each mobile but increases cell throughput and can be implemented in non-expensive mobiles with only one antenna. An illustration of the described case is shown in figure 2.13.

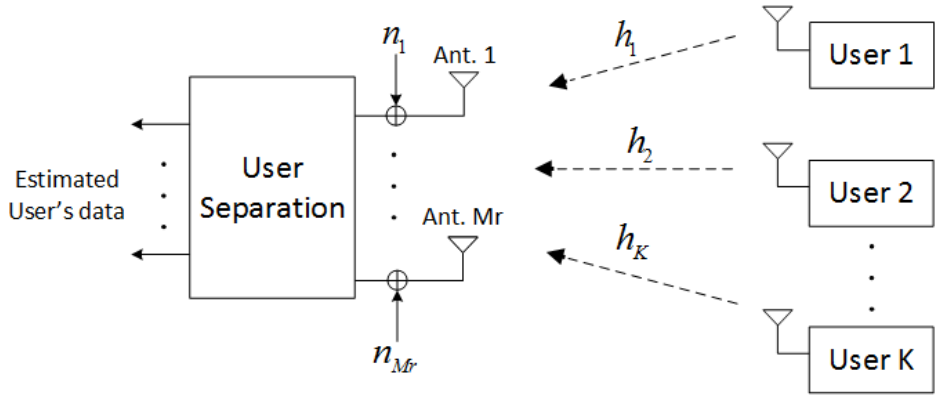


Figure 2.13: Uplink MU-MIMO transmissions

In the downlink, it is also possible to apply MU-MIMO but the implementation is not so trivial. The solution in the downlink is the use of beamforming technique, which will be described below, making that the desired symbol is sent to a given receiver with a constructive beam.

2.4 Beamforming

The main goals of beamforming techniques is to increase cell coverage and/or capacity. Beamforming techniques are used within many different technologies such as radar, sonar, seismology, radio astronomy, acoustics and wireless communications [54]. While spatial multiplexing and diversity work better if the antennas are far apart between them, with uncorrelated signals, beamforming works better if the antennas are closer together, and the signals sent or received are highly correlated. In the general case, transmit beamforming works by exploiting the interference patterns observed whenever the same signal is transmitted from two or more spatially separated transmission points.

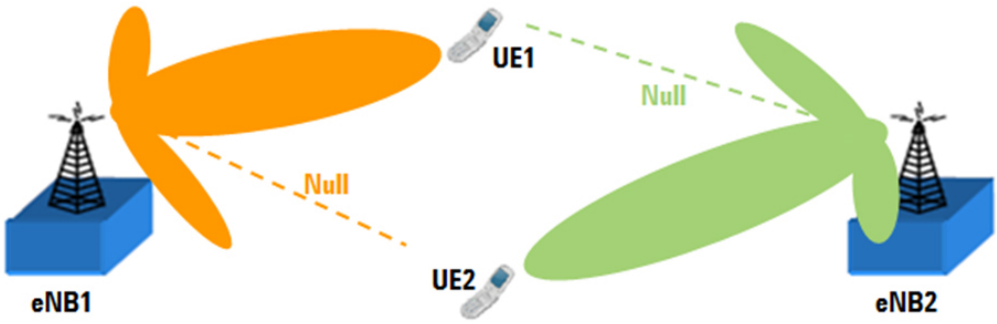


Figure 2.14: Beamforming technique [55]

As illustrated in figure 2.14, the transmitted signal is formatted in order to put a null in the interfered user and a main lobe in the desired user, just applying the suitable set of antenna weights it is possible to adjust amplitudes and phases of the transmitted signals. In beamforming we use an array of antennas, where each antenna sends the same symbol doing a phase shift according the CSI.

Chapter 3

Millimeter Wave and Massive MIMO Systems

The rapid increase of mobile data growth and the use of smartphones and other mobile data devices such as netbooks and ebook readers are creating unprecedented challenges for wireless services [31]. The current 4G systems including LTE and WiMAX already use advanced technologies in order to achieve spectral efficiencies close to the theoretical limits in terms of bits per second. However LTE has been deploying and is reaching maturity, but it may not be enough to support all requirements. 5G will need to be a paradigm shift that includes the use of much greater spectrum allocations at untapped mmWave frequency bands, highly directional beamforming antennas, longer battery life, lower outage probability, much higher bit rates in larger portions of the coverage area, lower infrastructure costs, and higher aggregate capacity for many simultaneous users [25].

So, the aim of this chapter is to discuss the main issues of mmWave, mMIMO, massive beamforming and follows with the different proposals for antenna architectures at higher frequencies. It also presents the fundamentals of channel modeling to be used in mmWave mMIMO systems and finishes with a briefly description of networks composed by small cells.

3.1 Millimeter Waves

Almost all mobile communication systems today use spectrum in the range of 300 MHz-3 GHz and in this section the reason why the wireless community should start looking at the 3-300 GHz spectrum for mobile broadband applications. Almost all commercial radio communications among AM/FM radio, satellite communication, GPS and Wi-Fi containing a narrow band of RF spectrum between 300 MHz - 3 GHz, that is usually called "sweet spot" due to its favorable propagation characteristics [56]. The global bandwidth shortage has motivated the exploration of the underutilized mmWave frequency spectrum, being this one of the key advantages, and represents a large amount of spectral bandwidth available for future broadband cellular communication networks [57].

The current 4G systems including LTE and WiMAX already use advanced technologies such as Orthogonal Frequency Division Multiplexing (OFDM), MIMO, multi-user diversity, link adaptation, turbo code in order to achieve spectral efficiencies close to theoretical limits in terms of bits per second per Hertz per cell, however it is not the sufficient for future necessities. mmWave has been widely used for long range point-to-point communication for many years. However, the antennas and components used in these systems including Power Amplifiers (PA), Low Noise Amplifiers (LNA), mixers, oscillators, synthesizers, modulators, demodulators are too big in size and consume too much power to be applicable in mobile communication [31] [56]. The steady advancement of semiconductor technologies (PAs, LNA's, mixers, antennas, ...) implemented with GaAs, SiGe, InP, GaN, or CMOS processes has triggered the possibility of utilizing mmWave frequencies for the next generation cellular data networks [58]. Recent studies suggest the combination of cost-effective CMOS technology, that can now operate well into the mmWave frequency bands, and high antenna gain at the mobile and the BS, promoting the viability of mmWave wireless communications [25]

With increasing of RF carrier frequency, the mmWave wavelength is reduced and this is a way to exploit polarization [59–61] and spatial processing techniques [34, 62–66], such as mMIMO and adaptive beamforming, in large antenna arrays. The operators want to create smaller cells, referred as pico and femto-cells, with ranges in the order of 10-200 m, in order to increase frequency reuse. Smaller cells are attractive for the mmWave spectral band and the large beamforming gains achievable with very large antenna arrays can conversely to mitigate the high mmWave path loss [26]. However backhaul become essential in this system to promote flexibility and quality of communications.

Recently the portion above 3 GHz has been explored for short-range and fixed wireless communications. For example Ultra-Wideband (UWB) in range of 3.1-10.6 GHz has been proposed to enable high data rate connectivity in personal area networks. mmWave systems for point-to-point communication already exist and they can achieve multi gigabit data rates at a distance of up to 1 km and in 2003 the Federal Communications Commission (FCC) announced that the 71-76 GHz, 81-86 GHz and 92-95 GHz frequency bands had become available to ultra-high-speed data communications including point-to-point wireless local area networks, mobile backhaul, and broadband [31]. This announcement became highly important because the backbone networks of 5G will move from copper and fiber to mmWave wireless connections, allowing rapid deployment and cooperation between BSs.

Another important discussion is the propagation characteristics of mmWave. Since the propagation loss in free-space depends on the frequency, higher frequencies are subject to a higher signal attenuation than lower frequencies. Moreover, signals at lower frequencies can penetrate easily through buildings, although the mmWave signals do not penetrate most solid materials very well. Rain and atmospheric absorption increase the propagation loss [25] [31]. For all these reasons high frequencies, historically, were ruled out for cellular usage mainly due to concerns regarding short-range and Non-Line-of-Sight (NLoS) coverage issues [28].

The figure 3.1 shows the rain attenuation and atmospheric absorption characteristics of mmWave propagation. The propagation through the atmosphere depends primarily on atmospheric oxygen, humidity, fog and rain as it can be seen in next Table 3.1.

The cell sizes in urban environments are on the order of 200 m, thus atmospheric absorption does not create significant additional path loss, particularly at 28 GHz and 38 GHz, moreover, the attenuation caused by atmospheric absorption is 0.012 dB for cells with 200 m at 28 GHz, 0.016 dB at 38 GHz, 0.04 dB between 80 GHz and 100 GHz, or 0.2 dB between 200 GHz and 240 GHz [25].

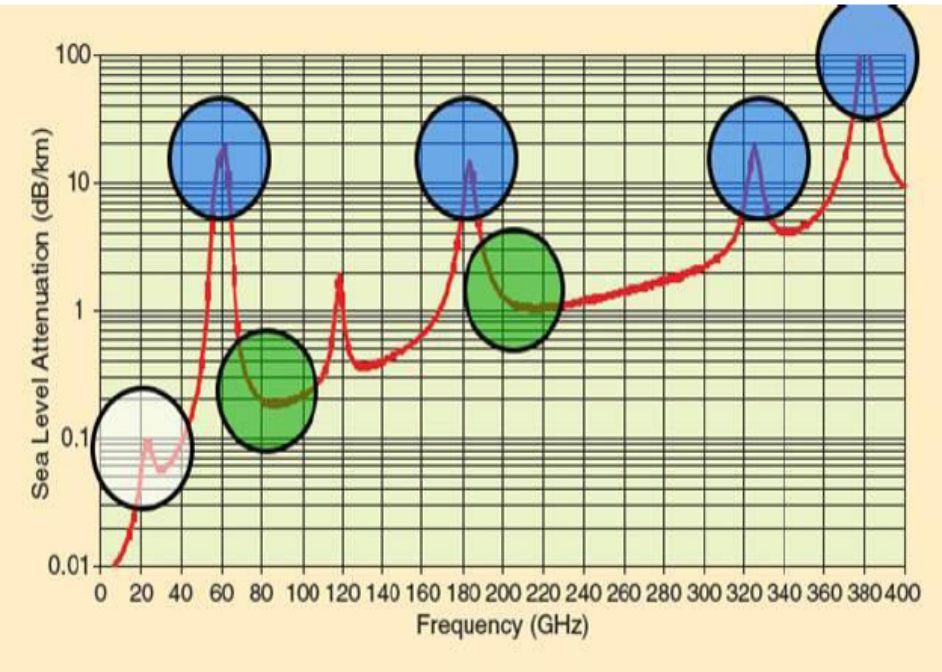


Figure 3.1: Atmospheric absorption across mmWave frequencies in dB/Km [67]

Effect	**Comments**	**Signal Loss (dB/Km)**
Oxygen	Sea level	0.22
Humidity	100% at 30C	1.8
Heavy Fog	10C, 1 gm/m (50m visibility)	2.2
Cloud Burst	25 mm/hr rain	10.7

Table 3.1: Signal Loss through Atmosphere [57]

To easily understand what is the practical performance of mmWave various channel measurements have been made [29, 68–70]. In both indoors and outdoors the practical results of channel measurement proved that this technology can really be implemented in the near future. In [68] measurements in different indoor environments such as laboratory offices, corridors and residences were made at 60 GHz, in conditions of Line-of-Sight (LoS) and NLoS with acceptable results. In [29, 69, 70], different outdoor measurements were made. Vehicular, University campus and big cities were the tested scenarios at 60 GHz, 38 GHz and 28

GHz, respectively. Sometimes difficulties were experienced and simple objects such as cars, windows or even people and trees have a major impact in mitigating the spread of mmWave.

3.2 Massive MIMO

MIMO technology has attracted much attention for wireless communications in the last years. The first approach was patented [71] in 1993, and the multiple transmit antennas were placed at one transmitter with the objective of improving the attainable link throughput. Basically, more antennas at the UEs and at the BSs create more DoF, which implies a better performance in terms of data rate or link reliability without an additional increase in bandwidth or transmit power [27].

A particular case of MIMO is mMIMO systems where the basic premise behind is to reap all the benefits of conventional MIMO, but on a much greater scale. In some communities mMIMO defines any MIMO configuration beyond the highest MIMO mode in current LTE (at present 8x8), other communities refer mMIMO simply to large number of antennas at the BSs from tens to hundreds or also thousands [72].

mMIMO is presented as an enabler for the future digital society infrastructure that will connect the Internet of people and Internet of Things with clouds and other network infrastructure. This only can happen because it is expected that mMIMO will increase the capacity 10 times or more and simultaneously improve the radiated energy efficiency in the order of 100 times [73]. Moreover, this system with a large number of antennas, simultaneously serve dozens of UEs without any penalty in user throughput and where the bandwidth efficiency is improved and transmit power is reduced, as it is proved by measurements in [74]. The multiplexing gains or diversity gains are improved due to the excessive DoF. The possibility to transmit simultaneously to several users and the flexibility to select what users are scheduled for reception at any given point in time is another possibility. These arrays of antennas help to focus energy with an extremely narrow beam reducing the inter-user interference even further by using, for example, ZF. However the number of terminals that can be simultaneously served is limited, not by the number of antennas, but rather by our inability to acquire CSI for an unlimited number of terminals [27]. New designs can be made extremely robust so that the failure of one or a few of the antenna units would not appreciably affect the system, because if an individual antenna fails the system is prepared to be adaptable.

Many different configurations and deployment scenarios for the actual antenna arrays used by a mMIMO system can be envisioned. The figure 3.2 illustrates several typical Antenna Array (AA), namely the linear AA, spherical AA, cylindrical AA, rectangular AA, and distributed AA. The linear AA is related to traditional MIMO systems that can only adjust the beam in the horizontal dimension [75]. This is an example of Two Dimensional (2D) that is mostly assumed in theoretical analysis. In order to exploit the vertical dimension other topologies may be considered. With these arrays it is possible to adjust both azimuth and elevation angles, and propagate signals in Three Dimensional (3D). The distributed antenna array is used inside buildings and in outdoor cooperation [74].

