Massive MIMO for 5G and beyond

Charul Thacker

le cnam

Massive MIMO for 5G and beyond

Project Report(ELE 209) by :Charul THACKER
Supervisor: Professor Pascal Chevalier

Date:01 December,2021

Contents

Abstract

Massive MIMO (multiple-input-multi-output) is one of the backbone technologies in 5G networks. Massive MIMO originated from the concept of multi-user MIMO. It consists of base stations (BSs) implemented with a large number of antennas to increase the signal strengths via adaptive beam forming and concurrently serving many users on the same time-frequency blocks. With Massive MIMO technology, there is a notable enhancement of both sum spectral efficiency (SE) and energy efficiency (EE) in comparison with conventional MIMO-based cellular networks. Resource allocation is an imperative factor to exploit the specified gains of Massive MIMO. Also,BS must have access to high-quality channel estimation that can be acquired via the uplink pilot transmission phase.

In particular, erroneous CSI is a substantial source of performance loss—about 2 to 3 dB loss is incurred even with the best CSI estimates in LTE-A. Accurate channel state information (CSI) is important for many candidate techniques of future wireless communication systems. However, acquiring CSI can sometimes be difficult, especially if the user equipment is mobile in which case the future channel realizations must be estimated/predicted.

The focus on this report is to study about 5G MIMO communication,implementation(beamforn and several algorithms that are widely used for Chanel estimation.

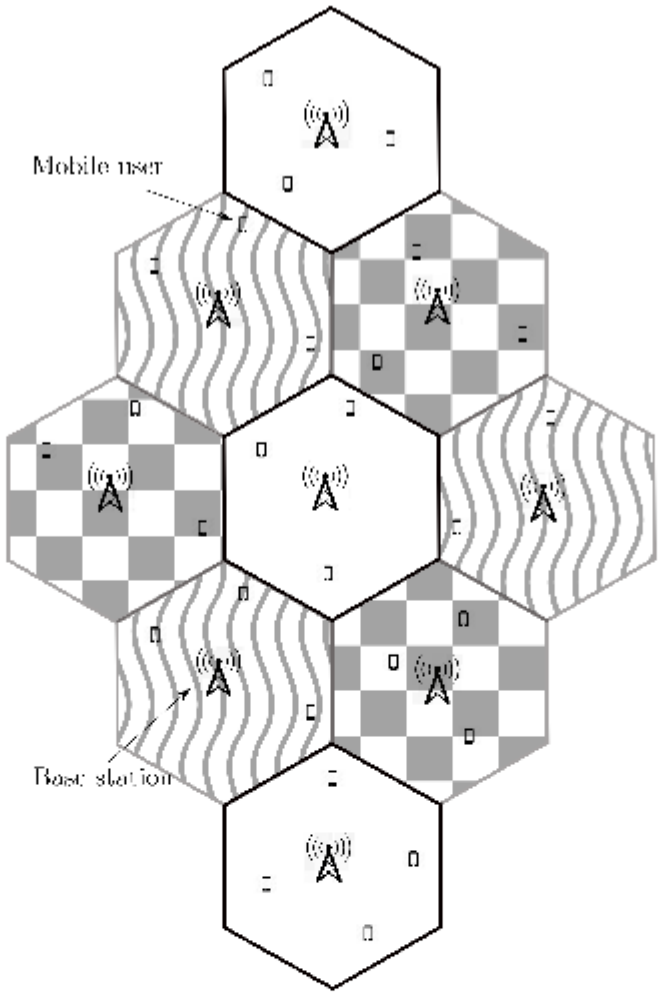
Chapter 1

Introduction

1.1 Cellular communication

Cellular communication is a fundamental technology where the transmitter and receiver communicate with nearby base stations (BSs) instead of directly with each other. This leads to wireless networks where transmissions take place over shorter distances and thereby more efficient utilization of the limited available frequency band for wireless communication. Figure 1 shows a cellular network in which the coverage area is divided into cells, where each cell has a fixed BS that the devices in the cell are connected to and provides service for them.

In addition, the frequency band can be divided into the frequency sub-bands, and each cell uses some of the sub-bands. The system also considers reusing the frequency sub-bands between the cells if sufficient distances separate them, which is called frequency reuse. Therefore, the frequency reuse is selected to balance between inter-cell interference and frequent reuse of the frequency bands. Note that in Figure 1, the cells with same pattern use the same frequency sub-bands. Hence, cellular technology helped wireless network designers to provide service for larger number of users in a given area and thereby accommodated the widespread usage of wireless communication. Ultimately, the idea of cellular technology resulted in the commercial implementation of wireless communication systems across the world.


Mobile user
Base station

Chapter 2

Massive MIMO

Massive MIMO originated as an extension of a multi-user MIMO system One of the greatest things in our era is that 5G is helping realize the Intelligent Internet of Everything (IIoE), bringing great changes to people's lives, many vertical industries and the entire society with making the world a better connected and digital one. Massive MIMO, as one of the core technologies of 5G, is key to meeting the high performance requirements and new service requirements of this amazing new era. Though Massive MIMO does offer great promises for highly capable 5G with wider bandwidth, more connections, lower latency and better reliability, realizing its full potentials requires effective responses to the challenges of network coverage, user experience, and network capability, which is relevant to all the mobile network operators and system vendors. After the Massive MIMO technology is introduced, the differentiation and flexibility of wireless network coverage in three-dimensional space have been greatly improved. The radio wave propagation model, user behavior and service distribution, beam management and beamforming are more complicated, flexible and difficult to measure. The location of problems in wireless networks, the effectiveness of response solutions, and the effectiveness and impacts of new functions become more complicated as the network scale increases. How to effectively predict, find, and evaluate the optimal solution in advance before the complicated real network encounters problems? While Massive MIMO enables 5G with much higher diversity and flexibility of network accessibility and capability in a three-dimensional space, the complexity of the network raises the questions in identifying network issues, offering effective solutions, and maximizing the benefits of the new technologies without paying too high a price.

2.1 Massive MIMO : A core technology for 5G and beyond

While the traditional radio devices often have just two, four, or maximum eight TRX channels, the radio devices powered by Massive MIMO technology can have 32 or 64 TRX channels, with up to 512 or even more antenna elements, which can lead to substantially higher capacity gain than traditional equipment. Furthermore, while the traditional devices focus more on coverage in horizontal dimension, Massive MIMO offers much better flexibility also in vertical dimension. Massive MIMO can exploit to a great extent the resources in space dimension and enable the users under the same base station to use the same time and frequency resources, which significantly enhances the network capacity without denser base stations and wider frequency bandwidth Taking synchronization signal and PBCH block (SSB) configuration as an example, SSB determines the basic coverage performance of the network. 4G broadcast channel is sent with a fixed wide beam, and its coverage does not change in most cases. However, 5G SSB can be configured with up to 7 (2.5 ms frame structure) or 8 (5 ms frame structure) beams according to frame structure. More SSB beams result in flexible configuration, i.e. multiple horizontal beams can be configured, or combination of horizontal and vertical beams can be configured. Different beams can be flexibly configured with different widths and heights, so that the 5G SSB beam configuration can support abundant scenarios and accurately meet differential coverage requirements. However, the increase in flexibility also brings a significant increases in configuration complexity. There are more than tens of thousands of combinations of antenna parameters configuration for 5G SSB beams. Here arises a huge technical problem on how to quickly and accurately find the configuration that is most suitable for the current scenario among tens of thousands of antenna parameters , and efficiently match the configuration with the change of scenarios and user behavior modes. Based on the quasi-orthogonal characteristics among multi-user channels, Massive MIMO can greatly improve the network capacity through SDMA. Due to the complexity of wireless channel propagation, and the randomness of user distribution and services, the design of a base station requires a well-performed algorithm for downlink transmission and uplink receiving to obtain a stable multi-user SDMA gain and anti-interference performance. Under the condition of a given number of antennas, the complexity of the Massive MIMO algorithm increases rapidly with the increase of the number of users and the maximum number of MU-MIMO multiplexing layers, which becomes one of the key technical difficulties affecting system capacity

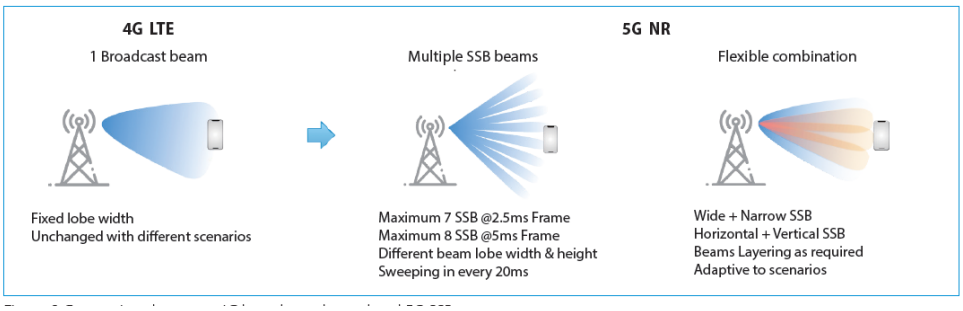


Figure 2.1: comparison between 4G and 5G broadcast

2.2 Multi-beam management and beamforming

5G SSB offers much better flexibility for broadcast channel coverage than 4G The broadcast control channel transmission and coverage of 4G – as the case of 2G and 3G, is through a wide beam.

