CHALMERS


# Throughput analysis of massive MIMO uplink with one-bit ADCs

SVEN JACOBSSON

Department of Signals and Systems
Chalmers University of Technology
Gothenburg, Sweden 2015
Master's Thesis EX006/2015

THESIS FOR THE DEGREE OF MASTER IN SCIENCE

# Throughput analysis of massive MIMO uplink with one-bit ADCs

SVEN JACOBSSON


CHALMERS

Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015

Master Thesis for the Master Programme "Communication Engineering"
Throughput analysis of massive MIMO uplink with one-bit ADCs
SVEN JACOBSSON



Technical report no. EX006/2015
Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

# Abstract

Massive MIMO is a disruptive technology that consists of equipping communication transceivers with an excessive number of antennas to enable unprecedented gains in throughput and radiated energy efficiency. From an industry perspective, massive MIMO will happen only if the required large antenna arrays can be built out of low-cost (and, hence, low-precision) components. Therefore it is of interest to investigate whether the above mentioned massive MIMO gains, which were theoretically derived under the assumption of ideal hardware, survive in the presence of significant impairments due to low-cost hardware solutions

The specific focus of this thesis is on low-precision analog-to-digital converters. Equipping base stations with a large number of high-precision converters would yield an intolerable increase in power consumption, especially for wideband systems operating at very large central frequencies (e.g., millimetre-wave systems). Therefore, radio-frequency solutions involving low-precision converters are attractive for systems equipped with large antenna arrays.

This thesis investigates the achievable uplink throughput in a massive MIMO system under the presence of one-bit converters. The base station is assumed not to have any *a priori* channel state information. Hence, uplink fading coefficients have to be estimated based on the one-bit quantised channel outputs. In this thesis, it is shown that simple signal processing techniques, such as least squares channel estimation and maximal-ratio combining, are sufficient to enable the use of high-order modulation schemes, as well as to support multiuser operation, despite of the nonlinearity introduced by the one-bit quantisers. Furthermore, a data-aided version of least squares channel estimation is shown to be capacity-achieving in the single-antenna case.

# Acknowledgements

This thesis is the concluding part of my studies at the Master's Programme in Communication Engineering at Chalmers University of Technology. The thesis work has been conducted at Ericsson Research in Gothenburg and at the department of Signals and Systems at Chalmers University of Technology.

I would like to express my gratitude towards everyone who has contributed to this work, especially to the following persons.

First, I would like to thank my two main supervisors, namely Ass. Prof. Giuseppe Durisi at Chalmers University of Technology and Dr. Mikael Coldrey at Ericsson Research. I would also like to thank my co-supervisors Dr. Fredrik Athley and Sven Petersson at Ericsson Research and Ass. Prof. Christoph Studer at Cornell University. Finally, I would like to thank Dr. Ulf Gustavsson at Ericsson Research and Vimar Björk at Ericsson for important input on hardware aspects.

Sven Jacobsson
Gothenburg, January 21, 2015

# Contents

# Acronyms

| | |
|---|---|
| ADC | Analog-to-Digital Converter |
| AGC | Automatic Gain Control |
| BS | Base Station |
| AWGN | Additive White Gaussian Noise |
| CDF | Cumulative Distribution Function |
| CSI | Channel State Information |
| CSIR | Channel State Information at the Receiver |
| CSIT | Channel State Information at the Transmitter |
| DAC | Digital-to-Analog Converter |
| DSP | Digital Signal Processing |
| EHF | Extremely High Frequency |
| ENOB | Effective Number Of Bits |
| FOM | Figure Of Merit |
| IID | Independent and Identically Distributed |
| ISSCC | IEEE International Solid-State Circuit Conference |
| LOS | Line-Of-Sight |
| LS | Least Squares |
| LTE | Long Term Evolution |
| MIMO | Multiple-Input Multiple-Output |
| MmWave | Millimetre-Wave |
| NLOS | Non-Line-Of-Sight |
| MRC | Maximal-Ratio Combining |
| OFDM | Orthogonal Frequency-Division Multiplexing |
| PDF | Probability Density Function |
| PMF | Probability Mass Function |
| PSK | Phase-Shift Keying |
| QAM | Quadrature Amplitude Modulation |
| QPSK | Quadrature Phase-Shift Keying |

| | |
|---|---|
| RF | Radio Frequency |
| SHF | Super High Frequency |
| SISO | Single-Input Single-Output |
| SIMO | Single-Input Multiple-Output |
| SINR | Signal-to-Interference-plus-Noise Ratio |
| SNDR | Signal-to-Noise-and-Distorsion-Ratio |
| SNR | Signal-to-Noise Ratio |
| TDD | Time-Division Duplex |
| UE | User Equipment |
| VLSI | Very-Large-Scale Integration |
| ZF | Zero-Forcing |

# Notation

In this thesis, scalars are denoted by lowercase letters, vectors are denoted by boldface lowercase letters and matrices are written as boldface uppercase letters, e.g.,

| | |
|---|---|
| $x$ | scalar, |
| $\boldsymbol{x}$ | vector, |
| $\boldsymbol{X}$ | matrix. |

No notational distinction is made between a random and a deterministic variable, i.e., $x$ is used to denote both a random variable and its realisation. The identity matrix of size $T \times T$ is written as $\boldsymbol{I}_T$. The superscripts $^T$ and $^H$ denotes transpose and Hermitian transpose respectively. The probability of an event is written as $\Pr\{\cdot\}$. The probability mass function of a discrete random variable $x$ is written as $P_x(\cdot)$ and the probability density function of a continuous random variable $y$ is written as $p_y(\cdot)$. The expectation function of a random variable is denoted as $\mathbb{E}_x[\cdot]$ where the subscript emphasises that the expectation is taken with respect to $x$. The distribution of a circularly symmetric complex Gaussian random variable with mean $\mu$ and variance $\sigma^2$ is denoted by $\mathcal{CN}(\mu,\sigma^2)$. Finally, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary part, respectively.

# Chapter 1

# Introduction

## 1.1 Background and motivation

A new generation of cellular networks has appeared approximately every tenth year since the release of Nordic Mobile Telephone in 1981, to the 4G network launched in recent years. The next generation cellular network, 5G, is a technology solution for year 2020 and beyond that has to meet the ever-increasing aggregate traffic demands from users — to enable fast and reliable access to information anywhere and at any time.

Given the increased traffic over mobile networks over the last years, high capacity demands are to be expected in the future. As only gradual improvements are expected for the Long Term Evolution (LTE) system embodying 4G, research activities on 5G technologies have emerged [1]. According to predictions, the total data traffic by year 2020 will have increased a thousand-fold compared to today's traffic volume [2], with more than 50 billion connected devices [3, 4]. To support this immense increase in demand in a sustainable and affordable way, tremendous improvements in spectral efficiency and energy efficiency are required. Therefore, 5G can not only be an evolution of existing solutions, but will also have to be complemented by new and possibly disruptive technologies. Five such technologies are presented in [5], with two of them being of particular interest to this work, namely massive multiple-input multiple-output (MIMO) and the move to millimetre-wave (mmWave) frequencies.

### 1.1.1 Massive MIMO

Increasing the capacity and reliability of wireless communication systems using multiple antennas has been an active research area since the last two decades. The use of multiple antennas is a part of the current LTE standard. Massive MIMO, or large-scale antenna systems, is a technology that has attracted a lot of recent research interest for use in 5G networks. Massive MIMO is a variation of multiuser MIMO, in which the number of antennas at the base station (BS) is scaled up by several orders of magnitude, such that the number of antenna elements exceeds by far the number of active users per time-frequency resource.

Massive MIMO offers big advantages over traditional multiuser MIMO and claims have been made about a ten-fold increase in capacity and hundred-fold improvement

in energy efficiency [6, 7]. The underlying reason for the improvements is that extensive spatial multiplexing is possible due to the large antenna array through beamforming, where the transmitted energy can be focused towards the intended user equipment (UE) with great precision into small regions in space.

Efficient spatial multiplexing is heavily dependent on the BS having accurate channel knowledge. In the uplink (from UE to BS), this is accomplished by having the UEs transmit pilots which the BS consequently use to estimate the fading coefficients. In the downlink (from BS to UE), the problem becomes more challenging. In the LTE standard, the BS transmits pilots to the UEs which perform channel estimation and feed back channel information to the BS. Unfortunately, this approach is not applicable in a massive MIMO setup as pilot sequences should be orthogonal between antennas so that the number of required pilots scales linearly with the array size. A better option is to consider a time-division-duplex (TDD) scheme where only the UEs are transmitting pilots and where the BS in the downlink instead relies on channel reciprocity between uplink and downlink [7]. Regardless, channel estimation is still one of the main challenges in massive MIMO as users in motion limit the coherence time of the channel during which estimation has to be performed. Thus, the coherence time limits the number of possible users as there is only a finite number of orthogonal pilot sequences. Reusing pilot sequences causes pilot contamination, which causes energy to be inadvertently directed towards unintended users [6]. Furthermore, the number of channel coefficients that have to be estimated scales with the number of antennas, which calls for increased signal processing at the BS.

As the size of the antenna array increases, so will the total power consumption of the required hardware components, such as power amplifiers, oscillators and converters. Luckily, the constraint on accurate and linear hardware in massive MIMO is relaxed due to the favourable law of large numbers so that, in the end, all that matters is their joint operation. Noise, fading, and hardware imperfections will to some extent average out when signals associated with a large number of antennas are combined in the air [7]. Thus, high-precision components used today could potentially be replaced by a large number of low-cost, low-precision components, thereby reducing the internal power consumption.

It is of interest to investigate whether the above mentioned massive MIMO gains, which were theoretically derived under the assumption of ideal hardware, survive under the presence of significant impairments due to low-cost hardware solutions. Although this question has been recently posed in the scientific literature [8], no definite answer is available yet.

### 1.1.2 Millimetre-wave

The spectrum ranging from approximately 300 MHz to 3 GHz available to today's cellular communication systems is becoming more crowded, with a scarce amount

of expected new available spectrum. There is, however, an enormous amount of spectrum still unused at frequencies in the super high frequency (SHF) band ranging from 3 to 30 GHz and in the extremely high frequency (EHF) band from 30 GHz to 300 GHz [9]. Frequencies in the EHF band are often referred to as mmWave as the wavelength, $\lambda$, ranges from 1 mm to 10 mm .

Operating at mmWave frequencies brings a lot of new research challenges and has long been deemed unsuitable for long-range cellular communications. The greatest concern has been the increased path loss incurred by communicating at higher frequency. According to Friis transmission equation the relation between the received power, $P_R$, and transmitted power, $P_T$, is given by

$$P_R \text{ [dBm]} = P_T \text{ [dBm]} + G_T \text{ [dB]} + G_R \text{ [dB]} - \text{PL [dB]} \tag{1.1}$$

where $G_T$ and $G_R$ are the antenna gains at the transmitter and receiver respectively and where PL is the free space path loss. If the transmitter and receiver are separated by a distance $d$, then the path loss is given by

$$\text{PL [dB]} = 20 \cdot \log_{10}\left(\frac{4\pi d}{\lambda}\right). \tag{1.2}$$

The free space path loss thus decreases with the wavelength by 20 dB per decade (or equivalently, increases with frequency by 20 dB per decade). However, for a fixed physical aperture, the antenna gains are proportional to the frequency squared. Hence, the increased path loss is compensated for by sending and receiving energy in narrow beams [10]. With a large antenna array, the narrow beams can be directed with great precision by adjusting the phase and amplitude at each antenna element, i.e., by beamforming.

Propagation issues such as higher penetration losses and attenuation due to rain, foliage, and atmospheric absorption are still a problem at higher frequencies. However, as an ultra-dense deployment of BSs is expected in 5G networks [4], the distance between the BS and UE will decrease, and the probability of a line-of-sight (LOS) channel will increase.

### 1.1.3 Analog-to-digital converters

Digital signal processing (DSP) is an integral part in all modern communication systems. In order to process data digitally, the analog baseband signal has to be mapped into digital domain, which requires conversion both in time and amplitude. The process of converting a continuous-time signal into a discrete-time signal is known as sampling, while converting a continuous amplitude into a discrete amplitude is referred to as analog-to-digital conversion, or simply as quantisation. The circuit that performs this operation, known as an analog-to-digital converter (ADC), is a necessary component in every system that includes DSP.

