CHALMERS
UNIVERSITY OF TECHNOLOGY

Joint communication and sensing in massive MIMO architectures with low-precision converters

Master's thesis in Communication Engineering

YIXIN HU

Department of Electrical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis

Joint communication and sensing in massive MIMO architectures with low-precision converters

YIXIN HU

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Division of Communication and Antenna Systems
Chalmers University of Technology
Gothenburg, Sweden 2020

Joint communication and sensing in massive MIMO architectures
with low-precision converters
YIXIN HU

Supervisor: Giuseppe Durisi, Chalmers University of Technology
Examiner: Giuseppe Durisi

Master's Thesis 2020
Department of Electrical Engineering
Division of Communication and Antenna Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 73 630 8331

Typeset in LaTeX
Gothenburg, Sweden 2020

Joint communication and sensing in massive MIMO architectures
with low-precision converters
YIXIN HU
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Ultra-reliable low latency communication (URLLC) is defined by the 3rd Generation Partnership Project (3GPP) as one of the three application scenarios in 5G. Together with massive multi-user multiple-input multiple-output (MU-MIMO) technology, the needs of real-time communication for applications such as autonomous driving and remote surgery are expected to be achieved. However, a large antenna array at the base station (BS) in a massive MIMO system leads to high energy consumption. To meet the stringent latency requirement as well as the power constraint, we investigate in this thesis the transmission of short packets with low-resolution analog-to-digital converters (ADCs) equipped at the BS antennas, where perfect channel state information (CSI) can not be assumed and the cost of pilot symbols, which are assigned to estimate the channel, needs to be considered.

In this thesis, we use methods from finite-blocklength information theory to analyse the performance of short packet transmission using low-resolution ADCs under an MU-MIMO uplink scenario. First, we investigate the quality of channel estimation under the assignment of different number of pilot symbols and transmit power. The results indicate that for each pilot length there exists an optimal power that results in the lowest estimation error. Next, we take into account the reliability of data transmission by considering the average power required to achieve a target packet error probability. We investigate the impact of low-resolution ADCs and find out that using 3-bit converters we are able to approach the performance obtained with infinite-resolution converters. We also evaluate the effect of power allocation. Specifically, we assign different power to pilot symbols and data symbols and show that performance improvements can be obtained when assigning more power to pilot symbols. Furthermore, we characterize the trade-off between the accuracy of channel estimation and the reliability of data transmission.

Keywords: massive MU-MIMO, URLLC, low-resolution ADCs, Bussgang decomposition, RCUs.

Acknowledgements

This thesis has been conducted at the Electrical Engineering Department of Chalmers University of Technology.

Foremost, I would like to express my most sincere appreciation to my supervisor and examiner Prof. Giuseppe Durisi, for proposing this interesting topic as well as offering patient guidance and encouragement all the time.

I would also like to thank Alejandro Lancho Serrano for helping me understand the theories and providing guidance in simulations, and Han Wu for providing learning resources related to my thesis topic. Assistance provided by Anzhong Hu through discussion on the simulation results was greatly appreciated. I am also grateful to Roman Sokolovskii who helped me get access to C3SE, which saved lots of simulation time.

Morever, I would like to express my gratitude to Yuning Li and Nahid Moheq who I studied with and gave me valuable ideas. Their company in the special period was a great support for completion of this report.

Yixin Hu, Gothenburg, September 2020

Contents

List of Figures

List of Tables

Acronyms

ADCs Analog-to-digital Converters.
AGC Automatic Gain Control.

BLMMSE Bussgang Linear Minimum Mean Square Error.
BS Base Station.

CSI Channel State Information.

MRC Maximal Ratio Combining.
MU-MIMO Multi-user Multiple-input Multiple-output.

PAT Pilot-assisted Transmission.

RCUs Random Coding Union Bound with parameter s.

UE User Equipment.
URLLC Ultra-reliable Low Latency Communication.

1

Introduction

1.1 Massive MU-MIMO

1.1.1 Background

In recent years, wireless technology and multimedia data services have been widely introduced in our lives and industries. Multiple-input multiple-output (MIMO) technology is generally considered a major breakthrough in the field of wireless communication, because it improves the capacity and spectrum utilization without additional bandwidth. As the name suggests, multiple antennas at both the transmitter and the receiver are used for signal transmission, compared to the single-input single-output (SISO) system where the transmitter operates with only one antenna as the receiver does.

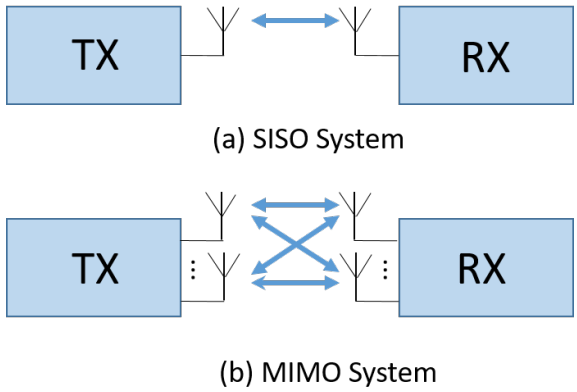

Figure 1.1: SISO and MIMO systems

The concept of MIMO was first used to refer to multiple input and output ports in electric circuits in the 50s of last century [2]. This term was later adopted by researchers from communication engineering, and the wireless channel became the reference point, with *multiple input* referring to the input signals to the channel, transmitted by multiple transmit antennas, and *multiple output* referring to the output signals from the channel, received by multiple receive antennas [3]. After MIMO technology in cellular system first being introduced in 2001 by Iospan Wireless Inc., and first being included in the WiMAX standard [4] in 2005, MIMO has been widely developed in wireless and mobile networks.

5G, which is known as the fifth-generation standard for cellular networks, is the

highest peak in the development of communication technology so far. Undoubtedly, the continuous evolution of 5G will greatly promote the improvement of technologies and applications such as the Internet of Things (IoT), smart factories, and autonomous driving.

As a key technology in 5G communication systems, massive multi-user multiple-input multiple-output (MU-MIMO) or large-scale MIMO is considered a promising method to further improve the spectral efficiency and energy efficiency. The idea of MU-MIMO is to increase the number of antennas at the BS dramatically as well as allowing several users to be served by the same BS at the same time, as shown in **Figure 1.2**. Large antenna arrays enable spatial processing which helps to eliminate the effects of interference and small-scale fading.

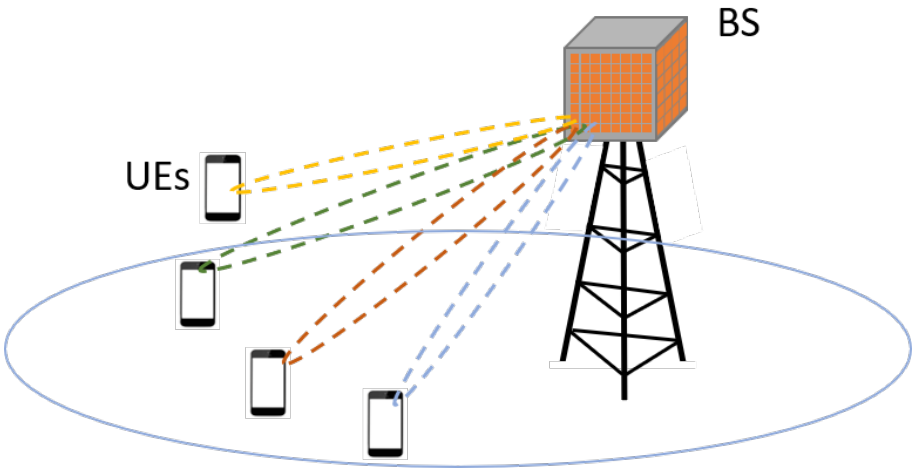

Figure 1.2: A single-cell massive MU-MIMO system where a base station with a large antenna array serves several users simultaneously (Each antenna is equipped with converters to transform signals from continuous time to discrete time).

1.1.2 Review of literatures

Massive MU-MIMO has become a research hotspot in recent years. Several approaches to investigating its potential and to apply it to practical systems have been studied.

One of the most important issues related to large antenna array is the increasing demand for power and hardware cost. Considering the MU-MIMO uplink scenario, a pair of analog-to-digital converter (ADC) hardware units is equipped at each antenna to generate the real and imaginary part of a received symbol. These ADC units are the dominant source of power consumption [5]. The consumed power grows exponentially with the ADC resolution, so a simple solution to achieve a reasonable power consumption is lowering the resolution of the converters [6]. However, using low resolution ADCs complicates the problems on channel estimation and symbol decoding. On the one hand, using low-resolution converters we can no longer assume perfect channel state information (CSI) at the BS, which can be acquired

using infinite-precision converters by observing the channel long enough. On the other hand, the quantization distortion is introduced inevitably, which makes channel estimation and data detection difficult.

A simple model was widely used where the quantization distortion is treated as additive white noise [7, 8], which is inaccurate because quantization is a nonlinear process where the noise is highly correlated with the input. In contrast, the decomposition method introduced by Bussgang [9] provides an exact linearization method, which can be performed on nonlinear devices with Gaussian inputs. It takes into account the spatial correlation by formulating the nonlinear quantization process as an equivalent linear function of the unquantized signal. Together with the minimum mean squared error (MMSE) channel estimation method, the so called Bussgang linear minimum mean squared error (BLMMSE) estimator is derived in [10] and proved to be accurate for 1-bit quantizers where the correlation matrix of the distortion can be calculated in closed form using the arcsine law [11].

In recent studies, the BLMMSE estimator is also generalized to analyse the performances of multi-bit quantizers. In [12] the correlation of distortion is approximated using a diagonal matrix by ignoring the temporal correlation, which is accurate when we transmit at low SNR or when the number of UEs is large, and even exact when the number of pilot symbols equals the number of UEs. Later in [13], the authors proposed two approximations for the correlation matrix of distortion. The *rounding approximation* is obtained by ignoring the overload distortion, which is accurate regardless of resolution. The *diagonal approximation* is based on the assumption that the off-diagonal elements are zero, which is accurate when the resolution is higher than 2 bits. In [13] the diagonal approximation is performed to evaluate the performance of BLMMSE estimator. Furthermore, it is shown that in the presence of a fronthaul constraint, the case of large antenna array lowers the effect of distortion ,i.e., more BS antennas with low-resolution ADCs perform better than fewer antennas with high-resolution ADCs.

1.2 URLLC

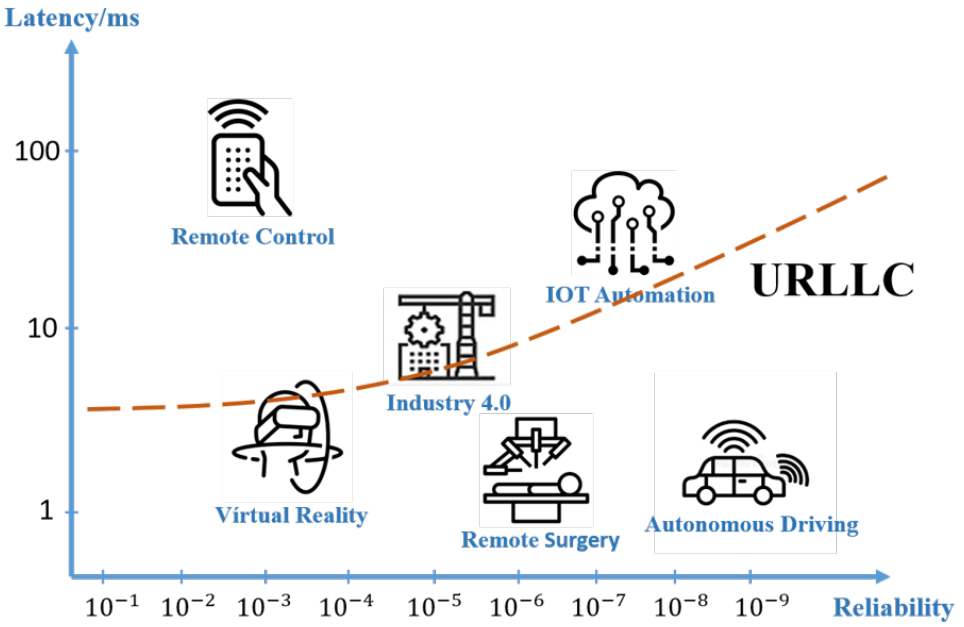

Figure 1.3: Reliability and latency requirements of URLLC applications[1].

5G puts forward higher requirements for data transmission compared to 4G, which includes not only speed but also improvements such as lower power consumption and delay.