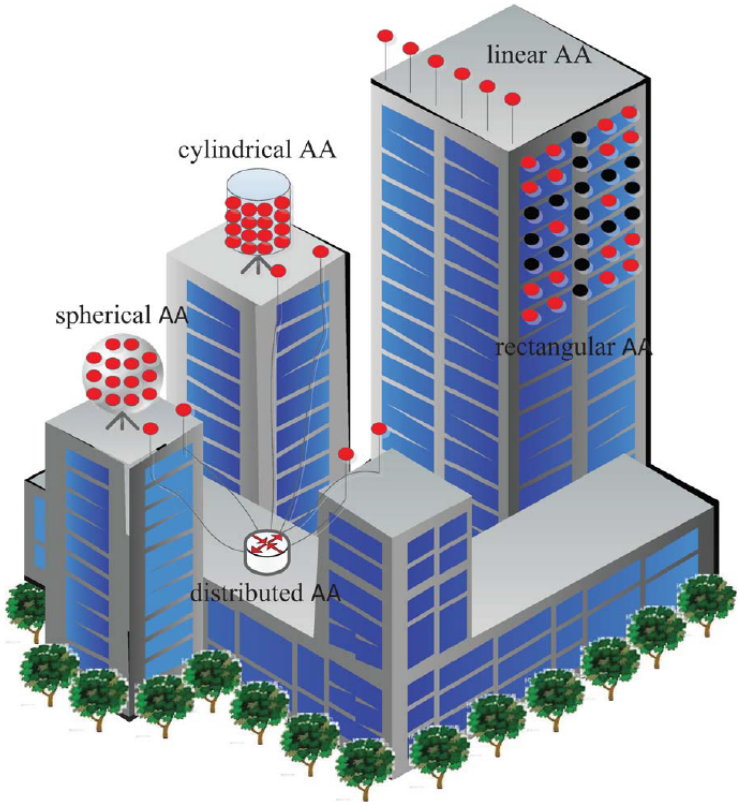


Figure 3.2: Antenna Array configurations mMIMO [74]

In mMIMO the BS needs to have good enough channel knowledge, on both the uplink and the downlink. In case of the uplink this could be easy by having the terminals sending pilots and based on which the BS estimates the channel responses to each of the terminals instead of the downlink that is more difficult. On the downlink, the BS sends out pilot waveforms that the terminals use to estimate the channel responses, and fed back to the BS. Unfortunately, for mMIMO systems this will not be feasible because the optimal downlink pilots should be mutually orthogonal between the antennas, thus the amount of time-frequency resources needed for downlink pilots scales with the number of antennas. Moreover, the number of channel responses that each terminal must estimate, again proportional to the number of base station antennas [75].

A main problem of conventional MIMO architectures in mMIMO implementations is the complexity of beamforming algorithms that make them impractical for commercial applications, further to perform wideband digital beamforming, each signal from/to an antenna element is usually divided into a number of narrowband signals and processed separately, which also adds a cost of digital signal processing. Therefore, a full digital signal implementation is simply unrealistic, because the excessive demand on real time signal processing for high antenna gains. However, some solutions are being proposed to overcome the issues of signal processing and will be discussed in section 3.4.

3.3 Massive Beamforming

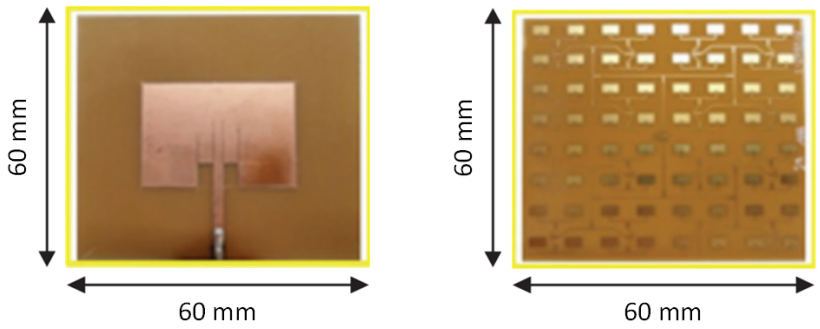
The possibility in the near future of the use the spectrum at high carrier frequencies open new possibilities. Higher frequencies are a significant challenge, at these frequencies the propagation is more hostile, in free-space propagation loss is higher and the wall penetration losses are higher. However, the antennas get smaller with high frequencies and open a possibility to assemble a large number of them in a small space. Each antenna consumes a portion of energy and all together increase the total power consumption for impracticable values. It is crucial to improve the energy performance and the key around this problem is the use of massive Beamforming.

The massive Beamforming with more antenna elements allow 3D beamforming and at the same time the beams get narrower. It becomes vital to transmit the signal in the appropriate direction, to maximize the received signal energy at the UE. This implementation provides several benefits, decreased interference which enables reduced overall transmission power in networks and extended service coverage which also provides high data rates.

The main challenge of massive Beamforming involves dealing with multi-user situations, gathering enough channel information, and dealing with multi-path and dispersive channels. Find a way to package all the antennas in a small space can be another problem which is not directly related with massive Beamforming techniques, however it is a necessary requirement to enable this approach.

3.3.1 Antenna Designs

With increasing of RF carrier frequency, the mmWave wavelength is reduced and the dimensions of the antennas are reduced too. In figure 3.3 it is possible to see two schematics of antennas, the first one represents a single antenna working at 3 GHz and the second one represents an array of antennas working at 30 GHz.

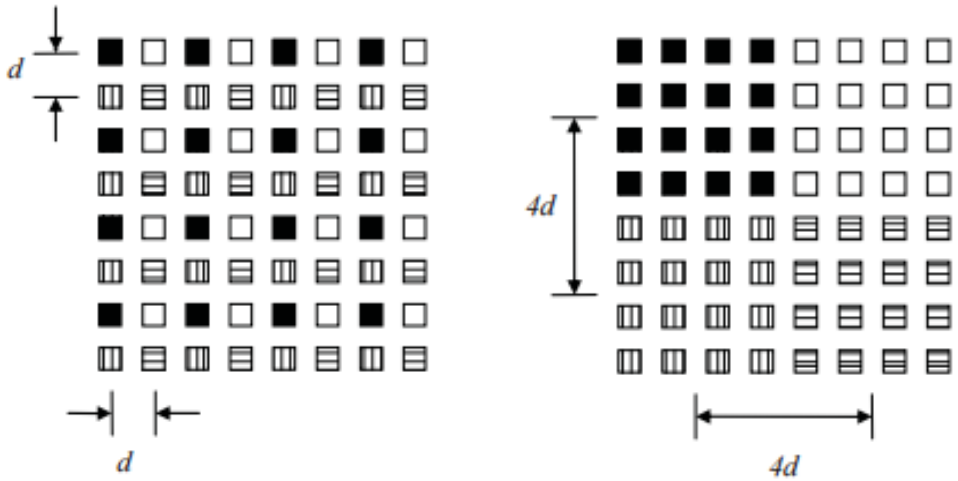


(a) Single Antenna 3 GHz (b) Array Antennas 30 GHz

Figure 3.3: Comparison antennas at different frequencies with the same dimensions [28]

As previously announced the reduced size of the antennas for high frequencies allows to assemble a large number of antennas in the same space. With increased transmission power after using antenna array, there is an opportunity to realize even longer range point-to-point

LoS links which are the targeted applications, such as those providing wireless connectivity between aircrafts and/or between aircrafts and ground vehicles or control stations [76].

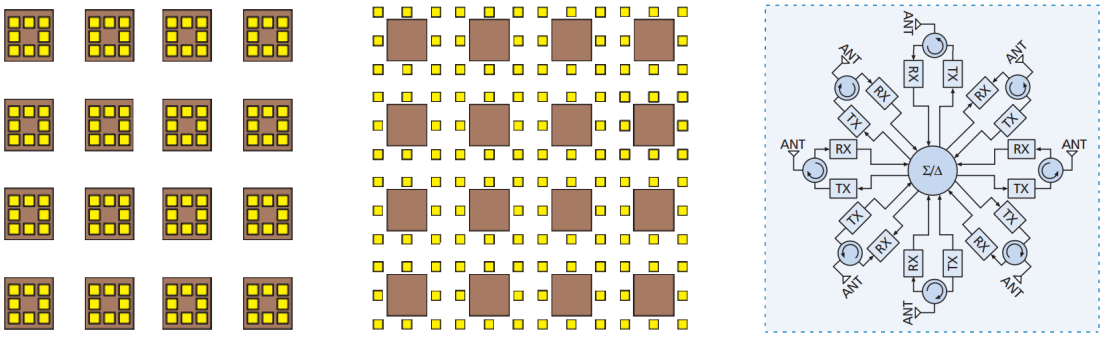


(a) Interleaved linear sub-arrays (b) Side-by-side linear sub-arrays

Figure 3.4: Hybrid Antenna Arrays [76]

To prevent grating lobes the antennas in the array must be placed closely together. This implies that each antenna cannot be large compared to the wavelength, but also not too small and isotropic since some gain is needed to overcome the large path loss [26]. However, at 72 GHz the required element spacing is only about 2 mm and with current Monolithic Microwave Integrated Circuits (MMIC) technology, the practical implementation of such a digital antenna array remains very difficult. In figure 3.4 two different ways are proposed to accommodate a hybrid receive antenna, assembled like figure 3.3*(b)*. These hybrid antennas are composed for 4 different sub-arrays of individual antennas and where each element in sub-array has its own RF chain [76]. The distance between adjacent elements in a sub-array is referred to as the sub-array element spacing and the distance between corresponding elements in adjacent sub-arrays is termed the sub-array spacing. In figure 3.4*(a)* the sub-array element spacing is $2d$ and the sub-array spacing is d, whereas in figure 3.4*(b)* the sub-array element spacing is d and the sub-array spacing is $4d$.

Another issue of the use of conventional MIMO architecture in a mMIMO implementation is the necessity to replicate multiple transmit/receive chains, one per each antenna. The analog components, such as the LNA or PA and the down or up converter associated with each antenna element, must be tightly packed behind the antenna element. For medium to high mmWave frequencies, antennas need to be located very close to their on-chip RF active components. At lower frequencies, the antennas can be located off-chip, although still in close proximity.



(a) High Frequencies (b) Lower Frequencies (c) Block diagram of the RF

Figure 3.5: Antenna Front-End Integration [26]

The figure 3.5 shows an array of antennas with 128 elements. The solution of figure 3.5*(a)* can be implemented at very high frequencies. 4x4 RF chips (brown squares) are used, each with 8 antennas in the package (yellow squares) because the insertion loss to bring the signal off-chip to an antenna would be too high. Instead, the solution of figure 3.5*(b)* will be used for lower frequencies. It is mounted nearby on a circuit board or on a different substrate, because at lower frequencies, the antennas are larger and not completely integrable in the chip package. In figure 3.5*(c)* a block diagram of the RF front-end with both TX and RX channels integrated within the same chip is shown, and connected to 8 antennas in either configuration *(a)* or *(b)* [26].

Each antenna unit uses extremely low power, in the order of mW that enables the use of renewable energies, like wind and solar, to be the source of electric energy for the BS.

3.4 Architectures for mmWave mMIMO Systems

In this Section an overview of different mMIMO mmWave based architectures is done. In Chapter 4 two of this architectures are evaluated and analyzed in detail. A major problem of conventional MIMO architecture in mMIMO implementations is the necessity to replicate multiple transmit/receive RF chains. It is difficult to have a fully dedicated RF chain for each antenna due the high cost and power consumption of some mixed signal components, like high-resolution analog-to-digital converters. This makes the conventional architecture used in current cellular systems, where precoding and combining are performed entirely in the digital baseband absolutely impracticable.

To contribute for the developments new types of hybrid structures were proposed [32,33]. The hybrid BF architectures use a much lower number of digital transceivers than the total antenna number, allowing it to be more practical and improving the efficiency cost. In the hybrid based architectures, the transmitter and receiver processing are divided between

the analog and digital domains. The approach proposed to reduce the transceiver number is via analog BF [77], where each transceiver is connected with multiple active antennas. Additionally the signal is phased on each antenna and controlled by analog phase shifters to obtain the maximal gain in the direction of the dominant channel paths. Thus, each transceiver generates one beam toward one user in analog BF. In relation to digital BF, over the transceivers, this can then be utilized to achieve multiple data stream precoding on top of analog BF, to enhance the performance [32]. The main difficulty in hybrid BF design is mainly due to the difficulty in CSI availability at BS transmitter side, because the mapping between transceivers and antennas makes channel estimation on both sides more complicated [78].

Different architectures have been designed, as the ones shown in figures 3.6, 3.7 and 3.8. Note that, N is the number of transceivers or RF chains and M represents the number of active antennas per transceiver.

3.4.1 Full-array architecture

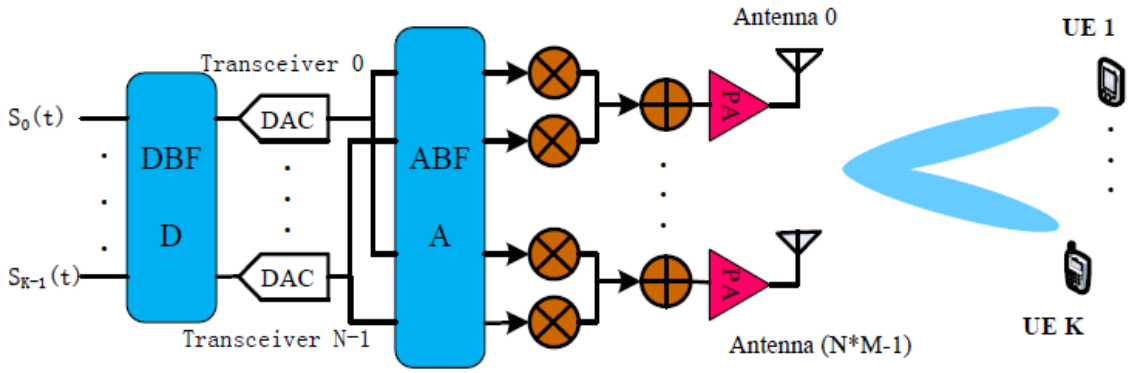


Figure 3.6: Hybrid BF: Full-array architecture [78]

In the first hybrid BF architecture shown in figure 3.6, each transceiver is connected with all antennas, in such a way that the transmitter signal on each of the N digital transceivers flows trough NM RF paths (mixer, power amplifier, phase shifter, etc) and summed up before being connected with each antenna element. Further, analog BF is performed over NM RF paths per transceiver, and digital BF can then be performed over N transceivers [32]. Obviously, the complexity of this architecture is rather high since the total number of RF paths is N^2M. However this architecture can achieve full BF gain.

3.4.2 Sub-array architecture

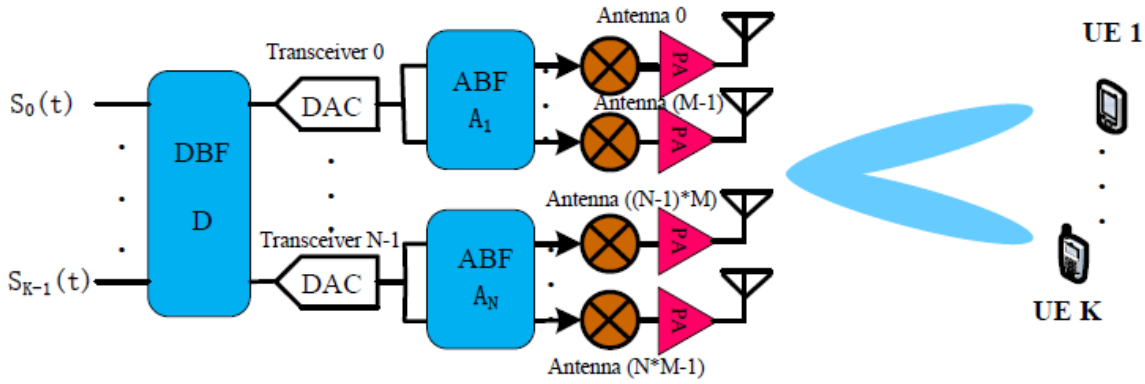


Figure 3.7: Hybrid BF: Sub-array architecture [78]

The second hybrid BF architecture is shown in figure 3.7. Here, each of the N transceivers is connected to M antennas. Analog BF is performed over only M RF paths in each transceiver, and digital BF is performed over N transceivers [32]. This architecture is more practical for BSs, where each transceiver is connected to a column of antennas. With active antennas on each RF path, BF can be performed by applying different phases to each column. The beamforming gain obtained with this structure is usually lower, but the complexity is reduced [78].