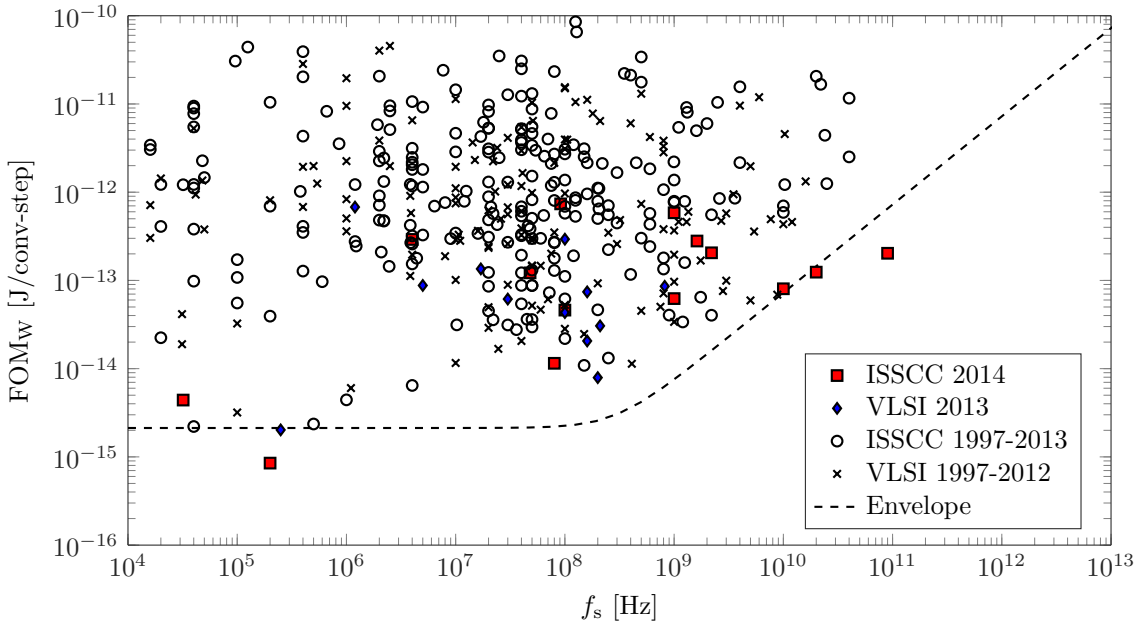



**Figure 1.1:** Walden's FOM as a function of the sample rate of the ADC.

The continuous amplitude signal is mapped into a discrete set of quantisation regions, where the mapping is determined by the conversion steps of the ADC. The number of possible conversion steps, $2^b$, is determined by the resolution of the ADC, which is measured in the number of bits, $b$, that are used to represent the quantised signal. However, any physical ADC introduces noise and distortion which reduces the effective resolution. The effective number of bits (ENOB) is a performance measure directly related to the signal-to-noise-and-distortion ratio (SNDR). It specifies the resolution of an ADC after these impairments have been taken into consideration, and is typically a few bits below the specified resolution [11]. A commonly adopted figure of merit (FOM) that relates the resolution and speed of an ADC to the power consumption is Walden's FOM, given by

$$\mathrm{FOM_W} = \frac{P_\mathrm{diss}}{2^\mathrm{ENOB} f_s} \quad [\mathrm{J/conv\text{-}step}] \tag{1.3}$$

where $P_\mathrm{diss}$ is the dissipated power and $f_s$ is the sampling rate at which the ADC operates. Walden's FOM is based on the empirical evidence that the effective resolution decreases with approximately one bit per doubling of the sampling rate [11, 12].

Figure 1.1 presents Walden's FOM for ADC designs presented at the IEEE International Solid-State Circuit Conference (ISSCC) and at the very-large-scale integration (VLSI) circuit symposium that has been collected in [13] from 1997 to 2014. Even though there exist modern high-precision ADCs that operate at high speed

(several designs presented at ISSCC 2014), the associated power dissipation is too high for many applications. The envelope in Figure 1.1 (based on the five best ADC designs) shows that the power consumption increases drastically for sampling rates higher than approximately 100 MHz. Thus, for wideband systems (e.g. mmWave systems), the ADC becomes a bottleneck in terms of power consumption.

As explained earlier, beamforming is a key technology to enable spatial multiplexing in massive MIMO and to overcome the increased path loss for mmWave frequencies. Beamforming can be performed in digital or analog domain. With digital beamforming, the baseband signals is multiplied with complex coefficients in digital domain at each antenna. In traditional MIMO systems, beamforming is performed digitally, which requires dedicated high-precision converters for each antenna element. This strategy is very costly for large bandwidths and large antenna arrays [14]. Commercial high-speed 8-bit converters with sampling rates in excess of 1 GHz consumes several Watts [15]. Thus, if each antenna in a mmWave massive MIMO system is connected to a high-precision ADC, the total power consumption of the converters alone would be substantial.

Therefore, it is of interest to study the performance of a massive MIMO system employing low-precision converters. Reducing the ADC resolution has the additional benefit of reducing the required accuracy/linearity of radio-frequency (RF) hardware. For example, using low-precision converters, the power amplifiers is allowed to be less linear.

Typically, as the received power per antenna may fluctuate widely, an automatic gain control (AGC) is required to fit the amplitude of the received signal to the ADC quantisation levels. In the special case of one-bit converters, i.e., the lowest possible ADC resolution, the AGC can be removed completely, as only the sign of the signal is important.

## 1.2 Previous work

Communication systems employing one-bit ADCs have been previously analysed in the context of low-power ultra-wideband systems [16, 17], and, more recently, in the context of millimetre-wave communication systems [18]. Communication over fading channels, where the channel coefficients are unknown to the receiver, is studied in [19, 20, 21]. Bayesian channel estimation techniques based on one-bit quantised outputs is considered in [22, 23].

In a timely contribution, which has greatly influenced the direction of this thesis, massive MIMO systems with one-bit converters are considered in [24].

Other contributions of particular importance to this work are referenced throughout the thesis.

## 1.3 Aim and outline

This work considers a massive MIMO system utilising one-bit ADCs at the BS. The main objective is to characterise achievable uplink throughput and to understand the limitations that the one-bit ADC entails. The channel is assumed to be unknown at both the transmitter and the receiver, so that the uplink fading coefficient has to be estimated based on a coarsely quantised channel output.

The aim of the thesis is to determine whether high throughputs are achievable with large antenna arrays, in spite of the one-bit quantisers. When this is established, it is further investigated whether multiple users can be served simultaneously. The thesis is organised as follows.

Chapter 2 considers the case when a single user with one transmit antenna is communicating with a BS equipped with an antenna array of varying size. The channel and system model is presented, and the operation of a one-bit quantiser is explained in detail. The concepts of achievable rate and channel capacity is defined. It is shown that when the channel is unknown to both the transmitter and receiver, a training based scheme involving least squares estimation is capacity-achieving in the single-antenna case. When the BS is equipped with a large antenna array, a system using a linear receiver is considered and achievable rates for different inputs are computed. Chapter 3 considers a one-bit massive MIMO system. It is shown that despite the one-bit converters, multiuser communication with high-order constellations is possible. Chapter 4 provides a discussion based on the obtained results and lists future research challenges. Finally, Chapter 5 summarises the results and concludes the thesis. Detailed mathematical derivations of some of the results in the thesis are provided in Appendices A and B.

# Chapter 2

# Single user uplink

## 2.1 Channel model

One of the main challenges in wireless communications is to abate the fading between the transmitter and receiver. Fading is the attenuation that affects the wireless signal while it propagates through the air interface and may vary with both time and frequency, as well as with geographical position.

Throughout this paper, a block-fading memoryless channel is considered. The channel is assumed to remain constant during a coherence interval of $T$ consecutive symbol transmissions before changing into a new independent realisation that again would remain constant for $T$ symbols. For a frequency-flat narrowband channel $T$ would correspond to the number of symbol durations in time for which the fading remains constant. For a frequency-selective channel under the assumption that orthogonal frequency-division multiplexing (OFDM) is used, $T$ is the number of subcarriers where the fading is constant. In general, $T$ can be considered as the number of "time-frequency" slots for which the fading is the same.

In this work, a single-carrier frequency-flat channel is considered, so that $T$ is the number of symbols in time for which the fading is constant as illustrated by Figure 2.1. Rayleigh fading is considered so that the random propagation coefficients are modelled as independent and identically distributed (IID) zero-mean circularly symmetric Gaussian random variables. This piece-wise constant random channel model can be seen as an approximation of a continuous fading model such as Jakes' [25]. In Jakes' model, the coherence time, $T_C$, is a measure of the minimum time required for the channel to become uncorrelated with itself and is given by

$$T_C = \sqrt{\frac{9}{16\pi f_D^2}} \quad \text{[seconds]}, \tag{2.1}$$

where $f_D$ is the Doppler shift of the channel that depends on the speed of light, c, the carrier frequency, $f_c$, and on the velocity of the user, $v$, according to

$$f_D = \frac{v f_c}{c} \quad \text{[Hz]}. \tag{2.2}$$

Hence, the coherence time reflects the mobility of the user. The coherence time is short in a high mobility scenario, and vice versa for low mobility. The number of

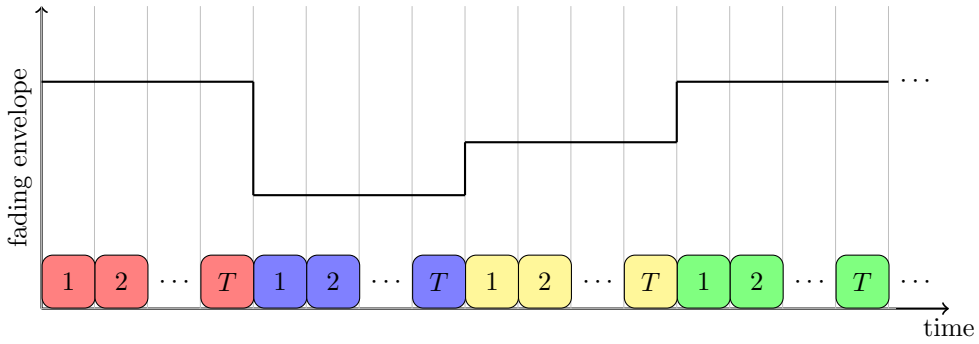



**Figure 2.1:** The random fading coefficients is constant during $T$ symbol periods in a block-fading channel before changing into a new independent random value. All $T$ symbols within a coherence interval experience the same fading.

symbols that can be transmitted during a constant fading interval depends on the coherence time of the channel and on the bandwidth, $B$, of the signal according to

$$T = T_C B, \tag{2.3}$$

since if symbols are transmitted at a higher rate, then more of them can be conveyed during a fixed coherence time.

For example, for a user in high mobility scenario, sitting in a vehicle that travels with a speed of 50 km/h, communicating at 1.9 GHz with a signal bandwidth of 20 kHz, the coherence interval is approximately $T = 100$ symbols. In contrast, for a possible 5G user in a small-cell, that is walking with a speed of 3 km/h communicating at the carrier frequency 30 GHz with a bandwidth of 100 MHz, the coherence interval is roughly $T = 5 \cdot 10^5$ symbols.

If the coherence interval is sufficiently large then the transmitter can send a large number of pilots to acquire channel state information (CSI) at the receiver (CSIR). In this case, it is common to study the performance in the so called perfect CSIR case, where full channel knowledge is assumed to be available for free, i.e. without having to send any pilots. This report mainly considers the case when no *a priori* CSI is available. The transmitter and the receiver are aware of the distribution of the fading but not of its realisation. This assumption is particularly important in a high mobility scenario where the cost of estimating the channel can not be neglected. Furthermore, this assumption is particularly important when low-resolution quantisers are employed at the receiver, since estimating the channel becomes increasingly challenging with coarsely quantised channel outputs.

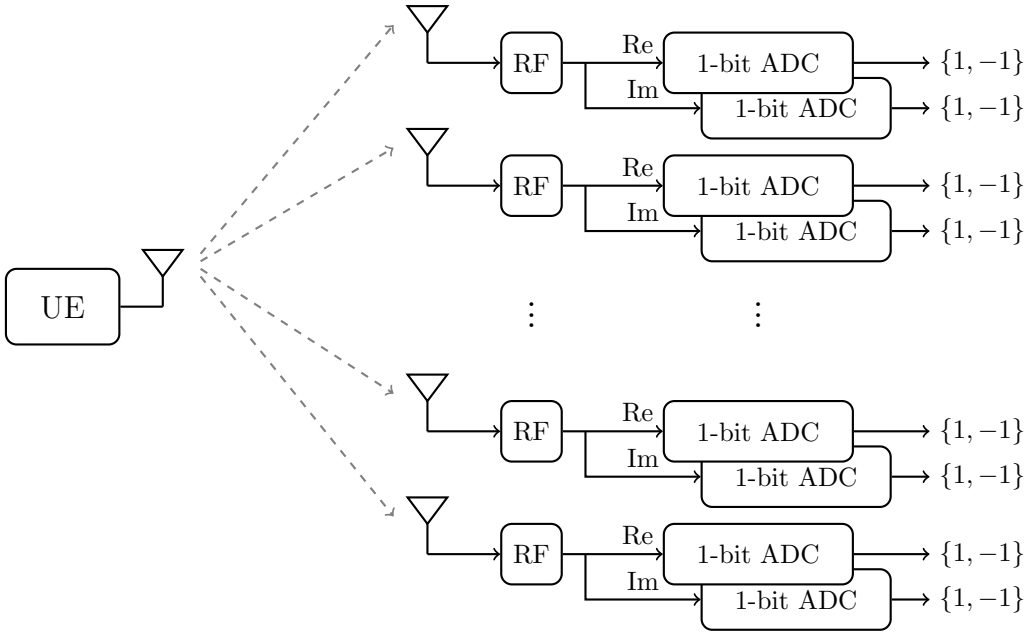



**Figure 2.2:** Single user uplink system model.

## 2.2 System model

Consider the single-cell uplink system with one-bit ADCs shown in Figure 2.2. A single-antenna UE communicates with a BS equipped with an array of $N$ antennas. The discrete-time complex baseband received signal in a coherence interval before quantisation, $\boldsymbol{Y} \in \mathbb{C}^{T \times N}$, is given by

$$\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{h} + \boldsymbol{W} \tag{2.4}$$

where $\boldsymbol{h} \in \mathbb{C}^{1 \times N}$ is the channel vector connecting the single-antenna user to the $N$ receiver antennas, so that $h_n \sim \mathcal{CN}(0,1)$ is the fading coefficient associated with the $n$th BS antenna. Furthermore, $\boldsymbol{W} \in \mathbb{C}^{T \times N}$ denotes the additive white Gaussian noise (AWGN), whose entries are IID zero-mean Gaussian random variables with unit variance, i.e. $w_{t,n} \sim \mathcal{CN}(0,1)$.