As defined by the 3rd Generation Partnership Project (3GPP)[14], there are three main scenarios in 5G: enhanced mobile broad band (eMBB), massive machine type communication (mMTC), and ultra-reliable low latency communication (URLLC), which are designed to satisfy requirements in different applications. eMBB aims at achieveing high data rates over a long time interval, which enables virtual reality and video streaming with error rate around 10^{-3} [15]. mMTC is designed for internet of things (IOT) service where lots of sensing devices need to be connected with a fixed uplink rate. Finally, URLLC refers to services that require low latency and highly reliable connections. As **Figure 1.3** shows, the support of URLLC is needed for specific applications. In terms of the real-time requirements in remote surgery, the target delay is normally lower than 10 ms which is difficult to achieve in 4G. Another use case is the motion control and factory automation in Industry 4.0, which requires end-to-end (E2E) latency of 1 ms and 10 ms respectively. The requirement is even higher for autonomous driving, where the delay needs to be as low as 1 ms with extremely high reliability of 99.9999% [16].

The use of massive MIMO technologies in URLLC was discussed in [17] which provides guidelines in the design of URLLC systems. As illustrated in this report, three main benefits can be obtained from the degrees of freedom which is created by large antenna array:

- High SNR links because of the array gain.
- Channel hardening, where the channel becomes almost deterministic due to spatial diversity. It improves the channel estimation accuracy and therefore relaxes the need for strong coding schemes.
- The capability to perform spatial division multiplexing.

1.2.1 Literatures review related to short-packet transmission

Due to the strict requirement of low latency, short Transmission Time Interval (TTI) is considered a necessary enabler for URLLC. A vast number of researches address short-packet transmission for URLLC systems. As stated in [18], asymptotic metrics such as ergodic and outage capacity, which are normally used under the assumption of large block length in the conventional communication systems, are no longer valid for short-packet transmissions. Nonasymptotic finite-blocklength information theory appears to be more appropriate.

It is described in [19] that the average error probability over an ensemble of codebooks, which are generated from a given probability distribution, can be upper bounded using a function of the number of codewords M, and the blocklength n. In [20], the author presented a so-called random-coding union (RCU) bound with parameter s (RCUs) which provides an upper bound on the average packet error probability $\epsilon^*(M, n)$. It is proved to be a powerful tool for performance analysis and is accurate for memoryless channels. In [18], the authors studied the maximum coding rate achievable, which is denoted by $R^*(n, \epsilon)$, based on the RCUs bound. Under the assumption of no a prior CSI, the authors analyze the performance of both PAT and non-coherent transmission where no pilot symbols are transmitted to estimate the channel. The results indicate that PAT is not energy efficient in the SISO scenario. In [21], the RCUs bound was applied to a massive MU-MIMO system with the set up of $M = 2^{30}$ and $n = 288$. With the assumption of no CSI, PAT scheme was performed to learn the channel. As the result shows, under a requirement on the target error probability, there exist an optimal number of pilots that minimize the average transmit power.

1.3 Objectives and methods

Unlike the previous works which were concentrated on either the effect of low-resolution ADCs or the performance analysis of short-packet transmission, we consider in this thesis both of the factors. Motivated by the potential application prospects of URLLC under the massive MU-MIMO scenario, we study the performances of uplink short packet transmission with low-resolution ADCs at the BS.

Specifically, we use Bussgang's theorem [22] to deal with the nonlinearity introduced by the low-resolution ADCs. With the help of Bussgang's decomposition, we managed to express the output of quantizers as a linear function of the input. This decomposition method is performed on pilot symbols and data symbols separately as introduced in [10]. As for short-packets transmission, the RCUs bound is used to upper-bound the average error probability. We consider also a simple power allocation scheme where we are aiming to get either better channel estimation or lower error probability for a given transmit power. Furthermore, the channel is assumed to be unknown and estimated using pilot.

1.4 Outline

This report is structured as follows. In the next chapter, we introduce in details the MU-MIMO system and finite-blocklength information-theory methods as well as signal processing methods used in this thesis. It is followed by Chapter 3, which provides the introduction to the massive MU-MIMO uplink transmission with low-precision quantizers at BS antennas. In Chapter 4, we discuss the performance of our system based on numerical results. We highlight the improvements when we use optimum power allocation and discuss if it is worth using it in the real systems. We conclude in Chapter 5 and introduce future works in Chapter 6.

2

Theory

In this chapter, we introduce the models and the tools we use in this thesis. Regarding the MU-MIMO wireless channel, we discuss the signal transmission and processing methods such as linear combining for symbol detection. As for short packet transmission, we introduce the random coding argument and the RCUs bound in finite-blocklength information theory. We also introduce pilot-assisted wireless transmission, which is a commonly used method for estimating the channel at the receiver. We present a power allocation algorithm to find the optimal number of pilots and the power assigned to them. We discuss also the uniform quantization and the automatic gain control unit, which mantains the quantization quality. To deal with the nonlinearity introduced by the quantizers, we discussed a popular method of linearization that is based on Bussgang's theorem and its application to MIMO systems.

2.1 MU-MIMO system

2.1.1 Wireless channels

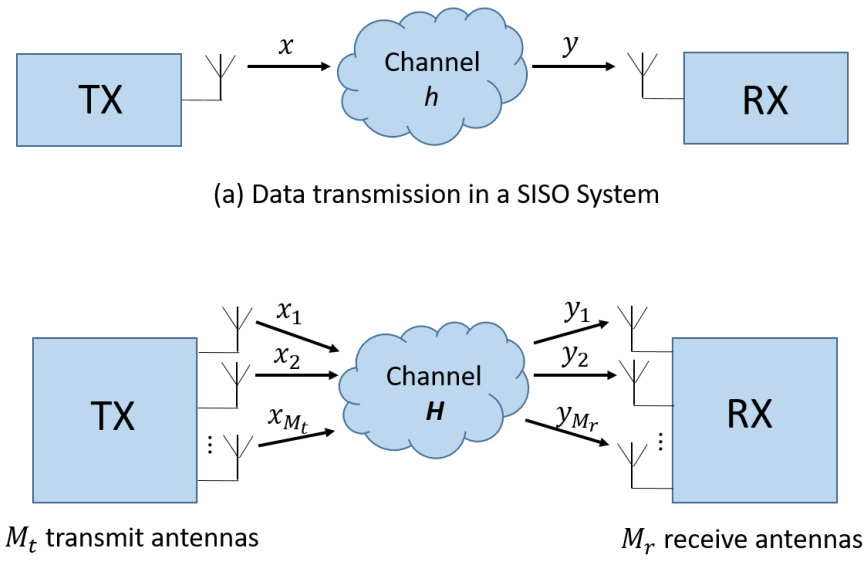

Figure 2.1: Data transmission in (a) a SISO system, (b) a MIMO system

We introduce in this section the signal transmission of different wireless communication systems based on the assumption that the data sequence is transmitted within a coherence block where the channel coefficients remain the same. **Figure 2.1** shows the data transmission in a SISO system and a MIMO system.

Let us start first with a simple SISO system. A data symbol x is transmitted through the channel denoted by coefficient h, and the antenna at the receiver end receives

$$y = hx + n, \tag{2.1}$$

where n denotes the additive noise, which in most cases Gaussian is distributed with known variance.

When we move to a MIMO system, where the transmitter is equipped with M_t antennas and the receiver is equipped with M_r antennas, the channel becomes a $M_r \times M_t$ matrix $\mathbf{H}$ with the coefficient $h_{i,j}$ denoting the channel between the ith receive antenna and the jth transmit antenna. Let x_j be the transmit symbol from the jth transmit antenna and y_i the received symbol at the ith receive antenna. The transmission can be formulated as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{M_r} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M_t} \\ h_{21} & h_{22} & \cdots & h_{2M_t} \\ \cdots & \cdots & \cdots & \cdots \\ h_{M_r1} & h_{M_r2} & \cdots & h_{M_rM_t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{M_t} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{M_r} \end{bmatrix} \tag{2.2}$$

or $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$ in vector form. The kth element of $\mathbf{y}$ is $y_k = \sum_{i=1}^{M_t} h_{ki}x_i + n_k$, which contains the signals from all transmit antennas Here, $\mathbf{n}$ is the $M_r \times 1$ noise vector with each of its element denotes the additive Gaussian noise at the corresponding antenna.

The uplink transmission of a MU-MIMO system with M antennas at the BS and U active single-antenna UEs can be modelled similarly to a standard MIMO system with U antennas transmitting independent signals. Under this setting, we have an $M \times U$ channel matrix $\mathbf{H}$ where the kth column of $\mathbf{H}$, which we denote by $\mathbf{h}_k$ is the $M \times 1$ channel vector between the kth UE and all M antennas at the BS.

2.1.2 Signal combining

Several different spatial processing methods can be implemented at the BS to obtain an estimate of the transmitted signals out of the received signals. As shown in **Figure 2.2**, we assume that the BS applies a linear filter, which is represented by the $M \times U$ matrix $\mathbf{W}$. By multiplying $\mathbf{y}$ with the combining matrix $\mathbf{W}$, the received signal $\mathbf{y}$ is separated into U streams, denoting by the combined signal $\mathbf{r}$:

$$\mathbf{r} = \mathbf{W}^H\mathbf{y} = \mathbf{W}^H\mathbf{Hx} + \mathbf{W}^H\mathbf{n}. \tag{2.3}$$

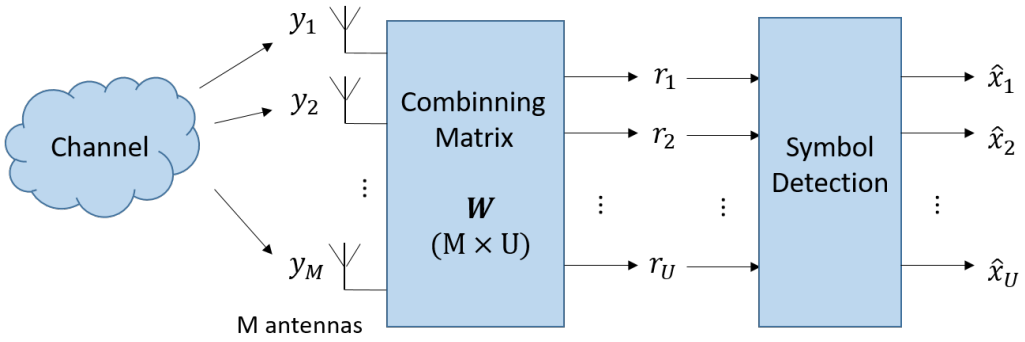

Figure 2.2: Signal combining and symbol detection at the BS

Assume that the channel $\mathbf{H}$ is known at the BS, we can apply maximum likelihood (ML) detection, where the BS searches over the set $\mathcal{X}$ that contains all the possible transmit sequences and choose the one that maximizes the likelihood function:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}\in\mathcal{X}} \mathrm{P}(\mathbf{y}|\mathbf{x}) = \arg \min_{\mathbf{x}\in\mathcal{X}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2, \tag{2.4}$$

The kth entry of $\mathbf{r}$ is

$$r_k = \mathbf{w}_k^H \mathbf{h}_k x_k + \sum_{i\neq k}^{U} \mathbf{w}_k^H \mathbf{h}_i x_i + \mathbf{w}_k^H \mathbf{n}, \tag{2.5}$$

where $\mathbf{w}_k^H$ denotes the kth column of $\mathbf{W}$. The first term in (2.5) represents the desired signal, and the other two terms represent the inter-user interference and additive noise, respectively. Hence, we can compute the signal-to-interference-plus-noise ratio(SINR) of the kth stream as

$$\mathrm{SINR}_k = \frac{\left|\mathbf{w}_k^H \mathbf{h}_k\right|^2}{\sum_{i\neq k}^{U} \left|\mathbf{w}_k^H \mathbf{h}_i\right|^2 + \|\mathbf{w}_k\|^2}. \tag{2.6}$$

Based on the knowledge of channel $\mathbf{H}$, there are two ways to find $\mathbf{W}$, which are maximal ratio combining (MRC) and zero-forcing (ZF).

2.1.2.1 Maximal ratio combining

The MRC approach aims to maximize the signal-to-noise ratio (SNR) of each signal stream in the absence of the inter-user interference, that is

$$\mathbf{w}_k = \mathrm{argmax} \frac{\left|\mathbf{w}_k^H \hat{\mathbf{h}}_k\right|^2}{\|\mathbf{w}_k\|^2}. \tag{2.7}$$

It holds that

$$\frac{\left|\mathbf{w}_k^H \hat{\mathbf{h}}_k\right|^2}{\|\mathbf{w}_k\|^2} \leq \frac{\|\mathbf{w}_k\|^2 \left\|\hat{\mathbf{h}}_k\right\|^2}{\|\mathbf{w}_k\|^2} = \left\|\hat{\mathbf{h}}_k\right\|^2, \tag{2.8}$$

thus we get the MRC receiver

$$\mathbf{w}_k = \hat{\mathbf{h}}_k, \tag{2.9}$$

which is $\mathbf{W}_{\text{MRC}} = \hat{\mathbf{H}}$ in the matrix form.

Using the MRC receiver, the received signal is just multiplied with the Hermitian transpose of the channel matrix. MRC receiver is simple and proved to be the optimum combiner for independent AWGN channels [23].