3.4.3 1-bit ADC architecture

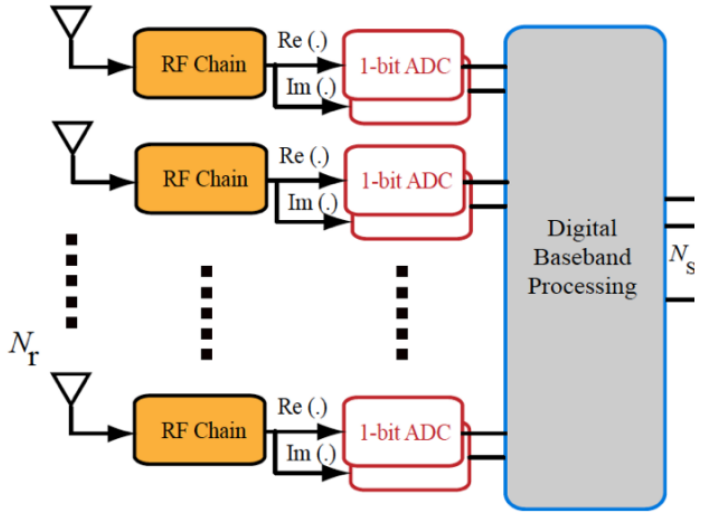


Figure 3.8: Hybrid BF: 1-bit ADC architecture [79]

In the third hybrid BF architecture shown in figure 3.8, the use of low resolution Analogue to Digital Converter (ADC) is proposed. In mmWave systems, the sampling rate of the ADCs scales up with the larger bandwidth and, at present, commercially ADCs with high speed and high resolution are too costly and take up a great amount of power. A solution can be the use of 1-bit ADC for each in-phase and quadrature baseband received signal, which reduces power consumption and cost. This architecture also simplifies the circuit complexity since automatic gain control may not be required. Unfortunately, there are a few disadvantages; firstly more RF chains are required in receivers; secondly, the capacity that can be achieved is also limited [33].

3.5 Millimeter Wave MIMO channel

Transmission loss is accounted for principally by free space loss. A general misconception among wireless engineers is that free-space propagation loss depends on frequency, therefore higher frequencies propagate less well than lower frequencies [31].

Carrier wavelengths in mmWave are one order of magnitude smaller than those for existing cellular and WiFi systems. As a result, the propagation and interference characteristics, limited by blockage of walls and furniture, are drastically different from our current experienced in wireless network design. This fact results in a different propagation geometry. Omni-directional transmission is essentially infeasible because of the increased propagation loss at smaller wavelengths. On the other hand, highly directive transmission and reception with electronically steerable beams can be achieved using compact antenna arrays.

However, this high free-space pathloss, characteristic of mmWave, leads to limited spatial selectivity or scattering and the large tightly-packed antenna arrays, characteristic of mmWave transceivers, lead to high levels of antenna correlation. A combination of tightly packed arrays in sparse scattering environments makes many of the statistical fading distributions used in conventional MIMO analysis inaccurate for mmWave channel modeling [34]. For all these reasons new channel models for mmWave have been discussed for the next generation [29,30].

3.5.1 Types of channels

There are three typical channel models that have been used for evaluating performance of wireless communications systems [74].

CBSM

Correlation-Based Stochastic Model (CBSM) has low complexity and it is mainly used for evaluating the theoretical performance of MIMO systems. However, it is somewhat simplistic and hence inaccurate for a realistic MIMO system. There are three different kinds of CBSM, i.e., the non-dispersive independent identically distributed (i.i.d.) Rayleigh fading model, the

non-dispersive correlated channel model, and the dispersive multi-path channel model. A summarized description of each is given below.

- **Non-dispersive i.i.d. Rayleigh channel model**: when this type of channel model is assumed no correlation existence between the transmit antennas and receive antennas. Consequently, the elements of the fast fading matrix are i.i.d. Gaussian variables.

- **Non-dispersive correlated Rayleigh channel model**: is a correlated channel model and it is used to characterize the Doppler-induced received signal correlation. The fast fading matrix of the correlated channel model is formed by the product of the correlation matrix and the standard complex-valued Gaussian matrix. The complex-valued Gaussian matrix describes the i.i.d. Rayleigh fading channel

- **Dispersive multi-path channel model**: this channel model can have different distributions Angle of Arrival (AoA)s from different UEs. Each UE is constituted by multiple independent paths, characterized by attenuation and AoA, from different directions. When UE are located at different angular positions, they can be separated according to their AoA.

GBSM

Geometry-Based Stochastic Model (GBSM) model is capable of accurately describing the realistic channel properties, and hence is more suitable for mMIMO channels. However, it must deal with an increased computational complexity, relatively to CBSM. According to the modeling of scattering propagation environments, the GBSM can be classified into *single-ring*, *twin-ring* and *elliptical models*. In mMIMO at higher frequencies *elliptical model* is the largest used and can be divided into 2D and 3D channel models. When BS is equipped with AA the angle of elevation is fixed and the 2D channel model is adequate for accurately evaluating the performance. Instead, when BS is equipped with a cylindrical or rectangular AA (see figure 3.2) the 3D channel model with an adjustable angle of elevation has to be considered.

PSM

Parametric Stochastic Model (PSM) has higher complexity than CBSM and lower accuracy than GBSM which represent a trade-off between the two channel models described above. PSM describes the signals on the receivers as a superposition of waves. There are models for conventional MIMO too, such as Double Directional Channel Model (DDCM) and Virtual Ray Model (VRM). Nevertheless, PSM is less studied in context of mMIMO requiring further research efforts, before it becomes a reality. PSM constitute a viable alternative, which are capable of reducing the complexity when compared with GBSM.

3.5.2 Millimeter Wave MIMO Channel model

In this subsection we describe the clustered mmWave channel model presented in [34]. Considering a single-user mmWave system with N_T transmit antennas and N_R receive antennas with N_{cl} scattering clusters where each cluster contribute with N_{ray} propagation paths to the channel matrix, the discrete-time narrowband channel $\mathbf{H}$ can be written as

$$\mathbf{H} = \gamma \sum_{i,l} \alpha_{il} \Lambda_r\big(\phi^r_{il}, \theta^r_{il}\big) \Lambda_t\big(\phi^t_{il}, \theta^t_{il}\big) \mathbf{a}_r\big(\phi^r_{il}, \theta^r_{il}\big) \mathbf{a}_t\big(\phi^t_{il}, \theta^t_{il}\big)^*, \tag{3.1}$$

where γ is a normalization factor such that $\gamma = \sqrt{N_T N_R / N_{cl} N ray}$. The α_{il} is the complex gain of the lth ray in the ith scattering cluster, whereas ϕ^r_{il} (θ^r_{il}) and ϕ^t_{il} (θ^t_{il}) are its azimuth (elevation) angles of arrival and departure, respectively. The transmit and receive antenna element gain at the corresponding angles of departure and arrival are given by the functions $\Lambda_t\big(\phi^t_{il}, \theta^t_{il}\big)$ and $\Lambda_r\big(\phi^r_{il}, \theta^r_{il}\big)$. The vectors $\mathbf{a}_r\big(\phi^r_{il}, \theta^r_{il}\big)$ and $\mathbf{a}_t\big(\phi^t_{il}, \theta^t_{il}\big)$ represent the normalized receive and transmit array response vectors at an azimuth (elevation) angle of ϕ^r_{il} (θ^r_{il}) and ϕ^t_{il} (θ^t_{il}) respectively. Finally, the α_{il} follow a complex Gaussian distribution, $\mathcal{CN}\big(0, \sigma^2_{\alpha,i}\big)$ where the $\sigma^2_{\alpha,i}$ represents the average power of the ith cluster.

An array response vector depends on the array antenna type. For a Uniform Linear Array (ULA) on the y-axis, with N_y elements, the array response vector is given by

$$\mathbf{a}_{ULA}\big(\phi\big) = \frac{1}{\sqrt{N_y}} \left[1, e^{j1kd\sin(\phi)}, e^{j2kd\sin(\phi)}, \ldots, e^{j(N_y-1)kd\sin(\phi)}\right]^T \tag{3.2}$$

For a Uniform Planar Array (UPA) in the yz-plane with N_y and N_z elements on the y and z axis respectively, the array response vector is given by

$$\mathbf{a}_{UPA}\big(\phi, \theta\big) = \frac{1}{\sqrt{N_y N_Z}} \left[1, \ldots, e^{jkd\left(m\sin(\phi)\sin(\theta) + n\cos\theta\right)}, \ldots, e^{\left(jkd(N_y-1)\sin(\phi)\sin(\theta) + (N_z-1)\cos\theta\right)}\right]^T \tag{3.3}$$

Where for the equations 3.2 and 3.3 $k = 2\pi/\lambda$, d is the inter-element spacing, $0 \leq m \leq N_y$ and $0 \leq n \leq N_z$.

3.6 Small Cell Networks

The combination of small cell geometries, as shown in figure 3.9, mmWave communications and arrays with a massive number of antennas will be essential to dramatically improve wireless access and throughput [26]. Wireless communication systems have been designed toward high spectral efficiency (SE) to support the explosively growing traffics. Both Small Cell Network (SCN) and mMIMO have been recognized as promising ways to provide high SE and Energy Efficiency (EE), which actually are two extreme ways to use the spatial resources, e.g., frequency reuse and spatial multiplexing.

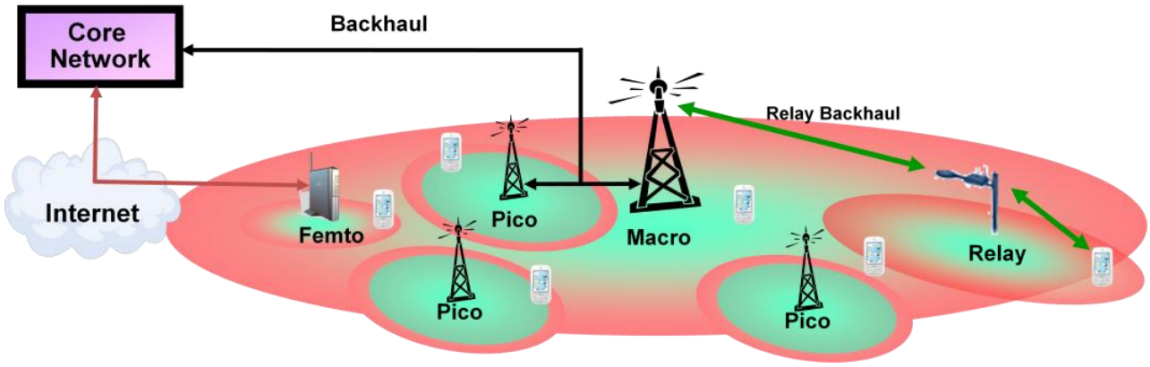


Figure 3.9: HetNet utilizing a mix of macro, pico, femto and relay BS [80]

In HetNets, small and low power BSs are overlaid within the macro network based on traffic and coverage area, representing a new paradigm for significantly expanding network capacity. This represents an attractive cost-effective solution to provide a uniform user experience [81]. The frequency reuse is increased through the creation of smaller cells, referred to as micro-cells, pico-cells and femto-cells. mmWave signals do not penetrate brick and other materials easily and which implies that mmWave cells may have to be located strictly in indoors and outdoors scenarios. Path loss, that is another critical problem in mmWave, can actually be beneficial in small-cell scenarios, because SCN limits inter-cell interference.

In traditional networks, the SE metric is considered to be the main performance indicator which measures how efficiently the frequency resources are utilized regardless of the efficient power consumption. On the other hand, EE is emerging as one of the key performance indicators for the next generation (5G) wireless communications systems due the current energy cost payable by operators for running their access networks [82].

Recently many valuable works analyzing the trade-off between EE-SE in mMIMO and SCN [81, 83–85]. The EE of SCN was analyzed in [81], and shown that increasing the BS density will improve the EE only when the circuit power consumption is less than a certain threshold. In [85] the relationship SE-EE was analyzed for mMIMO and it is demonstrated that the slope of the EE-SE curve only depends of the circuit power and bandwidth when the SE approaches zero. When the SE approaches infinity the relation depends on the number of data streams. However, it is in the [84] that the most interesting analysis was done. They

made a comparison between mMIMO and SCN to understand which are the performance of these systems. The conclusions demonstrate that when the number of cells is large, SCN is always more energy efficient than mMIMO. Instead, when the number of cells is small, mMIMO achieves higher EE than SCN, when the circuit power consumptions of mMIMO are much lower than SCN.

3.6.1 Cell Types

The HetNets, like is shown in figure 3.9, are composed by some different types of base stations or cells. The main differences between them are the coverage area, the transmitted power and their size. The existence of different cell types with different sizes and characteristics made a better deployment of the cellular systems possible. The smaller cells made the development and deployment of networks in the higher populated areas easier and cheaper. Sometimes, these cells are called green cells due the enhances in EE of the networks [82].

Macro-cells, micro-cells, pico-cells and femto-cells have the following main differences [86]:

- **Macro-cells**: is a conventional base station with 20 W of power and the range is about 1 Km to 20 Km. Macro-cell in hierarchical structure takes care of roaming mobiles.

- **Micro-cells**: is a conventional base station with 1 W to 5 W of power and the range is about 500 m to 2 Km. Micro-cells takes care of slow traffic (pedestrians and in-building subscribers).

- **Pico-cells**: the pico-cells are a small version of base stations, ranging in size from a laptop computer suitcase. They are designed to serve a very small area like a part of building, a street corner, railway stations etc. These are used to extend coverage in a indoor area where outdoor signals do not reach well or add network capacity in areas with very dense uses.

- **Femto-cells**: Femto-cells are also called home base stations and are typically designed for the use in a home or a small business. The current designs typically support 2 to 4 active mobile phones. The relatively reduced distance for the user leads to a higher received signal strength. This will result in an improvement in capacity through increase signal strength and reduced interference. Deploying femto-cells will enable more efficient uses of precious power and frequency resources.

When a cellular network is properly designed it is expected that the capacity increases as well as the coverage area in the cell-edges and consequently the user experience is better, which also severely contributed for the offloading of traffic from the macro-cells to the smaller cells. With the use of small cells the latency times in the downlink and uplink communications becomes lower, allowing fastest connections and also a better performance.

The Table 3.2 summarizes the main characteristics of each type of cell.

	Macro	**Micro**	**Outdoor Pico**	**Indoor Pico**	**Femto**
Key application	Large coverage areas	Medium coverage areas	Fit-in or extension	Indoor public areas	Indoor residential areas
Sectors	3	3	1	1	1
Number of users per cell	>200	Up to 200	Up to 100	Up to 50	<10
Mobility	High	High	High	Low	Low
Typical RF output power	10 W	4 W	1 W	500 mW	100 mW
Cost	High	Medium High	Medium	Low	Very Low

Table 3.2: Characteristics of the different base stations [87]

Chapter 4

Architectures and Processing Techniques for mmWave mMIMO Systems

As previously stated, a major problem of conventional MIMO architecture in a mMIMO implementation is the necessity to replicate multiple transmit/receive RF chains. It is difficult to have a full dedicated RF chain for each antenna due to the high cost and power consumption of some mixed signal components, like high-resolution analog-to-digital converters. This makes the conventional architecture used in current cellular systems, where precoding and combining are performed entirely in the digital baseband absolutely impracticable. Moreover, the large number of antennas have an impact in complexity of signal processing functions like channel estimation, precoding, combining and equalization.