In 5G era, the large-scale antenna array technology is introduced. Like the PDSCH service channel, through the cooperation of all antenna elements and RF transmission channels in the antenna array, 5G system provides SSB narrow beamforming capability. Multiple SSB narrow beams can be scanned and transmitted in time domain and space domain. In this way, SSB can achieve not only the same coverage performance as service channel, but also three-dimensional flexible coverage mode in horizontal and vertical dimensions. 5G, on the other hand, uses beamforming technologies based on a massive array of antenna elements and RF transmission channels to transmit multiple narrow beams of SSB in both time domain and space domain, in the same way of PDSCH transmission. This helps achieve the same coverage of SSB and of service channels and very flexible coverage in a threedimensional space.

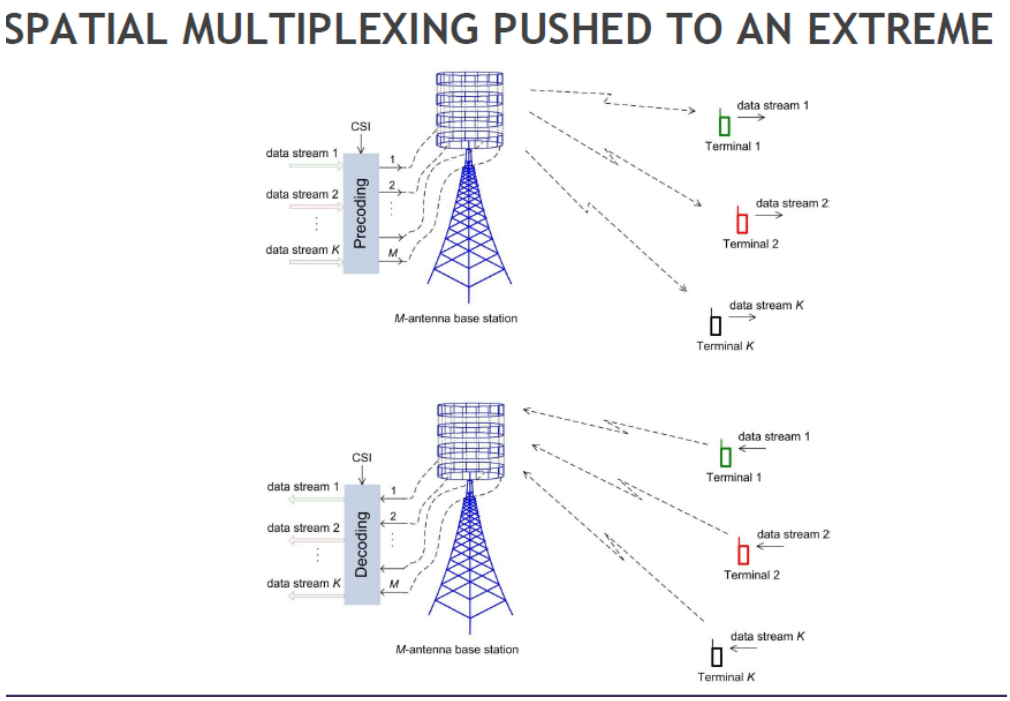


Figure 2.2: Massive MIMO serves all users over the same time/frequency resources

Chapter 3

Key Properties of Massive MIMO

3.1 Favorable Propagation

Favorable propagation is a phenomenon that appears when the propagation channels of two cellular users are mutually orthogonal. Favorable propagation helps the BS to cancel co-channel interference between users without having to design advanced algorithms for interference suppression. Consequently, it enhances the SE of both users. Let us assume that we have a single cell consisting of a BS that has M antennas and two single-antenna users. The vectors $g_1 \sim C\mathcal{N}(0, \mathrm{I}_M)$ and $\mathrm{g}_2 \sim C\mathcal{N}(0, \mathrm{I}_M)$ denote the channel responses of the two users over a narrowband channel. These vectors are circularly symmetric complex Gaussian distributed with zero mean and correlation matrix I_M and this channel model is known as independent and identically distributed (i.i.d) Rayleigh fading. In case the channel vectors are orthogonal, the inner product satisfies

$$\mathrm{g}_1^{\mathrm{H}}\mathrm{g}_2 = 0.$$

The BS can then separate the received signal from these two users without any loss in the desired signals. Let us assume x_1 and x_2 denote the data signals transmitted by these two users. The received signal at the BS is given by

$$\mathrm{y} = \mathrm{g}_1 x_1 + \mathrm{g}_2 x_2$$

Assuming the BS has perfect knowledge of both channel vectors, it can cancel the interference between the users by taking the inner product of the received signal y with the channel of the desired user. In addition, the noise effect is neglected for simplicity. For example, when considering user 1 , the inner product is

$$\mathrm{g}_1^{\mathrm{H}}\mathrm{y} = \|\mathrm{g}_1\|^2 x_1 + \mathrm{g}_1^{\mathrm{H}}\mathrm{g}_2 x_2 = \|\mathrm{g}_1\|^2 x_1,$$

that gives the desired signal of user one, since the part $\mathrm{g}_1^{\mathrm{H}}\mathrm{g}_2 x_2$ is zero thanks to orthogonality of the vectors in (1). This is an ideal situation for the BS, which is why it is called favorable propagation; however, this is not very likely to occur in practice or if the channel vectors are drawn from random distributions. However, in the case of Massive MIMO BSs, we can show that an approximate favorable propagation can happen asymptotically in the case of Rayleigh fading channels. It is defined as the inner product of the two normalized vectors satisfying :

$$\frac{\mathrm{g}_1^{\mathrm{H}}\mathrm{g}_2}{M} \to 0,$$

3.2 Channel Hardening

In this part, we define and explain the concept of channel hardening. Channel hardening refers to the fact that the channel is less susceptible to the small-scale fading effects and behaves more like a deterministic channel when utilizing all the antennas. Let us assume $g \sim C\mathcal{N}\left(0, \mathrm{I}_M\right)$ is the channel vector of an arbitrary user towards a Massive MIMO BS with M antennas, asymptotic channel hardening is defined as [10]

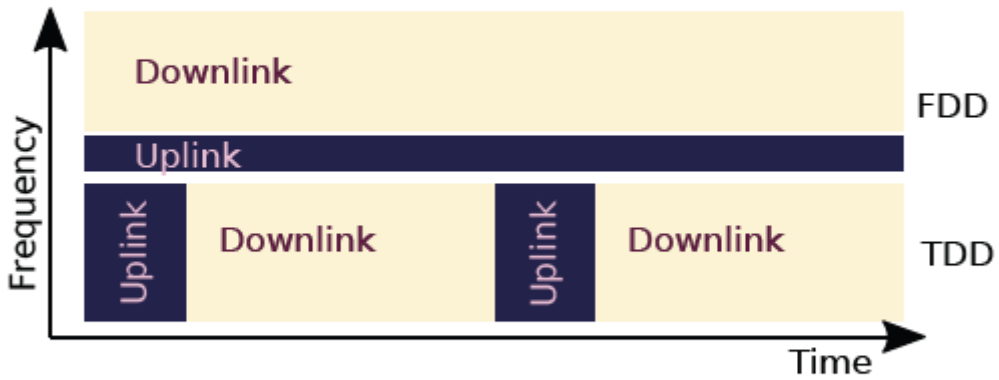
$$\frac{\|g\|^2}{\mathbb{E}\left\{\|g\|^2\right\}} \to 1$$

when $M \to \infty$ the convergence holds almost surely. Note that a squared norm in equation above when the BS processed the received signal, which is why its value is important for determining the communication performance. Asymptotic channel hardening implies that the value of $\|g\|^2$ is close to its mean value, so the variations are small. This phenomenon is an extension of the spatial diversity concept from conventional small-scale MIMO systems to the case of having a large number of antennas at the BSs. Channel hardening implies that the channel quality $\|g\|^2$ for a given channel realization is well approximated by the average channel quality $\mathbb{E}\left\{\|g\|^2\right\}$. Hence, if we want to select power coefficients based on the channel quality, we do not need to adapt them to the small-scale fading variations, but the same power can be used for a long time period. We consider channel hardening as one of the essential benefits of Massive MIMO systems, which helps us to propose practical power control schemes in the included papers in this thesis.

3.3 TDD and FDD for Massive MIMO

In order to process the uplink and downlink signals, each BS needs to estimate the channel vectors of its serving users in each channel coherence block. A coherence block is defined as the time-frequency block in which the fading channel is static.

In Massive MIMO, we assume that full statistical channel state information is available at the BSs. However, one should perform channel estimation at each BS, to obtain the instantaneous channel state information. Channel estimation is performed via pilot transmission . In the pilot transmission phase, each transmitter (e.g., a cellular user in uplink pilot transmission) sends one of the sequences from the set of predefined pilot signal sequences known by both the transmitter and receiver (e.g., the BS in the uplink pilot transmission). To estimate the channel from the transmitter, the receiver compares the signal with the "true" signal from the set. To support the pilot transmission of multiple transmitters in Massive MIMO systems, we generally prefer to have the same number of orthogonal pilot sequences as the number of transmitting antennas. It is also desirable to keep the pilot signals as short as possible to use most of the resources in a coherence block for data transmission. Pilot transmission for channel estimation in downlink and uplink of a Massive MIMO system requires a different number of pilot symbols. In the uplink, assuming thatwe have K single-antenna users, the system requires K pilot signal sequences to estimate the uplink channels. However, if the BS has M antennas, pilot transmission in the downlink require M pilot signals, where $\mathbf{M} >> \mathbf{K}$is normal in Massive MIMO systems.