2.1.2.2 Zero-forcing receiver

ZF is an algorithm that takes into account the effect of inter-user interference but ignores the additive noise, which is the opposite of what the MRC combiner does. The ZF combiner for the kth UE satisfies:

$$\begin{cases} \mathbf{w}_k^H \hat{\mathbf{h}}_k \neq 0 \\ \mathbf{w}_k^H \hat{\mathbf{h}}_i = 0 \quad , i \neq k \end{cases} \tag{2.10}$$

To eliminate the inter-user interference, (2.10) needs to hold for all k , which is enforced by setting $\mathbf{W}_{\text{ZF}} = \hat{\mathbf{H}}(\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1}$. Therefore, we have

$$\mathbf{r} = (\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^H \mathbf{y} = \mathbf{x} + (\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^H \mathbf{n}. \tag{2.11}$$

To guarantee that $(\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1}$ exists, the number of antennas M should be larger than the number of UEs U.

Since the ZF receiver is introduced with the effect of noise neglected, it performs well when the noise is relatively low, i.e., at high SNR.

2.1.3 LMMSE channel estimation

All the spatial processing methods just reviewed require $\mathbf{H}$ to be known to the receiver. However, it is very common in practical systems that no CSI is known at the receiver or the transmitter. Wireless channels in real life change dynamically due to lots of factors such as weather, moving obstacles and electromagnetic interferences, and therefore cannot be assumed as constant. Channel estimation methods are commonly applied in practical systems.

As the name suggests, linear minimum mean square error (LMMSE) is an estimation method that minimizes the MSE. It is also known as a special case of MMSE, where by definition of 'linear' we formulate the estimation as a linear transformation of the received data.

Denote by $\mathbf{h}$ the wireless channel, $\mathbf{x}$ the transmitted signal, $\mathbf{y}$ the received signal and $\mathbf{n}$ the additive noise, the channel estimation $\hat{\mathbf{h}}$ can be expressed as

$$\hat{\mathbf{h}} = \mathbf{U}\mathbf{y} = \mathbf{U}(\mathbf{h}\mathbf{x} + \mathbf{n}), \tag{2.12}$$

where $\mathbf{U}$ is the linear operator that needs to be optimized.

The MSE of channel estimation is $\mathbb{E}\left[(\hat{\mathbf{h}}-\mathbf{h})(\hat{\mathbf{h}}-\mathbf{h})^H\right]$. Therefore, the LMMSE algorithm can be written as follows:

$$\begin{aligned}\hat{\mathbf{h}} &= \underset{\hat{\mathbf{h}}=\mathbf{U}\mathbf{y}}{\operatorname{argmin}}\, \mathbb{E}\left[(\hat{\mathbf{h}}-\mathbf{h})(\hat{\mathbf{h}}-\mathbf{h})^H\right] \\ &= \underset{\hat{\mathbf{h}}=\mathbf{U}\mathbf{y}}{\operatorname{argmin}}\, \mathbb{E}\left[\hat{\mathbf{h}}\hat{\mathbf{h}}^H+\mathbf{h}\mathbf{h}^H-\hat{\mathbf{h}}\mathbf{h}^H-\mathbf{h}\hat{\mathbf{h}}^H\right] \\ &= \underset{\hat{\mathbf{h}}=\mathbf{U}\mathbf{y}}{\operatorname{argmin}}\, \mathbb{E}\left[\mathbf{U}\mathbf{y}\mathbf{y}^H\mathbf{U}^H+\mathbf{h}\mathbf{h}^H-\mathbf{U}\mathbf{y}\mathbf{h}^H-\mathbf{h}\mathbf{y}^H\mathbf{U}^H\right]\end{aligned} \tag{2.13}$$

We then set the derivative of (2.13) with respect to $\mathbf{U}$ to 0 in order to find the optimal $\mathbf{h}^H$:

$$\frac{\partial\mathbb{E}\left[\mathbf{U}\mathbf{y}\mathbf{y}^H\mathbf{U}^H\right]}{\partial\mathbf{U}}+\frac{\partial\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]}{\partial\mathbf{U}}-\frac{\partial\mathbb{E}\left[\mathbf{U}\mathbf{y}\mathbf{h}^H\right]}{\partial\mathbf{U}}-\frac{\partial\mathbb{E}\left[\mathbf{h}\mathbf{y}^H\mathbf{U}^H\right]}{\partial\mathbf{U}}=0, \tag{2.14}$$

where the second and the last term on the left side equal to 0 [24]. Thus,

$$\mathbb{E}\left[\mathbf{y}\mathbf{y}^H\mathbf{U}^H\right]-\mathbb{E}\left[\mathbf{y}\mathbf{h}^H\right]=0. \tag{2.15}$$

Let $\mathbf{C}_\mathbf{y}=\mathbb{E}\left\{\mathbf{y}\mathbf{y}^H\right\}$, $\mathbf{C}_{\mathbf{hy}}=\mathbb{E}\left\{\mathbf{h}\mathbf{y}^H\right\}$, we get

$$\mathbf{U}=\mathbf{C}_{\mathbf{hy}}\mathbf{C}_\mathbf{y}^{-1}, \tag{2.16}$$

combined with (2.12), the channel estimation can be formulated as

$$\hat{\mathbf{h}}=\mathbf{C}_{\mathbf{hy}}\mathbf{C}_\mathbf{y}^{-1}\mathbf{y}. \tag{2.17}$$

We observed from (2.17) that the LMMSE channel estimation requires both the received signal and the channel statistics.

2.1.3.1 Rayleigh fading channel

The ideal wireless transmission involves the transmission of a signal through a straight path between the transmitter and the receiver. It never happens in real life when there are lots of obstacles, which create more paths due to the reflection, refraction and scattering. The phenomenon that the signal power changes randomly due to the multi-path transmission is called fading.

We use in this thesis the Rayleigh fading model, where the magnitude of a signal when it reaches the receiving end varies in accordance with a Rayleigh distribution. This model applies to the environment with lots of obstacles, such as build-up urban area, so that no dominant line of sight transmission exists. A channel is Rayleigh faded when the real and imaginary part of its coefficients are independent and identically distributed (i.i.d) Gaussian distributed with zero mean and the same variance.

2.2 Short packet transmission

To satisfy the low latency requirement in URLLC, a packet within limited length is transmitted in one time-slot from a single-antenna UE. Conventional tools such as ergodic capacity and outage capacity are not suitable in the case of finite block-length. We introduce instead the results developed based on nonasymptotic, i.e., finite-blocklength information theory.

The reliability of short-packets transmission is determined by the packet error probability. In this section, we make use of a random coding argument to describe a non-asymptotic achievability bound that gives us an upper bound on the average error probability ϵ based on finite-blocklength information theory.

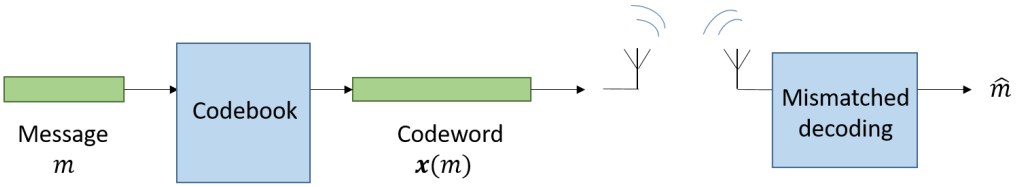

Figure 2.3: Diagram of message coding and decoding

2.2.1 Block fading channel

The block-fading channel is a common model in wireless communications for channels that vary slowly over the span of codewords with length N. The fading is considered flat during a block, and changes to a different fading state which is independent from the previous one from block to block. Despite its extreme simplification, it serves as a useful model to develop design criteria.

2.2.2 Random coding argument

Messages are not transmitted directly form the antennas. As shown in **Figure 2.3**, a codebook is used to map messages into codewords. To analyse the performance of a wireless system, instead of calculating the error probability for messages in each codebook, Shannon introduced a so called random coding argument[25] based on i.i.d codebook generation. According to this argument, the codebook is constructed randomly with each of the codeword randomly generated from a given probability distribution, which is proved to be 'good on average'. Therefore, we calculate the average error probability across all messages from all codebooks.

Assume each entrance of a transmitted codeword $\boldsymbol{x}$ is generated randomly from a determined probability distribution $p(x)$. We consider a (M, n_d, ϵ) code with rate $R = \frac{\log(M)}{n_d}$ where M denotes the number of messages, n_d denotes the block length and ϵ denotes the packet error probability. A message m from the message set $\{1, 2, 3, \cdots, M\}$ is mapped to a codeword $\boldsymbol{x}(m)$ of length n_d as the channel input. We are interested in finding a coding scheme for the triplets (M, n_d, ϵ) such that for

a determined M and n_d, the average error probability $\epsilon(M, n_d)$ does not exceed a desired value.

2.2.3 The RCUs bound

Consider the data transmission over a block-fading channel with blocklength n,

$$y_k = g\, x_k + w_k, \quad k = 1, 2, \cdots, n. \tag{2.18}$$

x_k and y_k denote the kth elements of the channel input $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ and channel output $\mathbf{y} = [y_1, y_2, \cdots, y_n]$, respectively. g is the channel gain.

The RCUs bound is derived in [20] based on mismatched decoding, where the receiver determines the estimation of the message by choosing the codeword which maximizes the decoding metric $q(\boldsymbol{x}(m), \boldsymbol{y})$:

$$\hat{m} = \underset{m}{\operatorname{argmax}}\, q(\boldsymbol{x}(m), \boldsymbol{y}). \tag{2.19}$$

The *generalized information density* is expressed as

$$i_s(\boldsymbol{x}, \boldsymbol{y}) = \log \frac{q(\boldsymbol{x}, \boldsymbol{y})^s}{\mathrm{E}\left[q(\boldsymbol{X'}, \boldsymbol{y})^s\right]}, \tag{2.20}$$

where $\boldsymbol{X'}$ denotes a different codeword from the same codebook.

Choose the decoding metric as:

$$q(\boldsymbol{x}, \boldsymbol{y}) = e^{-\|\boldsymbol{y} - \hat{g}\boldsymbol{x}\|^2}, \tag{2.21}$$

where $\hat{g}$ is the estimated channel gain, the mismatched scaled nearest-neighbor (SNN) decoder is obtained, the decoding rule in (2.19) can be written as

$$\hat{m} = \underset{m}{\operatorname{argmax}}\, e^{-\|\boldsymbol{y} - \hat{g}\boldsymbol{x}\|^2} = \underset{m}{\operatorname{argmin}}\, \|\boldsymbol{y} - \hat{g}\boldsymbol{x}\|^2, \tag{2.22}$$

where the decoder chooses the message m that minimize the Euclidean distance between the transmitted codeword and received codeword.

Error occurs when the message $\hat{m}$ picked by the decoder is different from the one we sent, i.e.$\Pr\{\hat{m} \neq m\}$. We consider the average error probability ϵ, which is averaged over all codebooks in which each codeword is generated randomly from i.i.d. Gaussian distribution. Based on the random coding union bounds (RCU), the average error probability ϵ is upper bounded by:

$$\epsilon \leq \mathrm{E}\left[min\left\{1, (M-1)\mathrm{P}\left\{q(\boldsymbol{X'}, \boldsymbol{Y}) \geq q(\boldsymbol{x}, \boldsymbol{Y}) \mid \boldsymbol{X}, \boldsymbol{Y}\right\}\right\}\right] \tag{2.23}$$

where

$$\mathrm{P}_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{X'}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x'}) = \mathrm{P}_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{P}_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x})\mathrm{P}_{\boldsymbol{X'}}(\boldsymbol{x'}). \tag{2.24}$$

We denote using $\boldsymbol{X}$ the transmitted codeword and $\boldsymbol{X'}$ a different codeword from the same codebook.

The moment-generating function for a codeword $\boldsymbol{x}$ can be written as $M(s) = \mathbb{E}[e^{s\boldsymbol{x}}]$. Apply the Chernoff's bound on $\mathbf{x}$, which is

$$\Pr\{\boldsymbol{x} \geq a\} \leq \inf_{s>0} e^{-(as - \log M(s))}, \tag{2.25}$$

to $\mathrm{P}\{q(\boldsymbol{X}', \boldsymbol{Y}) \geq q(\boldsymbol{X}, \boldsymbol{Y}) \mid \boldsymbol{X}, \boldsymbol{Y}\}$ in the RCU bound, we get

$$\mathrm{P}\{q(\boldsymbol{X}', \boldsymbol{Y}) > q(\boldsymbol{X}, \boldsymbol{Y})\} \leq \inf_{s>0} e^{-(s\, q(\boldsymbol{X}, \boldsymbol{Y}) - \log\left\{\mathbb{E}[e^{s\, q(\boldsymbol{X}', \boldsymbol{Y})}]\right\})}. \tag{2.26}$$

Setting

$$i_s(\boldsymbol{X}, \boldsymbol{Y}) = \log \frac{q(\boldsymbol{X}, \boldsymbol{Y})^s}{\mathrm{E}\left[q(\boldsymbol{X}', \boldsymbol{Y})^s \mid \boldsymbol{Y}\right]}, \tag{2.27}$$

the RCUs bound is derived from (2.23) as

$$\epsilon \leq \inf_{s>0} \mathbb{E}[e^{-(i_s(\boldsymbol{x}, \boldsymbol{y}) - \log(M-1))^+}], \tag{2.28}$$

where $(a)^+ \triangleq \max(0, a)$.