In order to accommodate mmWave with large antenna arrays is necessary to overcome the difficulties of signal processing, once some signal processing is done at the digital level and some left for the analog domain. Recently, some beamforming and/or combining/equalization schemes have been proposed for hybrid architectures [34,36,62,88]. In [34], a hybrid spatially sparse precoding and combining approach was proposed for mmWave mMIMO systems. The full-array architecture of mmWave channels was exploited to formulate the single-user multi-stream precoding/combining scheme as a sparse reconstruction problem. In [36] was considered a mmWave mMIMO system with the sub-array architecture for hybrid precoding scheme to achieve the near-optimal performance with low complexity. The authors in [62] proposed a simple precoding scheme, only based on the knowledge of partial channel information at both terminals, in the form of AoA and Angle of Departure (AoD). A joint design of transmit-receive mixed analog/digital beamformers that aims at maximizing the received average signal-to-noise-ratio SNR for SC systems, was proposed in [88].

In this Chapter we implement a hybrid sub-array based architecture for mmWave systems, which is denominated by Hybrid Architecture 2 (HA-2). The performance is compared against a hybrid full-array based architecture denominated by Hybrid Architecture 1 (HA-1). The HA-1 can approach close to optimal performance, but it involves very high computational complexity. For the full array architecture each RF chain is connected to all BS transmit antennas, which makes the analog precoder/beamformer complex to be implemented in hard-

ware. The HA-2 proposes a more realistic hybrid precoding architecture with low complexity, where each RF chain is connected to only a small set of transmit antennas. For the implemented HA-2, the hybrid precoder proposed in [36] is considered and at the receiver side, a new hybrid equalizer is designed to efficiently separate the spatial streams. The results have shown that the full-array architecture HA-1 can achieve better performance than sub-array one (HA-2) in almost all scenarios, as expected. However, the sub-array architecture can obtain acceptable performances (in some scenarios outperforms HA-1), with a lower computational complexity.

4.1 System Characterization

We consider a single-user mmWave system with N_t transmit antennas and N_r receive antennas, where the transmitter sends N_s data streams to the receiver, per time-slot. To enable multi-stream communication, the transmitter shown in figure 4.1 is equipped with N_t^{RF} transmit RF chains such that $N_s \leq N_t^{RF} \leq N_t$. For the HA-2 each RF chain is only connected to N_t / N_t^{RF} antennas in the transmitter and N_r / N_r^{RF} in the receiver, instead of HA-1 where the transmitter/receiver are connected to all antennas.

For both, we consider a block fading channel, that remains constant during each block but varies independently between different blocks, with size T. The following characterization is equally applied for both hybrid architectures. The received signal is given by

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N} \tag{4.1}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T] \in \mathbb{C}^{N_r \times T}$ denotes the received signal matrix, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T] \in \mathbb{C}^{N_t \times T}$ is the transmitted signal. $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_T] \in \mathbb{C}^{N_r \times T}$ is a zero mean Gaussian noise with variance σ_n^2. $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, which follows the clustered sparse mmWave channel discussed in Chapter 3. We assume perfect timing and frequency synchronization. Moreover, we assume that the channel $\mathbf{H}$ is known perfectly at both sides, transmitter and receiver.

As mentioned before, channel $\mathbf{H}$ is a result from the sum of all contributions of N_{cl} clusters and each will contribute to N_{ray} propagation paths for the channel matrix $\mathbf{H}$, which may be expressed as

$$\mathbf{H} = \mathbf{A}_r \mathbf{\Lambda} \mathbf{A}_t^H \tag{4.2}$$

$\mathbf{\Lambda}$ is a diagonal matrix, with entries (i, l) that corresponds to the path gains of the lth ray in the ith scattering cluster. The variables $\mathbf{A}_t = [\bar{\mathbf{a}}_t(\theta_{i,l}^t), \ldots, \bar{\mathbf{a}}_t(\theta_{N_{cl},N_{ray}}^t)]$ and $\mathbf{A}_r = [\bar{\mathbf{a}}_r(\theta_{i,l}^r), \ldots, \bar{\mathbf{a}}_r(\theta_{N_{cl},N_{ray}}^r)]$ are the matrix of array response vectors at the transmitter and receiver, whereas $\theta_{i,l}^r$ and $\theta_{i,l}^t$ are respectively, the azimuth AoA and AoD.

4.2 MmWave mMIMO Architecture 1

In Chapter 3 a general overview about this architecture was given. In this section we present in detail, the implemented one, including the design of beamformers/precoders and equalizer coefficients. The design of hybrid precoders are formulated as a sparsity constrained matrix reconstruction, that instead of directly maximizing mutual information, the near-optimal hybrid precoders can be found via an optimization of a sparse signal recovery with multiple measurement vectors. The algorithm takes an optimal unconstrained precoder as input and approximates it as linear combination of beam steering vectors. These vectors can be applied at RF and combined digitally at baseband. It is also present the extension of this sparse precoding approach to the receiver side. At the transmitter side a space-time encoder structure is employed, before the digital and analog precoders to increase the inherent diversity of the mmWave mMIMO system.

4.2.1 Transmitter Model Description

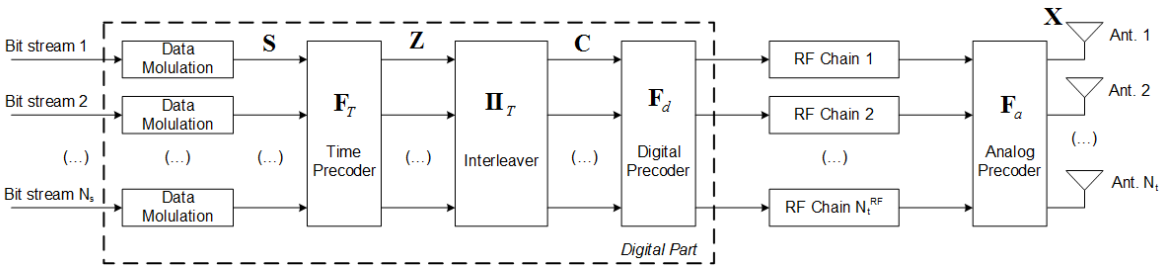


Figure 4.1: Transmitter block diagram for HA-1

The transmitter processing is decomposed into two parts, the digital and the analog, as shown in figure 4.1. These parts are modeled mathematically by $\mathbf{F}_a \in \mathbb{C}^{N_t \times N_t^{RF}}$ matrix for the analog part and $\mathbf{F}_d \in \mathbb{C}^{N_t^{RF} \times N_s}$ matrix for the digital one. The analog part is implemented using a matrix of analog phase shifters, due the hardware constraints, which force all elements of matrix $\mathbf{F}_a$ to have equal norm ($|\mathbf{F}_a(i,l)|^2 = N_t^{-1}$). The transmit total power constraint is enforced by normalizing $||\mathbf{X}||_F^2$ to $N_s T$.

The analog part is followed by the digital part and no other hardware-related constraints are placed on the baseband precoder, therefore the transmit signal is given by

$$\mathbf{X} = \mathbf{F}_a \mathbf{F}_d \mathbf{C}, \tag{4.3}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_T] \in \mathbb{C}^{N_s \times T}$ denotes a codeword constructed using a STBC. The STBC considered can be mathematically described by

$$\mathbf{z}_t = \mathbf{S}\mathbf{f}_t, \tag{4.4}$$

$$\mathbf{c}_t = \mathbf{\Pi}_t \mathbf{z}_t, \tag{4.5}$$

where $t = 1, \ldots, T$ denotes the time index, $\mathbf{f}_t \in \mathbb{C}^T$ denotes the column t of a T point DFT matrix ($\mathbf{F}_T = [\mathbf{f}_1, \ldots, \mathbf{f}_T]$). $\mathbf{\Pi}_t \in \mathbb{C}^{N_s \times N_s}$, $t = 1, \ldots, T$ is a random permutation matrix known at both, transmitter and the receiver sides and $\mathbf{S} = [s_{s,t}]_{1 \leq s \leq N_s, 1 \leq t \leq T} \in \mathbb{C}^{N_s \times T}$, with $s_{t,s}$, $t \in 1, \ldots, T$, $s \in 1, \ldots, N_s$ denoting a complex data symbol chosen from a QAM constellation with $\mathbb{E}[s_{s,t}]^2 = \sigma_s^2$, where $\sum_{s=1}^{N_s} \sigma_s^2 = N_s$. To compute codeword $\mathbf{C}$ we need to apply FFT transform to the rows of the symbol matrix $\mathbf{S}$ (see equation 4.4) and then permute each of the resulting T columns with a random permutation $\mathbf{\Pi}_t$, $t = 1, \ldots, T$ (see equation 4.5).

4.2.2 Receiver Model Description

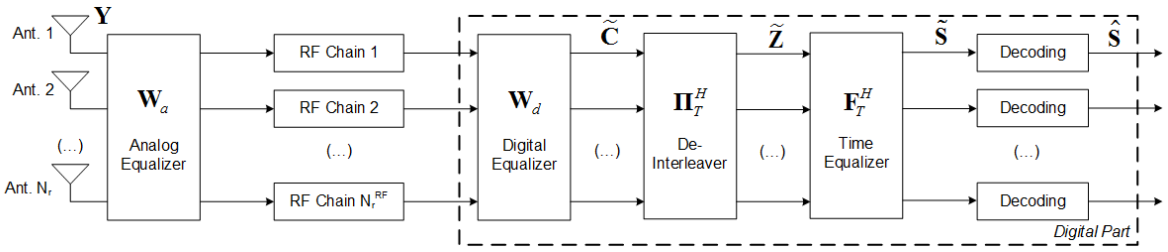


Figure 4.2: Receiver block diagram for HA-1

The receiver follows the design shown in figure 4.2. From equations 4.1 and 4.3 the received signal is given by

$$\mathbf{Y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_a\mathbf{F}_d\mathbf{C} + \mathbf{N}, \tag{4.6}$$

where ρ represents the average received power. The received signal is firstly processed through the analog phase shifters, modeled by the matrix $\mathbf{W}_a \in \mathbb{C}^{N_r^{RF} \times N_r}$, with N_r^{RF} processing RF chains, $N_s \leq N_r^{RF} \leq N_r$, and N_r the number of receive antennas. All elements of matrix $\mathbf{W}_a$ must have equal norm $|\mathbf{W}_a(i,l)|^2 = N_r^{-1}$ and then the baseband processing follows. The signal first passes through a linear digital filter $\mathbf{W}_d \in \mathbb{C}^{N_s \times N_r^{RF}}$, and follows the decoding of the STBC. The STBC decoding is obtained by

$$\tilde{\mathbf{Z}} = [\mathbf{\Pi}_1^H \tilde{\mathbf{c}}_1, \ldots, \mathbf{\Pi}_T^H \tilde{\mathbf{c}}_T], \tag{4.7}$$

$$\tilde{\mathbf{S}} = \tilde{\mathbf{Z}}\mathbf{F}_T^H. \tag{4.8}$$

More specifically, to obtain a soft estimate of the transmitted symbols we first invert the permutation applied at the transmitter side, see equation 4.7, and then apply the IDFT transform to the resulting matrix, as in equation 4.8.

4.2.3 Algorithm Description

Finally, we will detail the algorithm that takes an optimal unconstrained precoder as input and approximates it as a linear combination of beam steering vectors. In any case, the precoding problem consists of selecting the "best" N_t^{RF} array response vectors and finding their optimal baseband combination, to be applied at RF and combined at digitally baseband. We also detail how to extend this approach to the receiver side processing.

Hybrid Analog/Digital Precoding

We seek to design mmWave precoders $(\mathbf{F}_a, \mathbf{F}_d)$ that maximize the spectral efficiency. So, we start by examining the mutual information achieved by the hybrid precoders $(\mathbf{F}_a, \mathbf{F}_d)$ to obtain the channel's optimal unconstrained precoder $\mathbf{F}_{opt}$. To do so, define the channel's ordered SVD to be,

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{4.9}$$

where $\mathbf{V}$ is an $N_r \times rank(\mathbf{H})$ unitary matrix and $\mathbf{\Sigma}$ is a $rank(\mathbf{H}) \times rank(\mathbf{H})$ diagonal matrix of singular values arranged in decreasing order. Furthermore, defining the following two partitions of the matrices $\mathbf{\Sigma}$ and $\mathbf{V}$ as

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_2} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}, \tag{4.10}$$

where $\mathbf{\Sigma_1} \in \mathbb{C}^{N_s \times N_s}$ and $\mathbf{V}_1 \in \mathbb{C}^{N_t \times N_s}$. We note that the optimal unconstrained unitary precoder for $\mathbf{H}$ is simply given by

$$\mathbf{F}_{opt} = \mathbf{V}_1. \tag{4.11}$$

With $\mathbf{F}_{opt} \in \mathbb{C}^{N_t \times N_s}$, we defined the initial requirements to start building the algorithm presented in Table 4.1, which takes an optimal unconstrained precoder as input and approximates it as linear combination of beam steering vectors.

Algorithm: Precoding step by step

Require: $\mathbf{F}_{opt}$

$\mathbf{F}_a$ = Empty Matrix

$\mathbf{F}_{res} = \mathbf{F}_{opt}$

  for $i \leq N_t^{RF}$ **do**

    $\mathbf{\Psi} = \mathbf{A}_t^H \mathbf{F}_{res}$

    $k = \arg\max_{l=1,\ldots,N_{cl}N_{ray}} \left(\mathbf{\Psi}\mathbf{\Psi}^H\right)_{l,l}$

    $\mathbf{F}_a = \left[\mathbf{F}_a | \mathbf{A}_t^{(k)}\right]$

    $\mathbf{F}_d = \left[\mathbf{F}_a^H \mathbf{F}_a\right]^{-1} \mathbf{F}_a^H \mathbf{F}_{opt}$

    $\mathbf{F}_{res} = \frac{\mathbf{F}_{opt} - \mathbf{F}_a \mathbf{F}_d}{\left|\left|\mathbf{F}_{opt} - \mathbf{F}_a \mathbf{F}_d\right|\right|_F}$

  end for

$\mathbf{F}_d = \sqrt{N_S} \frac{\mathbf{F}_d}{\left|\left|\mathbf{F}_a \mathbf{F}_d\right|\right|_F}$

return $\mathbf{F}_a$, $\mathbf{F}_d$

Table 4.1: Precoder algorithm for HA-1

An overview about the code:

- **Steps 4 and 5:** the precoding algorithm starts by finding the column of matrix $\mathbf{A}_t = [\bar{\mathbf{a}}_t(\theta^t_{1,1}), \ldots, \bar{\mathbf{a}}_t(\theta^t_{N_{cl},N_{ray}})]$ along which the optimal precoder has the maximum projection.

- **Step 6:** after having found the best column, it then appends this column to the RF precoder $\mathbf{F}_a$.

- **Step 7:** least squares solution to $\mathbf{F}_d$ is calculated.

- **Step 8:** the contribution of the selected vector, found in steps 4/5, is now removed and the algorithm proceeds to find the column along which the residual precoding matrix $\mathbf{F}_{res}$ has the largest projection.

- **Step 9:** the process continues until all N_t^{RF} beamforming vectors have been selected.

- **Step 10:** ensures that the transmit power constraint is exactly satisfying.

At the end of the N_t^{RF} iterations, the algorithm would have constructed an $\mathbf{F}_a \in \mathbb{C}^{N_t \times N_t^{RF}}$ matrix, and found the optimal $\mathbf{F}_d \in \mathbb{C}^{N_t^{RF} \times N_s}$ which minimizes $||\mathbf{F}_{opt} - \mathbf{F}_a\mathbf{F}_d||_F^2$.