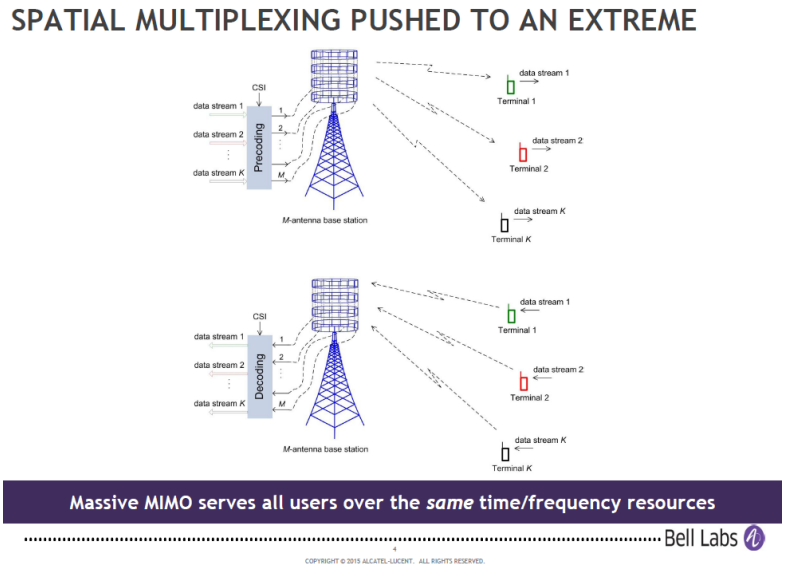


Figure 3.1: FDD need training time proportional to 2M+K whereas for TDD it is proportional to K

TDD refers to separating the uplink and downlink transmissions in the time domain while using the whole bandwidth; assuming that both happen in the same coherence time, channel reciprocity holds. It means that the channel is the same in both directions. Hence, by doing channel estimation in one direction (i.e., uplink here), the estimated channel is valid for the other direction (i.e., downlink) as well. Therefore, in TDD Massive MIMO systems, we require K pilot sequences only. Hence, channel estimation does not depend on M. In FDD, the uplink and downlink transmission occur simultaneously but in different frequency bands. Hence, due to the different frequency bands for uplink and downlink, the channel reciprocity does not hold . Consequently, we need to estimate the channels separately for each direction. Therefore, we require both uplink and downlink pilots for channel estimation in FDD. In the downlink, we need M pilot signal sequences and an additional M signals for reporting back the estimated channel to the BS in the uplink. Besides, we need K pilots for uplink channel estimation. In total, assuming the resources are equally decided between uplink and downlink, FDD needs $(2\mathbf{M}+1)/2$ pilot signals. The time-frequency separation of these two protocols is illustrated in Figure above, please note that in FDD there is usually many 100 MHz between the uplink and downlink. One can see that channel estimation overhead in TDD Massive MIMO is substantially smaller than in FDD Massive MIMO, and it is not scaled with M. Therefore, **TDD is a preferable duplexing mode for Massive MIMO systems**. On the other hand, low latency requirements

could be easier to handle in an FDD system. If, for example, an automatic control system needs a small piece of data within a short time frame, but the system has just switched to an uplink slot, then there is a good chance that the information will be invalid by the time the system reaches its down link slot(4G system use FDD).

Chapter 4

Orthogonal frequency division multiplexing:

4.1 An overview

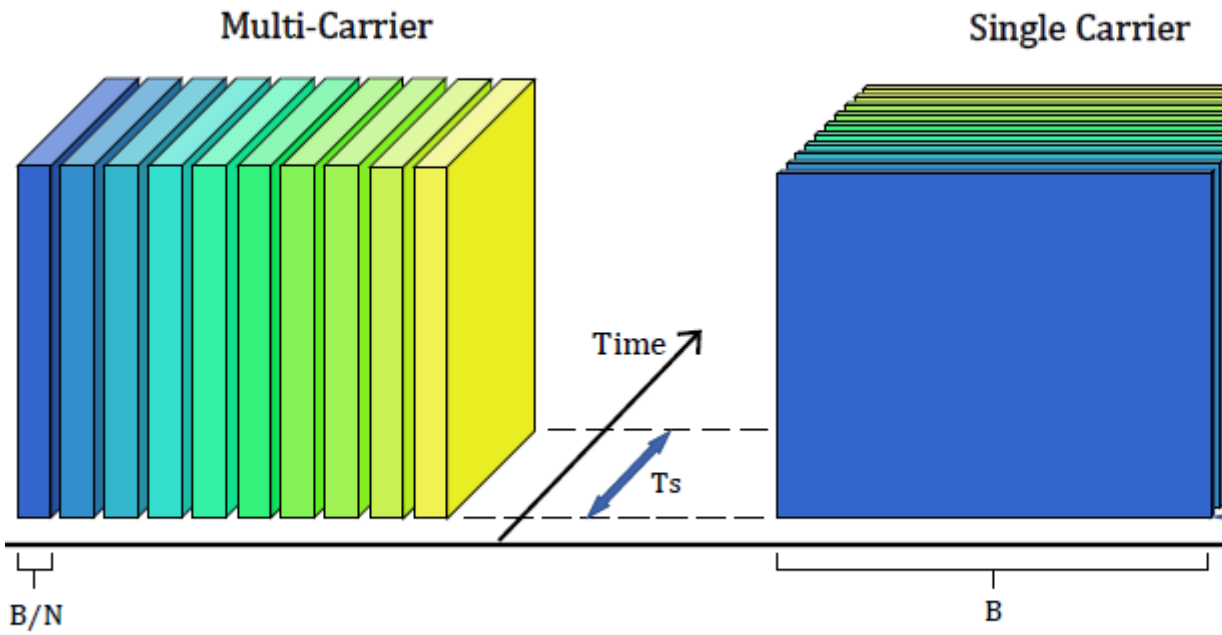
In an Orthogonal Frequency Division Multiplexing (OFDM) system a broadband signal is created as several narrow band signals that are super positioned into one time limited signal, denoted one OFDM symbol, before being transmitted over the radio channel. Each of these narrow band signals, often referred to as sub carriers, can then be used to encode separate pieces of information, or messages. By adjusting the frequency band of the narrow band signals based on the time duration of the OFDM symbol, modulated narrow band signals can be made orthogonal over the symbol time such that, under ideal assumptions, the messages encoded on different sub carriers will not interfere with each other. Under realistic assumptions, the system and receiver can be designed to ensure that interference between subcarriers is very small, if the transmitter and receiver are synchronized in time and frequency with sufficient accuracy. Likewise, the system is often designed to ensure that interference between subsequent OFDM symbols in time, Inter-Symbol Interference (ISI), can be considered negligible.

In an OFDM system, a single subcarrier over the duration of a single OFDM symbol is referred to as a time-frequency resource or simply resource. Just as the channel changes over time, it may also change over frequency. A channel can either be flat fading, with constant channel gain (although different phase) over all subcarriers, or it can be frequency selective, in which case the gain varies over different subcarriers. At the base station, a scheduling algorithm will be used to assign resources to each user equipment within the system. Most utilized scheduling algorithms are based on some CSI, which may consist of the complex-valued channel gains or simply a Channel Quality Index (CQI). The CQI may include

information of the Signal to Noise Ratio (SNR) of the subcarrier, or simply information on which subcarriers have channels that are above a given SNR threshold. The scheduler will aim to schedule messages on the resources where the channels of a user are good.

4.2 OFDM and GFDM for Massive MIMO

the extension of MIMO/BLAST techniques to OFDM schemes is simple, perhaps with additional pre-processing and/or employing adaptive loading schemes. However, OFDM signals carry high envelope fluctuations and a high Peak-to-Average Power Ratio (PAPR) leading to amplification drawbacks. Because of this, numerous techniques for reducing the envelope fluctuations of OFDM signals have been proposed. However, these techniques require an increased number of signalprocessing tasks, especially on the transmitter side, and possibly some signal distortion when a nonlinear (NL) signal processing is employed, such as the amplitude clipping technique. Despite the additional complexity, the PAPR reduction techniques do not achieve a null PAPR or a value near zero, essential to maximizing the efficiency of RF signal power amplification. Thus, even for the most sophisticated PAPR reduction techniques, the transmitted signals still have PAPR higher than those for SC signals based on similar constellations, which makes an efficient amplification difficult. Single Carrier (SC) modulations, combining block transmission techniques and FDE, are an alternative approach for broadband wireless systems. Like in OFDM modulations, a Cyclic Prefix (CP) is appended to data blocks, long enough to cope with the channel length. The received signal is converted to the frequency domain, equalized in the frequency domain, and then transformed back to the time domain. A simple time-frequency domaincomparison between SC and Multi Carrier (MC) modulations is presented in Figure.