Consider the i.i.d.Gaussian codebook , $x \sim \mathcal{CN}(0, \rho)$ where ρ denotes the transmit power, the generalized information density can be formulated as:

$$i_s(\boldsymbol{x}, \boldsymbol{y}) = -s \left\|\boldsymbol{y} - \hat{g}\boldsymbol{x}\right\|^2 + s\frac{\|\boldsymbol{y}\|^2}{1 + s\rho\left|\hat{g}\right|^2} + n\log(1 + s\rho\left|\hat{g}\right|^2), \tag{2.29}$$

where $\hat{g}$ denotes the estimation of the channel gain g.

2.3 Pilot assisted wireless transmission

We assume in this thesis a block-fading channel which remains constant during the transmission of a codeword, and the BS has no CSI in advance. The channel needs to be estimated for the BS to perform signal processing. We assume that this is done by using pilot sequences that are known at both the transmitter and the receiver.

Due to the characteristics of block-fading channel in our case, channel estimation is performed for each block and pilot sequence is added to each packet. The more pilot symbols we transmit, the better channel estimation we will get.

However, pilot sequence contains no information about the message. Therefore, the resources such as time and power that we used to transmit pilot symbols are actually the resources taken away from the transmission of the information symbols. When it comes to short-packet transmission, the design of pilots needs special attentions because the amount of resources is fixed per block. The amount of pilots and the power allocated to pilot symbols affect the system performance, which is

measured by the accuracy of channel estimator and the reliability of data transmission.

To solve this interesting problem, we ask the following questions:
- How to design the pilots optimally?
- How many pilot symbols should we insert in each packet?
- How much power should we assign to pilots?
- How much performance improvement should we expect and is it reasonable enough compared to the increase in implementation complexity?

For the discrete memoryless channel, the probability of receiving $\boldsymbol{y} = [y_1, y_2, \cdots, y_n]$ given that the input to the channel is $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]$ can be written as

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{i=1}^{n} P(y_i \mid x_i). \tag{2.30}$$

2.3.1 Power allocation

We introduce in this section an optimal power allocation between pilot symbols and data symbols. We know from [10] that dramatic improvement in spectrum efficiency can be achieved by power allocation in the 1-bit massive MIMO scenario. Assuming similar performance for multi-bit cases makes sense. Appropriate resource allocation is needed especially when we are under a power constraint, i.e., average power ρ_{ave}.

The idea of the optimization problem is to minimize the average transmit power by choosing the training length and the power of pilot symbols properly. Instead of transmitting pilot symbols and data symbols using the same power, we assign different power to different parts in a block, i.e., ρ_p for pilot symbols and ρ_d for data symbols where $\rho_p \neq \rho_d$. The length of the pilot symbols is taken into account as well. Aiming at the lowest power consumption under the requirement of target error probability ϵ, the optimization problem can be formulated as follows:

$$\begin{aligned} &\underset{N,\epsilon}{\text{minimize}} && \rho_{ave} \\ &\text{subject to} && \tfrac{\rho_p \cdot n_p + \rho_d \cdot n_d}{n_p + n_d} = \rho_{ave} \\ & && \text{error probability} \leq \epsilon \end{aligned} \tag{2.31}$$

2.4 Quantization

To map the received analog signals with infinite-precision complex values to a smaller set of finite values for data processing purposes, quantizers are widely used in wireless systems. In our case, with two ADCs per BS antenna, quantization is applied separately to the real and imaginary part of the complex input y. Denoting by $\mathcal{Q}$ the quantization operator, we have $\mathcal{Q}(y) = \mathcal{Q}(\Re\{y\}) + j\mathcal{Q}(\Im\{y\})$. For a vector $\mathbf{y}$, quantization is applied to each of its elements.

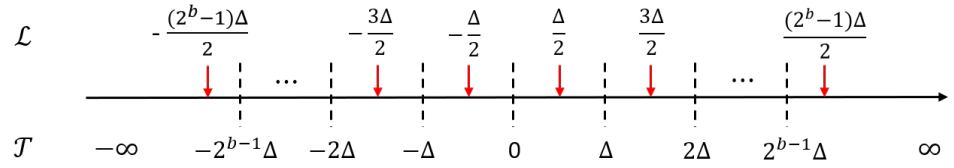

Figure 2.4: A symmetric uniform quantizer with b-bits resolution.

We assume in this thesis symmetric uniform quantizers with step size Δ and b-bit resolution as shown in **Figure 2.4**, where the red arrows denote the quantization levels and the dashed lines denote the thresholds. The number of quantization levels is $L = 2^b$. The quantizer is charaterized by $\mathcal{L} = \{\ell_1, \ell_2, \cdots, \ell_{L-1}\}, \ell_1 < \ell_2 < \cdots < \ell_{L-1}$ which is the set of quantization levels, and $\mathcal{T} = \{\tau_0, \tau_1, \cdots, \tau_L\}, -\infty = \tau_0 < \tau_1 < \cdots < \tau_L = \infty$ which is the set of $L+1$ thresholds. The set of complex-valued ADC outputs at each antenna is therefore $\mathcal{X} = \mathcal{L} \times \mathcal{L}$. Specifically, we have $\ell_i = \Delta(i - \frac{L}{2} + \frac{1}{2})$ and $\tau_i = \Delta(i - \frac{L}{2})$. The quantization rule is set as

$$\mathcal{Q}(\Re\{y\}) = \begin{cases} -\frac{(2^b-1)\Delta}{2}, & -\infty < \Re\{y\} < -2^{b-1}\Delta \\ (i - 2^{b-1} + \frac{1}{2})\Delta, & (i-1-2^{b-1})\Delta < \Re\{y\} < (i-2^{b-1})\Delta \\ \frac{(2^b-1)\Delta}{2}, & 2^{b-1}\Delta < \Re\{y\} < \infty \end{cases} \quad (2.32)$$

This rule holds also for $\mathcal{Q}(\Im\{y\})$.

2.4.1 Automatic gain control

In a practical wireless system, due to the effect of multi-path fading and interference from strong unwanted signals, the power of the received signal at each antenna can vary dynamically over a wide range, which makes the quality of quantization uncontrollable. For instance, let us denote the amplitude of a signal as A. No matter how far it is from the range of quantization level, if $A \gg \frac{(2^b-1)\Delta}{2}$ or $A \ll -\frac{(2^b-1)\Delta}{2}$, the quantization result $\mathcal{Q}(A)$ will always be $\pm\frac{(2^b-1)\Delta}{2}$, leading to a large error named *overload* distortion. On the other hand, an input signal that is really small compared to the quantization interval Δ leads to a constant output.

To improve the quality of quantization and prepare for further processing, Automatic Gain Control (AGC) unit is necessary in wireless systems. The AGC unit adjusts the received signal in order to bring its level to an acceptable level for accurate quantization. Furthermore, knowledge of the power level of the received signal is needed to perform AGC. We will show in the next chapter an AGC algorithm based on our choice of pilot and channel estimation method.

2.5 Linearization using Bussgang's theorem

A practical system normally consists of several nonlinear components such as quantizers and nonlinear amplifiers, where the output can not be formulated as a linear

function of the input, making the analysis complicated. We will introduce an exact linearization method that greatly simplifies the signal processing.

Named after Bussgang, the theorem regarding to the correlation analysis of Gaussian distributed signals as inputs to a nonlinear process was first reported in 1952. It provides a useful tool to deal with nonlinearities in a system and has been considered as a landmark. The decomposition that we are about to discuss is an extension of the results in the original paper and has been widely used in various studies as a tool for analysing the effect of nonlinearities in wireless systems. Analysis and extensions in this section are based mainly on [9], [26].

2.5.1 Bussgang's theorem regarding crosscorrelations

Theorem 1 (A Theorem regarding Cross-correlation). *For two Gaussian signals, the crosscorrelation function taken after one of them has undergone nonlinear amplitude distortion is identical, except for a factor of proportionality, to the crosscorrelation function taken before the distortion.*

Consider $f_1(t)$ and $f_2(t)$ as two Gaussian random signals. $f_1(t)$ is the input to a memoryless distortion device with distortion function U, where the output is denoted as $F_1(t) = U[f_1(t)]$. Denote using ϕ the cross-correlation function of two signals, we can formulate **Theorem 1** as:

$$\phi_{F_1 f_2}(\tau) = g\phi_{f_1 f_2}(\tau). \tag{2.33}$$

where g is the scale factor.

The theorem can be proved as follows:

The amplitudes of $f_1(t)$ and $f_2(t)$ at anytime t can be assumed as zero-mean Gaussian random variables x_1 and x_2, respectively. The probability density functions of x_1 and x_2, are:

$$p(x_1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_1^2}{2}\right), \quad p(x_2) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_2^2}{2}\right). \tag{2.34}$$

The expectation of the product of x_1 and x_2, which is denoted using $\bar{m}$, can be calculated as

$$\begin{aligned} \bar{m} &= \mathbb{E}\{x_1 x_2\} \\ &= \iint_{-\infty}^{\infty} x_1 x_2\, p(x_1, x_2)\, dx_1 dx_2 \\ &= \lim_{T\to\infty} \frac{1}{2T}\int_{T}^{-T} f_1(t) f_2(t)\, dt. \end{aligned} \tag{2.35}$$

where $p(x, y)$ represents the joint distribution the two random variables and can be expressed by a function of $\bar{m}$ as:

$$p(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\bar{m}^2}}\exp(-\frac{x_1^2 + x_1^2 - 2\bar{m}x_1x_2}{2(1-\bar{m}^2)}). \tag{2.36}$$

The second equation holds because the ensemble average and the time average in a stationary process are identical.

The last term in (2.35) is also known as the cross-correlation function of two signals, thus we have

$$\phi_{f_1 f_2}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{T}^{-T} f_1(t) f_2(t)\, dt = \bar{m}. \tag{2.37}$$

Next, we assume that the signal $f_1(t)$ undergoes a nonlinear device characterized by U with output $F_1(t) = U[f_2(t)]$, and the signal $f_2(t)$ undergoes a similar device characterized by V with output $F_2(t) = V[f_2(t)]$. The cross-correlation function of $F_1(t)$ and $F_2(t)$ can be formulated as

$$\phi_{F_1 F_2}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{T}^{-T} F_1(t) F_2(t)\, dt. \tag{2.38}$$

Just like x_1 and x_2, the amplitudes of signal $F_1(t)$ and $F_2(t)$ can be treated as random variables $U(x_1)$ and $V(x_2)$. Similar to (2.25), we get

$$\phi_{F_1 F_2}(\tau) = \iint_{-\infty}^{\infty} U(x_1) V(x_2)\, p(x_1, x_2)\, dx_1 dx_2 = \mathbb{E}\left\{U(x_1) V(x_2)\right\}. \tag{2.39}$$

For the case of one of the signal undergoes nonlinear distortion device as stated in **Theorem 1**, let x_2 be the signal without distortion. Therefore we can set $U(x_1) = x_1$ and rewrite (2.39) as

$$\begin{aligned} \phi_{f_1 F_2}(\tau) &= \iint_{-\infty}^{\infty} x_1 V(x_2)\, p(x_1, x_2)\, dx_1 dx_2 \\ &= \frac{1}{2\pi\sqrt{1-\bar{m}^2}} \iint_{-\infty}^{\infty} x_1 V(x_2) \exp\left(-\frac{x_1^2 + x_1^2 - 2\bar{m} x_1 x_2}{2(1-\bar{m}^2)}\right) dx_1 dx_2. \end{aligned} \tag{2.40}$$

Since

$$\int_{-\infty}^{\infty} x_1 \exp(-\frac{x_1^2 - 2\,\bar{m}\, x_1 x_2}{2(1-\bar{m}^2)}) dx_1 = \bar{m}\sqrt{2\pi(1-\bar{m}^2)}\, x_2 \exp\left(\frac{\bar{m}^2\, x_2^2}{2(1-\bar{m}^2)}\right), \tag{2.41}$$

(2.40) can be simplify as

$$\phi_{f_1 F_2}(\tau) = \frac{\bar{m}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x_2\, V(x_2) \exp\left(-\frac{x_2^2}{2}\right) dx_2. \tag{2.42}$$

Combined with (2.37), we get the ratio

$$\phi_{f_1 F_2}(\tau) / \phi_{f_1 f_2}(\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x_2\, V(x_2) \exp(-\frac{x_2^2}{2})\, dx_2 \tag{2.43}$$

which is the ensemble average over $x_2\, V(x_2)$. Denote this ensemble average using a constant g, we finally prove that $\phi_{F_1 f_2}(\tau) = g\, \phi_{f_1 f_2}(\tau)$.