The vector $\boldsymbol{x} \in \mathbb{C}^{T \times 1}$ denotes the transmitted signal during a coherence interval so that $x_t$ is the input at time $t \in \{1, \ldots, T\}$. The input is drawn from a constellation subject to the average power constraint

$$\mathbb{E}_{\boldsymbol{x}}\left[||\boldsymbol{x}||^2\right] \leq T\rho \tag{2.5}$$

where $\rho$ denotes the average signal-to-noise ratio (SNR) per antenna. The UE is assumed to be equipped with a sufficiently high-precision digital-to-analog converter (DAC), so that the quantisation of the transmitted signal is ignored.

The real and imaginary components of the received signal at each antenna are quantised separately by an ADC of one-bit precision, so that there are in total $2N$ quantisers required at the BS. As the quantised signal at any antenna is represented by only one bit per dimension, the possible quantisation outcomes are $\mathcal{R} = \{1 + j, -1 + j, -1 - j, 1 - j\}$. Define the joint operation of all $2N$ quantisers as $\mathcal{Q}(\cdot) : \mathbb{C}^{T\times N} \to \mathcal{R}^{T\times N}$ so that the quantised signal can be written as

$$\boldsymbol{R} = \mathcal{Q}(\boldsymbol{Y}) = \mathcal{Q}(\boldsymbol{x}\boldsymbol{h} + \boldsymbol{W}), \tag{2.6}$$

where $\mathcal{Q}(\cdot)$ is defined such that each real and imaginary components of the quantised signal are

$$\Re\{r_{t,n}\} = \text{sign}\{\Re\{y_{t,n}\}\} = \text{sign}\{\Re\{x_t h_n + w_{t,n}\}\} \tag{2.7}$$

$$\Im\{r_{t,n}\} = \text{sign}\{\Im\{y_{t,n}\}\} = \text{sign}\{\Im\{x_t h_n + w_{t,n}\}\} \tag{2.8}$$

where $y_{t,n}$ and $r_{t,n}$ are the received signals at antenna $n$ and time $t$, before and after quantisation. The probability of observing a particular quantised outcome is the same as the probability of observing $y_{t,n}$ in the corresponding quadrant. For example, the probability of the outcome $r_{t,n} = 1 + j$ equals the probability of both the real and imaginary part of $y_{t,n}$ being greater than zero, i.e.,

$$\begin{aligned} P_{r_{t,n}\,|\,x_t,h_n}(1 + j \,|\, x_t, h_n) &= \Pr\{\Re\{x_t h_n + w_{t,n}\} \geq 0\} \\ &\quad \cdot \Pr\{\Im\{x_t h_n + w_{t,n}\} \geq 0\} \\ &= \Pr\{\Re\{w_{t,n}\} \geq -\Re\{x_t h_n\}\} \\ &\quad \cdot \Pr\{\Im\{w_{t,n}\} \geq -\Im\{x_t h_n\}\} \\ &= \Pr\left\{\sqrt{2}\Re\{w_{t,n}\} \geq -\sqrt{2}\Re\{x_t h_n\}\right\} \\ &\quad \cdot \Pr\left\{\sqrt{2}\Im\{w_{t,n}\} \geq -\sqrt{2}\Im\{x_t h_n\}\right\}, \end{aligned} \tag{2.9}$$

where $\sqrt{2}\Re\{w_{t,n}\} \sim \mathcal{N}(0,1)$ and $\sqrt{2}\Im\{w_{t,n}\} \sim \mathcal{N}(0,1)$. This follows because $w_{t,n} \sim \mathcal{CN}(0,1)$. Define

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-v^2/2\right) dv \tag{2.10}$$

as the cumulative distribution function (CDF) of a standard normal random variable. The conditional probability of the outcome $r_{t,n} = 1 + j$ in (2.9) can be written as

$$P_{r_{t,n}\,|\,x_t,h_n}(1 + j \,|\, x_t, h_n) = \Phi\left(\sqrt{2}\Re\{x_t h_n\}\right) \Phi\left(\sqrt{2}\Im\{x_t h_n\}\right). \tag{2.11}$$

Similarly, the conditional probabilities of the remaining outcomes can also be ex-

pressed in terms of the CDF of a standard normal random variable as

$$P_{r_{t,n}\,|\,x_t,h_n}(-1+j\,|\,x_t,h_n) = \Phi\Big(-\sqrt{2}\Re\{x_t h_n\}\Big)\,\Phi\Big(\sqrt{2}\Im\{x_t h_n\}\Big)\,, \tag{2.12}$$

$$P_{r_{t,n}\,|\,x_t,h_n}(-1-j\,|\,x_t,h_n) = \Phi\Big(-\sqrt{2}\Re\{x_t h_n\}\Big)\,\Phi\Big(-\sqrt{2}\Im\{x_t h_n\}\Big)\,, \tag{2.13}$$

$$P_{r_{t,n}\,|\,x_t,h_n}(+1-j\,|\,x_t,h_n) = \Phi\Big(\sqrt{2}\Re\{x_t h_n\}\Big)\,\Phi\Big(-\sqrt{2}\Im\{x_t h_n\}\Big)\,. \tag{2.14}$$

Hence, the probability mass function (PMF) of an outcome can be written as

$$P_{r_{t,n}\,|\,x_t,h_n}(r_{t,n}\,|\,x_t,h_n) = \Phi\Big(\sqrt{2}\Re\{r_{t,n}\}\Re\{x_t h_n\}\Big)\,\Phi\Big(\sqrt{2}\Im\{r_{t,n}\}\Im\{x_t h_n\}\Big)\,. \tag{2.15}$$

It follows from the assumption that the channel is uncorrelated across all antennas that the PMF of the received signal over all antennas at time $t$ is

$$P_{\boldsymbol{r}_t\,|\,x_t,\boldsymbol{h}}(\boldsymbol{r}_t\,|\,x_t,\boldsymbol{h}) = \prod_{n=1}^{N} \Phi\Big(\sqrt{2}\Re\{r_{t,n}\}\Re\{x_t h_n\}\Big)\,\Phi\Big(\sqrt{2}\Im\{r_{t,n}\}\Im\{x_t h_n\}\Big)\,, \tag{2.16}$$

where $\boldsymbol{r}_t = x_t\boldsymbol{h} + \boldsymbol{w}_t$ is the $t$th row of $\boldsymbol{R}$. Furthermore, as the noise is IID, it follows that the PMF of the received signal during the entire coherence interval can be written as

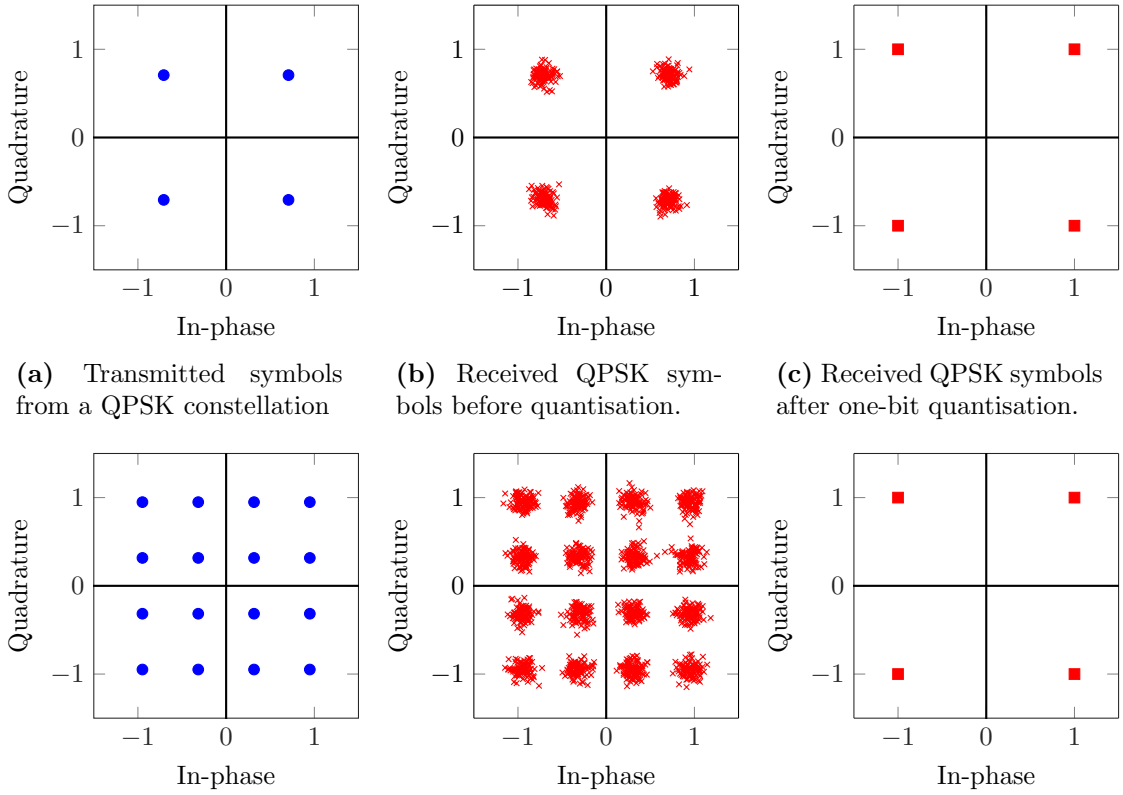
$$P_{\boldsymbol{R}\,|\,\boldsymbol{x},\boldsymbol{h}}(\boldsymbol{R}\,|\,\boldsymbol{x},\boldsymbol{h}) = \prod_{n=1}^{N}\prod_{t=1}^{T} \Phi\Big(\sqrt{2}\Re\{r_{t,n}\}\Re\{x_t h_n\}\Big)\,\Phi\Big(\sqrt{2}\Im\{r_{t,n}\}\Im\{x_t h_n\}\Big)\,. \tag{2.17}$$

The probabilistic relation between the input and output, called the channel law, is obtained from (2.17) by the law of total probability, i.e., by averaging out the channel coefficients $\boldsymbol{h}$:

$$\begin{aligned} P_{\boldsymbol{R}\,|\,\boldsymbol{x}}(\boldsymbol{R}\,|\,\boldsymbol{x}) &= \mathbb{E}_{\boldsymbol{h}}\big[P_{\boldsymbol{R}\,|\,\boldsymbol{x},\boldsymbol{h}}(\boldsymbol{R}\,|\,\boldsymbol{x},\boldsymbol{h})\big] \\ &= \mathbb{E}_{\boldsymbol{h}}\left[\prod_{n=1}^{N}\prod_{t=1}^{T} \Phi\Big(\sqrt{2}\Re\{r_{t,n}\}\Re\{x_t h_n\}\Big)\,\Phi\Big(\sqrt{2}\Im\{r_{t,n}\}\Im\{x_t h_n\}\Big)\right]. \end{aligned} \tag{2.18}$$

The channel law in (2.18) and the conditional PMF in (2.16) is key when computing the achievable rates for different signalling schemes in the no *a priori* CSI case and perfect CSIR case, respectively. As there is no closed-form expression for the expectation over $\boldsymbol{h}$ in (2.18), this expectation is computed by averaging over a large number of random channel realisations, i.e., by Monte Carlo methods.

To gain further understanding in the operation of the one-bit quantiser, refer to Figure 2.3, which depicts the transmitted and the received signal at an arbitrary antenna, before and after quantisation for QPSK and 16QAM inputs. Independently




**(a)** Transmitted symbols from a QPSK constellation

**(b)** Received QPSK symbols before quantisation.

**(c)** Received QPSK symbols after one-bit quantisation.

**(d)** Transmitted symbols from a 16QAM constellation.

**(e)** Received 16QAM symbols before quantisation.

**(f)** Received 16QAM symbols after one-bit quantisation.

**Figure 2.3:** Input signals from a QPSK or 16QAM constellation along with the received signal before and after quantisation at any antenna. The SNR is $\rho = 10$ dB, for several random noise realisations and a unit gain fading coefficient.

of the input distribution, the received signal $y_{t,n}$ at a given antenna $n$ and channel use $t$, is mapped to the ADC outcome $r_{t,n}$ in the corresponding quadrant. Consequently, all information about the amplitude is lost in the quantisation and all that can be inferred from the quantised signal is in which quadrant the transmitted signal ended up in after being affected by fading and noise. Thus, as shown in Figure 2.3, any high-order constellation collapses to only four points.