Hybrid Analog/Digital Combining

Assuming the hybrid precoders $\mathbf{F}_a\mathbf{F}_d$ are fixed, we seek to design hybrid combiners $\mathbf{W}_a\mathbf{W}_d$ that minimize MMSE between the transmitter and processed received signals. The algorithm of Table 4.2, requires the full-digital $\mathbf{W}_{MMSE}$ that is given by

$$\mathbf{W}_{MMSE} = \frac{1}{\sqrt{\rho}}\left(\mathbf{F}_d^H\mathbf{F}_a^H\mathbf{H}^H\mathbf{H}\mathbf{F}_a\mathbf{F}_d + \frac{\sigma_n^2 N_s}{\rho}\mathbf{I}_{N_s}\right)^{-1}\mathbf{F}_d^H\mathbf{F}_a^H\mathbf{H}^H, \tag{4.12}$$

where $\mathbf{W}_{MMSE} \in \mathbb{C}^{N_s \times N_r}$ and is similar to the algorithm presented in Table 4.1 for the transmitter.

Algorithm: Combining step by step

Require: $\mathbf{W}_{MMSE}$

$\mathbf{W}_a$ = Empty Matrix

$\mathbf{W}_{res} = \mathbf{W}_{MMSE}$

for $i \leq N_r^{RF}$ **do**

$\Psi = \mathbf{A}_r^H \mathbb{E}[\mathbf{y}\mathbf{y}^H]\mathbf{W}_{res}$

$k = \arg\max_{l=1,\ldots,N_{cl}N_{ray}} \left(\mathbf{\Psi}\mathbf{\Psi}^H\right)_{l,l}$

$\mathbf{W}_a = \left[\mathbf{W}_a | \mathbf{A}_r^{(k)}\right]$

$\mathbf{W}_d = \left[\mathbf{W}_a^H \mathbb{E}[\mathbf{y}\mathbf{y}^H]\mathbf{W}_a\right]^{-1} \mathbf{W}_a^H \mathbb{E}[\mathbf{y}\mathbf{y}^H]\mathbf{W}_{MMSE}$

$\mathbf{W}_{res} = \frac{\mathbf{W}_{MMSE} - \mathbf{W}_a\mathbf{W}_d}{\left|\left|\mathbf{W}_{MMSE} - \mathbf{W}_a\mathbf{W}_d\right|\right|_F}$

end for

return $\mathbf{W}_a$, $\mathbf{W}_d$

Table 4.2: Combining algorithm for HA-1

In the algorithms design it was assumed perfect knowledge of the CSI, i.e. matrix $\mathbf{H}$, which may be not feasible for practical systems. Moreover, in FDD based systems the receiver should fed back the matrix $\mathbf{H}$ to the transmitter in order to compute matrices $\mathbf{F}_a$ and $\mathbf{F}_d$. More details about these algorithms can be found in [34].

4.3 MmWave mMIMO Architecture 2

In this Section we present in detail the mmWave mMIMO HA-2 implemented with sub-array architecture (see figure 3.7), where each RF chain is connected to only an independent subset from the total number of antennas [36]. As mentioned before, this sub-array architecture is less complexed than the previous one and for that reason it is a more realistic solution for practical implementation. Then, we also describe an iterative hybrid precoding scheme based on the SIC principle to achieve near-optimal performance with low complexity. We consider each antenna array one by one, which means that the precoding for each antenna array is connected to a specific RF chain. For the first antenna array, we use the digital precoder to control the amplitude. After that, the analog precoder is applied to adjust the phase and then, the optimization problem for the first antenna array is solved with low complexity. The SIC approach is applied after the optimization problem for the first array. So, before solving the optimization problem for the second array of antennas we first eliminate the contribution of the first array from the total achievable capacity. Generalizing, before optimizing the capacity for an array we need to eliminate all contributions from the above arrays. The only exception is for the first array, where obviously there is not any contribution to be removed. In figure 4.3 a diagram to easily understand the concept is shown.

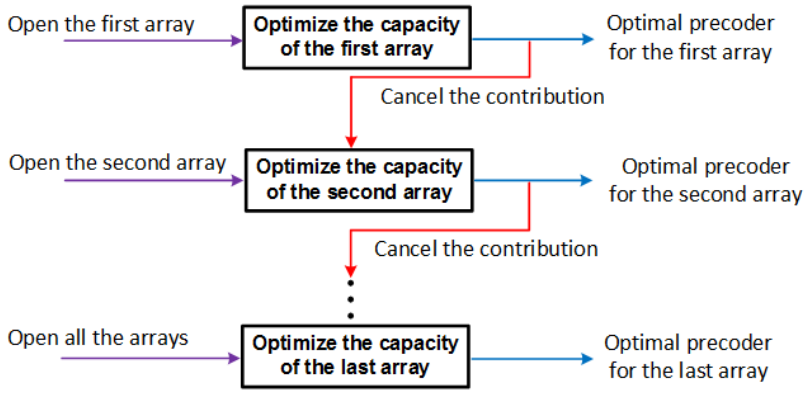


Figure 4.3: Diagram of the proposed iterative hybrid analog/digital precoding [36]

4.3.1 Transmitter Model Description

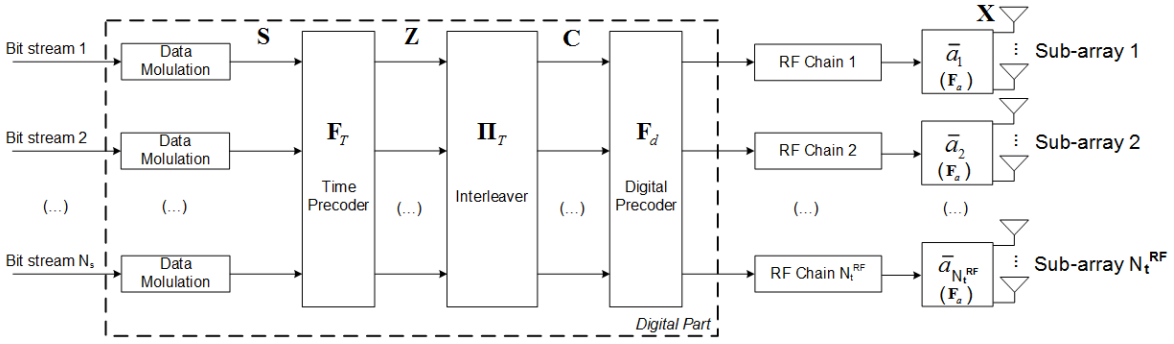


Figure 4.4: Transmitter block diagram for HA-2

As it can be seen from figure 4.4 the main difference of the HA-2 transmitter regarding the one implemented for HA-1 (see figure 4.1) is that the each RF chain is only connected to a set of antennas. Firstly, after the modulation is applied a FFT transformation at rows of the symbol matrix $\mathbf{S} \in \mathbb{C}^{N_s \times T}$ and then the resulting matrix $\mathbf{Z} \in \mathbb{C}^{N_s \times T}$ is randomly permuted, as detailed in subsection 4.2.1. Following, the matrix $\mathbf{C} \in \mathbb{C}^{N_s \times T}$ in the baseband is precoded by the digital precoder $\mathbf{F}_d \in \mathbb{C}^{N_t^{RF} \times N_s}$, which is a diagonal matrix given by

$$\mathbf{F}_d = diag\big[d_1, d_2, \ldots, d_n\big] \tag{4.13}$$

where $d_n \in \mathbb{R}$ for $n = 1, 2, \ldots, N_t^{RF}$. The main reason of digital precoding is essentially to perform power allocation.

The digital part is followed by the analog part where the streams pass through the corresponding RF chain, and the digital domain signal from each RF chain is delivered to only N_t / N_t^{RF} phase shifters to perform the analog precoding. The analog precoding $\mathbf{F}_a \in \mathbb{C}^{N_t \times N_t^{RF}}$, can be denoted by the analog weighting vector $\bar{\mathbf{a}}_n \in \mathbb{C}^{(N_t/N_t^{RF})\times 1}$, whose elements have the same amplitude but different phases. The analog precoder matrix comprising N_t^{RF} analog weighting vectors $\{\bar{\mathbf{a}}_m\}_{m=1,\ldots,N_t^{RF}}$ is given by

$$\mathbf{F}_a = \begin{bmatrix} \bar{\mathbf{a}}_1 & 0 & \ldots & 0 \\ 0 & \bar{\mathbf{a}}_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \bar{\mathbf{a}}_{N_t^{RF}} \end{bmatrix}_{N_t \times N_t^{RF}} . \tag{4.14}$$

After the analog precoding, each data stream is finally transmitted by a sub-array of antennas.

4.3.2 Receiver Model Description

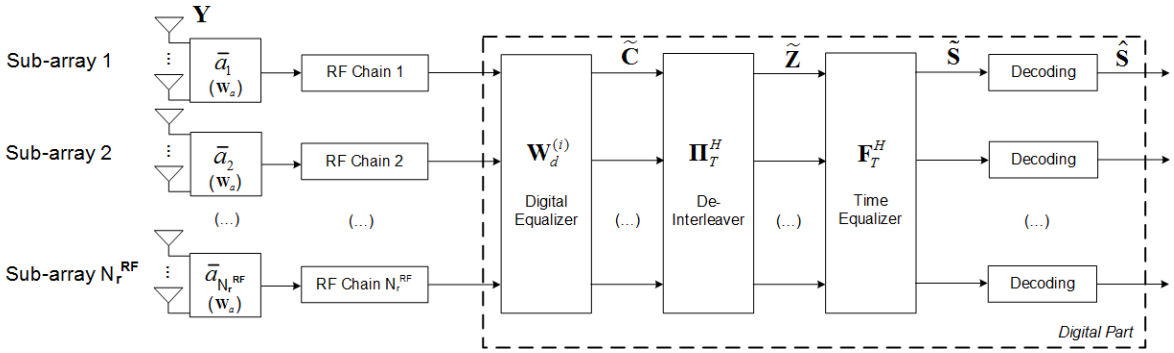


Figure 4.5: Receiver block diagram for HA-2

The receiver follows the design shown in figure 4.5, and the individual sub-arrays of receive antennas (N_r / N_r^{RF} per RF chain) is processed independently. The received signal can be written as

$$\mathbf{Y} = \rho \mathbf{H} \mathbf{F}_a \mathbf{F}_d \mathbf{C} + \mathbf{N}, \tag{4.15}$$

where ρ is the average received power and $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_T] \in \mathbb{C}^{N_r \times T}$ is additive white Gaussian noise with variance σ_n^2. For each time slot T, a vector $\mathbf{y}_t = \left[y_{t,1}, y_{t,2}, \ldots, y_{t,N_r}\right]^T$ is extracted from the matrix $\mathbf{Y}$ and processed through the analog phase shifters, modeled by the matrix $\mathbf{W}_a \in \mathbb{C}^{N_r^{RF} \times N_r}$, then the baseband processing follows. The signal firstly passes through a linear filter $\mathbf{W}_d \in \mathbb{C}^{N_s \times N_r^{RF}}$, which can again be generalized as $\mathbf{W}_d = diag\left[d_1, d_2, \ldots, d_n\right]$ where $d_n \in \mathbb{R}$ for $n = 1, 2, \ldots, N_r^{RF}$, such as in transmitter side. However, before obtaining the soft estimation of the transmitted symbols, we first have to invert the permutation applied at the transmitter side, see equation 4.7, and then apply the IDFT transform, as in equation 4.8 to the resulting matrix $\hat{\mathbf{S}} \in \mathbb{C}^{N_s \times T}$.

4.3.3 Algorithm Description

In this section, we present a low-complexity SIC-based hybrid precoding to achieve near-optimal performance, for the hybrid mmWave mMIMO based system discussed in the previous section. According to the system model, the overall hybrid precoding matrix $\mathbf{P}$ can be represented as $\mathbf{P} = \mathbf{F}_a\mathbf{F}_d = diag\{\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_{N_t^{RF}}\} \times diag\{d_1, \ldots, d_{N_t^{RF}}\}$.

There are three constraints for the $\mathbf{P}$ design. Firstly, $\mathbf{P}$ should be a block diagonal matrix similar to the form of $\mathbf{F}_a$ as shown in equation 4.14. We have $\mathbf{P} = diag\{\bar{\mathbf{p}}_1, \ldots, \bar{\mathbf{p}}_{N_t^{RF}}\}$, where $\bar{\mathbf{p}}_n = d_n\bar{\mathbf{a}}_n \in \mathbb{C}^{(N_t/N_t^{RF})\times 1}$ is the non-zero vector of the nth column $\mathbf{p}_n$ of $\mathbf{P}$. Secondly, the non-zero elements of each column of $\mathbf{P}$ should have the same amplitude. Moreover, the amplitude of non-zero elements of the analog precoding matrix $\mathbf{F}_a$ is fixed to $1/\sqrt{N_t}$. The last constraint, is related with the fact of the Frobenius norm of $\mathbf{P}$ should satisfy $||\mathbf{P}||_F \leq N_t^{RF}$ to meet the total transmit power constraint. However, these constraints are very difficult to be solved.

So, based on special diagonal structure of hybrid precoding matrix $\mathbf{P}$, we can observe that the precoding on each sub-array of antennas is independent. Therefore, we decompose the total achievable capacity into a series of sub-optimization problems. We start by designing the digital-analog precoders and then we present the the analog-digital equalizer implemented.

Hybrid Analog/Digital Precoding

The aim is to design mmWave precoders $(\mathbf{F}_a, \mathbf{F}_d)$ that maximize the performance for each sub-array. To obtain this precoders we apply the SIC idea previously described, which proceeds in steps and decompose the global optimization problem in several subproblems. Let us define the total achievable rate after m decompositions as

$$\mathbf{T}_m = \mathbf{I}_N + \frac{\rho}{\sigma^2}\mathbf{H}\mathbf{P}(1:m)\mathbf{P}(1:m)^H\mathbf{H}^H, \tag{4.16}$$

where $\mathbf{P}(1:m)$ denotes the first m columns of $\mathbf{P}$ and $\mathbf{T}_m \in \mathbb{C}^{N_r\times N_r}$. Let us also describe the matrix $\mathbf{G} \in \mathbb{C}^{N_t\times N_t}$

$$\mathbf{G} = \mathbf{H}^H\mathbf{T}_m^{-1}\mathbf{H}. \tag{4.17}$$

After each sub-optimization, we must find the remaining achievable rate $\mathbf{T}_m$, and update the matrix $\mathbf{G}$. Then, we define matrix $\mathbf{Q}$ of size $M \times M$ as a sub-matrix of $\mathbf{G}$ by only keeping the rows and columns of $\mathbf{G}$ for the sub-array of antennas that will be optimized. Denote $M = N_t/N_t^{RF}$ as a number of antennas for each RF chain. Define the ordered SVD of the Hermitian matrix $\mathbf{Q}$ as $\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{\Sigma}$ is an $M \times M$ diagonal matrix containing the singular values of $\mathbf{Q}$ in a decreasing order, and $\mathbf{V}$ is an $M \times M$ unitary matrix. We obtain $\mathbf{v}_1$ extracting the first column of $\mathbf{V}$ with size $M \times 1$. The analog precoder for an individual sub-array of antennas is given by

$$\bar{\mathbf{a}}_m = e^{jangle(\mathbf{v}_1)}, \tag{4.18}$$

where $angle(\mathbf{v}_1)$ denotes the phase vector of $\mathbf{v}_1$.