2.5.2 Bussgang's theorem over a single device

When it comes to a device with input $f_1(t)$ and output $F_1(t)$, we can simply replace x_2 with x_1 and rewrite (2.33) as

$$\phi_{F_1 f_1}(\tau) = g\, \phi_{f_1 f_1}(\tau). \tag{2.44}$$

A consequence of Eq.(2.44) is that the output signal can be decomposed as

$$U(x_1) = gx_1 + d, \tag{2.45}$$

where d is the distortion that is uncorrelated to x_1. It is known as the *Bussgang decomposition* which indicates that the output of a non-linear device with Gaussian input is a scaled version of the input plus the uncorrelated distortion d. $d = U(x_1) - gx_1$ is a non-linear function of x, hence d is not Gaussian distributed.

2.5.3 Bussgang decomposition in MU-MIMO systems

In recent studies, Bussgang's theorem has been widely used to model the nonlinear devices in MIMO communication systems. We introduce here the extension of Bussgang decomposition to MIMO systems.

2.5.3.1 The complex-valued version of Bussgang's theorem

It was proved by Minkoff [27] that the Bussgang's theorem can be extended to the complex Gaussian signals. That is, for the complex Gaussian variables $x_1 \sim \mathcal{CN}\left(0, \sigma_{x_1}^2\right)$ and $x_2 \sim \mathcal{CN}\left(0, \sigma_{x_2}^2\right)$. x_1 undergoes a distortion where the output is denoted by $y = U(x_1)$. Denote using C the cross-correlation function, $C_{yx_2} \triangleq \mathbb{E}\{yx_2^*\}$ and $C_{x_1x_2} \triangleq \mathbb{E}\{x_1x_2^*\}$, it holds that

$$C_{yx_2} = gC_{x_1x_2}. \tag{2.46}$$

2.5.4 Extension to MIMO systems

Assume that there are M antennas at the BS, and $\mathbf{x} = [x_1, x_2, \cdots, x_M]^T$ is the vector of length M denoting the received signal at a time instant with each of its element corresponds to a single antenna. Characterize using U the distortion. The output can be expressed element-wise as

$$\mathbf{z} = U(\mathbf{x}) = \begin{bmatrix} U(x_1) \\ \vdots \\ U(x_M). \end{bmatrix} \tag{2.47}$$

We call it a MIMO extension since the distortion function takes multiple inputs and provide multiple outputs.

It is easy to conclude that each output is a distorted version of only the input

with the same index. We can then simplify the Bussgang decomposition operator $\mathbf{G}$ as a diagonal matrix $\mathbf{G} = \text{diag}([g_1, g_2, \cdots, g_M]^T)$. It holds that

$$\mathbf{C_{yx}} = \mathbf{G}\mathbf{C_x}. \tag{2.48}$$

and we can obtain the Bussgang decomposition

$$\mathbf{z} = \mathbf{G}\mathbf{x} + \mathbf{d} = \begin{bmatrix} g_1 x_1 \\ \vdots \\ g_M x_M \end{bmatrix} + \mathbf{d} \tag{2.49}$$

Lack of correlation between $\mathbf{x}$ and $\mathbf{d}$ can be proved as

$$\mathbb{E}\left[\mathbf{d}\mathbf{x}^H\right] = \mathbb{E}\left[(\mathbf{y} - \mathbf{G}\mathbf{x})\mathbf{x}^H\right] = \mathbf{C_{yx}} - \mathbf{G}\mathbf{C_y} = \mathbf{0}_{M\times M} \tag{2.50}$$

The Bussgang gain matrix $\mathbf{G}$ can be computed in for a broad range of nonlinearities as long as we know $\mathbf{C_y}$. We will indicate in the next chapter two ways of finding $\mathbf{G}$ according to different ADC resolutions. For 1-bit ADCs, $\mathbf{C_y}$ can be calculated in closed form using the arcsin law. For multi-bit ADCs, the covariance of distortion $\mathbf{d}$ is formulated in order to get $\mathbf{C_y}$.

3

System Model

In this thesis, we consider MU-MIMO uplink transmission over the i.i.d. Rayleigh-fading channel where the channel realizations between transmit-receive antenna pairs are assumed to be independent. There is no CSI at either the UEs or BS, and pilots will be used for the BS to learn the channel. The channel is assumed to remain constant for a coherence block of N symbols in time.

As shown in **Figure 3.1**, a BS equipped with B antennas communicates with $U \ll B$ single-antenna UEs simultaneously. The UEs transmit independently their own data over the fading channel, and the resulting signal is detected by B BS antennas. The signal received at each of the antenna is quantized by a pair of converters that generate the in-phase and quadrature component respectively. The quantized signal is then used to perform channel estimation, followed by linear combining, and decoding. We put before ADCs the radio frequency (RF) circuits to indicate the preprocessing of band-pass signals that occur before sampling and quantization.

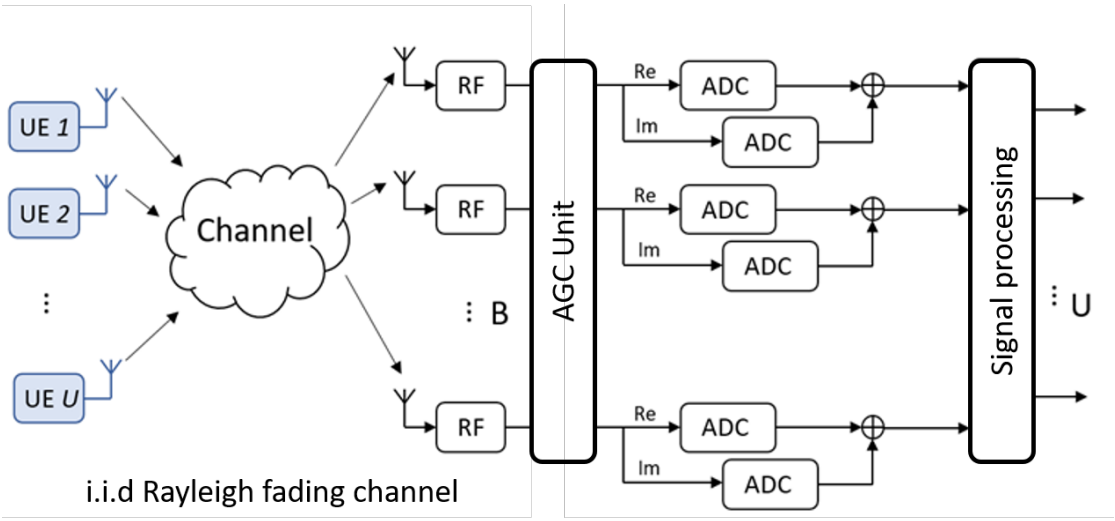

Figure 3.1: Massive MU-MIMO uplink system with low-precision converters at the BS.

We denote by $\mathbf{X} \in \mathbb{C}^{U \times N}$ the transmitted signal from U UEs, each of length N. $\mathbf{H} \in \mathbb{C}^{B \times U}$ is the realization of the block-fading channel and $\mathbf{Y} \in \mathbb{C}^{B \times N}$ is the received signal at B antennas where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]$. The B-dimensional received signal at time k is

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{z}_k, \tag{3.1}$$

where $\mathbf{z}_k$ denotes the additive white Gaussian noise (AWGN) at the antennas.

The parameters and notations used in this section are summarized in the following tables:

Table 3.1: Parameters

Parameter	Definition
$\mathbf{X}/\mathbf{x}$	Transmitted symbols matrix/vector
$\mathbf{Y}/\mathbf{y}$	Received symbols matrix/vector
$\mathbf{r}$	Decoded symbols
$\mathbf{P}$	Pilot matrix
$\mathbf{H}/\mathbf{h}$	Channel realization matrix/vector
$\mathbf{Z}/\mathbf{z}$	AWGN noise matrix/vector
U	Number of UEs
B	Number of antennas
N	Block length
b	resolution of ADCs (bit)
Δ	step size of ADCs
$\mathcal{Q}$	Quantization operator
$\mathcal{L}$	Quantization level set
$\mathcal{T}$	Quantization threshold set
$\mathbf{C}_a/\mathbf{C}_{a,b}$	Covariance matrix of a/Cross correlation matrix of a and b
$\mathbf{A}$	Scaling matrix for quantization
$\mathbf{G}$	Bussgang gain
ϵ	packet error probability

Table 3.2: Notations

Notation	Definition
$\mathrm{vec}\,(\mathbf{A})$	Vectorization of matrix $\mathbf{A}$, with the columns stacked on top of one another
$\|\mathbf{a}\|$	ℓ_2-norm of vector $\mathbf{a}$
$\mathbf{A} \otimes \mathbf{B}$	The Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\Phi(x)$	The CDF of the standard normalized distribution $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} \mathrm{d}u$
$\mathbb{E}\{x\}$	The expected value of x

3.1 Channel estimation

PAT is performed to estimate the channel. n_p pilot symbols are inserted at the beginning of each transmitted sequence by each UE. Putting them together we will get a $U \times n_p$ pilot matrix $\mathbf{P}$ which will then be analysed by the BS to learn the channel. As the information of pilots is assumed to be known at the BS, we can therefore separate the transmission of pilots and data symbols as two distinct stages, which means the the coherence block is divided into two parts which refer to channel estimation and data transmission, respectively. We will present in this section the transmission and processing of the pilots.

The received signal at all BS antennas can be written in matrix form as

$$\mathbf{Y}_p = \sqrt{\rho_p}\mathbf{H}\mathbf{P} + \mathbf{Z}, \tag{3.2}$$

where $\mathbf{Y}_p$ is a $B \times U$ matrix and $\mathbf{Z}$ is the additive white Gaussian noise matrix with independent entrances $\mathbf{z}_{k,b} \sim \mathcal{CN}(0,1)$. We denote here using ρ_p the transmit power, as all the pilot symbols are of unit power, ρ_p is also the SNR.

For simplicity, we vectorize $\mathbf{Y}_p$ by stacking its columns, that is

$$\mathrm{vec}\,(\mathbf{Y}_p) = \mathbf{y}_p = \left[y_{p_{1,1}}, \cdots, y_{p_{B,1}}, y_{p_{1,2}}, \cdots, y_{p_{B,2}}, \cdots y_{p_{1,U}}, \cdots, y_{p_{B,U}}\right]^T. \tag{3.3}$$

Rewriting (3.2) in vector form, we have

$$\mathbf{y}_p = \sqrt{\rho_p}\tilde{\mathbf{P}}\mathbf{h} + \mathbf{z}, \tag{3.4}$$

where $\tilde{\mathbf{P}} = \mathbf{P}^T \otimes \mathbf{I}_B$, $\mathbf{h} = \mathrm{vec}\,(\mathbf{H})$, $\mathbf{z} = \mathrm{vec}\,(\mathbf{Z})$.

Due to the effect of fading channel and random noise, the amplitudes of received signals may vary a lot. To match the dynamic range of the quantizers which is normally assumed to be equal on all antennas as well as minimize the quantization error, automatic gain control is performed where the received signals are first scaled by a matrix $\mathbf{A}$ before quantized. To make the average receiving power at each antenna equal to $1/B$, the scaling matrix A is designed as

$$\mathbf{A} = \frac{1}{\sqrt{B}}\mathrm{diag}(\mathbf{C}_{\mathbf{y}_p})^{-1/2}, \tag{3.5}$$

where $\mathbf{C}_{\mathbf{y}_p}$ is the covariance matrix of $\mathbf{y}_p$, which is equal to

$$\mathbf{C}_{\mathbf{y}_p} = \mathbb{E}\left(\mathbf{y}_p\mathbf{y}_p^H\right) = \rho_p\mathbf{P}^T(\mathbf{P}^T)^H \otimes \mathbf{I}_B + \mathbf{I}_{n_pB}. \tag{3.6}$$

To get rid of interferences during channel estimation, we make the pilot sequences from different UEs mutually orthogonal to get $\mathbf{P}\mathbf{P}^H = U\mathbf{I}_{n_p}$. This can be achieved by picking the first U columns of the $n_p \times n_p$ DFT matrix. Therefore,

$$\mathrm{diag}(\mathbf{C}_{\mathbf{y}_p}) = (\rho U + 1)\mathbf{I}_{n_pB} \tag{3.7}$$

$$\mathbf{A} = \sqrt{\frac{\rho_p U + 1}{B}} \mathbf{I}_{n_p B}. \tag{3.8}$$

Performing quantization on the scaled signal, we get

$$\mathbf{y}_q = \mathcal{Q}(\mathbf{A}\mathbf{y}_p). \tag{3.9}$$

As stated before, quantization is a nonlinear process that introduces unneglectable distortion. The distortion caused by finite-resolution quantizers is correlated with its input, which can not be treated as additive noise. However, applying Bussgang decomposition we can find a linear operator $\mathbf{G}_p$ and represent the quantized signal as the sum of a linear function of the unquantized signal $\mathbf{y}_p$ and a distortion $\mathbf{d}$, i.e.,

$$\mathbf{y}_q = \mathbf{G}_p(\mathbf{A}\mathbf{y}_p) + \mathbf{d}. \tag{3.10}$$

To minimize the power of distortion $\mathbf{d}$ and make the distortion uncorrelated with the received signal $\mathbf{y}_p$, $\mathbf{G}_p$ is designed as

$$\mathbf{G}_p = \frac{\Delta}{\sqrt{\pi}} \mathrm{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_p}\mathbf{A})^{-1/2} \sum_{i=1}^{L-1} \exp(-\Delta^2 (i - L/2)^2 \mathrm{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_p}\mathbf{A})^{-1}) \tag{3.11}$$

with the effect of scaling matrix $\mathbf{A}$ taken into consideration. We use Δ to denote the step size of the quantizers. Due to the nonlinearity of the quantization operator $\mathcal{Q}$, the distortion $\mathbf{d}$ is not Gaussian distributed.