## 2.3 Channel capacity

Consider the discrete-time system model in Figure 2.4. The source emits a uniformly distributed random message $m$ that is passed as input to an encoder to produce a

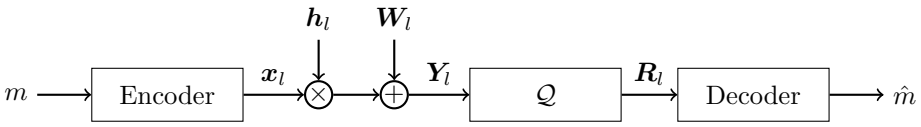



**Figure 2.4:** Single-user system model including encoding and decoding

length-$L$ sequence $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L\}$ of channel inputs, where $\boldsymbol{x}_l$ is the input in the $l$th coherence interval. The channel corrupts this sequence through the fading processes and the additive noise, as explained in Section 2.2. The received quantised sequence $\{\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_L\}$ is passed to a decoder which makes an estimate, $\hat{m}$, about the message that was transmitted.

Channel coding is performed over $L$ coherence intervals of duration $T$ symbols each, so that the length of a codeword is $LT$ symbols. By coding over multiple fading realisations, the intervals of favourable fading compensate for those of unfavourable fading. According to Shannon's channel coding theorem [26], it is possible to construct a code operating at an achievable rate, $R$, and communicate reliably, so that the error probability $\Pr(\hat{m} \neq m)$ tends to zero as the length of the code goes to infinity (so that the code experiences infinitely many channel realisations). The highest possible rate that achieves reliable communication is defined as the channel capacity.

Assume that the transmitted symbol at any time is drawn from a constellation $\mathcal{X}$, such that the transmitted signal during the entire coherence interval is $\boldsymbol{x} \in \mathcal{X}^{T\times 1}$. If the inputs are distributed according to $P_{\boldsymbol{x}}(\boldsymbol{x})$ such that the average power constraint in (2.5) is satisfied, then the achievable rate is

$$R(\rho) = \frac{1}{T}\, \mathcal{I}(\boldsymbol{x}; \boldsymbol{R}) \tag{2.19}$$

where $\mathcal{I}(\boldsymbol{x}; \boldsymbol{R})$ is the mutual information between the channel input $\boldsymbol{x} \in \mathcal{X}^{T\times 1}$ and the channel output $\boldsymbol{R} \in \mathcal{R}^{T\times N}$ given by

$$\mathcal{I}(\boldsymbol{x}; \boldsymbol{R}) = \sum_{\boldsymbol{x}\in\mathcal{X}^{T\times 1}} \sum_{\boldsymbol{R}\in\mathcal{R}^{T\times N}} P_{\boldsymbol{x}}(\boldsymbol{x})\; P_{\boldsymbol{R}\,|\,\boldsymbol{x}}(\boldsymbol{R}\,|\,\boldsymbol{x})\; \log_2 \frac{P_{\boldsymbol{R}\,|\,\boldsymbol{x}}(\boldsymbol{R}\,|\,\boldsymbol{x})}{P_{\boldsymbol{R}}(\boldsymbol{R})}, \tag{2.20}$$

where

$$P_{\boldsymbol{R}}(\boldsymbol{R}) = \mathbb{E}_{\boldsymbol{x}}\big[P_{\boldsymbol{R}\,|\,\boldsymbol{x}}(\boldsymbol{R}\,|\,\boldsymbol{x})\big] = \sum_{\boldsymbol{x}\in\mathcal{X}^{T\times 1}} P_{\boldsymbol{x}}(\boldsymbol{x})\; P_{\boldsymbol{R}\,|\,\boldsymbol{x}}(\boldsymbol{R}\,|\,\boldsymbol{x}) \tag{2.21}$$

is the output probability of the received quantised signal during a coherence interval. The normalisation by $T$ in (2.19) ensures that the rate is expressed in bits per channel use, or equivalently, in bits/s/Hz. In this thesis, the throughput of a system is measured by the achievable rate. Hence, the terms throughput and achievable rate are used interchangeably.

The channel capacity is the supremum of the achievable rates over all input distributions that satisfy the average power constraint in (2.5), i.e.,

$$C(\rho) = \sup_{p_{\boldsymbol{x}}(\boldsymbol{x})} \frac{1}{T}\, \mathcal{I}(\boldsymbol{x}; \boldsymbol{R}). \tag{2.22}$$

Note that the capacity-achieving distribution does not have to be discrete. In such a case, the sum over all inputs in (2.20) needs to be replaced with an integral over the support of the probability density function (PDF) $p_{\boldsymbol{x}}(\boldsymbol{x})$. Also note that by assuming a particular signalling scheme, where the input is chosen from a certain constellation (e.g. QPSK or 16QAM), and computing the achievable rate, a lower bound on the channel capacity is obtained because of (2.22).

If the channel fading realisation is somehow known perfectly at the receiver then the capacity does not depend on $T$, as there is no cost involved in learning the channel. The capacity under the assumption of perfect CSIR is

$$C_{\text{CSIR}}(\rho) = \sup_{p_{x_t}(x_t)} \mathcal{I}(x_t; \boldsymbol{r}_t \,|\, \boldsymbol{h}), \tag{2.23}$$

where $\mathcal{I}(x_t; \boldsymbol{r}_t \,|\, \boldsymbol{h})$ is the conditional mutual information between the channel input and output when $\boldsymbol{h}$ is known. If $x_t \in \mathcal{X}$ is a discrete random variable, then

$$\mathcal{I}(x_t; \boldsymbol{r}_t|\boldsymbol{h}) = \mathbb{E}_{\boldsymbol{h}} \left[ \sum_{x_t \in \mathcal{X}} \sum_{\boldsymbol{r}_t \in \mathcal{R}^{N\times 1}} P_{x_t}(x_t) P_{\boldsymbol{r}_t|x_t,\boldsymbol{h}}(\boldsymbol{r}_t|x_t,\boldsymbol{h}) \log_2 \frac{P_{\boldsymbol{r}_t|x_t,\boldsymbol{h}}(\boldsymbol{r}_t|x_t,\boldsymbol{h})}{P_{\boldsymbol{r}_t|\boldsymbol{h}}(\boldsymbol{r}_t|\boldsymbol{h})} \right]. \tag{2.24}$$

Since conditioning reduces entropy (information can not hurt) and since $\boldsymbol{x}$ and $\boldsymbol{h}$ are independent, the capacity with perfect CSIR is an upper bound on the capacity when no *a priori* CSI is available. The upper bound, when the coherence interval increases, is generally tight in the case of no quantiser (hereafter referred to as the *infinite precision* case), as knowledge about the channel can be learned by transmitting a large amount of training symbols. Whether this is true also in the case of one-bit quantisers, will be investigated for of a single-antenna receiver in Section 2.3.1.

### 2.3.1 Capacity in the single receive antenna case

With a single antenna at the receiver, (2.6) is reduced to a single-input single-output (SISO) channel that can be written as

$$\boldsymbol{r} = \mathcal{Q}(\boldsymbol{y}) = \mathcal{Q}(\boldsymbol{x}h + \boldsymbol{w}), \tag{2.25}$$

where $\boldsymbol{y}$ and $\boldsymbol{r}$ is the received signal on the single antenna, before and after quantisation, during the coherence interval.

### Channel state information at the receiver

In the perfect CSIR case, the capacity is achieved with QPSK signalling [27]. The capacity can be computed from (2.24) with QPSK inputs as

$$C_{\text{CSIR}}(\rho) = 2 - \mathbb{E}_h\Big[\mathcal{H}\Big(\Phi\Big(\Re\{h\}\sqrt{2\rho}\Big)\Big) + \mathcal{H}\Big(\Phi\Big(\Im\{h\}\sqrt{2\rho}\Big)\Big)\Big] \tag{2.26}$$

where

$$\mathcal{H}(\epsilon) \triangleq -\epsilon \log_2(\epsilon) - (1-\epsilon)\log_2(1-\epsilon) \tag{2.27}$$

denotes the binary entropy function. As there are only four possible outcomes at any time in the single-antenna case, the capacity is obviously upper bounded by 2 bits per channel use, which is achieved when the SNR tends to infinity. In [21] it is shown that for one-bit quantisation the capacity in the low SNR regime is reduced by a factor $2/\pi$ (corresponding to a power penalty of approximately 2 dB) compared to the infinite precision case. It is also shown that this loss can be fully recovered by allowing for asymmetric quantisers (non-zero quantisation thresholds).

When the effect of the quantiser is ignored (in the infinite precision case) the channel capacity with perfect CSIR is given by the well known formula [28]:

$$C_{\text{CSIR}}^{\text{inf}}(\rho) = \mathbb{E}_h\big[\log(1 + |h|^2\,\rho)\big]\,. \tag{2.28}$$

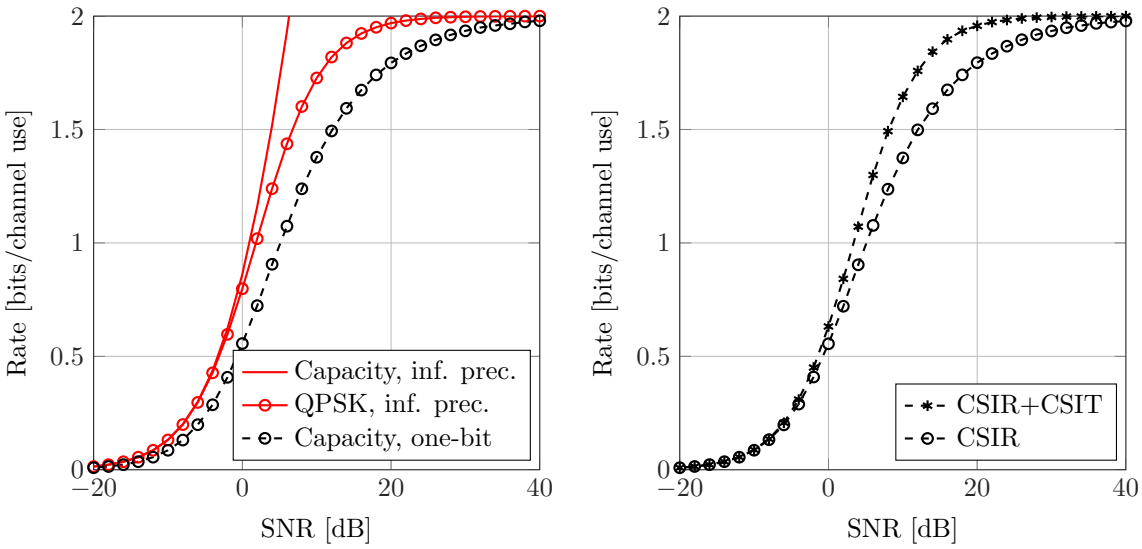
The capacity in the perfect CSIR case is plotted in Figure 2.5a for one-bit quantisation and with infinite precision case. For low SNR, the capacity in the quantised case suffers the power loss of 2 dB. For high SNR, the quantised capacity saturates at 2 bits per channel use, while the infinite precision capacity grows unrestrained. Also shown is the rate with QPSK in the infinite precision case.

### Channel state information at the transmitter

In the case when also perfect CSI at the transmitter (CSIT) is available, the capacity (see capacity definition with CSIT in, e.g., [28]) is achieved by transmitting rotated QPSK symbols [18]. The result is quite intuitive as if the channel is known to the transmitter, then it can be accounted for, by sending a rotated version of the prospected QPSK symbol, so that the channel rotates it back to its default phase. The capacity in this case is given by

$$C_{\text{CSIT}}(\rho) = 2 - 2\,\mathbb{E}_h[\mathcal{H}\left(\Phi(|h|\,\sqrt{\rho})\right)]\,. \tag{2.29}$$

The capacity when also perfect CSIT is available, is shown in Figure 2.5b. A slightly higher rate is obtained (again as additional information can not hurt). In the low SNR case, however, there is no significant improvement compared to only perfect CSIR as the additive noise is the dominating impairment.




**(a)** Comparison between infinite precision capacity and one-bit quantised capacity with CSIR. Also shown is the QPSK rate for infinite precision.