To sum up, the pseudo-code of the implemented hybrid analog/digital precoding scheme is described in Table 4.3

Algorithm: Precoding step by step

Require: $\mathbf{H}$, N_t, N_t^{RF}

$\mathbf{P} = 0_{N_t \times N_t^{RF}}$

$M = \frac{N_t}{N_t^{RF}}$

for $1 \leq m \leq N_t^{RF}$

$\mathbf{G} = \mathbf{H}^H \mathbf{T}_m^{-1} \mathbf{H}$

$\mathbf{Q} = \mathbf{G}\Big(M\big(m-1\big) : Mm, M\big(m-1\big) : Mm\Big)$

$\mathbf{v}_1$ =The first right-singular vector of $\mathbf{Q}$

$\bar{\mathbf{a}}_m = e^{jangle(\mathbf{v}_1)}$, $\mathbf{b}_m = \begin{bmatrix} \mathbf{0}_{M(m-1)\times 1} \\ \bar{\mathbf{a}}_m \\ \mathbf{0}_{M(N-1)\times 1} \end{bmatrix}$

$d_m = \frac{Re\left[\mathbf{v}_1^H \bar{\mathbf{a}}_m\right]}{M}$

$\mathbf{p}_m = d_m \mathbf{b}_m$

end for

$\mathbf{F}_a = \left[\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{N_t^{RF}}\right]$

$\mathbf{F}_d = diag\big(d_1, d_2, \ldots, d_{N_t^{RF}}\big)$

$\mathbf{P} = \left[\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N_t^{RF}}\right]$

return $\mathbf{F}_a$, $\mathbf{F}_d$, $\mathbf{P}$

Table 4.3: Precoder algorithm for HA-2

Hybrid Analog/Digital Combining

We present in detail the algorithm implemented at the receiver side, i.e., the hybrid digital-analog equalizer. In the design of equalizer we assume a decoupled joint transmitter-receiver optimization problem, i.e., we consider fixed precoders $\mathbf{F}_a$ and $\mathbf{F}_d$ in the equalizer design. We also assume perfectly channel knowledge, to find the near-optimal hybrid combiners $\mathbf{W}_a$ and $\mathbf{W}_d$. Firstly, it is necessary to define $\mathbf{H}_{eq} \in \mathbb{C}^{N_r \times N_s}$ as

$$\mathbf{H}_{eq} = \mathbf{H}\mathbf{F}_a\mathbf{F}_d \tag{4.19}$$

and also rewriting the equation 4.16

$$\mathbf{T}_m = \mathbf{I}_N + \frac{\rho}{\sigma^2}\mathbf{H}_{eq}\mathbf{W}_{opt}(1:m)\mathbf{W}_{opt}(1:m)^H\mathbf{H}_{eq}^H \tag{4.20}$$

where $\mathbf{W}_{opt} = \mathbf{W}_d\mathbf{W}_a$ is a matrix $N_s \times N_r$.

After N_r^{RF} iterations, we use a ZF equalizer to efficiently remove the interference described as

$$\mathbf{W}_{ZF} = \left(\mathbf{H}_o^H \mathbf{H}_o\right)^{-1} \mathbf{W}_{opt}, \tag{4.21}$$

where $\mathbf{H}_o = \mathbf{W}_{opt} \times \mathbf{H}_{eq}$ and $\mathbf{W}_{ZF} \in \mathbb{C}^{N_s \times N_r}$. To sum up, the pseudo-code of the proposed hybrid analog/digital equalizer scheme is described in Table 4.4

Algorithm: Combining step by step

Require: $\mathbf{H}$, $\mathbf{F}_a$, $\mathbf{F}_d$, N_r, N_r^{RF}

$\mathbf{W}_{opt}$=$\mathbf{0}_{N_r \times N_r^{RF}}$

$M = \frac{N_r}{N_r^{RF}}$

$\mathbf{H}_{eq} = \mathbf{H}\mathbf{F}_a\mathbf{F}_d$

for $1 \leq m \leq N_r^{RF}$

$\mathbf{G} = \mathbf{H}_{eq}^H \mathbf{T}_m^{-1} \mathbf{H}_{eq}$

$\mathbf{Q} = \mathbf{G}\Big(M\big(m-1\big) : Mm, M\big(m-1\big) : Mm\Big)$

$\mathbf{v}_1$ =The first right-singular vector of $\mathbf{Q}$

$\bar{\mathbf{a}}_m = e^{jangle(\mathbf{v}_1)}$, $\mathbf{b}_m = \begin{bmatrix} \mathbf{0}_{M(m-1)\times 1} \\ \bar{\mathbf{a}}_m \\ \mathbf{0}_{M(N-1)\times 1} \end{bmatrix}$

$d_m = \frac{Re\left[\mathbf{v}_1^H \bar{\mathbf{a}}_m\right]}{M}$

$\mathbf{w}_m = d_m \mathbf{b}_m$

end for

$\mathbf{W}_a = \left[\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{N_r^{RF}}\right]$

$\mathbf{W}_d = diag\big(d_1, d_2, \ldots, d_{N_r^{RF}}\big)$

$\mathbf{W}_{opt} = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{N_r^{RF}}\right]$

Return: $\mathbf{W}_a$, $\mathbf{W}_d$, $\mathbf{W}_{opt}$

Table 4.4: Combining algorithm for HA-2

4.4 Performance Results

This section presents a set of performance results for both implemented hybrid receiver architectures HA-1 and HA-2. It is considered a clustered channel model with $N_{cl} = 8$ clusters, each with $N_{ray} = 10$ rays, with Laplacian distributed azimuth angles of arrival and departure. The average power of all N_{cl} clusters is the same. The antenna element spacing is assumed to be half-wavelength. The channel remains constant during a block, with size $T = 32$, and takes independent values between blocks. In the following, results for two different precoders will be present: hybrid based precoders ($\mathbf{F}_a$ and $\mathbf{F}_d$) and fixed hybrid random precoders that are independent of the instantaneous channel realization, i.e. the phases of each phase shifter

are set to a random value uniformly distributed between $-\pi$ and π, and the digital part of the equalizer ($\mathbf{W}_d$) is selected uniformly at random from the set of unitary matrices.

We will compare the performance of the presented schemes for three different scenarios, with the parameters described in Table 4.5.

	Transmit RF chains	**Transmit Antennas**	**Receive RF chains**	**Receive Antennas**	**Number streams**
Scenario 1	4	64	4	16	4
Scenario 2	8	128	8	32	8
Scenario 3	16	256	16	64	16

Table 4.5: General parameters of the different scenarios

The Scenario 1 can be seen as the baseline scenario, and the other two scenarios are a scaling of the first. Note that the parameters of Scenario 2 are two times higher than the ones of Scenario 1, and for Scenario 3 the parameters are two times and four times higher than Scenario 2 and 1.

The metric performance considered is the average BER, which is presented as a function of the E_b/N_0, with E_b denoting the average bit energy and N_0 denoting the one-sided noise power spectral density. We consider $\sigma_1^2 = \ldots, = \sigma_{N_s}^2 = 1$ and then the average E_b/N_0 is identical for all streams $s \in 1, \ldots, N_s$.

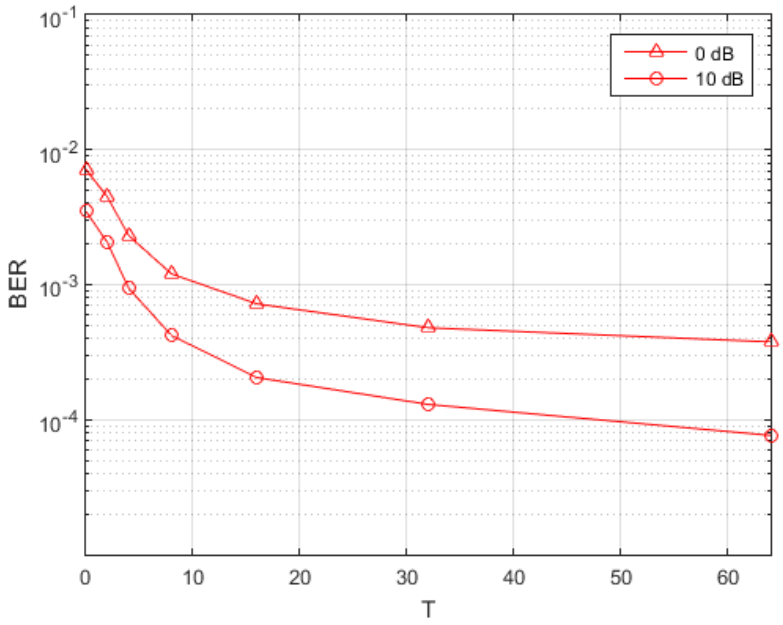


Figure 4.6: Performance of the proposed hybrid equalizer as a function of time slots (T) for HA-1 and Scenario 3

Let us start the analysis of the implemented analog-digital based architectures by verifying the impact of the parameter T. In figure 4.6 are presented the BER results for Scenario 3 and different values of the parameter T. These results were obtained for an $E_b/N_0 = 0$ dB and $E_b/N_0 = 10$dB, for the hybrid precoders of the HA-1. From figure 4.6 we verify that the decay of the BER with T is fast until $T \approx 32$ and for $T > 32$ the additional gains obtained by increasing the value of variable T are marginal. Therefore, all the following results are obtained for $T = 32$.

4.4.1 Scenario 1

This scenario is the one that requires the smallest number of antennas and RF chains involved in processing. We start by presenting the performance for both architectures HA-1 and HA-2 using hybrid random precoders as described above. As it can be seen from figure 4.7 HA-1 outperforms HA-2 as expected, since each RF chain is connected to all transmit antennas and therefore the diversity gain is higher (or in terms of capacity the achievable rate is higher). Taking into account the average BER at the target 10^{-3} the gap between both architectures is close to 6 dB. Figure 4.8 shows the performance comparison between the same architectures but now using hybrid precoders presented in Sections 4.2.3 and 4.3.3. For this case the HA-1 only outperforms the HA-2 for low SNR regime, for E_b/N_0 larger than -15 dB the HA-2 outperforms the HA-1. Therefore, using the hybrid precoders seems that in some cases the HA-2 outperforms HA-1 contrarily to the case of using random precoders.

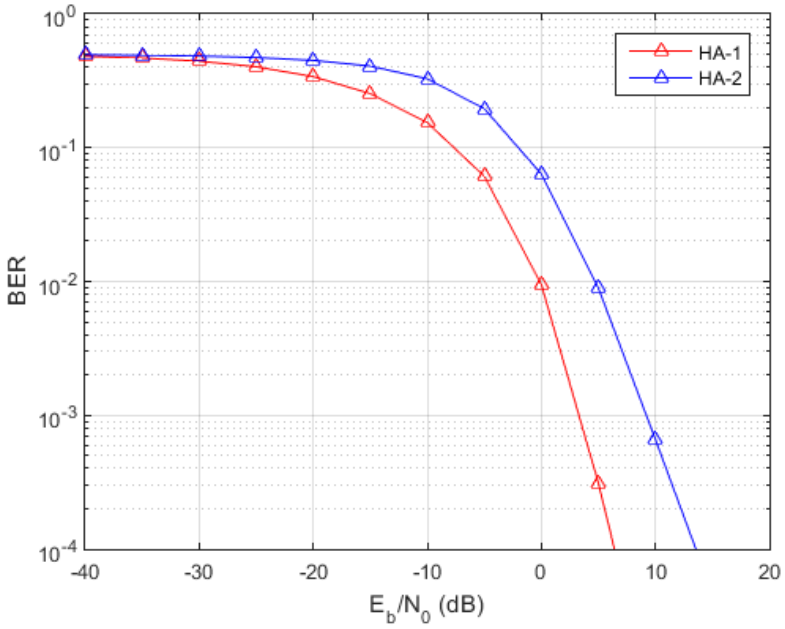


Figure 4.7: Performance comparison of both architectures considering hybrid random precoders, for Scenario 1

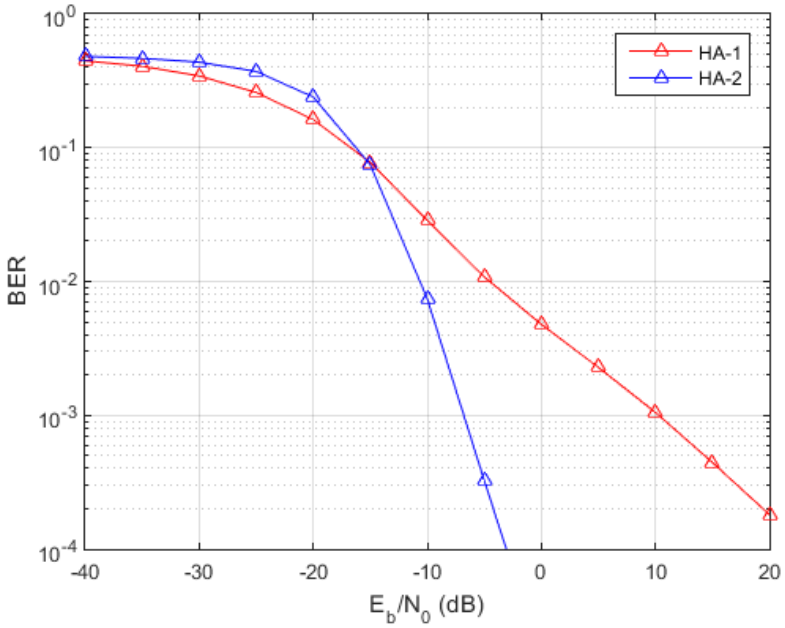


Figure 4.8: Performance comparison of both architectures considering hybrid precoders, for Scenario 1

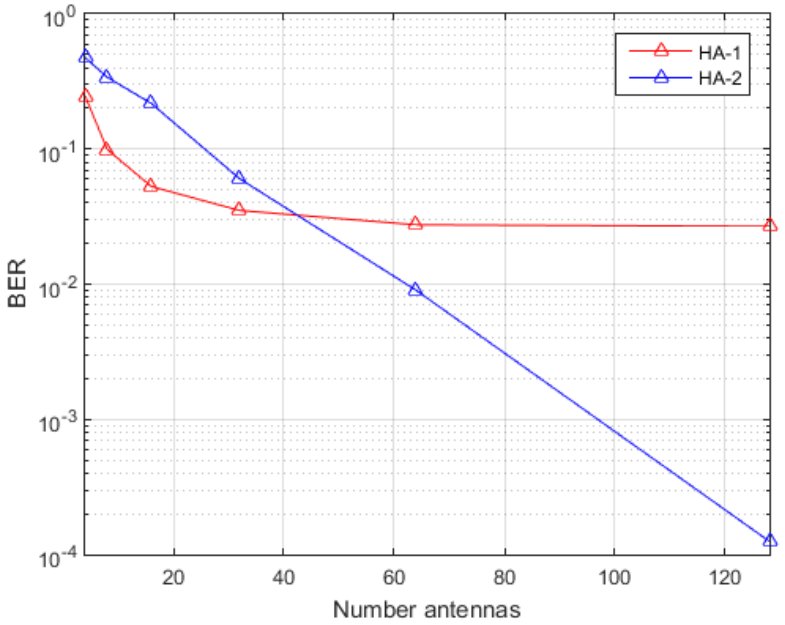


Figure 4.9: BER performance as a function of the number of transmit antennas considering hybrid precoders, for Scenario 1

To try to understand the previous results obtained for hybrid precoders, we also evaluate the performance impact with the number of transmit antennas for each architecture. The same values for the parameters described in Table 4.5 are considered, except the number of transmit antennas that varies from 4 to 128. The results were obtained for a $E_b/N_0 = -10dB$. From the results we can see that HA-1 outperforms HA-2 up to approximately 40 transmit antennas. After this point HA-2 clearly outperforms HA-1. For more than $N_t = 40$ the increasing gain obtained for HA-1 is marginal, while for HA-2 the performance still improve with the number of transmit antennas. This can be explained because both sides have 4 RF chains and this means that for the first set of antennas (N_t=4) each RF chain has just one antenna for the HA-2, while HA-1 has RF chains connected for all antennas. When we increase the number of antennas and keep the number of RF chains constant, we basically increase the number of antennas in each sub-array that contribute to achieve the near-optimal optimization.