We evaluate in this thesis the quality of channel estimation by the mean squared error (MSE), which is $\mathbb{E}\left\{\left\|h_{i,j} - \widehat{h}_{i,j}\right\|^2\right\}$. The linear minimum mean square error (LMMSE) channel estimator that minimize the MSE can be obtained as:

$$\hat{\mathbf{h}} = \mathbf{C}_{\mathbf{h}\mathbf{y}_q} \mathbf{C}_{\mathbf{y}_q}^{-1} \mathbf{y}_q, \tag{3.12}$$

where $\mathbf{C}_{\mathbf{h}\mathbf{y}_q}$ and $\mathbf{C}_{\mathbf{y}_q}$ will be given next. With our choice of linear operator $\mathbf{G}_p$, the distortion $\mathbf{d}$ is also uncorrelated with the channel realization $\mathbf{h}$. Hence, we can express the cross-correlation between channel realization $\mathbf{h}$ and the quantized signal $\mathbf{y}_q$ as:

$$\mathbf{C}_{\mathbf{h}\mathbf{y}_q} = \sqrt{\rho_p} \mathbf{G}_p \mathbf{A} \tilde{\mathbf{P}}^H, \tag{3.13}$$

which can be calculated using the above formulas. The covariance matrix of the quantized signal can be expressed as:

$$\mathbf{C}_{\mathbf{y}_q} = \rho_p \mathbf{G}_p^2 \mathbf{A}^2 \tilde{\mathbf{P}} \tilde{\mathbf{P}}^H + \mathbf{G}_p^2 \mathbf{A}^2 \mathbf{I}_{n_p B} + \mathbf{C}_{\mathbf{d}}, \tag{3.14}$$

where computing $\mathbf{C}_{\mathbf{d}}$ is often difficult. For 1-bit ADCs, $\mathbf{C}_{\mathbf{y}_q}$ can be computed in closed form using the arcsin law:

$$\mathbf{C}_{\mathbf{y}_q} = \frac{2}{\pi}(\arcsin(\mathbf{S}^{-1/2} \Re\left\{\mathbf{C}_{\mathbf{y}_p}\right\} \mathbf{S}^{-1/2}) + j \arcsin(\mathbf{S}^{-1/2} \Im\left\{\mathbf{C}_{\mathbf{y}_p}\right\} \mathbf{S}^{-1/2})) \tag{3.15}$$

where $\mathbf{S} = \mathrm{diag}(\mathbf{C}_{\mathbf{y}_p})$.

There is no such closed form for multi-bit quantizers. However, we can obtain an easy-to-compute approximation of $\mathbf{C_d}$ by modeling the quantization distortion as a white process in both space and time [7]. The off-diagonal elements of $\mathbf{C_d}$ are assumed to be zero, and the diagonal matrix $\mathbf{C_d^{diag}}$ is given by

$$\begin{aligned}\mathbf{C_d^{diag}} &= \text{diag}(\mathbf{C}_{\mathbf{y}_q}) - \mathbf{G}\text{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_p}\mathbf{A})\mathbf{G} \\ &= \frac{\Delta^2}{2B}(L-1)^2\mathbf{I}_{n_pB} - \mathbf{G}\text{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_p}\mathbf{A})\mathbf{G} \\ &\quad - 4B\Delta^2\sum_{i=1}^{L-1}(i-\frac{L}{2})\Phi\left(\sqrt{2B}\Delta\text{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_p}\mathbf{A})^{-\frac{1}{2}}(i-\frac{L}{2})\right).\end{aligned} \tag{3.16}$$

Then $\mathbf{C}_{\mathbf{y}_q}$ can be obtained by substituting (3.16) into (3.14).

With both $\mathbf{C}_{\mathbf{h}\mathbf{y}_q}$ and $\mathbf{C}_{\mathbf{y}_q}$ computed, we can easily get the estimated channel $\hat{\mathbf{h}}$ from (3.12). Furthermore, we can reshape $\hat{\mathbf{h}}$ to a matrix by taking every B elements as a column and put aside each other. We denote in matrix form using $\hat{\mathbf{H}} = \text{unvec}(\hat{\mathbf{h}})$ the Bussgang LMMSE channel estimator.

3.2 Data transmission

When it comes to the data transmission stage, we utilize the estimated channel to help decode data symbols. Similarly, we assume all U users transmit data symbols to the BS independently at the same time. The B-dimensional received signal at the kth channel use can be written as

$$\mathbf{y}_k = \sqrt{\rho_d}\mathbf{H}\mathbf{x}_k + \mathbf{z}_k, \tag{3.17}$$

where ρ_d denotes the power of data symbols, $\mathbf{x}_k = [x_{k,1}, \cdots, x_{k,U}] \in \mathbb{C}^U$ denotes the transmitted symbols from U UEs at the time instant k. Using Gaussian codebooks, $x_{k,u}$ are drawn independently from $\mathcal{CN}(0,1)$ distribution. $\mathbf{z}_k$ is the B-dimensional AWGN vector at time k with $\mathbf{z}_{k,b} \sim \mathcal{CN}(0,1)$ for each antenna at the BS. $\mathbf{H} \sim \mathbb{C}^{B\times U}$ is the channel realization matrix. Since pilot symbols and data symbols stay inside one coherence block, the channel matrix $\mathbf{H}$ is the same as the one for pilot transmission.

As introduced in the previous section, automatic gain control is required and the received symbols need to be scaled before quantization to match the dynamic range of the ADCs. The scaling matrix $\mathbf{A}$ for data symbols is

$$\mathbf{A} = \frac{1}{\sqrt{B}}\text{diag}(\mathbf{C}_{\mathbf{y}_k})^{-1/2}. \tag{3.18}$$

Since we have our codewords $\mathbf{x}_k$ from the Gaussian codebook, it holds that $\mathbb{E}\left\{\mathbf{x}_k\mathbf{x}_k^H\right\} = \mathbf{I}_U$. Therefore,

$$\mathbf{C}_{\mathbf{y}_k} = \rho_d\mathbf{H}\mathbf{H}^H + \mathbf{I}_B. \tag{3.19}$$

Performing the Bussgang decomposition, we can formulate the quantized signal as

$$\mathbf{y}_q^k = Q(\mathbf{A}\mathbf{y}_k) = \sqrt{\rho_d}\mathbf{G}\mathbf{A}\mathbf{H}\mathbf{x}_k + \mathbf{G}\mathbf{A}\mathbf{z}_k + \mathbf{d}, \tag{3.20}$$

where the Bussgang operator

$$\mathbf{G} = \frac{\Delta}{\sqrt{\pi}}\text{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_k}\mathbf{A})^{-1/2}\sum_{i=1}^{L-1}\exp(-\Delta^2(i-L/2)^2\text{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_k}\mathbf{A})^{-1}) \tag{3.21}$$

is diagonal for Gussian inputs. For the case of 1-bit quantizers, by setting $L = 1$, we can easily obtain

$$\mathbf{G} = \frac{\Delta}{\sqrt{\pi}}\text{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_k}\mathbf{A})^{-1/2}. \tag{3.22}$$

After quantization, signal processing methods are performed to get an estimation of the transmitted symbols. We use a linear combiner $\mathbf{W}$ to detect the data symbols from different users. By multiplying the signal with $\mathbf{W}^H$, the quantized signal is separated into U streams:

$$\mathbf{r}_k = \mathbf{W}^H\mathbf{y}_q^k = \sqrt{\rho_d}\mathbf{W}^H\mathbf{G}\mathbf{A}\mathbf{H}\mathbf{x}_k + \mathbf{W}^H\mathbf{G}\mathbf{A}\mathbf{z}_k + \mathbf{W}^H\mathbf{d} \tag{3.23}$$

$\mathbf{r}_k = [r_{k,1}, r_{k,2}, \cdots, r_{k,U}]^T$, where the uth entrance of $\mathbf{r}_k$ is the detected symbol that corresponds to the symbol transmitted by the uth UE.

For a transmitted symbol $x_{k,u}$, the corresponding symbol after combining is

$$r_{k,u} = \sqrt{\rho_d}\mathbf{w}_u^H\mathbf{G}\mathbf{A}\mathbf{h}_u x_{k,u} + \sqrt{\rho_d}\mathbf{w}_u^H\sum_{i\neq u}^{U}\mathbf{G}\mathbf{A}\mathbf{h}_i x_{k,i} + \mathbf{w}_u^H\mathbf{G}\mathbf{A}\mathbf{z}_k + \mathbf{w}_u^H\mathbf{d} \tag{3.24}$$

where $\mathbf{w}_u$, $\mathbf{h}_u$ denote the uth column of matrices $\mathbf{W}$ and $\mathbf{H}$, respectively. The first term in this equation is our desired signal, the second term denotes the multi-user interference after linear combining and the last two terms denote the additive noise and quantization distortion.

$\mathbf{W}$ is obtained on the basis of the estimated channel $\hat{\mathbf{H}}$. We formulate two types of commonly used combiners as

$$\mathbf{W} = \begin{cases}\mathbf{G}\mathbf{A}\hat{\mathbf{H}}, & \text{MRC}\\ \mathbf{G}\mathbf{A}\hat{\mathbf{H}}((\mathbf{G}\mathbf{A}\hat{\mathbf{H}}^H)\mathbf{G}\mathbf{A}\hat{\mathbf{H}}^H)^{-1}, & \text{ZF}\end{cases} \tag{3.25}$$

and use MRC in our simulation.

3.3 Packet error probability

We measure the performance of short-packet transmission in this thesis based on the average packet error probability. Error occurs when the message decoded from the received packet is different from the one we sent. As introduced in **Section 2.2.3**, the average error probability upper bounded by

$$\epsilon \leq \inf_{s>0}\mathbb{E}[e^{-(i_s(\boldsymbol{x},\boldsymbol{y})-\log(M-1))^+}] \tag{3.26}$$

is obtained based on the random coding argument and average across the codewords from all possible codebooks. The bound is calculated by averaging over all

the $e^{-(i_s(\boldsymbol{x},\boldsymbol{y})-\log(M-1))^+}$ we get.

We write the right side of (3.24) into two terms as

$$r_{k,u} = \sqrt{\rho_d} g_u x_{k,u} + n_{k,u}. \tag{3.27}$$

The first term denotes the part that contains information of the symbol we want, where we denote using $g_u = \mathbf{w}_u \mathbf{G}\mathbf{A}\mathbf{h}_u$ the channel gain corresponds to the uth UE. The second term $n_{k,u} = \sqrt{\rho_d}\mathbf{w}_u^H \sum_{i\neq u}^{U} \mathbf{G}\mathbf{A}\mathbf{h}_i x_{k,i} + \mathbf{w}_u^H \mathbf{G}\mathbf{A}\mathbf{z}_k + \mathbf{w}_u^H \mathbf{d}$ represents the combination of interference, distortion and noise.

Performing mismatched SNN decoding, we look for the codeword which, after scaling, has the smallest Euclidean distance from the output of the combiner, that is

$$\hat{\mathbf{x}}_u = \underset{\mathbf{x}}{\operatorname{argmin}} \left\| \mathbf{r}_u - \sqrt{\rho_d}\hat{g}_u \mathbf{x}_u \right\|^2. \tag{3.28}$$

Here $\hat{g}_u = \mathbf{w}_u \mathbf{G}\mathbf{A}\hat{\mathbf{h}}_u$ is the estimated channel gain. Consider the i.i.d.Gaussian codebook where we get the codewords from, the generalized information density can be formulated as:

$$i_s(\mathbf{x}_u, \mathbf{r}_u) = -s \left\| \mathbf{r}_u - \sqrt{\rho_d}\hat{g}_u \mathbf{x}_u \right\|^2 + s\frac{\left\|\mathbf{r}_u\right\|^2}{1 + s\rho_d \left|\hat{g}_u\right|^2} + n_d \log(1 + s\rho_d \left|\hat{g}_u\right|^2). \tag{3.29}$$

Note here each i_s corresponds to one codeword received. Hence for every single realization of the block-fading channel, we obtain as many i_s as the number of UEs.

Since $\log(M-1)$ in (3.26) is a constant as M is pre-defined in the system, the more i_s we collect, the tighter this bound will be. Testing over large amount of realizations we can get a reliable bound and analyse the performance of our system.

4

Simulation Results

In this chapter, we evaluate the performances of a single-cell massive MIMO uplink transmission and present the numerical results in four sections. As stated before, a PAT scheme is utilized in order to learn the channel, in which resources are taken away from data symbols to the transmission of pilots. Therefore we put our focus on the trade-off between channel estimation and data transmission.