The overall implementation complexity, as well as the achievable performance, is similar for SC schemes with FDE and OFDM schemes. However, the signal-processing load is more intense at the receiver for the SC case. This situation, combined with the lower envelope fluctuations of SC signals, makes them more suitable for uplink transmission (i.e., the transmission from the User Equipment (UE) to the Base Station (BS)), while the OFDM schemes prevail as a better choice for the downlink transmission (i.e., the transmission from the BS to the UE)In most of the cases, a linear FDE is used at the receiver, although it was already proved that NL equalizers can have significantly better performance than linear equalizers . For this reason, it is beneficial to design NL equalizers for SC-FDE schemes. Among several different NL equalizers, DFE is especially attractive due to its good performance-complexity trade-off. A hybrid time–frequency SC-FDE employing a frequency-domain feed forward filter and a time-domain feedback filter were proposed. Although this scheme can provide better performance than a linear FDE, it can suffer from error propagation as in the conventional time-domain version, especially if the feedback filter has a large number of taps.In terms of the achievable data rate, which has been proven tight when the channel hardens, SC and OFDM transmission are equivalent in massive MIMO. Due to channel hardening, all tones

of the OFDM transmission have equally good channels, therefore the advantage of using OFDM and employing the water filling method across frequencies results in little gain. The PHY layer radio access network emerges as critical design for the envisioned service architecture. If synchronism is needed, the transceivers need to operate with a common clock for their processing. Furthermore, when we consider OFDM, the waveform detection process is free of crosstalk only when orthogonality is assured. Both aspects are related, such that some type of synchronization is often required to establish orthogonality. However, as soon as the orthogonality is broken (for example, due to random channel access or multi-cell operation), the signal distortion grows extremely in OFDM. In fact, 5G is expected to be centered on OFDM-based schemes. Nonetheless, waveforms like Discrete Fourier Transform spread OFDM (DFT-S-OFDM), Block-Windowed Burst OFDM (BWBOFDM), and Generalized Frequency Division Multiplexing (GFDM) include SC-FDE as special cases, with the associated advantages, especially for the uplink transmission . Moreover, massive MIMO is not restricted to 5G, and the advantages of SC-FDE-based mm MIMO schemes apply to other scenarios.

4.3 GFDM

1 GFDM principles

Generalized Frequency Division Multiplexing (GFDM) is a flexible multicarrier modulation scheme. The modulation is performed block by block, where each GFDM data block consist of certain number of subcarriers and subsymbols. By setting the number of subcarriers and the number of subsymbols to 1, GFDM allows single-carrier frequency domain equalization (SC-FDE) and CPOFDM as its special cases, respectively. Furthermore, pulse shaping with a prototype filter $g_{0,0}(m)$ is another flexibility in GFDM to reduce out-of-band (OOB) emissions. In contrast to linear convolution used in FBMC, GFDM brings circular convolution into play. Let $g_{k,n}$ denote the pulse shape corresponding to the data symbol $s_{k,n}$, that is transmitted at subcarrier n and time k, it can be written as

$$g_{k,n}[m] = g_{0,0}\left[(m - kN_{\text{sub}}) \bmod (N_{\text{sub}}K)\right] \cdot e^{j2\pi \frac{n}{N_{\text{sub}}} m}$$

where K denotes the number of sub-symbols within a GFDM block. Thus the time domain signal $x(m)$ of a GFDM block is expressed as

$$x[m] = \sum_{n=0}^{N_{\text{sub}}-1} \sum_{k=0}^{K-1} g_{k,n}[m] s_{k,n}$$

Optionally, cyclic prefix (CP) and cyclic suffix (CS) can be added in the GFDM data block.

Chapter 5

MIMO Implementation

Digital beamforming, provides the highest flexibility in terms of the possible beamforming algorithms that can be employed. This comes from the fact that, by digitally manipulating the signal, it is possible to adjust the phase and amplitude of each signal that feeds an antenna element. This scenario, shown in Figure, requires each antenna connected to the baseband through a dedicated mixer, a DAC, a filter, and an amplifier, i.e., an entire RF chain. This turns the implementation of digital beamforming in a massive MIMO architecture with hundreds of antennas in an expensive and challenging task due to the high power consumption, complexity, and cost.

Digital beamforming structure is shown above. Millimeter wave frequencies will probably be operated together with massive antenna arrays to overcome propagation attenuation. This makes a fully digital user separation not feasible, since the amount of energy required for all the analog-to-digital and digital-toanalog conversions would be huge. A possible solution is to attribute to each user an own radio-frequency beamforming matrix, involving users to be separated in time rather than frequency. In other words, it may be difficult to have a fully digital beamformer that is sub-carrier dependent due to hardware complexity constraints, while it is simpler to have a unique beamformer for the whole available Radio Frequency (RF) bandwidth that can change at the time slot rate .

On the other hand, the analog beamforming is a simpler and cheaper variant as the baseband is connected to the multiple antennas through phase shifters. These will be connected to each antenna so that it is possible to adjust only the signal phase. Phase shifters reduce the hardware limitations, allowing for low complexity implementations. However, the performance of fully analog beamforming techniques is limited, and it is usually only used for single-stream transmission. These constraints make it very difficult to form multiple beams, tune the sidelobes with accuracy, or steer the nulls.

The analog beamforming is a simpler and cheaper variant as the baseband is

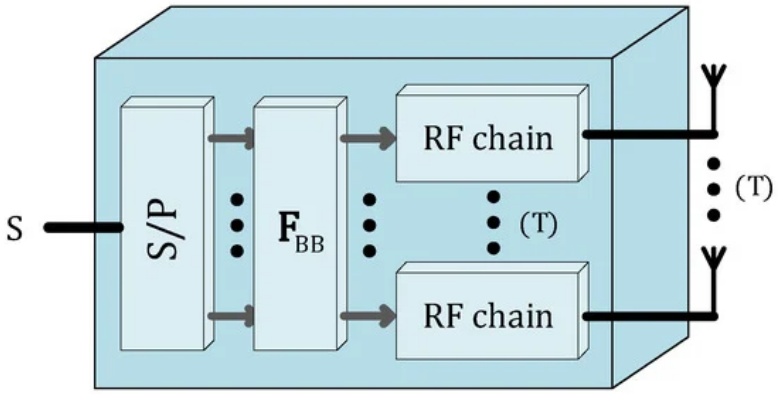


Figure 5.1: Digital beamforming

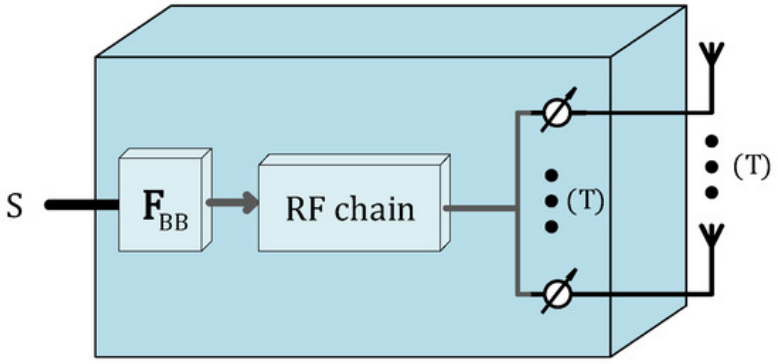


Figure 5.2: Analog beamforming structure

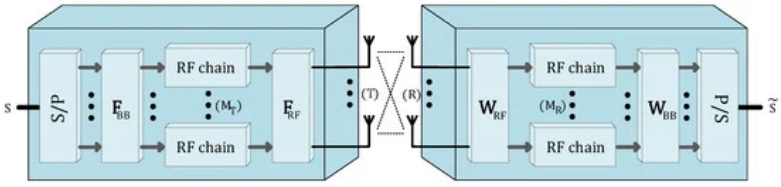


Figure 5.3: Hybrid MIMO transmission system structure

connected to the multiple antennas through phase shifters. These will be connected to each antenna so that it is possible to adjust only the signal phase. Phase shifters reduce the hardware limitations, allowing for low complexity implementations. However, the performance of fully analog beamforming techniques is limited, and it is usually only used for single-stream transmission. These constraints make it very difficult to form multiple beams, tune the sidelobes with accuracy, or steer the nulls.

5.1 Fully connected and subconnected Hybrid beamforming architecture

The need for a suitable signal processing scheme for massive MIMO sparked the study and analysis between the digital and analog implementations trade-offs . Hybrid beamforming have emerged as an approach which combines the best of both worlds by using a small number of RF chains and connect them to the antenna array through, for example, a stage of analog phase shifters . This approach is motivated by the fact that the number of up-down conversion chains is only lower-limited by the number of data streams that are to be transmitted. This solution has attracted the attention from the academy and even the industry . However, there is still the challenging task to design optimal hybrid beamforming schemes, which have a complex nature due to the non-convex constant modulo constrain imposed by the analog phase shifters . In addition to the non-convex constrain, phase shifters are usually controlled digitally and have a discrete resolution. These conditions create a large number of possible combinations of beamforming weights and phases to be optimized, which carries a considerable computational complexity.

Regarding the implementation of hybrid beamforming with analog and digital domains, two types of architectures have emerged: fully connected and subconnected architectures shown in Figures below, respectively. In the fully connected architecture, each RF chain is connected to all of the transmitter antennas through an analog device (switch, phase shifter, etc.). This architecture enables a bigger number of signal combinations and adjustments , but the optimization of

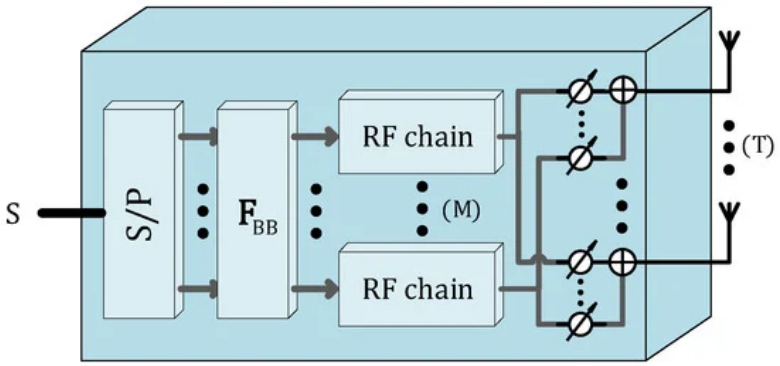


Figure 5.4: Hybrid beamformer with a fully-connected structure.