4.4.2 Scenario 2

In Scenario 2 all parameters increased two times in relation to Scenario 1. Comparing the results obtained for this scenario and for random and hybrid precoders (figures 4.10 and 4.11, respectively) with the ones obtained for Scenario 1 (figures 4.7 and 4.8), we can see that a performance improvement of HA-1 and a performance degradation of HA-2. We can also see that for this scenario and considering random precoders HA-1 outperforms HA-2 for all E_b/N_0 range and for hybrid precoders there is a point (approximately -3 dB) from which HA-2 outperforms HA-1. From figure 4.10 (random precoders), we can see that the gap between HA-1 and HA-2, at a target BER of 10^{-3}, is 10.5 dB. The gap for the case where the hybrid precoders are used (figure 4.11) is 5.2 dB.

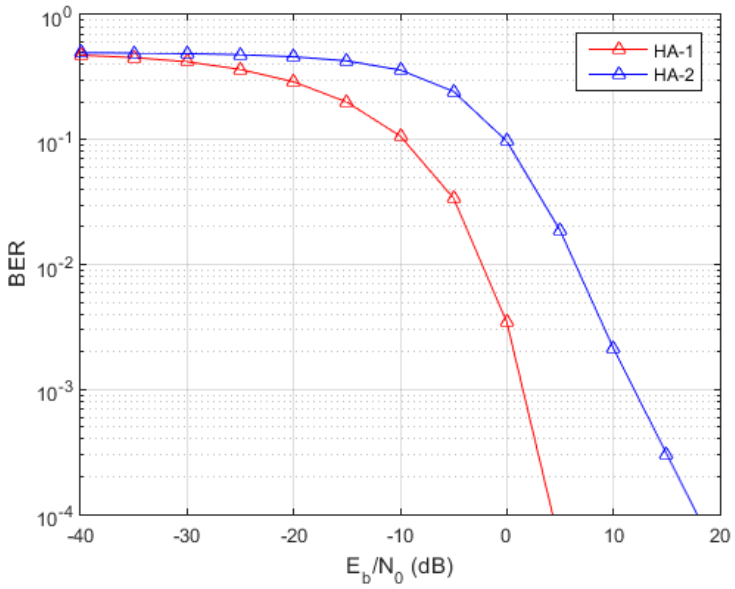


Figure 4.10: Performance comparison of both architectures considering hybrid random precoders, for Scenario 2

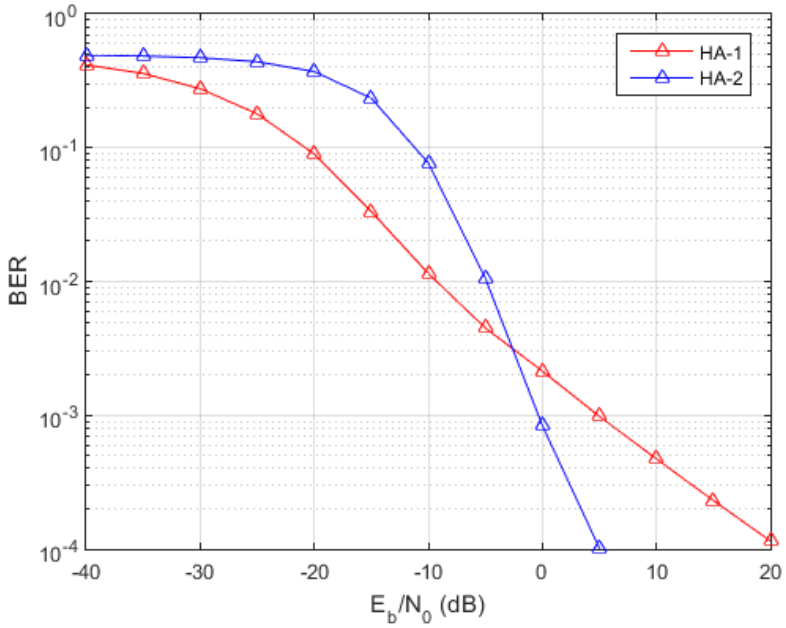


Figure 4.11: Performance comparison of both architectures considering hybrid precoders, for Scenario 2

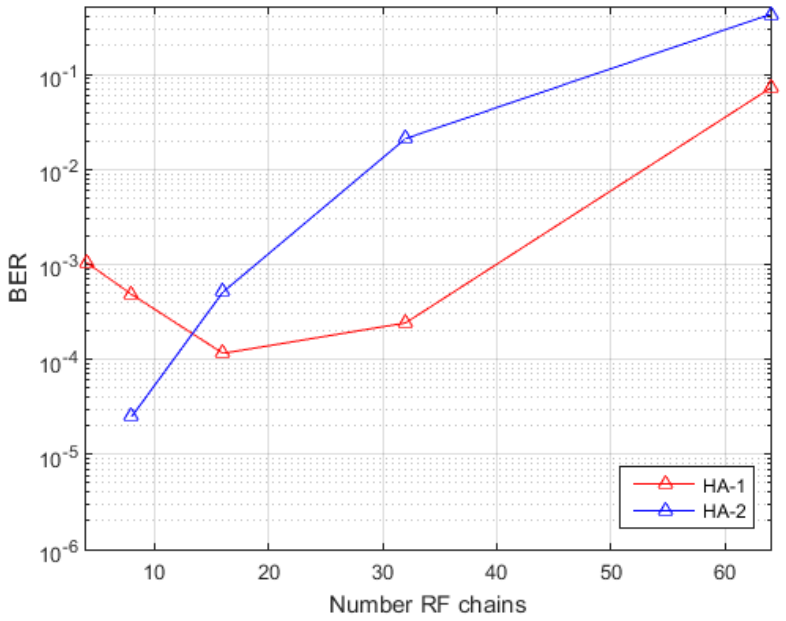


Figure 4.12: BER performance as a function of the number of RF chains considering hybrid random precoders, for Scenario 2

In Scenario 1, we analyzed the influence in the number of antennas in both architectures. In this scenario and to try to understand the performance degradation of HA-2 when compared to Scenario 1, we evaluated the BER performance as function of number of RFs chains as shown in figure 4.12. From this figure we can see that for HA-1 there is an optimal number of RF chains that is approximately 13, less or larger this number the performance degrades. For HA-2 we can see that the performance reduces as the number of RFs chain increases. With the increase of RFs chains HA-2 reduces the number of antennas for each sub-array and for that reason the performance is reduced.

4.4.3 Scenario 3

In Scenario 3 the number of antennas and RF chains in transmitter and receiver sides increases two and four times in relation with Scenario 2 and 1, respectively. We also obtained results for random and hybrid precoders as in the Scenarios 1 and 2. From figures 4.13 and 4.14 we can see that HA-1 outperforms HA-2 for both random and hybrid precoders and for all E_b/N_0 range. These results confirm the main conclusion of the previous scenarios. In fact HA-2 is highly sensitive to the number of RF chains since increasing it the number of antennas processed at each RF chain is reduced and impacts a lot on the system performance. From figure 4.13, with random precoders, we can see that the gaps between HA-1 and HA-2, at a target BER of 10^{-3}, is 15.8 dB. The gap for the case where de hybrid precoders are used (figure 4.14) is 13.8 dB.

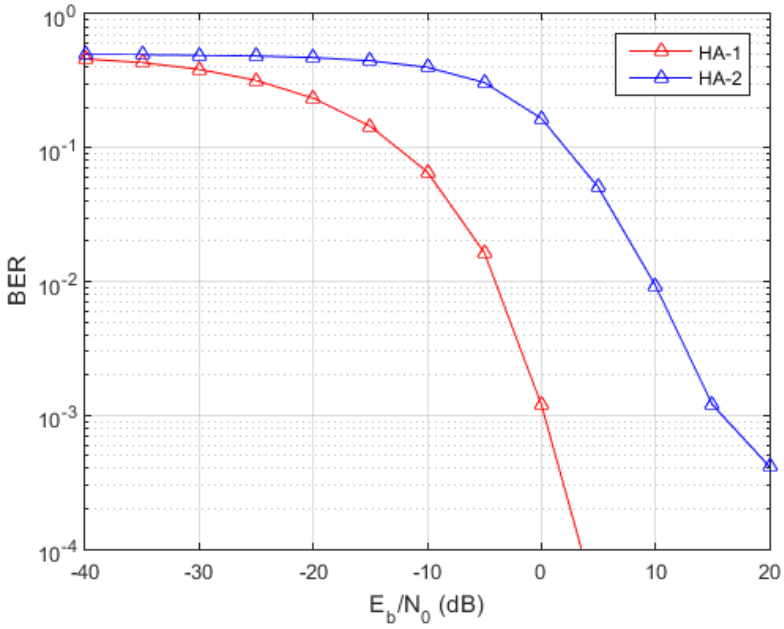


Figure 4.13: Performance comparison of both architectures considering hybrid random precoders, for Scenario 3

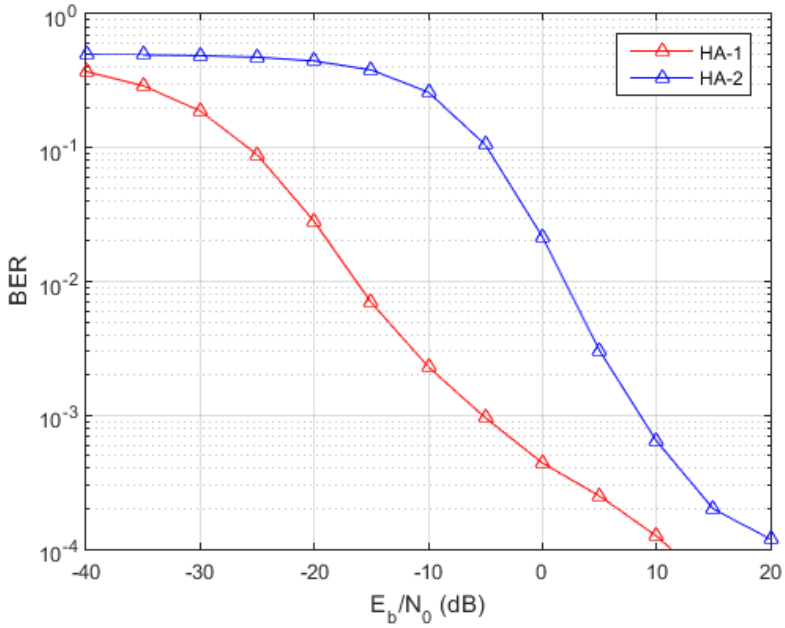


Figure 4.14: Performance comparison of both architectures considering hybrid precoders, for Scenario 3

Chapter 5

Conclusions and Future Work

The paradigm shift from 4G to 5G wireless system is driven by the expected huge growth in user bit rates and overall system throughput. The conventional wireless system is nearly reaching maturity and will not be able to support the new requirements. MmWave communications and mMIMO are seen as keys enabling elements for future 5G wireless communication. The adoption of mmWave transmission is mainly due to the huge bands available. Moreover, the small wavelength means small antennas, allowing transmitters and receivers to have a large number of antenna elements. This results in many challenges to be solved related to high propagation free-space losses, absorption mainly due to obstacles, small diffraction effects, different signal processing techniques, hardware complexity, cost, etc. Therefore, there is many open problems that should be solved in the incoming years. In this dissertation we contributed to understand part of these problems by implementing and comparing two different hybrid architectures and processing techniques for mmWave mMIMO systems.

5.1 Conclusions

This dissertation started by presenting the evolution of mobile communications from the first to 4th generation. Then a briefly discussion about the forthcoming for 5th generation was given. The progresses made over the years are remarkable, for mobile communications. To be truthful, mobile phones invaded our lives and they came to stay. Nowadays, these devices are much more than voice communication equipment, they became our diaries, notebooks, in other words, they are our personal assistant in "Anywhere, anything, anytime" in life and at work. Just see, 1st generation voice transmissions were analog, while the 2nd had moved to digital. With the 3rd generation it became possible to access multimedia data services, which changed the way we look at mobile phones. Nowadays, it is easily to access for everything almost instantaneously with LTE. It is predictable that in the near future mobile phones will increase their importance, serving us more and better services. Virtual Reality, Immerse or Tactile Internet and Multi-person videoconferencing which are now hard to imagine, will be a common reality.

In the 2nd Chapter we have seen the ways of the use of multiple antenna systems in order to achieve more throughput and coverage. Many of techniques studied are part of the current systems, exploiting diversity, spatial multiplexing and beamforming, usually on both sides of the system according to the requirements.

In the 3rd Chapter, the mmWave and mMIMO was described which is a promising approach to achieve the multi Gigabit per second required by future wireless systems. Small cells are also one of the key solutions to increase the capacity. Moreover, we identified relevant hybrid architectures for mmWave mMIMO systems, where the number of antennas is lower than the number of RF chains and the processing is done in two different domains, analog and digital. This allows the use of a massive number of antennas to perform highly directional beams. It becomes vital to transmit the signal in the appropriate direction to maximize the received signal energy and overcome the issues of high path loss in mmWave.

The complexity remains one of the major difficulties to overcome in conventional systems for the next generation. So, in the 4th Chapter, two different hybrid architectures were considered for mmWave mMIMO systems that can be part of the solution. The first one was designed for a full-array architecture where each RF chain is connected to all antennas, the second one is a sub-array architecture where each RF chain is only connected to a small set of antennas.

We evaluated these architectures for different parameters/scenarios and the main conclusions are as follows:

- For a small number of RF chains the performance of HA-2 is better than HA-1 for medium to high E_b/N_0 range. For higher numbers of RF chains HA-1 outperforms HA-2 for all E_b/N_0 range.
- The HA-2 is high sensitive to the number of RF chains, since increasing this number results in degradation of the system performance.
- To mitigate the previously problem the number of antennas per RF chain should be increased.

5.2 Future Work

Concerning the future work, we suggest the following improvements on the simulation platform:

- We considered a single-user scenario, but use hundreds of antennas for one user can not be a realistic scenario. Considering multi-user scenarios in future research, can potentially increase the interest in lower complexity hybrid architectures.
- In this dissertation it was assumed perfect CSI at both sides. It would be useful to evaluate the implemented architectures under imperfect CSI.
- It would also be interesting to extend the processing from 2D scenario to 3D scenario, where the elevation angles of the AoA and AoD are taken into account.
- The optimization problem in both architectures was done in terms of capacity, it would be interesting to formulate the optimization problem in terms of BER since the results are in function of BER.