We start by illustrating the accuracy of channel estimation. In Section 3.1 we show how the number of pilots n_p and the transmit power ρ_p affect the MSE. In Section 3.2 we illustrate with and without power allocation, the minimum SNR required to achieve the target error probability $\varepsilon < 10^{-3}$ according to different training lengths. After that, in Section 3.3 we plot the MSE as a function of SNR to characterize the trade-off between channel estimation and data transmission. At last, according to the results we got in Section 3.2 and 3.3, we plot the MSE as a function of n_p.

The parameters of simulation can be found in **Table 4.1**.

Table 4.1: System parameters

Parameter Name	Value
Number of antennas at the BS	100
Number of single-antenna UEs	10
Blocklength	288 symbols
Message payload	30 bits
Target error probability	$\epsilon < 10^{-3}$

Substituting (3.6) into (3.5) we can rewrite the scaling matrix as

$$\mathbf{A} = \frac{a}{\sqrt{B\rho_p}}\mathbf{I}_{n_pB} \tag{4.1}$$

where

$$a\mathbf{I} = \mathrm{diag}(\mathbf{P}^T\mathbf{P}^{TH} + \mathbf{I}), \tag{4.2}$$

which means the main diagonal entries of $\mathbf{A}$ are the same and depend only on transmit power ρ_p. (3.16) and (3.21) can be simplified where

$$\mathrm{diag}(\mathbf{A}\mathbf{C}_{\mathbf{y}_p}\mathbf{A}) = \mathbf{A}\mathrm{diag}(\mathbf{C}_{\mathbf{y}_p})\mathbf{A} = \frac{1}{B}. \tag{4.3}$$

Hence, neither $\mathbf{G}_p$ nor $\mathbf{C_d}$ depend on the transmit power or training length, but depend on the resolution of the quantizers, as we show in **Table 4.2**:

Table 4.2: Main diagonal entries of $\mathbf{G}_p$ and $\mathbf{C_d}$ for different ADC resolutions

Resolution	$\mathbf{G}_p$	$\mathbf{C_d}$
2 bits	~ 0.8352	$\sim 9.3021 \times 10^{-4}$
3 bits	~ 0.8764	$\sim 3.8404 \times 10^{-4}$
10 bits	~ 0.9164	$\sim 2.3224 \times 10^{-4}$
Infinite	1	0

4.1 Accuracy of channel estimation

We present first the MSE of channel estimation, $\text{MSE} = \mathbb{E}\left\{\left\|h_{i,j} - \hat{h}_{i,j}\right\|^2\right\}$, which is averaged over all the channel coefficients, as a function of training length n_p for ADCs with different resolutions. The power of pilot symbols is set as $\rho_p = 0dB$. As we expected, the more pilot symbols we transmit, the better channel estimation we will get. However, the trend of accuracy improvement does no go linearly with the increase of pilot length. MSE decreases rapidly as n_p increases from 10 to 40, but flattens out gradually which implies a limited improvement.

Increasing the resolution of quantizers is another option for better estimation. Great improvement is observed by upgrading 1-bit quantizers with 1 more bit. Using 3-bit ADCs, the quality gap of channel estimation is less than $3dB$ from the quality obtained with infinite-precision quantizers.

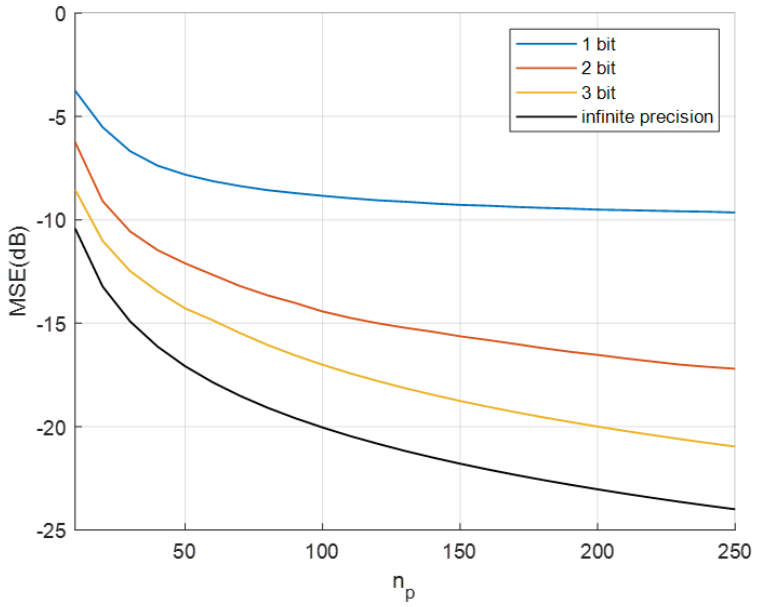

Figure 4.1: Channel estimation accuracy vs pilot length n_p for ADC with 1, 2, 3, inf bits, SNR $= 0dB$.

Specifically for 1-bit ADCs, the MSE of the Bussgang channel estimator as a function of the pilot power for different training length n_p is shown in **Figure 4.2**. The accuracy of channel estimation eventually converges to a certain level with the increase of SNR, and it converges faster when we increase the length of pilot sequence within the range of 100 symbols.Moreover, for $n_p \geq 100$ there exists an optimal SNR that gives the lowest MSE. However, large training length means lots of power taken away from data symbols. As will be presented later, when $N = 288$, there is no reason to make $n_p > 50$, especially when we perform power allocation.

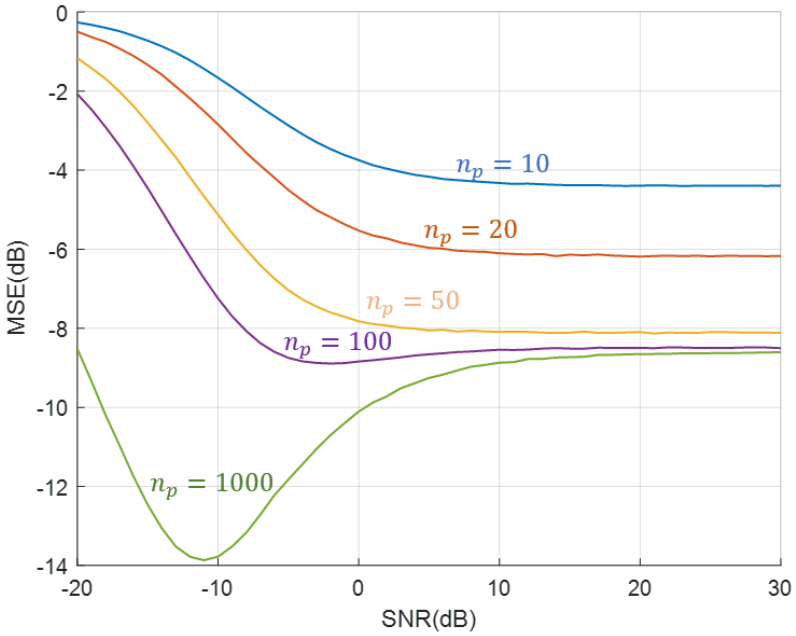

Figure 4.2: Channel estimation accuracy vs SNR with different pilot length n_p for 1-bit ADC.

4.2 Optimal PAT scheme

We illustrate the optimization of pilots by means of minimizing the average transmit power $\rho_{ave} = \frac{\rho_p \cdot n_p + \rho_d \cdot n_d}{n_p + n_d}$ under the requirement of a target error probability $\varepsilon < 10^{-3}$. The scenarios with and without power allocation are both considered. By power allocation we refer to the method stated in Section 2.3.1, where the transmit power of data symbols and pilots can be different.

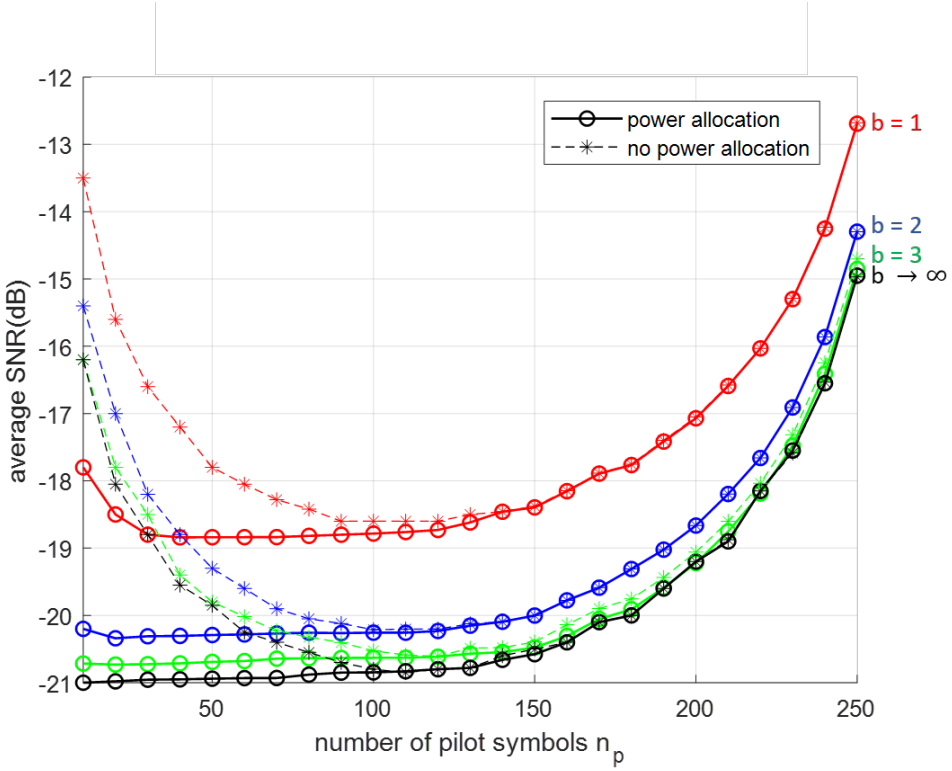

Figure 4.3: Average SNR required as a function of trainning length n_p to achieve $\varepsilon < 10^{-3}$ under condition of the resolution set 1, 2, 3, inf bits, with and without power allocation

The dash lines in **Figure 4.3** represent the results we obtained without power allocation, where $\rho_p = \rho_d$ is assumed, while the solid lines present the performances with power allocation methods.

We observe from the dash lines that without any power allocation, the uplink SNR decreases with the increase of the training length until $n_p = 120$. While for $n_p > 120$, the required SNR increases rapidly because as we set larger training length for channel estimation, there are fewer remaining symbols in a block which leads to a weaker encoder for the data payload, thus the performance loss offsets the accuracy of channel estimation. We have approximately $2.3dB$ performance gap between 1-bit ADCs and infinite resolution ADCs, and the gap became as small as $0.6dB$ for 2-bit ADCs. Using the quantizers with 3-bit resolution, we are able to approach infinite-resolution performance.

It can be easily observed that power allocation improves the performances significantly. For instance, to achieve the same error probability using 3-bit ADCs, we can save up to 4.8 dB using power allocation with 10 pilot symbols. For infinite-resolution quantizers, the optimal number of pilot symbols that results in the lowest average power is 10, which is the same as the number of UEs. The lower the resolution is, the larger the training length is needed. We have the optimal $n_p \approx 40$, $n_p \approx 20$, and $n_p \approx 13$ for 1-bit ,2-bit and 3-bit resolution cases respectively. The reason is that a larger number of pilot symbols are needed for better channel estimation to reduce the effect of quantization distortion. Again with 3-bit ADCs, the performance is fairly close to the case of infinite-resolution ADCs. The performance gap between the cases with and without power allocation became smaller with the increase of training length. For $n_p > 130$ we can hardly see any performance improvement.

Compare either the dash lines or the solid lines, we conclude that the performance improvement achieved by increasing resolution from 1 bit to 2 bits is apparent, but for more than 3 bits, the improvement becomes smaller and might be offset by higher power consumption. Combined with the results showed in **Figure4.2**, we are actually far from the SNR which gives us the best channel estimation under each pilot length.

4.3 Channel estimation vs data transmission

In this section, we consider the trade-off between channel estimation accuracy and data transmission reliability under the requirement of error probability $\varepsilon < 10^{-3}$. For each n_p and the required average SNR in **Figure 4.3**, the MSE can be easily obtained as we have both n_p and ρ_p. We depicted in **Figure 4.4** the MSE as a function of the SNR required to achieve an error probability $\varepsilon < 10^{-3}$.