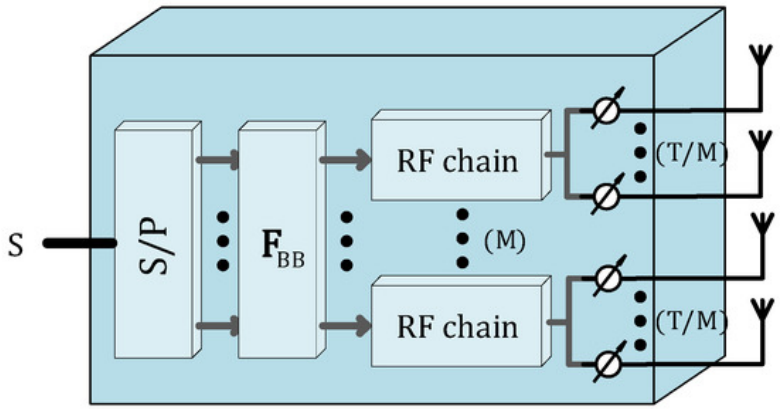


Figure 5.5: Hybrid beamformer with a sub-connected structure.

the digital and analog precoder can have a high computation complexity.

In the sub-connected architecture, each RF chain connects only a subset of antennas . When compared with the fully connected counterpart, sub-connected architectures allow a smaller number of phase shifters. Thus, the power consumption is reduced, and the computational complexity is also lower. It is also important to mention that subconnected architecture can be implemented in a dynamic or in a fixed way. In the dynamic sub-connected case, each RF chain can dynamically connect to a different set of antennas, and, in the fixed sub-connected one, each RF chain is always physically connected to the same set of antennas.

As the number of RF chains is increased, the efficiency of these power amplifiers becomes crucial to satisfy the energy efficiency requirements in massive MIMO systems. Thus, the use of NL power amplifiers in RF amplification stage allows for minimizing power consumption, being particularly important for high bit rate scenarios and critical at Gigabit data rates and above. However, to avoid signal

distortion at NL amplifiers, the transmitted signal should have a low PAPR, which is difficult to achieve, especially in OFDM signals or SC signals using high order constellations and/or highly selective filtering. It is then essential to employ recoding and PAPR reduction techniques. However, as stated before, PAPR reduction techniques have limitations and are usually associated with an increased processing complexity and/or nonlinear distortion. Furthermore, using such power-efficient amplifier constraints can require the use of precoding techniques, whose complexity and power consumption might compromise the energy efficiency of the entire system. It was shown that SC modulation can, in theory, achieve near-optimal sum rate performance in massive MIMO systems operating with a low-transmit-powerto- receiver-noise-power ratios, distinct from the channel power delay profile and with an equalization-free receiver. As already mentioned, SC modulation maintains an

almost constant envelope, yielding an optimal PAPR performance. Conventional MIMO systems can employ both NL precoding and linear precoding techniques without favoritism, although NL methods such as lattice-aided methods and dirty-paper-coding have better performance sacrificing the implementation low complexity. A solution for the PAPR problem is given, where a quasi-constant symbol decomposition is performed prior to the amplification stage. This system was further developed, and the impact of different symbol decomposition was evaluated. Contrary to the conventional MIMO, massive MIMO systems can use linear precoders, such as MRT, MMSE, and ZF in order to reduce the implementation complexity [20].

ZF is a well-known combining algorithm which can also be applied in digital beamforming and can nullify the multiuser interference in a multi-user MIMO system. In this case, the ZF matrix FZF is calculated directly from the propagation channel $\mathbf{H}$ as follows:

$$\mathbf{F}_{BB}^{\mathrm{ZF}} = \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H\right)^{-1}$$

MMSE technique is an alternative where the principle of this method is to minimize the MMSE between the actual transmitted data and the received signal. The matrix associated with this optimization is:

$$\mathbf{F}_{BB}^{\mathrm{MMSE}} = \left(\mathbf{H}^H\mathbf{H} + 2\sigma_n^2\mathbf{I}\right)^{-1}\mathbf{H}^H$$

where σ_n^2 is the noise variance. Contrarily to ZF, the MMSE receiver promotes a better combination of interference reduction and noise enhancement, since it is designed to minimize the total noise.

The maximum ratio technique was employed in the transmission in [79] and studied as a receiving technique in [80]. The signals from all the antenna elements are weighted with respect to their SNR, being the optimum weights matched to the wireless channel. The amplitude is changed, and the phase of the individual

signals must be adjusted, thus requiring an individual RF chain and phasing circuit for each antenna element. MR provides an output SNR equal to the sum of the individual SNRs, which produces the best statistical reduction of fading of any known linear diversity technique [81]. In MR, precoding matrix will be calculated as:

$$\mathbf{F}_{BB}^{\mathrm{MRC}} = \frac{\mathbf{H}^{H}}{T}$$

Chapter 6

Channel estimation

6.1 Need of channel estimation

In coherent demodulation schemes, channel estimation and equalization represent two key aspects in wireless communication systems. The channel estimation procedure allows us to know how is the propagation environment of our transmitted signal, and then, compensate its effects by equalization, which is crucial for achieving reliable communication with high data rates. The random matrix theory shows that the effects of small-scale fading, interference and noise can be effectively suppressed when the number of antennas is very large. Hence, using linear detection schemes, such as maximum ratio combining (MRC), will provide an optimal performance. However, when the number of antennas is not large enough, equalizers based on zero-forcing (ZF) or minimum mean square error (MMSE) criterion must be used in order to keep the performance. However, these equalizers require a matrix inversion which is a very hard operation in terms of complexity, increasing the delay of the link. Furthermore, channel estimation is also a challenging task when the number or antennas is very large, due to high number of different channels that must be estimated. This fact requires a long training period, where the BS and the UEs must exchange a great amount of pilot-sequences, decreasing the overall efficiency of the system.

6.2 mmWave

5G will use two key frequency ranges to support heterogeneous services. The first band corresponds to the frequencies below 6GHz, whose objective is to provide a full coverage in urban, suburban, and rural areas; and to support massive MTC and transportation services. The second band corresponds to the spectrum above 6GHz, known as mmWave , which will be used to provide ultra high broad-band

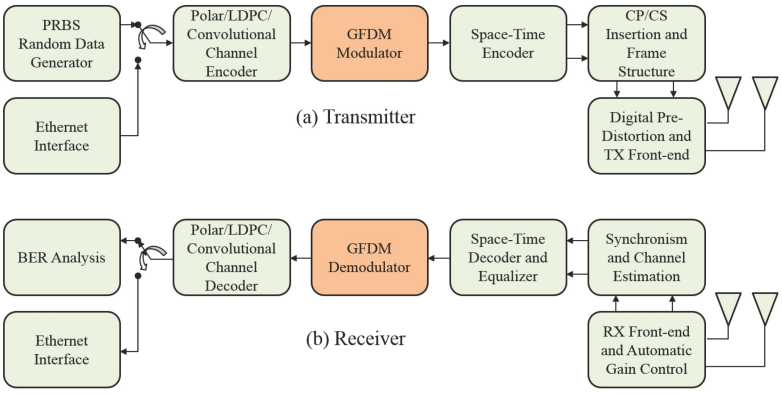


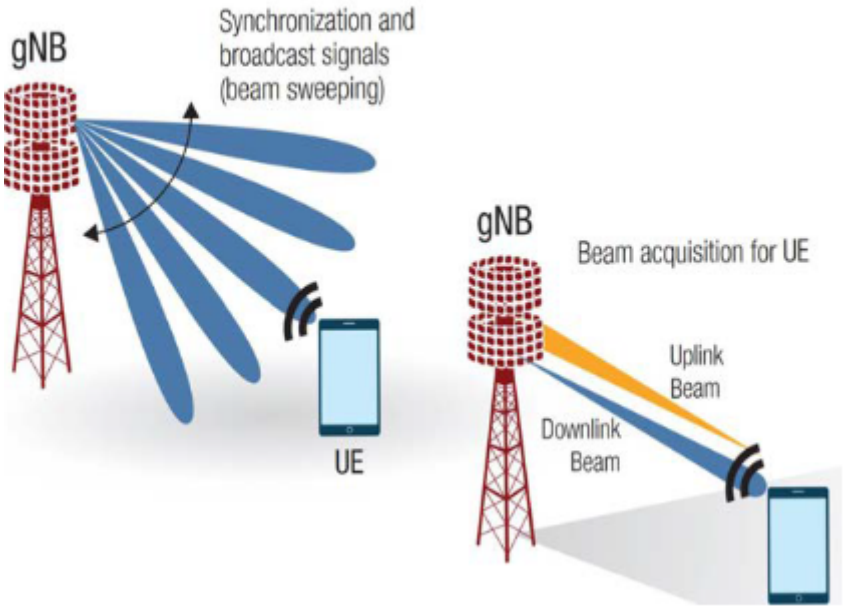
Figure 6.1:

services. However, the transmission at these bands suffers from new physical effects that were not so severe in the traditional low-frequency bands, such as a significantly higher path loss and susceptibility to blockage, among others. Note that in mmWave, large antenna arrays can be easily manufactured due to the small wavelengths.

Directional links will take advantage of the beam-forming gain in order to improve the link budget and provide an acceptable communication quality. However, it additionally requires a fine alignment of the beams among the BS and UEs. The process of seeking the best beam for either BS or UEs is known as beam management . we provide an example where the BS transmits some synchronization sequences using some beams with predefined directions and time slots. Each UE measures the received energy of each transmitted beam and they feed their best beam back to the BS. Then, the same process is performed again exchanging the roles among BS and UEs in order to seek the best beams for each UE. Once discovered the initial beams for each element of the network, a radio link can be established. However, these beams only correspond to a coarse estimation of the relative location among the BS and UEs. Later, the channel estimation process is not only needed in order to perform the precoding/postcoding equalization, but also it is required to improve the direction of the beams. Furthermore, if any UE is moving, a continuous location tracking is crucial in the overall performance of the link.