**(b)** Comparison between one-bit quantised capacity with CSIR and with/without CSIT.

**Figure 2.5:** Capacity comparison in the SISO case.

### No *a priori* channel state information

The no *a priori* CSI case is studied in [19, 20] where it is proven that on-off QPSK is capacity achieving. During each coherence interval, the transmitter sends QPSK symbols with probability $\eta$ or remains silent with probability $1-\eta$ to save power. Hence, symbols in other coherence intervals can be transmitted with higher power while still satisfying the average power constraint. The duty-cycle $\eta$ is given by

$$\eta = \begin{cases} \dfrac{\rho}{\rho_c}, & \rho \leq \rho_c \\ 1, & \rho > \rho_c \end{cases} \tag{2.30}$$

where $\rho_c$ is the threshold SNR after which standard QPSK becomes optimal. The capacity in the no *a priori* CSI case is shown to be

$$C(\rho) = \begin{cases} \dfrac{\rho}{\rho_c} R_{\text{QPSK}}(\rho_c), & \rho \leq \rho_c \\ R_{\text{QPSK}}(\rho), & \rho > \rho_c \end{cases} \tag{2.31}$$

where $R_{\text{QPSK}}(\rho)$ is the achievable rate with QPSK. The critical SNR is found by solving the following optimisation,

$$\rho_c = \arg\max_{\rho} \ \frac{1}{\rho} R_{\text{QPSK}}(\rho) \quad \rho \geq 0. \tag{2.32}$$

In general, as the coherence time increases, the threshold SNR decreases so that the benefit of using on-off QPSK instead of standard QPSK becomes less apparent. In Appendix A, it is shown that the achievable rate with QPSK symbols can be computed as

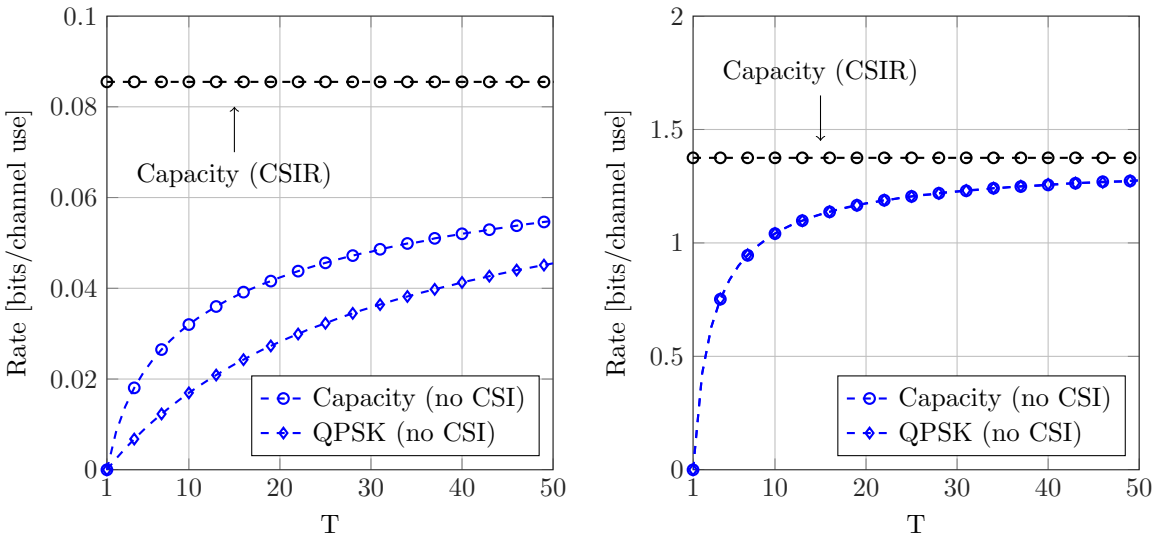
$$\begin{aligned} R_{\text{QPSK}}(\rho) &= \frac{2}{T} \sum_{k=0}^{T} \binom{T}{k} \gamma(k,\, T-k) \log_2 \Big( 2^T \gamma(k,\, T-k) \Big) \\ &= 2 + \frac{2}{T} \sum_{k=0}^{T} \binom{T}{k} \gamma(k,\, T-k) \log_2 \Big( \gamma(k,\, T-k) \Big), \end{aligned} \tag{2.33}$$

where

$$\gamma(\alpha,\, \beta) \triangleq \mathbb{E}_g \Big[ \Phi\left(-g\sqrt{\rho}\right)^{\alpha} \Phi\left(g\sqrt{\rho}\right)^{\beta} \Big], \tag{2.34}$$

and where $g \sim \mathcal{N}(0,1)$. Thus, instead of having to evaluate all input and output combinations in (2.20), the achievable rate with QPSK signalling can be computed using (2.33) with a complexity that scales linearly with $T$. The no *a priori* CSI capacity is subsequently found by solving the optimisation problem in (2.32) to get the capacity in (2.31). The capacity as a function of $T$ is plotted in Figure 2.6a for $\rho = -10$ dB and in Figure 2.6b for $\rho = 10$ dB. The achievable rate with standard QPSK signalling, i.e., without optimising the power is shown as well. Finally, the figure shows also the perfect CSIR capacity, which is independent of $T$.

As $T$ increases, the capacity in the no *a priori* CSI case approaches the perfect CSIR capacity. As SNR increases, the convergence is faster. This is expected as with higher SNR, channel estimation becomes easier, as less noise is acting on pilots. Furthermore, for high SNR, transmitting QPSK symbols without optimising the power achieves the capacity. For low SNR this is not the case. Thus, standard QPSK (no power optimisation) is capacity-achieving for high SNR. Note that for $T = 1$ the capacity is zero. Indeed, the receiver can not infer which symbol is transmitted, since the fading could have shifted it to any quadrant.




**(a)** $\rho = -10$ dB. **(b)** $\rho = 10$ dB.

**Figure 2.6:** Capacity with and without *a priori* CSI for low and high SNR as a function of the coherence interval $T$.

### 2.3.2 Capacity in the multiple-receive-antenna case

Consider again the single-input multiple-output (SIMO) system model in (2.6).

#### Channel state information at the receiver

In the infinite precision case when perfect CSIR is available, the capacity of the SIMO channel is

$$C_{\text{CSIR}}^{\text{inf}}(\rho) = \mathbb{E}_{\boldsymbol{h}}\left[\log_2 \det\left(\boldsymbol{I}_N + \rho \boldsymbol{h}^H \boldsymbol{h}\right)\right]. \tag{2.35}$$

In the case of one-bit quantisation and perfect CSIR, the capacity-achieving distribution is unknown. In [17], it is shown that QPSK is optimal in the low-SNR regime and that the power penalty of 2 dB compared to the infinite precision case carries through. However, QPSK is not optimal for high SNR.

In Figure 2.7, the achievable rate with QPSK and 16QAM for a quantised SIMO system with 5 receive antennas is compared with the corresponding rates in the infinite precision case. Also shown is the capacity in the infinite precision case computed from (2.35). The achievable rates with QPSK and 16QAM is computed by considering all inputs and outputs in (2.24) and averaging the mutual information with respect to a large number of channel realisations.

As shown in the figure, 16QAM outperforms QPSK in the high-SNR regime. Thus, in contrast to the single-antenna case, 16QAM symbols can successfully be

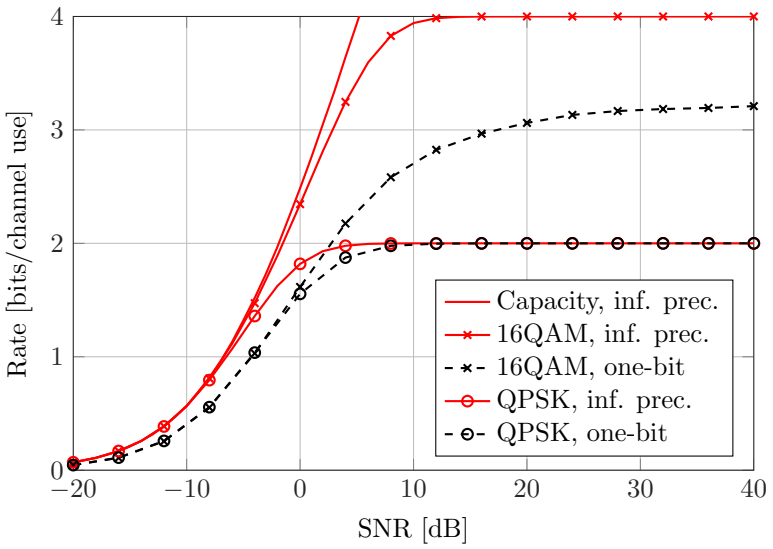



**Figure 2.7:** Infinite precision capacity and achievable rates with QPSK and 16QAM for infinite precision and one-bit quantisation. Perfect CSIR is available in a $1 \times 5$ SIMO system.

distinguished to achieve a rate higher than 2 bits per channel use, which suggests that amplitude information in the input distribution can be retained when more receive antennas are added.

### Channel state information at the transmitter

The case when also CSIT is available is studied in [18] where it is shown that the high-SNR capacity grows logarithmically with the number of receive antennas. The optimal input distribution for a given SNR and a given number of antennas is found by a cutting plane algorithm [29].

### No *a priori* channel state information

The no *a priori* CSI case is studied in [19], in which some properties of the mutual information are stated. However, the capacity-achieving distribution is unknown. Furthermore, computing the achievable rates becomes quickly computationally demanding as the number of possible outputs in (2.20) grows exponentially with both $T$ and $N$. This problem is cleverly avoided in [24] which considers a massive MIMO system where the received signal on all antennas is mapped to a scalar using a linear receive filter. Such a receiver requires some channel knowledge, which can be obtained by channel estimation in a training based scheme.

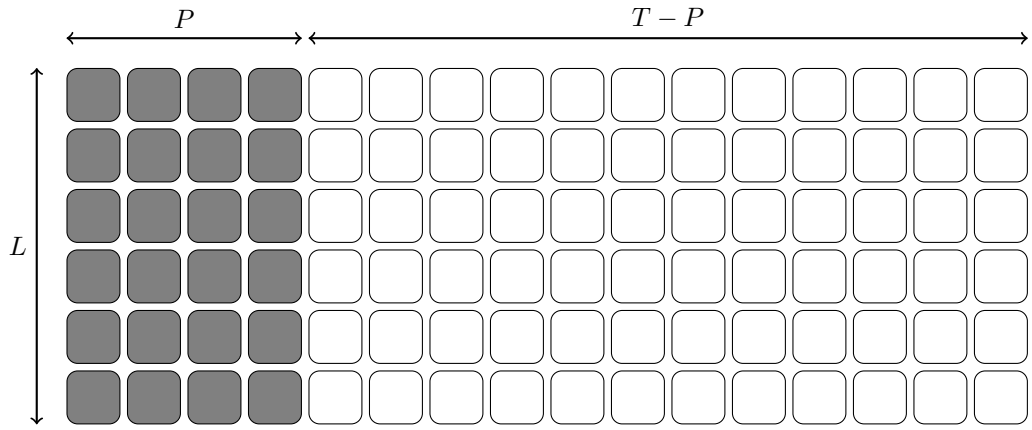



**Figure 2.8:** A codeword spanning $L$ coherence intervals of $T$ symbols each in a training based communication scheme. The first $P = 4$ time slots in each coherence interval are in this illustration reserved for pilot symbols (grey blocks) while the remaining time slots are used to convey data symbols (white blocks).

## 2.4 Training based channel estimation

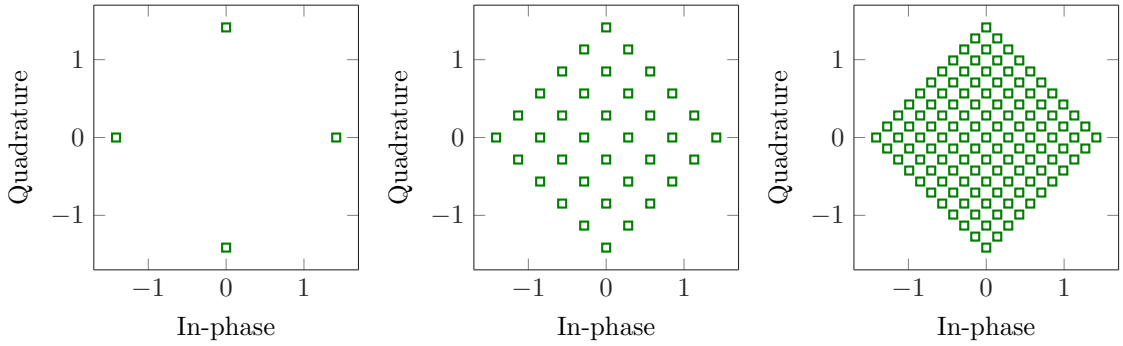
When the channel is not known *a priori* at the receiver, CSI can be acquired by dedicating a part of the coherence interval to transmitting a training sequence (pilots), which are used to estimate the channel. Assume that the first $P$ time slots in each coherence interval are reserved for pilots, as illustrated by Figure 2.8. Let $\boldsymbol{x}^{(p)} = [x_1, \ldots, x_P]^T$ be the corresponding pilot vector (known to the receiver). Furthermore, let $\boldsymbol{x}^{(d)} = [x_{P+1}\ , \ldots,\ x_T]^T$ be the data vector (unknown to the receiver) containing the remaining entries of $\boldsymbol{x}$. The resulting 1-bit quantised received signals for the pilot and data sequences in the SISO case are denoted $\boldsymbol{r}^{(p)}$ and $\boldsymbol{r}^{(d)}$ respectively.

There are many possible techniques that can be used to perform channel estimation. A relatively simple way to estimate the channel at the receiver on the basis of the pilot symbols is to use least squares (LS) estimation which is explained below for the single-antenna case.

### 2.4.1 Least squares channel estimation

LS estimation has the advantage of requiring little signal processing compared to many other methods. This makes it favourable in a massive MIMO setting. The LS approach to produce an estimate $\hat{h}$ of the channel is to minimise the squared error between the received pilots and the assumed noiseless data, i.e.,

$$\hat{h} = \arg\min_{h} ||\boldsymbol{r}^{(p)} - \boldsymbol{x}^{(p)}h||^2. \tag{2.36}$$




**(a)** 1 pilot. **(b)** 5 pilots. **(c)** 10 pilots.

**Figure 2.9:** Set of possible LS estimation outcomes for a training sequence with varying length consisting of QPSK pilots.