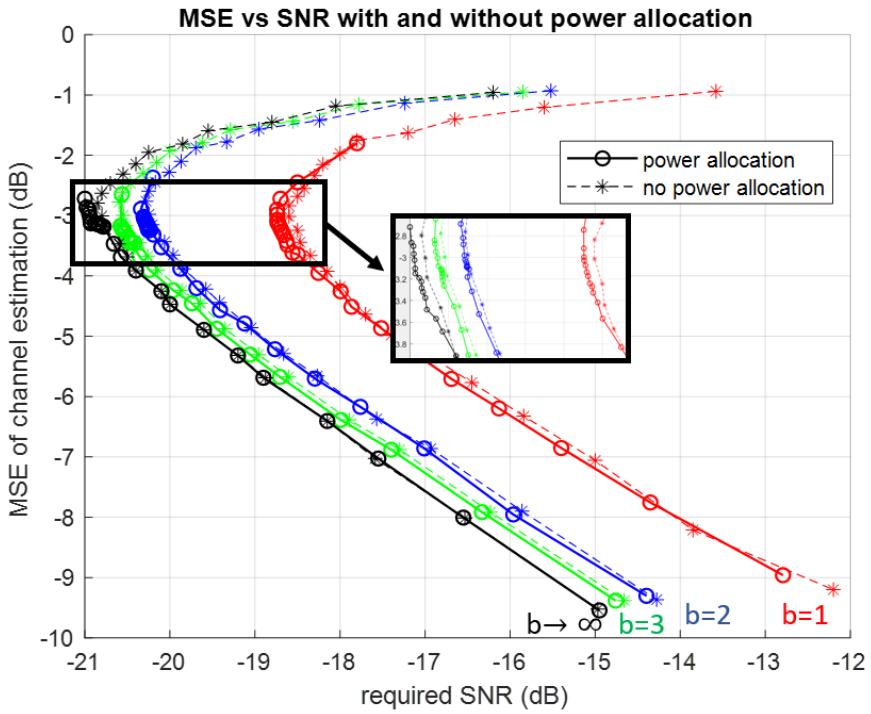

Figure 4.4: MSE of channel estimation as a function of SNR under the condition of $\varepsilon < 10^{-3}$ and resolution set 1,2,3,inf bits, with and without power allocation.

The shape of the curves looks like a 90-degree clockwise rotation of the curves in **Figure 4.3** because the MSE decreases with the increase of n_p as analysed in Section 3.1. With power allocation, as there is more power assigned to pilot symbols, we can always get a better channel estimation. As observed, lots of the MSEs we got concentrated around $-3dB$ to $-4dB$ for all the resolutions within a small range of required SNR, which correspond to $n_p < 120$ in **Figure 4.3**. No notable improvement in channel estimation can be achieved until we increase the number of pilot symbols dramatically. We can also observe small gaps around $-3dB$ which indicates that with power allocation, less power is consumed to achieve the same MSE. If we are under a power constraint, i.e., $-18.6dB$ for 1-bit ADCs, we can achieve the same reliability but a more accurate channel estimation by using more pilot symbols, i.e., $n_p = 160$ for MSE $< -3.7dB$ instead of $n_p = 20$ for MSE $= -2.36dB$.

4.4 Channel estimation based on PAT scheme

We look back to the relationship between channel estimation and the length of pilots. Unlike Section 4.1 where only pilot symbols are considered, results we obtained here shows the estimation accuracy required to transmit the data reliably.

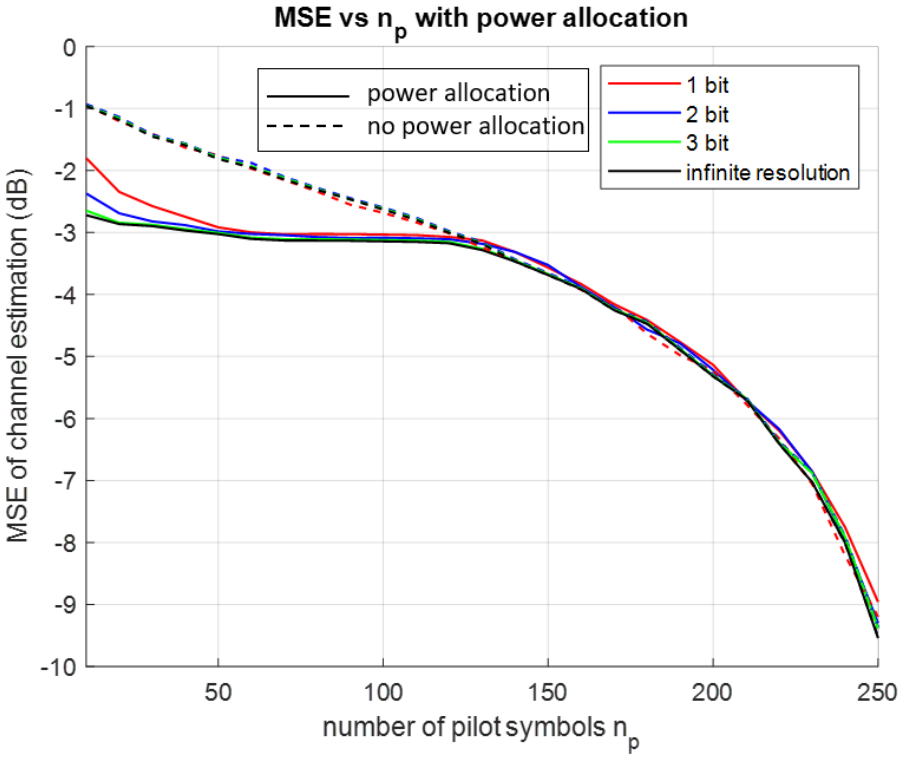

Figure 4.5: Channel estimation pilot length n_p to achieve $\varepsilon < 10^{-3}$ under condition of the resolution set 1,2,3,inf bits, with and without power allocation

In **Figure 4.5**, each pair of MSE and n_p corresponds to a required SNR which can be found in **Figure 4.3** or **Figure 4.4**, to achieve the target error probability $\varepsilon < 10^{-3}$. With solid lines and dashed lines refer to the scenarios with and without power allocation respectively, it's easy to see that the channel estimation is better with power allocation when $n_p < 120$. For $n_p > 130$, there is hardly any improvement observed.

For the case of power allocation, lower ADC resolution corresponds to higher requirement on estimation accuracy, which is reasonable because better channel estimation is needed to overcome larger quantizer distortion. While for the case without power allocation, we have a quite different observation where no matter how we increase the resolution, the quality of channel estimation does not differ a lot for different training lengths. The MSEs we find for 1-bit ADCs can be used as a reference for multi-bit cases, to identify the required transmit power for target error probability. This indicates that the problem of achieving the target error probability equals to that of achieving the target channel estimation accuracy, which will greatly reduce the computing time.

5
Conclusion

5.1 Conclusion

We analyzed in this thesis the performances of short packet transmission in the massive MU-MIMO uplink scenario over a flat Rayleigh-fading channel with low-resolution ADCs at the BS. We focused on the trade-off between channel estimation accuracy and data transmission reliability in order to design the optimal PAT scheme. We used Bussgang decomposition to characterize the quantization process and nonasymptotic bound on the error probability, which is more appropriate for finite blocklength. A simple power allocation method is performed to minimize the average power consumed under a target error probability.

The numerical results show that with fairly low-precision quantizers, i.e., 3-bit ADCs, the performance we can achieve is close to the infinite-precision case. There exist an optimal training length if no power allocation is performed or the resolution of quantizers is as low as 1 to 2 bits. Furthermore, we can use as many pilot symbols as the number of UEs to get the optimal performance if we increase the resolution together with power allocation. The performance improvement by power allocation is notable when we use a limited number of pilot symbols. By increasing pilot length, we can get better channel estimation with slightly more consumed power. However, there is no significant improvement if a lot more pilot symbols are used and power is invested. Furthermore, we found that if power allocation is not available, the problem of achieving the target error probability might be equal to the problem of achieving the target channel estimation accuracy, which leads to a reduction in simulation time.

5.2 Future work

The normal reliability requirement in URLLC is higher than 99.999%, indicating the target error probability $\epsilon \leq 10^{-5}$. Our current simulation is based on $\epsilon \leq 10^{-3}$ due to the long running time and lots of set-up cases. To get more realistic simulation, higher number of realizations need to be tested.

Moreover, the i.i.d Rayleigh fading channel is too ideal to be true in real wireless systems, more realistic channel models need to be studied. Local scattering spatial correlation model is worth to be considered as the spatial correlation caused by scattering around the UEs is an important characteristic of massive MIMO.

Bibliography

[1] S. R. Pokhrel, J. Ding, J. Park, O. S. Park, and J. Choi. "Towards enabling critical mMTC: a review of URLLC within mMTC". *IEEE Access*, 8:131796–131813, 2020.

[2] P. Elias, A. Feinstein, and C. Shannon. "A note on the maximum flow through a network". *IRE Transactions on Information Theory*, 2(4):117–119, 1956.

[3] P. F. Driessen and G. J. Foschini. "On the capacity formula for multiple input-multiple output wireless channels: a geometric interpretation". In *1999 IEEE International Conference on Communications (Cat. No. 99CH36311)*, volume 3, pages 1603–1607 vol.3, 1999.

[4] B. Kamali. *The IEEE 802.16 standards and the WiMAX technology*, pages 189–258. 2018.

[5] R. H. Walden. "Analog-to-digital converter survey and analysis". *IEEE Journal on Selected Areas in Communications*, 17(4):539–550, 1999.

[6] J. Singh, O. Dabeer, and U. Madhow. "On the limits of communication with low-precision analog-to-digital conversion at the receiver". *IEEE Transactions on Communications*, 57(12):3629–3639, 2009.

[7] Sven Jacobsson, Giuseppe Durisi, Mikael Coldrey, and Christoph Studer. "Linear precoding with low-resolution DACs for massive MU-MIMO-OFDM downlink". *IEEE Transactions on Wireless Communications*, 18(3):1595–1609, 2019.

[8] L. Fan, S. Jin, C. Wen, and H. Zhang. "Uplink achievable rate for massive MIMO systems with low-resolution ADC". *IEEE Communications Letters*, 19(12):2186–2189, 2015.

[9] Julian J. Bussgang. "Crosscorrelation functions of amplitude-distorted Gaussian signals". *Tech. Rep. 216, Research Lab. Electron*, 1952.

[10] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu. "Channel estimation and performance analysis of one-bit massive MIMO systems". *IEEE Transactions on Signal Processing*, 65(15):4075–4089, 2017.

[11] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath. "Uplink performance of wideband massive MIMO with one-bit ADCs". *IEEE Transactions on Wireless Communications*, 16(1):87–100, 2017.

[12] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer. "Throughput analysis of massive MIMO uplink with low-resolution ADCs". *IEEE Transactions on Wireless Communications*, 16(6):4038–4051, 2017.

[13] Yasaman Ettefagh, Sven Jacobsson, Anzhong Hu, Giuseppe Durisi, and Christoph Studer. "All-digital massive MIMO uplink and downlink rates under a fronthaul constraint". pages 416–420, 2019.

[14] L. A. Sonkusare and S. N. Dhage. "Analysis of LTE UE RF parameters for 3GPP specification". In *2015 International Conference on Computers, Communications, and Systems (ICCCS)*, pages 82–86, 2015.

[15] Madyan Alsenwi, Nguyen H. Tran, Mehdi Bennis, Shashi Raj Pandey, Anupam Kumar Bairagi, and Choong Seon Hong. "Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: a deep reinforcement learning based approach". *IEEE Transactions on Wireless Communications*, 2020.

[16] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh. "5G URLLC: design challenges and system concepts". In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pages 1–6, 2018.

[17] P. Popovski, Č. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A. Bana. "Wireless access in ultra-reliable low-latency communication (URLLC)". *IEEE Transactions on Communications*, 67(8):5783–5801, 2019.

[18] Johan Östman, Giuseppe Durisi, Erik G. Ström, Mustafa C. Coşkun, and Gianluigi Liva. "Short packets over block-memoryless fading channels: Pilot-Assisted or Noncoherent Transmission?". *IEEE Trans. Commun.*, 2018, to appear.

[19] Y. Polyanskiy, H. V. Poor, and S. Verdu. "Channel coding rate in the finite blocklength regime". *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.

[20] A. Martinez and A. G. i Fàbregas. "Saddlepoint approximation of random-coding bounds". In *2011 Information Theory and Applications Workshop*, pages 1–6, 2011.

[21] J. Östman, A. Lancho, and G. Durisi. "Short-packet transmission over a bidirectional massive MIMO link". In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 1399–1403, 2019.

[22] J. Mo and R. W. Heath. "High SNR capacity of millimeter wave MIMO systems with one-bit quantization". In *2014 Information Theory and Applications Workshop (ITA)*, pages 1–5, 2014.

[23] S. M. Alamouti. "A simple transmit diversity technique for wireless communications". *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, 1998.

[24] S. Trampitsch. "Complex-valued data estimation". pages 7–9, 2013.

[25] C. E. Shannon. "A mathematical theory of communication". *The Bell System Technical Journal*, 27(3):379–423, 1948.

[26] Özlem Tuğfe Demir and Emil Björnson. "The Bussgang decomposition of nonlinear systems: basic theory and MIMO extensions". *IEEE Signal Processing Magazine*, 2020.

[27] J. Minkoff. "The role of AM-to-PM conversion in memoryless nonlinear systems". *IEEE Transactions on Communications*, 33(2):139–144, 1985.