6.3 CSI

The channel estimation procedure allows us to know how is the propagation environment of our transmitted signal, and then, compensate its effects by equalization,



which is crucial for achieving reliable communication with high data rates. In order to obtain the channel state information (CSI), the most used method is based on pilot symbol assisted modulation (PSAM), where the transmitter exclusively sends a known preamble or pilot-sequences, and the receiver can obtain the CSI through some minimization criterion, such as least squares (LS) or MMSE. These methods have some significant advantages, namely small error and low-complexity. However, the transmission of pilot sequences reduces the overall efficiency of the system.

6.4 Time-Varying nature of channel

The multipath effect is a phenomenon that causes multiple versions of the transmitted signal to arrive at the receiver at different time delays. Reflecting objects and scatterers in the transmission environment generate multiple versions of the transmitted signal. Each of the paths will have different characteristics, such as amplitude, phase, arrival time, and angle of arrival. The multiple signals may constructively or destructively add up at the receiver, thus creating the rapid fluctuations in the received signal envelope. The time variations appear to be unpredictable to the user of the channel. Therefore it is reasonable to characterize the time-variant multipath channel statistically.

6.5 Doppler Shift

Due to the relative motion between the transmitter and the receiver, each multipath wave is subjected to a shift in frequency. The frequency shift of the received signal caused by the relative motion is called the Doppler shift. It is proportional to the speed of the mobile unit. It is given by:

$f_d = v * f_c \cos a / c$

Where, fc = transmitted frequency
v=velocity of vehicle
a= incident angle
c=speed of light

The Doppler shift in a multipath propagation environment spreads the bandwidth of the multipath waves within the range of

$$f_c \pm f_d max$$

where $f_d max$ is the maximum doppler shift, given by

$$f_d max = v\frac{f_c}{c}$$

6.6 Frequency selective fading

If the transmitted signal bandwidth is greater than the channel coherence bandwidth, the spectral components of the transmitted signal with a frequency separation larger than the coherence bandwidth are faded independently. This phenomenon is known as frequency selective fading. In wide band systems, the transmitted signals usually undergo frequency selective fading.

6.7 Rayleigh fading distribution

The Rayleigh fading describes the statistical time varying nature of received envelope of a flat fading signal, or the envelope of an individual multipath component. The Rayleigh distribution has a probability density function (pdf) given by

$$p(r) = \begin{cases} r/\sigma^2 \exp\left(-\frac{r^2}{2\sigma^2}\right) (0 \leq r \leq \infty) \\ (0 \leq r \leq \infty) \end{cases}$$

Where σ is the RMS value of the received signal before envelope detection.

6.8 Ricean fading distribution

When a dominant stationary signal component is present among the multipath components, the fading envelope is Ricean. The random multipath components are superimposed on a dominant signal such as Line-of-sight path. At the output, the effect is of adding a dc component to random multipath. The Ricean distribution degenerates to Rayleigh distribution when the dominant component fades away.

$$p(r) = \begin{cases} r/\sigma^2 \exp\left(-\frac{r^2+A^2}{2\sigma^2}\right) I\left(\left(\frac{AR}{\sigma^2}\right)(0 \leq r \leq \infty)\right. \\ 0(r < 0) \end{cases}$$

The parameter A denotes the peak amplitude of the dominant signal. The I (.) is the modified Bessel's function of first kind and zero-order. The parameter K is defined as the ratio between deterministic signal power and the variance of multipath. It is given by $K = \frac{A^2}{2\sigma^2}$

Chapter 7

Kalman filter

The Kalman filter is essentially a recursive Wiener filter which uses a state space model and a state vector (as a prior) to account for all prior knowledge of the channel statistic and previous states. These are combined with new measurement information through a Baysian estimation . The Kalman filter has an advantage to the Wiener filter that, due to the use of the state space model, the Kalman filter can account for past channel measurements at lower complexity than that of the Wiener filter. It can also handle known time varying properties of linear dynamic models and signal statistics. The Kalman filter was developed by Rudolf Kalman . It takes a Bayesian approach to estimation, utilizing a priori information of the channel in the form of a state space model of the channel's dynamics and a prior estimate of the state vector. Whenever a measurement of the channel is available then an a posteriori channel estimate can be computed.

Massive multiple-input multiple-output (MIMO) in frequency division duplex (FDD) systems , suboptimal Kalman filter is suggested to estimate channels based on nonorthogonal pilots. By introducing a fixed grid of beams, the system generates sparsity in the channel vectors seen by each user, which then estimates its most relevant channels based on unique pilot codes for each beam. Downlink time division duplex (TDD) channels are estimated based on uplink pilots. By using a predictor antenna, which scouts the channel in advance, the desired downlink channel can be estimated using pilot-based estimates of the channels before and after it (in space). Results indicate that, with the help of Kalman smoothing, predictor antennas can enable accurate CSI for TDD downlinks at vehicular velocities of 80 km/h.

7.1 Mathematical presentation of the filter

The Kalman filter requires a state space model of the signal that is to be estimated, which in this case consists of narrow band channels (subcarriers) represented by complex numbers. The state space model represents the assumed statistics of the small scale fading of each subcarriers. For this section we shall assume that these models are known, and then in next modelling Sections , will go into the issues with estimating models. Let us assume that a vector of N complex-valued zero mean narrow band channels $\mathbf{h}_\tau \in \mathbb{C}^{N\times 1}$ can be described by a discrete time wide sense stationary AR state space model

$$\begin{aligned} \mathbf{x}_{\tau+1} &= \mathbf{A}\mathbf{x}_\tau + \mathbf{B}\mathbf{w}_\tau, \\ \mathbf{h}_\tau &= \mathbf{C}\mathbf{x}_\tau, \\ \mathbf{Q} &= E\left[\mathbf{w}_\tau \mathbf{w}_\tau^*\right], \\ \Pi &= E\left[\mathbf{x}_\tau \mathbf{x}_\tau^*\right], \end{aligned} \tag{7.1}$$

which describes its evolution over time. Here, $\mathbf{w}_\tau \in \mathbb{C}^{N\times 1}$ is a zero mean vector of white Gaussian noise with time independent autocorrelation $\mathbf{Q}$, denoted process noise, and $\mathbf{x}_\tau \in \mathbb{C}^{Nn_{AR}\times 1}$ is the state vector which is assumed to be independent of $\mathbf{w}_\tau$. The matrices $\mathbf{A} \in \mathbb{C}^{Nn_{AR}\times Nn_{AR}}, \mathbf{B} \in \mathbb{C}^{Nn_{AR}\times N}$ and $\mathbf{C} \in \mathbb{C}^{N\times Nn_{AR}}$ are time independent state space matrices and n_{AR} is the model order of an AR model that is used to represent the temporal correlation of a single channel component.The integer time index τ represents the OFDM-symbols when pilots may be sent. In a situation where $\mathbf{x}_\tau$ is perfectly known, the extrapolation of the state vector one step into the future, $\mathbf{x}_{\tau+1}$, would consist of two terms: The term $\mathbf{A}\mathbf{x}_\tau$, which would be known, and the term $\mathbf{B}\mathbf{w}_\tau$, for which only the second order moments, represented by $\mathbf{Q}$, would be known. This last term represents what is new and unknown at time $\tau + 1$ in the state vector $\mathbf{x}_{\tau+1}$.

As the process noise is assumed to be zero mean Gaussian, the Maximum Likelihood (ML) estimate of the second term would be an all zero vector, and the one step prediction error of $\mathbf{h}_{\tau+1}$ would become $\mathbf{C}\mathbf{B}\mathbf{w}_\tau$. As was shown in [15, 16], the one step prediction error for band limited signals can be forced to be zero. The next state vector must then be perfectly described by a number of past state vector and so $\mathbf{Q} = 0$ in such a state space model.

Furthermore, let us assume that a $\mathbf{y}_\tau \in \mathbb{C}^{K\times 1}$ represents a vector of pilot measurements at time τ at K transmission resources where only known pilots were transmitted. The entries of $\mathbf{y}_\tau$ may represent measurements at different pilot bearing time-frequency resources and/or at different receive antennas. The measurement process is modeled by

$$\mathbf{y}_\tau = \Phi_\tau \mathbf{h}_\tau + \mathbf{n}_\tau \tag{7.2}$$

where $\Phi_\tau \in \mathbb{C}^{K\times N}$ is a pilot matrix and $\mathbf{n}_\tau \in \mathbb{C}^{K\times 1}$ represents the sum of noise and interference, e.g. from base stations not considered in the cluster, which is assumed zero mean with covariance matrix $\mathbf{R} = E\left[\mathbf{n}_\tau \mathbf{n}_\tau^*\right]$ of full rank. It will, unless otherwise specified, be denoted measurement noise.

It is worth noting that the estimations that are evaluated here are solely based on pilot measurements of the channel. These estimates could be further improved upon by using the transmitted data symbols as well. In such a case, the pilot based channel estimate will first be used to determine which symbol has been sent, e.g. through ML detection. As only a discrete number of symbols can be sent (the number is determined by the modulation format), the ML estimated symbol can then be used as a pilot to re-estimate the channel. This process can be repeated iteratively if desired.

The aim is now to produce an MSE optimal estimate $\hat{\mathbf{x}}_\tau$ of the state vector $\mathbf{x}_\tau$ (and thereby of $\mathbf{h}_\tau = \mathbf{C}\mathbf{x}_\tau$), based on the measurement $\mathbf{y}_\tau$ and on all other available relevant information.