The solution is found by setting the derivative of the norm on the right-hand side of (2.36) to zero, which yields

$$\hat{h} = \left( \left( \boldsymbol{x}^{(p)} \right)^H \boldsymbol{x}^{(p)} \right)^{-1} \left( \boldsymbol{x}^{(p)} \right)^H \boldsymbol{r}^{(p)}. \tag{2.37}$$

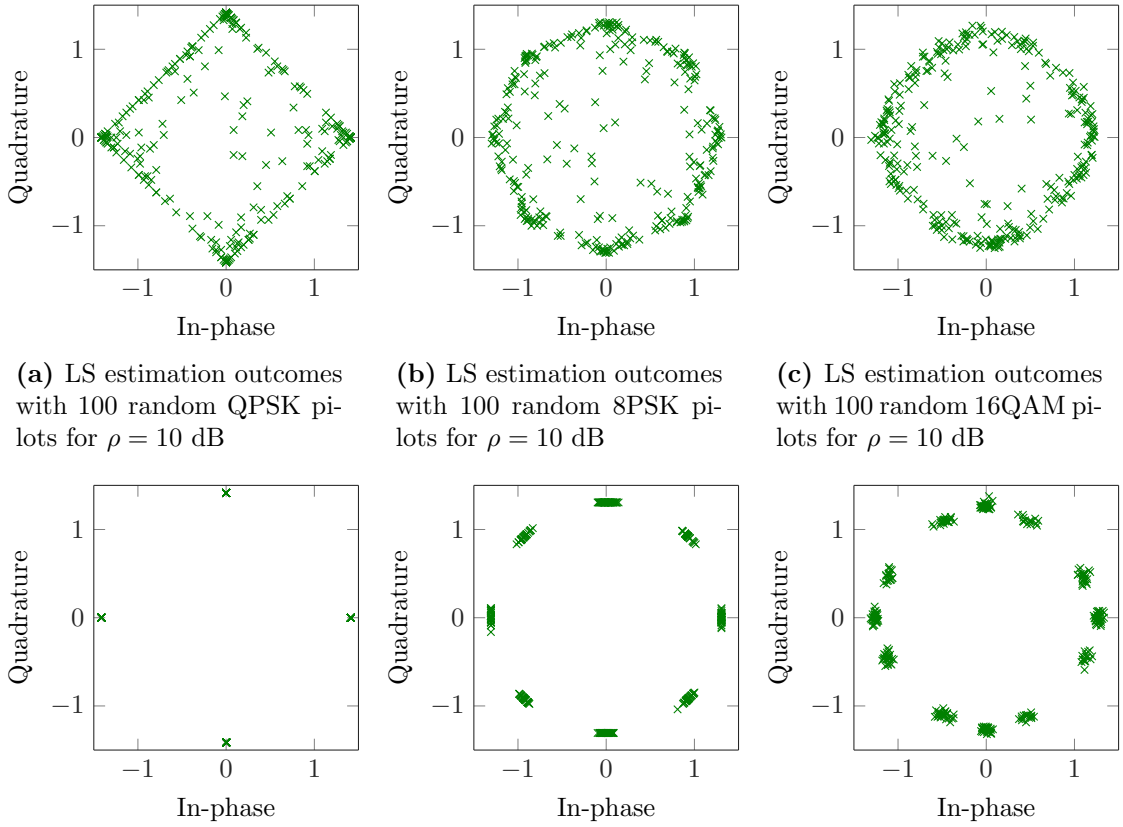
Finally, if all pilots are transmitted with the same power, which is the case for QPSK symbols, then the LS estimate reduces to

$$\hat{h} = \frac{1}{P\sqrt{\rho}} \sum_{i=1}^{P} \left( x_i^{(p)} \right)^H r_i^{(p)}. \tag{2.38}$$

If the inputs $\{x_i^{(p)}\}$ belong to a finite cardinality alphabet (such as QPSK and 16QAM), then so will the set of possible LS estimation outcomes since, there are only finitely many different received signals. In Appendix B it is shown that, with QPSK pilots, the LS estimator can return $(P+1)^2$ possible estimation outcomes, distributed on a rectangular grid as shown in Figure 2.9 for different lengths of the training sequence.

From the figure, it becomes apparent that no matter how many pilots are transmitted, the channel can never be properly estimated with this method, as there are an infinite number of possible fading realisations outside the discrete grid. Thus, the assumption of perfect CSIR appears unreasonably optimistic when one-bit quantisers are used.

To understand what information about the fading channel that can actually be learned using LS estimation, refer to Figure 2.10, which displays some LS estimation outcomes. A training sequence consisting of 100 random pilots drawn from various constellations have been used to estimate the channel. As SNR increases, the number




**(a)** LS estimation outcomes with 100 random QPSK pilots for $\rho = 10$ dB

**(b)** LS estimation outcomes with 100 random 8PSK pilots for $\rho = 10$ dB

**(c)** LS estimation outcomes with 100 random 16QAM pilots for $\rho = 10$ dB

**(d)** LS estimation outcomes with 100 random QPSK pilots for $\rho = 100$ dB

**(e)** LS estimation outcomes with 100 random 8PSK pilots for $\rho = 100$ dB

**(f)** LS estimation outcomes with 100 random 16QAM pilots for $\rho = 100$ dB

**Figure 2.10:** LS estimation outcomes using 100 random pilots.

of estimation outcomes decreases. For example, with QPSK pilots in the high-SNR regime, there are only four distinct outcomes. This means that the phase of the fading channel can be estimated with 90° precision. Indeed, when the phase of the channel is larger than −45° but less than 45°, no matter which QPSK symbol was transmitted, the quantised output will appear in the same quadrant. Therefore, the best guess is that the phase of the fading coefficient is 0°, which also is the outcome of (2.38). Similarly, when the phase of the fading is larger than 45° but smaller than 135°, the resulting channel estimate from (2.38) is 90°. Obviously, in the presence of noise, there are more possible outcomes because also the noise causes phase rotations.

In general, using high-order constellations, the phase of the fading channel can be estimated, in the high-SNR regime, with a granularity that equals the phase difference between the pilot symbols. However, the amplitude of the fading channel can not be estimated using the LS approach. Thus, all that can be learned about the one-bit

quantised fading channel from the LS estimate, is (to some extent) the phase of the fading channel.

### 2.4.2 Achievable rate with training for a single antenna

Using training and LS channel estimation with $P$ pilots, the achievable rate is

$$\frac{T-P}{T}\,\mathcal{I}(x_t; r_t \,|\, \hat{h}) \tag{2.39}$$

where $\hat{h}$ is the LS estimate from (2.37) using $P$ pilots, and where $x_t \in \mathcal{X}$ and $r_t \in \mathcal{R}$ are the transmitted and received signals at time $t \in \{P+1,\ \ldots,\ T\}$. The scaling by $T-P$ reflects the cost of sending $P$ pilots to estimate the channel (only $T-P$ time-slots within the coherence interval are available for data transmission). The normalisation by $T$ in (2.39) is again necessary for the rate to be measured in bits per channel use.

In Appendix B, it is shown that the rate with training and LS estimation in (2.39), assuming QPSK inputs, can be expressed as

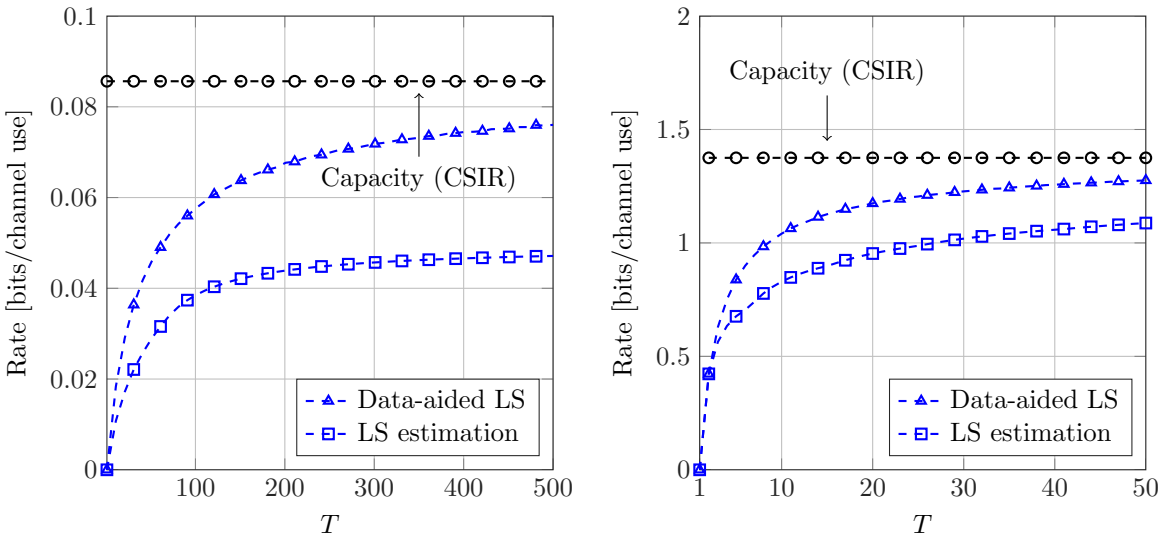
$$R^{\mathrm{LS}}_{\mathrm{QPSK}}(\rho) = 2\left(\frac{T-P}{T}\right)\left(1 - \sum_{l=0}^{P}\binom{P}{l}\gamma(l,\,P-l)\;\mathcal{H}\left(\frac{\gamma(l+1,\,P-l)}{\gamma(l,\,P-l)}\right)\right). \tag{2.40}$$

The achievable rate assuming QPSK and LS estimation is a lower bound on the achievable rate with QPSK. However, as stated by the following lemma, LS estimation is the optimal strategy when $T = 2$.

**Lemma 1.** *Training and LS channel estimation with one pilot and one data symbol achieves the QPSK rate in (2.33) for $T = 2$.*

*Proof.* See Appendix B. □

Encouraged by this result, a data-aided version of LS estimation is considered, which can be explained as follows. Assume that a code spanning a large number of coherence intervals is constructed. Furthermore, assume that the first slot in each coherence interval is used to transmit a pilot. If the first two symbols (pilot and data) are encoded with a rate lower than $\mathcal{I}(x_2; r_2|\hat{h})$, where $\hat{h}$ is the estimate based on the pilot, and if the length of the code goes to infinity, then $x_2$ can be decoded correctly. Proceed now to code the third symbol in each coherence interval with a rate less than $\mathcal{I}(x_3; r_3|\hat{h})$, where $\hat{h}$ is the estimate based on the pilot and the first data symbol. Once more, $x_3$ can be decoded correctly as the code length goes to infinity, and can be treated as a pilot to refine the estimate of $h$. By repeating this procedure for all $T$ symbols in the coherence interval, the achievable rate with QPSK can be achieved, as stated in the following theorem




**(a)** $\rho = -10$ dB. **(b)** $\rho = 10$ dB.

**Figure 2.11:** Achievable rate with LS estimation and data-aided LS estimation, for low and high SNR as the duration of the coherence interval varies.

**Theorem 2.** *Training and data-aided LS estimation with only one pilot achieves the QPSK rate in (2.33) for all $T$.*

*Proof.* see Appendix B. □

Theorem 2 states that the QPSK rate can be achieved with a training-based scheme. As the QPSK rate in the no *a priori* CSI case is capacity achieving for high SNR, it can be concluded that data-aided LS estimation is capacity-achieving in the high-SNR regime. For low SNR, the power optimisation in (2.31) has to be applied.

The achievable rate with LS estimation and data-aided LS estimation is plotted in Figure 2.11 for $\rho = -10$ dB and $\rho = 10$ dB. For the achievable rate with LS estimation (non data-aided) the number of pilots is optimised for all $T$. Note that the data-aided LS rate approaches the perfect CSIR rate both at low and high SNR. Hence, in spite of the fact that only limited channel knowledge about the phase of the fading can be acquired through LS estimation, the perfect CSIR rate can still be approached. However, a longer coherence interval is required for the achievable rate with data-aided LS estimation to converge in the low-SNR regime.

More pilots are required at lower SNR as the fading coefficients have to be estimated based on a noisy channel. Hence, there is a large rate degradation by using standard LS estimation compared to data-aided LS estimation (which only uses one pilot). As SNR increases, the gap decreases. Note that for $T = 2$, data-aided LS and standard LS estimation yields the same rate.