Bibliography

[1] Vashu Kumar and Vipin Kumar. A new generation wireless mobile network-5g. *International Journal of Computer Applications*, 70(20), 2013.

[2] The history of mobile phones from 1973 to 2008. [Available online at] `http://www.knowyourmobile.com/nokia/nokia-3310/19848/history-mobile-phones-1973-2008-handsets-made-it-all-happen`. [Accessed] 2016-05-28.

[3] M. L. Roberts, M. A. Temple, R. F. Mills, and R. A. Raines. Evolution of the air interface of cellular communications systems toward 4g realization. *IEEE Communications Surveys Tutorials*, 8(1):2–23, First 2006.

[4] T. Emery. Cell phones: Hot/cold affair. *The Cincinnati Enquirer*, First 2004.

[5] Pankaj Sharma. Evolution of mobile wireless communication networks-1g to 5g as well as future prospective of next generation communication network. *International Journal of Computer Science and Mobile Computing*, 2(8):47–53, 2013.

[6] 5G globalization or regionalization. [Available online at] `https://siliconjoralemon.wordpress.com/2015/06/22/5g-globalization-or-regionalization/`. [Accessed] 2016-06-06.

[7] V. Pereira, T. Sousa, P. Mendes, and E. Monteiro. Evolution of mobile wireless: From voice calls to ubiquitous multimedia group communications. *2nd International Working COnference on Performance Modelling and Evaluation of Heterogenous Networks*, Seventh 2004.

[8] Christopher Cox. *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications*. John Wiley & Sons, 2012.

[9] Global mobile subscribers and market share by technology. [Available online at] `http://www.4gamericas.org/en/resources/statistics/statistics-global/`. [Accessed] 2016-05-31.

[10] Jorge Aido. MIMO Processing Techniques for Heterogeneous Cellular Systems. Master's thesis, Universidade de Aveiro, the Portugal, 2015.

[11] Ajay R Mishra. *Cellular technologies for emerging markets: 2G, 3G and beyond*. John Wiley & Sons, 2010.

[12] Farooq Khan. *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[13] H. Holma and A. Toskala. *LTE for UMTS - OFDMA and SC-FDMA Based Radio Access*. John Wiley Sons Ltd, New York, NY, USA, 2009.

[14] Summary of 3gpp standards releases for lte. [Available online at] `http://www.3gpp.org/technologies/keywords-acronyms/98-lte`. [Accessed] 2016-06-02.

[15] Alícia João. Equalization and precoding techniques for future MC-CDMA systems. Master's thesis, Universidade de Aveiro, the Portugal, 2014.

[16] G. A. Abed, M. Ismail, and K. Jumari. The evolution to 4g cellular systems: Architecture and key features of lte-advanced networks. *International Journal of Computer Networks and Wireless Communications*, 2(1):21–26, 2012.

[17] Ian F Akyildiz, David M Gutierrez-Estevez, and Elias Chavarria Reyes. The evolution to 4g cellular systems: Lte-advanced. *Physical Communication*, 3(4):217–244, 2010.

[18] Hetnet/small cells. [Available online at] `http://www.3gpp.org/hetnet`. [Accessed] 2016-06-24.

[19] Highlights of 3gpp release 12. [Available online at] `http://www.unwiredinsight.com/2014/highlights-of-3gpp-release-12`. [Accessed] 2016-06-24.

[20] Evolution of lte in release 13. [Available online at] `http://www.3gpp.org/news-events/3gpp-news/1628-rel13`. [Accessed] 2016-06-25.

[21] C. Sharma. *The Past, Present and Future of the Mobile Industry Evolution*. MobileFuture Forward, Seattle, USA, 2015.

[22] Peter Olanders. Network evolution towards 5g. Ericsson AB, 2015.

[23] D. Warren and C. Dewar. Understanding 5g: Perspectives on future technological advancements in mobile. GSMA Intelligence, December 2014.

[24] W. Tong and Z. Peiying. 5G a technology vision. [Available online at] `http://www1.huawei.com/en/about-huawei/publications/winwin-magazine/hw-329304.htm`. [Accessed] 2016-05-31.

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

[26] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino. Millimeter-wave massive mimo: the next wireless revolution? *IEEE Communications Magazine*, 52(9):56–62, September 2014.

[27] Fredrik Rusek, Daniel Persson, Buon Kiong Lau, Erik G Larsson, Thomas L Marzetta, Ove Edfors, and Fredrik Tufvesson. Scaling up mimo: Opportunities and challenges with very large arrays. *Signal Processing Magazine, IEEE*, 30(1):40–60, 2013.

[28] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar. Millimeter-wave beamforming as an enabling technology for 5g cellular communications: theoretical feasibility and prototype results. *IEEE Communications Magazine*, 52(2):106–113, February 2014.

[29] Y. Azar, G. N. Wong, K. Wang, R. Mayzus, J. K. Schulz, H. Zhao, F. Gutierrez, D. Hwang, and T. S. Rappaport. 28 ghz propagation measurements for outdoor cellular communications using steerable beam antennas in new york city. In *2013 IEEE International Conference on Communications (ICC)*, pages 5143–5147, June 2013.

[30] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir. Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. *IEEE Transactions on Antennas and Propagation*, 61(4):1850–1859, April 2013.

[31] Z. Pi and F. Khan. An introduction to millimeter-wave mobile broadband systems. *IEEE Communications Magazine*, 49(6):101–107, June 2011.

[32] S. Han, C. l. I, Z. Xu, and C. Rowell. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5g. *IEEE Communications Magazine*, 53(1):186–194, January 2015.

[33] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath. Mimo precoding and combining solutions for millimeter-wave systems. *IEEE Communications Magazine*, 52(12):122–131, December 2014.

[34] Omar El Ayach, Sridhar Rajagopal, Shadi Abu-Surra, Zhouyue Pi, and Robert W Heath. Spatially sparse precoding in millimeter wave mimo systems. *Wireless Communications, IEEE Transactions on*, 13(3):1499–1513, 2014.

[35] Xinyu Gao, Linglong Dai, Shuangfeng Han, Robert W Heath Jr, et al. Energy-efficient hybrid analog and digital precoding for mmwave mimo systems with large antenna arrays. *arXiv preprint arXiv:1507.04592*, 2015.

[36] Linglong Dai, Xinyu Gao, Jinguo Quan, Shuangfeng Han, and I Chih-Lin. Near-optimal hybrid analog and digital precoding for downlink mmwave massive mimo systems. In *Communications (ICC), 2015 IEEE International Conference on*, pages 1334–1339. IEEE, 2015.

[37] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, and H Vincent Poor. *MIMO wireless communications*. Cambridge university press, 2007.

[38] Ian Poole. Mimo formats - siso, simo, miso, mu-mimo. [Available online at] `http://www.radio-electronics.com/info/antennas/mimo/formats-siso-simo-miso-mimo.php`. [Accessed] 2016-05-07.

[39] Michail Matthaiou. Mimo systems in wireless networks. *Signal Processing Group*, 2011.

[40] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE-the UMTS long term evolution*. Wiley Online Library, 2015.

[41] A. Sezgin. *Space-Time Codes for MIMO systems Quasi-Orthogonal design and concatenation*. PhD thesis, Technische Universitt Berlin, Jun 2005.

[42] Manar Mohaisen, YuPeng Wang, and KyungHi Chang. Multiple antenna technologies. *arXiv preprint arXiv:0909.3342*, 2009.

[43] Alamouti space-time block coding. [Available online at] `http://www.nutaq.com/blog/alamouti-space-time-block-coding`. [Accessed] 2016-06-06.

[44] Yi Huang and Kevin Boyle. *Antennas: from theory to practice*. John Wiley & Sons, 2008.

[45] Khaled Fazel and Stefan Kaiser. *Multi-Carrier and Spread Spectrum Systems*. Wiley Online Library, 1997.

[46] A. Silva and A. Gameiro. Multiple antenna systems lecture. Technical report, Universidade de Aveiro, 2016.

[47] Branka Vucetic and Jinhong Yuan. *Space-time coding*. John Wiley & Sons, 2003.

[48] Alain Sibille, Claude Oestges, and Alberto Zanella. *MIMO: from theory to implementation*. Academic Press, 2010.

[49] Gustavo Anjos. Mimo processing techniques for 4g systems. Master's thesis, Universidade de Aveiro, the Portugal, 2013.

[50] Mohinder Jankiraman. *Space-time codes and MIMO systems*. Artech House, 2004.

[51] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[52] Sergio Verdu. *Multiuser detection*. Cambridge university press, 1998.

[53] D. Gesbert, M. Kountouris, R. W. Jeath Fr., C. Chae, and T. Salzer. From single user to multiuser communications: Shifting the mimo paradigm. [Available online at] `http://www.eurecom.fr/~gesbert/papers/TutorialMUMIMOv3.p`. [Accessed] 2016-06-15.

[54] Agilent Technologies Application Note. Verify and visualize your td-lte beamforming signals. 2012.

[55] Design and test challenges for 8x8 mimo devices for lte advanced. [Available online at] `https://www.wirelessdesignmag.com/article/2013/12/design-and-test-challenges-8x8-mimo-devices-lte-advanced`. [Accessed] 2016-06-13.

[56] F. Khan, Z. Pi, and S. Rajagopal. Millimeter-wave mobile broadband with large scale spatial processing for 5g mobile communication. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1517–1523, Oct 2012.

[57] Prasanna Adhikari. Understanding millimeter wave wireless communication. *VP of Business Development for Network Solutions, Loea Corporation, San Diego*, 2008.

[58] W. Hong, Y. G. Kim, K. h. Baek, and Y. Lee. Design and testing of a millimeter-wave beam-steering mesh-grid array for 5th generation (5g) mobile communication handset devices. In *Radio Science Meeting (Joint with AP-S Symposium), 2014 USNC-URSI*, pages 282–282, July 2014.

[59] Nicholas P Lawrence, Brian W-H Ng, Hedley J Hansen, and Derek Abbott. Analysis of millimeter-wave polarization diverse mimo capacity. In *2014 39th International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz)*, pages 1–2. IEEE, 2014.

[60] Qiang Xu, Houjun Sun, Jianyong Yin, and Xin Lv. A millimeter wave variable polarization monopulse frontend based on microstrip array antenna. In *Microwave Technology and Computational Electromagnetics, 2009. ICMTCE. International Conference on*, pages 264–267, Nov 2009.

[61] Qiao Xiaolin, Jin Tao, Qi Xiaohui, Zhang Min, Yuan Shuqing, and Zhang Qunxing. Anti-millimeter wave polarization agile active jamming. In *Microwave and Millimeter Wave Technology, 2007. ICMMT'07. International Conference on*, pages 1–4. IEEE, 2007.

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

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

[64] André BJ Kokkeler and Gerard JM Smit. Digitally assisted analog beamforming for millimeter-wave communication. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 1065–1070. IEEE, 2015.

[65] Xinyu Gao, Linglong Dai, Chau Yuen, and Zhaocheng Wang. Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive mimo systems. 2015.

[66] Ahmed Alkhateeb, Geert Leus, and Robert W Heath. Limited feedback hybrid precoding for multi-user millimeter wave systems. *IEEE Transactions on Wireless Communications*, 14(11):6481–6494, 2015.

[67] T. S. Rappaport, J. N. Murdock, and F. Gutierrez. State of the art in 60-ghz integrated circuits and systems for wireless communications. *Proceedings of the IEEE*, 99(8):1390–1436, Aug 2011.

[68] N. Moraitis and A. D. Panagopoulos. Millimeter wave channel measurements and modeling for indoor femtocell applications. In *2015 9th European Conference on Antennas and Propagation (EuCAP)*, pages 1–6, May 2015.

[69] E. Ben-Dor, T. S. Rappaport, Y. Qiao, and S. J. Lauffenburger. Millimeter-wave 60 ghz outdoor and vehicle aoa propagation measurements using a broadband channel sounder. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–6, Dec 2011.

[70] J. N. Murdock, E. Ben-Dor, Y. Qiao, J. I. Tamir, and T. S. Rappaport. A 38 ghz cellular outage study for an urban outdoor campus environment. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 3085–3090, April 2012.

[71] A.J. Paulraj and T. Kailath. Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (dtdr), September 6 1994. US Patent 5,345,599.

[72] Berthold Panzner, Wolfgang Zirwas, Stefan Dierks, Mads Lauridsen, Preben Mogensen, Kari Pajukoski, and Deshan Miao. Deployment and implementation strategies for massive mimo in 5g. In *Globecom Workshops (GC Wkshps), 2014*, pages 346–351. IEEE, 2014.

[73] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. Massive mimo for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, February 2014.

[74] Kan Zheng, Long Zhao, Jie Mei, Bin Shao, Wei Xiang, and Lajos Hanzo. Survey of large-scale mimo systems. *Communications Surveys & Tutorials, IEEE*, 17(3):1738–1760, 2015.

[75] Erik Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas Marzetta. Massive mimo for next generation wireless systems. *Communications Magazine, IEEE*, 52(2):186–195, 2014.

[76] X. Huang, Y. J. Guo, and J. D. Bunton. A hybrid adaptive antenna array. *IEEE Transactions on Wireless Communications*, 9(5):1770–1779, May 2010.

[77] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen. Design considerations for 60 ghz cmos radios. *IEEE Communications Magazine*, 42(12):132–140, Dec 2004.

[78] Shuangfeng Han, Chih Lin I, C. Rowell, Z. Xu, Sen Wang, and Zhengang Pan. Large scale antenna system with hybrid digital and analog beamforming structure. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 842–847, June 2014.

[79] R. Heath, A. Alkhateeb, J. Mo, and N. Gonzlez-prelcic. Millimeter wave mimo precoding/combining: Challenges and potential solutions, 2014.

[80] Qualcomm. Lte advanced: Heterogeneous networks. [Available online at] `https://www.qualcomm.com/documents/lte-heterogeneous-networks`. [Accessed] 2016-06-23.

[81] C. Li, J. Zhang, and K. B. Letaief. Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations. *IEEE Transactions on Wireless Communications*, 13(5):2505–2517, May 2014.

[82] H. Pervaiz, L. Musavian, and Q. Ni. Energy and spectrum efficiency trade-off for green small cell networks. In *2015 IEEE International Conference on Communications (ICC)*, pages 5410–5415, June 2015.

[83] W. Liu, S. Han, and C. Yang. Energy efficiency comparison of massive mimo and small cell network. In *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, pages 617–621, Dec 2014.

[84] Wenjia Liu, Shengqian Han, Chenyang Yang, and Chengjun Sun. Massive mimo or small cell network: Who is more energy efficient? In *Wireless Communications and Networking Conference Workshops (WCNCW), 2013 IEEE*, pages 24–29, April 2013.

[85] Z. Xu, S. Han, Z. Pan, and I. Chih-Lin. Ee-se relationship for large-scale antenna systems. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 38–42, June 2014.

[86] RK Jain, Sumit Katiyar, and NK Agrawal. Hierarchical cellular structures in high-capacity cellular communication systems. *arXiv preprint arXiv:1110.2627*, 2011.

[87] D. Chang. Acceleraating global wimax adoption: The move to picocell and femtocell base satations. [Available online at] `http://www.embedded.com/print/4009436`. [Accessed] 2016-06-27.

[88] J. Nsenga, A. Bourdoux, and F. Horlin. Mixed analog/digital beamforming for 60 ghz mimo frequency selective channels. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–6, May 2010.