The state space model assumed above in equation 1 represents the a priori information, along with an estimate of the state space vector $\hat{\mathbf{x}}_{\tau-1}$ and an error covariance matrix of this estimate

$$\mathbf{P}_{\tau-1} = E\left[\hat{\mathbf{x}}_{\tau-1}\overline{\mathbf{x}}_{\tau-1}^*\right] \tag{7.3}$$

where the estimation error $\overline{\mathbf{x}}_{\tau-1} = \mathbf{x}_{\tau-1} - \hat{\mathbf{x}}_{\tau-1}$ is uncorre lated to the estimate. As the state space vector is a sum of weighted independent complex-valued white Gaussian zero mean vectors, through (1), it is in itself a zero mean complex-valued white Gaussian vector. Therefore, in the case that no other estimate is available, the best a priori guess of the state space vector would be its mean value, i.e. an all zero vector and $P_{\tau-1} = \Pi$. In fact, this is a way to initiate the filter. Other options for initiating the Kalman filter are discussed in Paper I.

Based on the a priori information and the measurement (2), an a posteriori MSE estimate is given by the recursive set of matrix difference equations, know as the Kalman equations

$$\hat{\mathbf{x}}_\tau = \mathbf{A}\hat{\mathbf{x}}_{\tau-1} + \mathbf{K}_\tau\left(\mathbf{y}_\tau - \mathbf{J}_\tau \mathbf{A}\hat{\mathbf{x}}_{\tau-1}\right) \tag{7.4}$$

$$\mathbf{P}_\tau = (\mathbf{I} - \mathbf{K}_\tau \mathbf{J}_\tau)\,(\mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*) \tag{7.5}$$

$$\mathbf{K}_\tau = (\mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\,\mathbf{J}_\tau^*\,(\mathbf{R} + \mathbf{J}_\tau\,(\mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\,\mathbf{J}_\tau^*)^{-1} \tag{7.6}$$

$$\hat{\mathbf{h}}_\tau = \mathbf{C}\hat{\mathbf{x}}_\tau \tag{7.7}$$

where $\mathbf{J}_\tau = \Phi_\tau \mathbf{C}$ and * denotes the conjugate transpose. The matrix $\mathbf{K}_\tau \in Nn_{AR} \times K$ is known as the Kalman gain and through (3.4) it adjusts the a prior estimate, which is given in form of the one step prediction based on the previous state vector estimate $\mathbf{A}\hat{\mathbf{x}}_{\tau-1}$, by weighting the error of the one step prediction of the measurement, that is produced by using $\hat{\mathbf{h}}_\tau = \mathbf{C}\mathbf{A}\hat{\mathbf{x}}_{\tau-1}$ which is given by $\mathbf{y}_\tau - \Phi_\tau \hat{\mathbf{h}}_\tau = \mathbf{y}_\tau - \mathbf{J}_\tau \mathbf{A}\hat{\mathbf{x}}_{\tau-1}$.

The focus here will be on the information that we gain from equations (4)(7) For more information on Bayesian inference have to be studied.

The Kalman filter is optimal in the se nse that it we ights all available information in order to get the best estimate (with respect to the MSE). Therefore, the conclusions that can be drawn from equations (4)-(7) are in accordance with what most people intuitively would expect.

If the previous estimate of the state vector is very inaccurate (compared to the the power of the measurement noise and the process noise), then $\mathrm{P}_{\mathrm{f}-1}$ will be large and, by (4) and (6), the estimate of the state vector will be based almost purely on the measurement. If the opposite is true and the state vector estimate is accurate, then the Kalman gain will depend on the ratio between the powers of the measurement noise $\mathbf{n}_\tau$ and the process noise $\mathbf{w}_\tau$.

When the process noise is large compared to the measurement noise, then the a posteriori estimate will be based on the measurement to a larger extent and if the measurement noise is large compared to the process noise, the a posteriori estimate will be primarily based on the a priori information. Both a large measurement noise vector and a large process noise vector will increase the error of the a posterior estimate, through (5)-(6). Likewise, a large error in the prior estimate will in general cause a large error in the posterior estimate.

7.2 Predictions and smoothing

Some hypothesis is needed for how channel components that are not directly related to timefrequency resources with pilots are related to the measurements y_τ. In a Kalman filter, the dynamic state space model (1) fulfills that purpose. Wherever

no direct measurements are available, model based ex trapolation or interpolation is used.

When there is no available pilot measurement at time step τ, e.g. if the channel that is to be estimated is in the future or if no uplink pilots are transmitted during e.g. a downlink frame in a TDD system, this can be described by setting the pilot matrix Φ_τ to an all zero matrix. The n equation (2) gives $\mathbf{y}_\tau = \mathbf{n}_\tau$ and since $\mathbf{J}_\tau = \Phi_\tau \mathbf{C} = 0$, the Kalman gain will by (6) be an all zero matrix. Then the estimated state vector obtained only from an extrapolation of $\hat{\mathbf{x}}_{\tau-1}$ and the corresponding error covariance matrix become

$$\begin{aligned}\hat{\mathbf{x}}_{\tau|\tau-1} &= \mathbf{A}\hat{\mathbf{x}}_{\tau-1} \\ \mathbf{P}_{\tau|\tau-1} &= \mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*\end{aligned} \tag{7.8}$$

Here, the notation $\tau_1 \mid \tau_2$ is introduced to denote an estimate of the state vector at time τ_1 given measurements up until time τ_2.

The one step prediction in (8) is simple to extend to a multistep prediction by continuing to assume that the pilot matrix is an all zero matrix. We then obtain

$$\begin{aligned}\hat{\mathbf{x}}_{\tau|\tau-m} &= \mathbf{A}^m \hat{\mathbf{x}}_{\tau-m}, \\ \mathbf{P}_{\tau|\tau-m} &= \mathbf{A}^m \mathbf{P}_{\tau-m} \left(\mathbf{A}^*\right)^m + \sum_{i=0}^{m-1} \mathbf{A}^i \mathbf{B}\mathbf{Q}\mathbf{B}^* \left(\mathbf{A}^*\right)^i\end{aligned} \tag{7.9}$$

The corresponding channel prediction estimate and prediction error covariance matrix of the channel estimate are given by

$$_{\tau|\tau-m} = \mathbf{C}\hat{x}_{\tau|\tau-m!} \tag{7.10}$$and

$$\Gamma_{\tau|\tau-m} = E\left[\left(\mathbf{h}_\tau - \mathbf{h}_{\tau|\tau-m}\right)\left(\mathbf{h}_\tau - \mathbf{h}_{\tau|\tau-m}\right)^*\right] = \mathbf{C}\mathbf{P}_{\tau|\tau-m}\mathbf{C}^4 \tag{7.11}$$respectively. A great advantage to the Kalman filter, as compared to other linear filters is that the error covariance matrix of the state vector is part of the "package deal", i.e. it is calculated as part of the Kalman equations (4)-(7). This information can be used for example in a multi-antenna transmit precoding stage to ensure that poor channel estimates have less impact on the final solution than accurate channel estimates.

Prediction is required when you need to estimate the channel before you have access to measurements of it. It will in general result in a prediction that is worse than the filter estimate (7). If one on the other hand has the opportunity to wait with estimating the channel until some more pilot measurements are available, then a smoothing estimate $\hat{\mathbf{h}}_{\tau|\tau+m}$ for $m > 0$ can be obtained. One use of smoothing is to improve upon the accuracy of a filter estimate. A nother use, that will be the focus of Paper V, is to obtain channel estimates at time steps τ at which no pilots are available, by using both past and future measurements.

There are two standard ways to perform channel smoothing. One option is by extending the state vector and including all future states up until the point when there are no more measurements available.

A second option is to use two filters.The basic idea is that the first filter calculates the filter or prediction estimate of the channel based on all available measurements up until the time τ of the estimate by $(4) - (7)$. Then a second filter uses a time reversed state model and performs a backward recursion to estimate the state vector of the time reversed system at time $\tau + 1$ based on all future available pilot measurements. This is then extrapolated through the time inverted state space model into a one step backwards prediction, similar to the forward prediction (8). Through this, we obtain a channel estimate based on the future measurements up until a given time T which we denote $\hat{\mathbf{h}}_\tau \mid \tau + 1, \ldots, \tau + T$. Assuming that this estimate has a covariance matrix $\Gamma_{\tau|\tau+1,\ldots,\tau+T}$, the combined smoothed channel estimate is given by an MSE optimal weighting of the backward and forward estimates

$$_{\tau|\tau+T} = \Gamma_{\tau|\tau+T}\left(\Gamma^{-1}_{\tau|\tau}\hat{\mathbf{h}}_\tau + \Gamma^{-1}_{\tau|\tau+1,\ldots,\tau+T}\hat{\mathbf{h}}_{\tau|\tau+1,\ldots,\tau+T}\right), (7.12) \text{with}$$

$$\Gamma_{\tau|\tau+T} = \left(\Gamma^{-1}_{\tau|\tau} + \Gamma^{-1}_{\tau|\tau+1,\ldots,\tau+T} - \Pi^{-1}\right)^{-1} (7.13) (\text{The term } \Pi^{-1}, \text{ which represents prior information of the channel is included in both } \Gamma^{-1}_{\tau|\tau} \text{ and } \Gamma^{-1}_{\tau|\tau+1,\ldots,\tau+T}, \text{ and hence one of these must be removed in (13).})$$

1 Comments on quality and the use of stationary filters

The complexity of the Kalman filter has been investigated , assuming that the state model (1) is set up on diagonal form. From this work, we see that the highest complexity is related to calculating the Kalman gain (6) and the error covariance matrix (5). Fortunately, an important property of the Kalman filter is that for any stable time invariant system (1), with time invariant Φ_τ in the measurement equation (2) the Kalman filter converges such that $\mathbf{P}_{\tau-1} \to \mathbf{P}_\tau$, when $\tau \to \infty$.