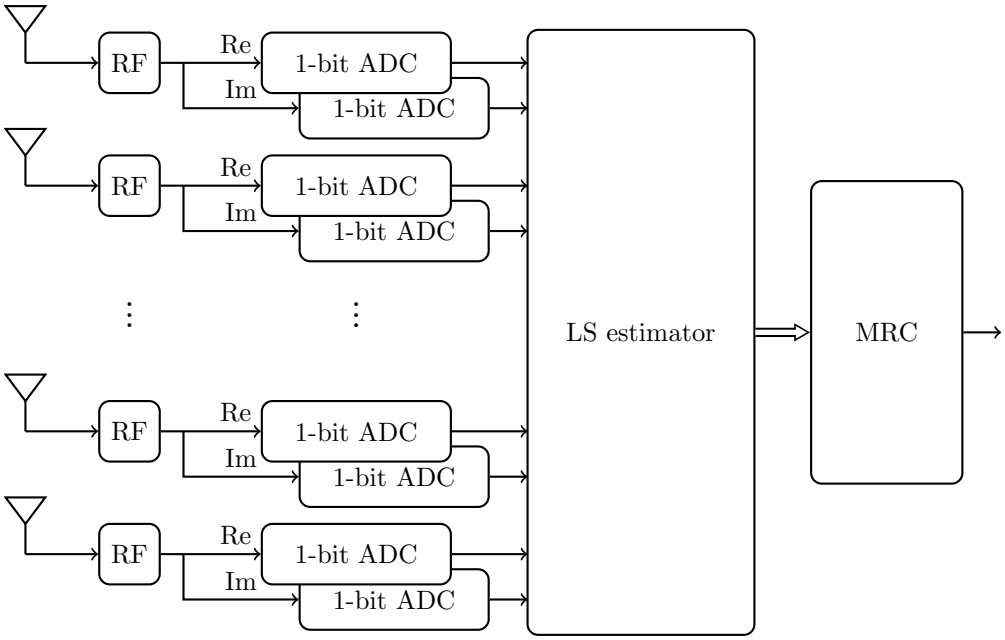



**Figure 2.12:** Single-user BS receiver with a large antenna array and MRC. The transmitted message from a single user is combined using MRC to obtain a soft estimate of the transmitted signal.

### 2.4.3 Maximal-ratio combining

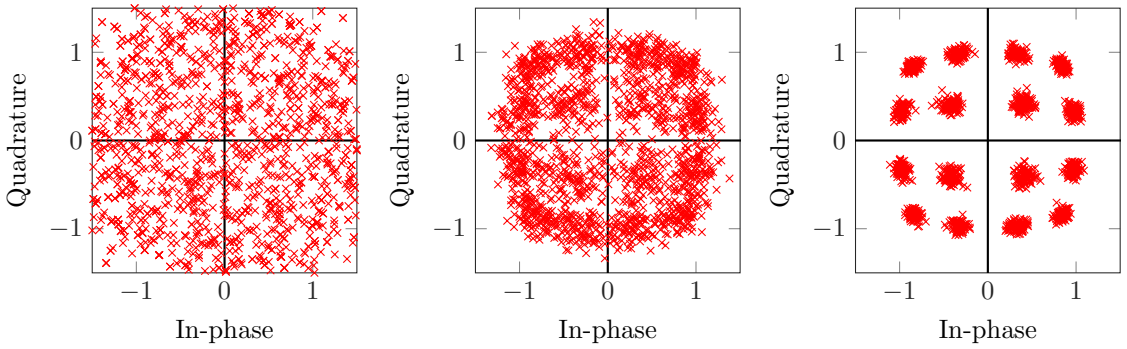
Finally, consider the case when the number of receive antennas grows large. Channel estimation has to be performed at every antenna, as the fading coefficient is different for each path. Hence, the required amount of signal processing scales with the array size. In the multiple-antenna case, the LS estimate, $\hat{\boldsymbol{h}} \in \mathbb{C}^{1\times N}$, can be written as

$$\hat{\boldsymbol{h}} = \left(\left(\boldsymbol{x}^{(p)}\right)^H \boldsymbol{x}^{(p)}\right)^{-1} \left(\boldsymbol{x}^{(p)}\right)^H \boldsymbol{R}^{(p)}, \tag{2.41}$$

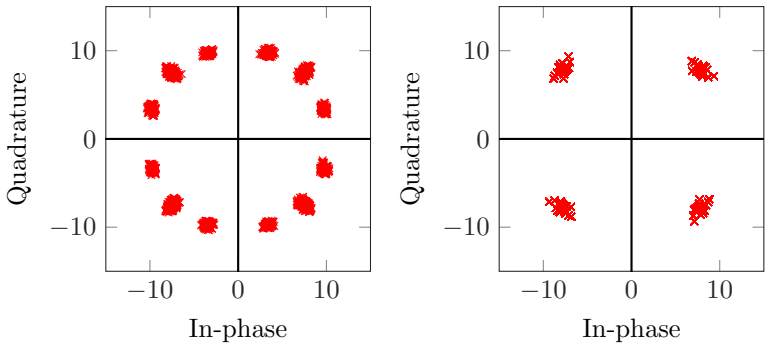
where $\boldsymbol{R}^{(p)} \in \mathcal{R}^{P\times N}$ is the received pilot sequence at all antennas.

Increasing the number of antennas enables mitigation of fading and noise by diversity combining. The idea behind diversity combing is to send the same message over several independent channel realisations and combine them in such a way that the effect of the fading is reduced. A common method to achieve diversity is by maximal-ratio combining (MRC), which produces a soft estimate of the transmitted signal at time $t \in \{P+1,\ \ldots,\ T\}$, denoted $\tilde{x}_t$, that is the weighted sum of the received signal on all branches, i.e.,

$$\tilde{x}_t = \frac{\boldsymbol{r}_t \hat{\boldsymbol{h}}^H}{||\hat{\boldsymbol{h}}||^2}. \tag{2.42}$$




**(a)** 4 antennas, $\rho = 0$ dB. **(b)** 40 antennas, $\rho = 0$ dB. **(c)** 400 antennas, $\rho = 0$ dB.

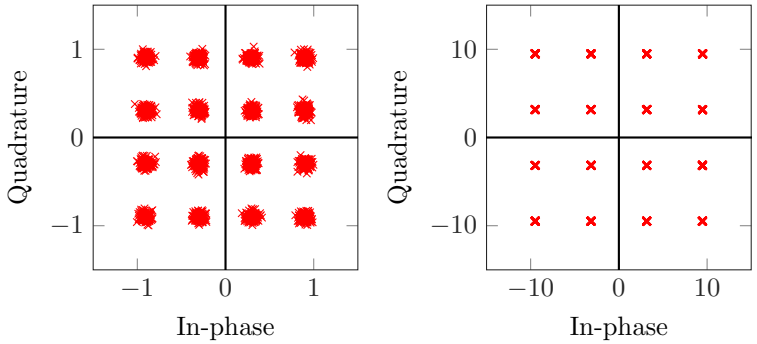



**(d)** 400 antennas, $\rho = 20$ dB. **(e)** 400 antennas, $\rho = 20$ dB, correlated fading coefficients.

**Figure 2.13:** Received signal after MRC for 16QAM inputs, $P = 20$ pilots have been used to estimate the channel.

Thus, if $\hat{\boldsymbol{h}} = \boldsymbol{h}$, the SNR after MRC is the sum of the SNR on the individual branches so that the aggregate SNR increases linearly with the array size without any increase in the transmit power. This increase in SNR is called array gain. MRC is often also referred to as matched filtering in the literature. In Figure 2.12, the one-bit BS receiver under the assumption that LS estimation is used together with MRC for a single user is depicted. The MRC block relies on the receiver having accurate CSI so that all received signals are added up coherently (constructively). It is therefore particularly important that the receiver knows the phase of the fading at each branch [28]. Fortunately, this information is provided (with some accuracy) by the LS estimator. Figure 2.13 shows the MRC output for varying number of receive antennas and SNR. The channel estimate used by the MRC is based on $P = 20$ pilots. Note that, as the size of the antenna array increases, the 16QAM constellation becomes progressively distinguishable.

Amplitude modulation is made possible by the large antenna array, solely because




**(a)** 400 antennas, $\rho = 0$ dB. **(b)** 400 antennas, $\rho = 20$ dB.

**Figure 2.14:** Received signal after MRC for 16QAM inputs in the infinite precision case, $P = 20$ pilots have been used to estimate the channel.

of the AWGN. The noise is more likely to flip symbols with lower power, e.g., the inner constellation points in 16QAM. Thus, when combining the received signal on all antennas with MRC, these symbols will be added up less constructively. Hence, the absolute value of the MRC output will be smaller. As a consequence, if the SNR is increased such that the noise will no longer cause any flips on the inner symbols, amplitude modulated symbols can no longer be distinguished as shown in Figure 2.13d.

The reason for the receiver being able to distinguish the phase of the transmitted signal also in the high-SNR case, is due to the fact that the fading is uncorrelated on all antennas. The zero-mean phase error caused by quantisation per antenna, is averaged out by the large antenna array. On the contrary, if the fading coefficients are fully correlated for all antennas, then increasing the array size make no difference for high SNR, as any constellation will collapse to QPSK, as shown in Figure 2.13e.

Finally, consider the same MRC receiver in the infinite precision case where $\boldsymbol{r}_t$ in (2.42) is replaced with $\boldsymbol{y}_t = x_t\boldsymbol{h}+\boldsymbol{w}_t$, and where the estimation in (2.41) is carried out based on unquantised outputs. The MRC output for $N = 400$ antennas is depicted in Figure 2.14. Obviously, as the SNR increases, the symbols becomes more distinguishable. However, note that for $\rho = 0$ dB SNR, there is not a substantial difference between the constellation diagram for the infinite precision case in Figure 2.14a and for the one-bit case in Figure 2.13c.

### 2.4.4 Achievable rate with training for multiple antennas

The achievable rate in a multiple-antenna system with LS estimation and MRC is

$$\frac{T-P}{T}\,\mathcal{I}(x_t;\tilde{x}_t\,|\,\hat{\boldsymbol{h}}), \tag{2.43}$$

where $\hat{\boldsymbol{h}}$ is the LS estimate in (2.41) based on $P$ pilots, and where $x_t$ and $\tilde{x}_t$ is the input and MRC output at time $t \in \{P+1, \ldots, T\}$. The conditional mutual information between a discrete input and the MRC output given the LS estimate is

$$\mathcal{I}(x_t; \tilde{x}_t \,|\, \hat{\boldsymbol{h}}) = \mathbb{E}_{\hat{\boldsymbol{h}}} \left[ \sum_{x_t, \tilde{x}_t} P_{x_t}(x_t) \; P_{\tilde{x}_t \,|\, x_t, \hat{\boldsymbol{h}}}(\tilde{x}_t \,|\, x_t, \hat{\boldsymbol{h}}) \; \log_2 \frac{P_{\tilde{x}_t \,|\, x_t, \hat{\boldsymbol{h}}}(\tilde{x}_t \,|\, x_t, \hat{\boldsymbol{h}})}{P_{\tilde{x}_t \,|\, \hat{\boldsymbol{h}}}(\tilde{x}_t \,|\, \hat{\boldsymbol{h}})} \right]. \quad (2.44)$$

To compute the mutual information, the probability of all possible MRC outcomes has to be considered. For a discrete input, both the set of channel estimates and quantised received signals have finite cardinality. Hence, the set of MRC outputs in (2.42) has also finite cardinality. However, instead of tracing the domain of MRC outcomes which depend on the input distribution, a more flexible approach is adopted, where the mutual information is approximated as suggested in [24]. The output domain is further discretised to a rectangular grid. The MRC output is replaced with a discrete random variable $\tilde{x}^{\Delta}$ whose PMF is approximated by Monte Carlo analysis, i.e., by considering a large number of channel realisations and summing the number of outcomes in each region on the grid. The resulting mutual information after the approximation is

$$\mathcal{I}(x_t; \tilde{x}_t \,|\, \hat{\boldsymbol{h}}) \approx \mathbb{E}_{\hat{\boldsymbol{h}}} \left[ \sum_{x_t, \tilde{x}_t^{\Delta}} P_{x_t}(x_t) \; P_{\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}) \; \log_2 \frac{P_{\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}})}{P_{\tilde{x}_t^{\Delta} | \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | \hat{\boldsymbol{h}})} \right]. \quad (2.45)$$

The discretisation of the output domain can actually be seen as a type of hard decoding, where outputs that fall in a particular region are mapped to the same output. Thus, the approximation in (2.45) is actually a lower bound by the data-processing inequality [26].

The achievable rate as a function of SNR for QPSK and 16QAM inputs is shown in Figure 2.15. The number of antennas is fixed to $N = 400$, the coherence interval is set to $T = 1000$ symbols, and the number of pilots has been optimised for each SNR. The mutual information has been approximated using (2.45) and by considering 200 random fading realisations. For each fading realisation, the probability $P_{\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}})$ is evaluated by considering 200 random noise vector realisations for each channel realisation and input.

Finally, the probability $P_{\tilde{x}_t^{\Delta} | \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | \hat{\boldsymbol{h}})$ is found by averaging over all inputs, i.e.,

$$P_{\tilde{x}_t^{\Delta} | \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | \hat{\boldsymbol{h}}) = \mathbb{E}_{x_t} \left[ P_{\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}) \right] = \sum_{x \in \mathcal{X}} P_{x_t}(x_t) P_{\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}}(\tilde{x}_t^{\Delta} | x_t, \hat{\boldsymbol{h}}). \quad (2.46)$$

From Figure 2.15 it is clear that 16QAM outperforms QPSK for SNR larger than approximately $-20$ dB. Furthermore, the maximal 16QAM rate of 4 bits per channel use rate is obtained in spite of the one-bit quantisers.