Moreover, it will converge regardless of whether the initial error covariance matrix is known or not, i.e. even if P_0 is not an accurate representation of the second order statistics of the estimation error of $\hat{h}_0$ the filter will converge, although at a slower rate [10]. Therefore a large part of the on line complexity can be reduced by calculating both the error covariance filter and through that also the Kalman gain off-line. The error covariance matrix of the stationary filter $\mathbf{P}_f$ can be found by setting $\mathbf{P}_f = \mathbf{P}_{\tau-1} = \mathbf{P}_\tau$ in (5), which gives a discrete time algebraic Riccati equation

$$\mathbf{P}_f = (\mathbf{I} - \mathbf{K}_f \mathbf{J}_\tau)(\mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)$$

(7.14)

$$\mathbf{K}_f = (\mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\mathbf{J}_\tau^*(\mathbf{R} + \mathbf{J}_\tau(\mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\mathbf{J}_\tau^*)^{-1} \tag{7.15}$$

Figures 1 and 2 illustrate the convergence of Kalman filters for two different types of AR models. The se depict the differe nce between the covariance matrix of the one step prediction error when the filter error covariance matrix is calculated through the filter recursions $(4) - (6)$ and when it is found by the stationary solution to the Riccati equation, where P_p denotes the error covariance matrix of the one step prediction of the stationary filter given by

$$\mathbf{P}_p = \mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*.$$

We can see that the filter converges fairly quickly, especially with a Doppler spectrum that is relatively flat. As the AR model has to be based on training data, that same training data set can be used to ensure that the filter converges off-line before it is to be used.

In figure 1,the maximum norm of$P|1Pp$ normalized by the max norm of Pp, where Pp is the one step prediction error that obeys the algebraic Riccati equation. Results are shown for different SNR of the pilot measurements by (2). The system in (1) is a fourth order AR model for a single channel tap h with the poles in 0.82± 0.29i and 0.70±0.10i, which gives an almost flat Doppler spectrum.

In figure 2,the maximum norm of $P|1Pp$ normalized by the max norm of Pp. Results are shown for different SNR of the pilot measurements by (2). The system in (1) is a fourth order AR model for a single channel tap h with the poles in 0.91± 0.35i and 0.86±0.33i, which gives a Doppler spectra similar to the Jakes spectrum but not band limited.

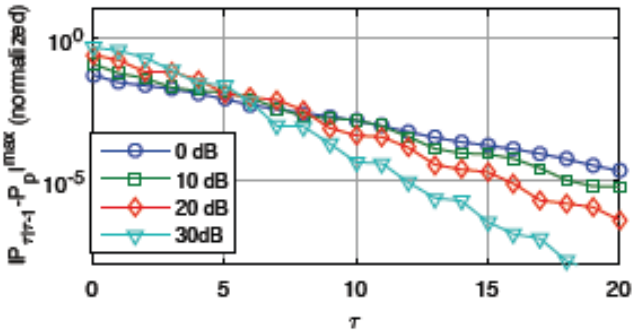


Figure 7.1:

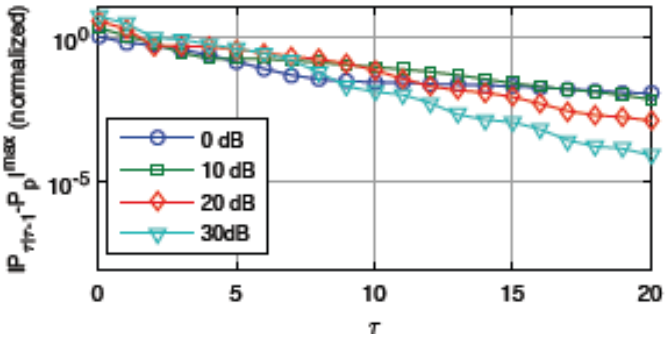


Figure 7.2:

7.3 Estimation

The model (1) must be estimated before the Kalman filter can be used. The work here is based on AR modeling based on the Yule Walker equations,as this was shown to give the best estimation performance . In general the accuracy of the model will improve the more information that is available. If the channel is approximately wide sense stationary the n the more past channel data that is included in the training data, the better the AR model will be. For stationary (non-moving) users, this assumption is in general valid over a long time window. Channels to stationary or very slow moving users are also fairly easy to estimate (and predict) with very good accuracy.

The need for more advanced estimators arises primarily when users are mobile. For these users, the small scale fading can only be considered a wide sense stationary process while the shadow fading statistics (the numbers and powers of the contributing multipaths) remains relatively constant. Through the measurements used in Papers I and II it could be observed that for pedestrian users and at a carrier frequency of 2.66GHz, over a time of approximately on second, the assumption of a wide sense stationary system is relatively sound.

This section provides a brief overview of how to estimate the parameters of the state space model in (1).

7.4 Estimation of the parameters of the state space matrices of one channel element

The AR model that represents the small scale fading is estimated based on training data. What is considered training data may vary. When a new user enters a system, then the channels can be estimated by e.g. a ML estimate. These can then be used as training data to find the autocorre lation function of each individual component in the channel vector, $h_{i,\tau}$

$$R_h(t) = E\left[h_{i,\tau} h^*_{i,\tau-t}\right] \tag{7.16}$$

Here, $h_{i,\tau}$ is an individual element of the channel vector $\mathbf{h}_\tau$. Alternatively, all available pilot and data measurements can be demodulated and used to estimate the channels. The sequence of estimated channels can then be further smoothed and used as training data to find an AR model that in turn is used to filter the original data. The relation between an AR model of one channel tap $h_{i,\tau}$ of order n_{AR}

$$h_{i,\tau} = -a_1 h_{i,\tau-1} - \ldots - a_{n_{AR}} h_{i,\tau-n_{AR}} + v_\tau \tag{7.17}$$

and the autocorrelation function (3.16) is given by the Yule-Walker equations

$$-Ya = z, \tag{7.18}$$

where $\mathbf{a}$ is a vector of the AR coefficients $\mathbf{a}^T = [a_1, \ldots, a_{n_{AR}}]$ and

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} R_h(0) & R_h^*(1) & \ldots & R_h^*(n_{AR}-1) \\ R_h(1) & R_h(0) & \ldots & R_h^*(n_{AR}-2) \\ \vdots & \vdots & \ddots & \ldots \\ R_h(n_{AR}-1) & R_h(n_{AR}-2) & \ldots & R_h(0) \end{bmatrix} \\ \mathbf{z}^T &= \begin{bmatrix} R_h(1) & R_h(2) & \vdots & R_h(n_{AR}) \end{bmatrix} \end{aligned} \tag{7.19}$$

By inverting the matrix on the left hand side, the problem of estimating a can be solved, so that the variance of the one step prediction error, which equals the driving noise term v_τ in (17), is minimized. The poles of the AR process (17), i.e. the roots of the polynomial

$$z^{n_{AR}} + a_1 z^{n_{AR}-1} + \ldots + a_{n_{AR}}$$

represent peaks in the Doppler spectrum of the estimated model. A pole close to the unit circle represe nts a very narrow-band peak, which could for example be a strong LOS component, whereas a pole further inside the unit circle will provide a wider peak. For a strictly band limited signal, all poles will be placed on the unit circle.

The small scale fading could also be modelled by an Autoregressive Moving Average (ARMA) process. Then the zeros of such a model could be used to suppress some parts of the Doppler spectrum. However, while the AR parameters can be found by a closed form solution of the linear problem (18), estimating the parameters of an ARMA model based on training data of the past channel realizations pose a non linear estimation problem, which may be solved by an iterative optimization algorithm. ML estimation is the most commonly used framework .

Bibliography

[1] S. Yang and L. Hanzo, "Fifty years of mimo detection: The road to large-scale mimos," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.

[2] A. Puglielli, A. Townley, G. LaCaille, V. Milovanović, P. Lu, K. Trotskovsky, A. Whitcombe, N. Narevsky, G. Wright, T. Courtade, *et al.*, "Design of energy- and cost-efficient massive mimo arrays," *Proceedings of the IEEE*, vol. 104, no. 3, pp. 586–606, 2015.

[3] P. Zhang, J. Chen, X. Yang, N. Ma, and Z. Zhang, "Recent research on massive mimo propagation channels: A survey," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 22–29, 2018.

[4] T. L. Marzetta and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[5] J. Zhang, Y. Zhang, Y. Yu, R. Xu, Q. Zheng, and P. Zhang, "3-d mimo: How much does it meet our expectations observed from channel measurements?," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1887–1903, 2017.

[6] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive mimo: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[7] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE communications magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[8] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive mimo systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[9] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, "Channel estimation for orthogonal time frequency space (otfs) massive mimo," *IEEE Transactions on Signal Processing*, vol. 67, no. 16, pp. 4204–4217, 2019.

[10] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive mimo has unlimited capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, 2017.

[11] M. H. Alsharif, A. H. Kelechi, M. A. Albreem, S. A. Chaudhry, M. S. Zia, and S. Kim, "Sixth generation (6g) wireless networks: Vision, research activities, challenges and potential solutions," *Symmetry*, vol. 12, no. 4, p. 676, 2020.