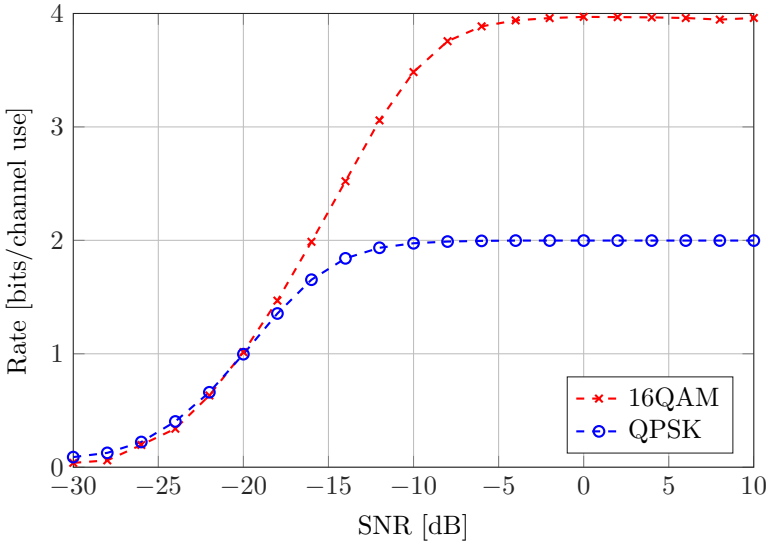



**Figure 2.15:** Achievable rate with MRC from (2.43) versus SNR with $N = 400$ antennas and a coherence interval of $T = 1000$ symbol transmission. The number of pilots is optimised for every SNR.

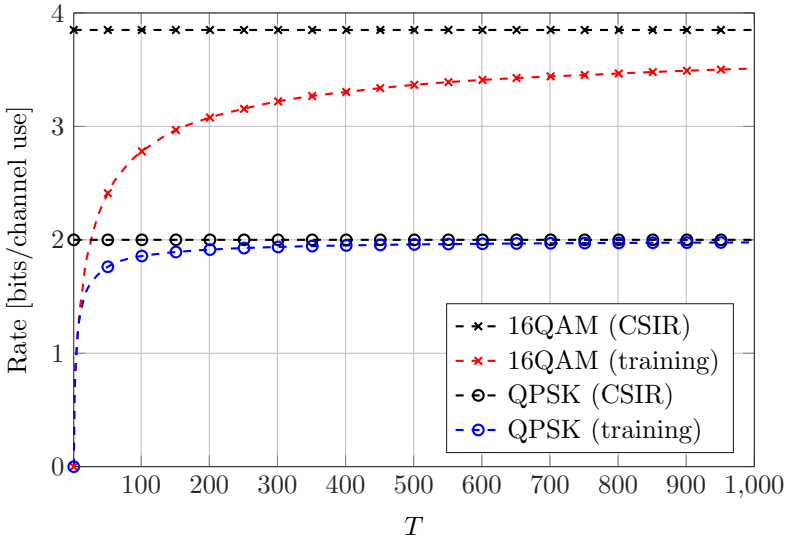



**Figure 2.16:** Achievable rate with MRC from (2.43) versus $T$ with $N = 400$ antennas for $\rho = -10$ dB. The number of pilots is optimised for every $T$.

It is also interesting to study how the achievable rate depends on the coherence interval. This investigation is shown in Figure 2.16 for the case when the SNR is fixed to 10 dB. Also shown is the perfect CSIR rate, where the receiver knows the channel perfectly, so that no training symbols has to be sent. The QPSK rate in the training

based scheme converges to the perfect CSIR rate quickly. For 16QAM, the rate is still increasing after $T = 1000$ symbols. Thus, there is a considerable rate loss in the high-mobility case.

To understand the gain associated with the large antenna array, the performance is compared to the single-antenna case. When $\rho = -10$ dB, the achievable rate with 16QAM for a receiver using LS estimation and MRC is approximately 3.5 bits per channel use, which is over forty times higher than the capacity in the single-antenna case. Thus, through the array gain due to the large number of antennas, much higher throughputs are possible without increasing the radiated transmit power. Recall from Section 1.1.1, that another benefit of using large antenna arrays is the possibility to simultaneously serve multiple users. Whether this holds true for one-bit massive MIMO is investigated next.

# Chapter 3

# Massive MIMO uplink

## 3.1 System model

Consider the case when the same BS in the uplink serves $K = 20$ single-antenna users in the same time-frequency resource as depicted in Figure 3.1.

The received signal on any antenna is the superposition of the transmitted signal from all users. The discrete-time complex baseband received quantised signal within a coherence interval can be written as

$$\boldsymbol{R} = \mathcal{Q}(\boldsymbol{Y}) = \mathcal{Q}(\boldsymbol{X}\boldsymbol{H} + \boldsymbol{W}), \tag{3.1}$$

where $\boldsymbol{H} \in \mathbb{C}^{K\times N}$ is the channel matrix connecting the $K$ users to the $N$ BS antennas, so that $h_{k,n} \sim \mathcal{CN}(0,1)$ is the fading coefficient between the $k$th user and the $n$th BS antenna. The matrix $\boldsymbol{X} \in \mathbb{C}^{T\times K}$ denotes the transmitted signals from all users during the coherence interval so that $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]$ where $\boldsymbol{x}_k \in \mathcal{X}^{T\times 1}$ is the transmitted signal from the $k$th user. Each user is subject to the power constraint in (2.5), so that the matrix $\boldsymbol{X}$ is subject to the average power constraint

$$\mathbb{E}_{\boldsymbol{X}}\left[\mathrm{tr}\left\{\boldsymbol{X}\boldsymbol{X}^H\right\}\right] \leq KT\rho, \tag{3.2}$$

where $\rho$ is the average SNR per user per antenna. Thus, the average received power on any antenna is $K$ times higher than in the single user scenario. The interference from other users on any antenna is a zero-mean random process with a variance that depends on the transmit power and the number of users.

Typically, for the infinite precision case (neglecting the effect of the quantisers), as the number of BS antennas grow large, the channel associated with the different users are almost orthogonal. Thus, many users can be served simultaneously. Whether the same holds under one-bit quantisation is investigated next.

## 3.2 Linear receivers

The first part of the coherence interval is again reserved for pilots symbols so that the channel input can be separated as

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(p)} \\ \boldsymbol{X}^{(d)} \end{pmatrix} \tag{3.3}$$

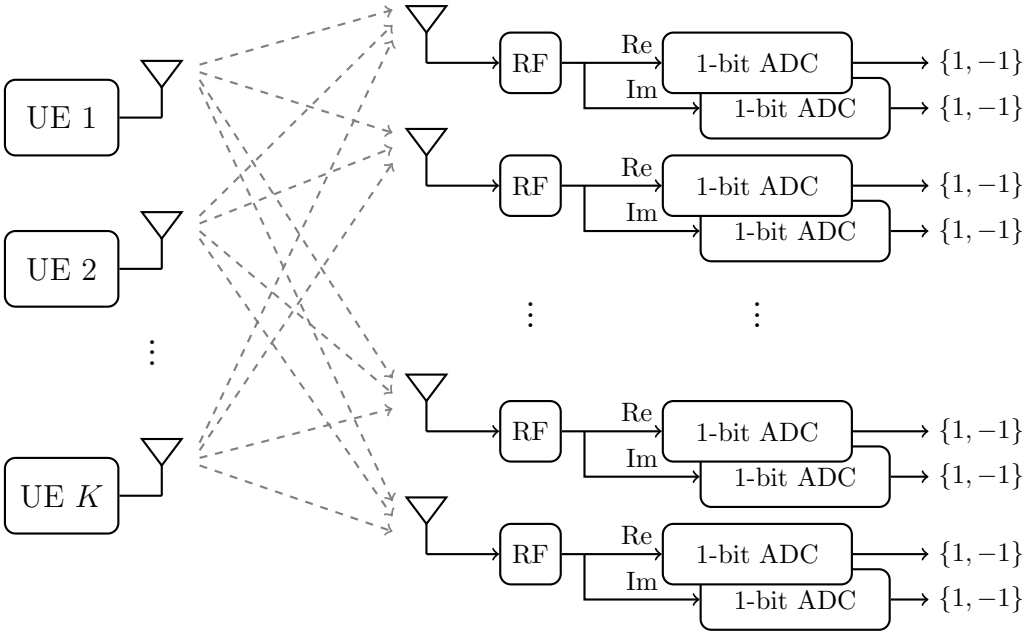



**Figure 3.1:** Massive MIMO uplink system model.

where $\boldsymbol{X}^{(p)} \in \mathcal{X}^{P \times K}$ contains the pilots from all users and $\boldsymbol{X}^{(p)} \in \mathcal{X}^{T-P \times K}$ contains the data symbols from all users. Exactly as before, the pilots are used to obtain a channel estimate through LS estimation. As stated in Section 1.1.1, the pilots from different users have to be orthogonal to avoid pilot contamination. One way to realise this is by letting only one user transmit pilots at any time. Denote the channel response associated with the $k$th user as $\boldsymbol{h}_k \in \mathbb{C}^{1 \times N}$. The $k$th row of the LS channel matrix estimate, $\hat{\boldsymbol{H}}$, is given by

$$\hat{\boldsymbol{h}}_k = \left( \left( \boldsymbol{x}_k^{(p)} \right)^H \boldsymbol{x}_k^{(p)} \right)^{-1} \left( \boldsymbol{x}_k^{(p)} \right)^H \boldsymbol{R}_k^{(p)}, \tag{3.4}$$

where $\boldsymbol{x}_k^{(p)}$ is the transmitted pilots from the $k$th user, so that $\boldsymbol{X}^{(p)} = [\boldsymbol{x}_1^{(p)}, \ldots, \boldsymbol{x}_K^{(p)}]$. Similarly, $\boldsymbol{R}_k^{(p)}$ corresponds to the received pilots over all antennas from the $k$th user. Assume that all users send the same number of pilots. If a total of $P$ pilots are sent, then only $P/K$ of them originated from the $k$th user, as only one user transmit pilots at any time. Hence, all users can transmit their pilots at $K = 20$ times the power (corresponding to a 13 dB increase) and still meet the power constraint in (3.2).

During the data transmission phase, the procured channel estimate is used in a linear receiver to distinguish the messages sent from different users as shown in Figure 3.2. Two linear receivers are considered in this work, namely MRC and zero-forcing (ZF).

With MRC in the multiuser case, a soft estimate of the data symbol sent from

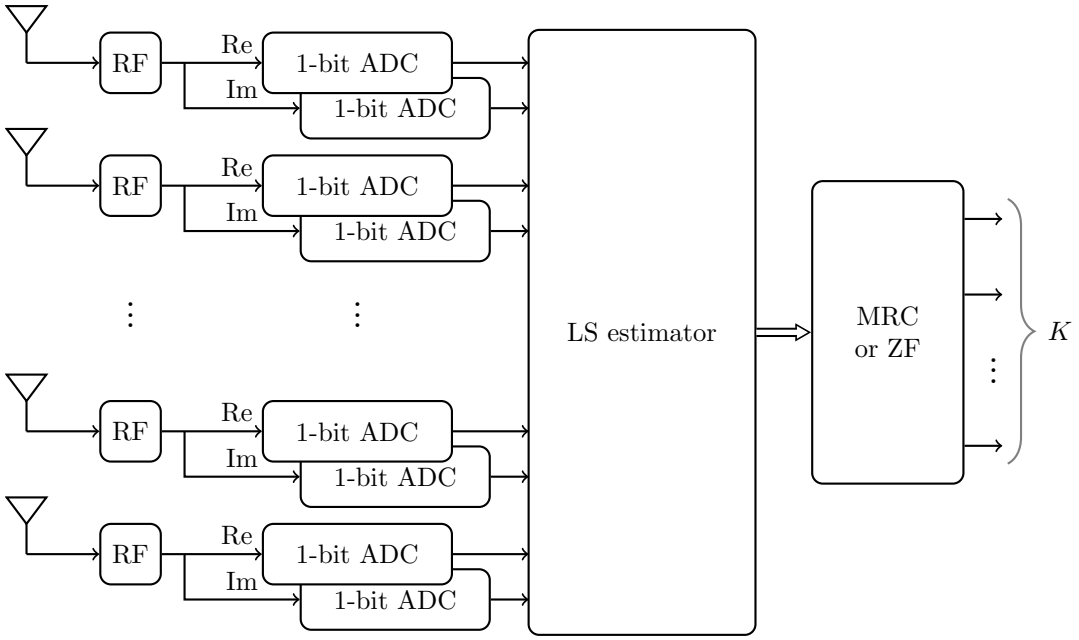



**Figure 3.2:** Massive MIMO receiver with LS estimation. The message from a several users is separated using MRC or ZF.

user $k = 1, \ldots, K$ at time $t \in \{P+1, \ \ldots, \ T\}$ is obtained as follows:

$$\tilde{x}_{t,k} = \frac{\boldsymbol{r}_t \hat{\boldsymbol{h}}_k^H}{||\hat{\boldsymbol{h}}_k||^2}, \tag{3.5}$$

where $\boldsymbol{r}_t$ is the one-bit quantised version of the superposition of the transmitted signal from all users at time $t$. Using this technique, the sought after signal is combined coherently while the zero-mean interference and AWGN is combined incoherently, analogously to the AWGN in the single user case. It is important that the phase of the fading channel corresponding to the interfering users are independent over a large number of antennas. Otherwise, there is a risk that interference is combined constructively. For the Rayleigh fading channel considered in this work, the phases of the fading channels are indeed independent.

Another approach is to actively try to force the interference from other users to zero with a ZF receiver. The soft ZF estimate is obtained as

$$\tilde{\boldsymbol{x}}_t = \boldsymbol{r}_t \hat{\boldsymbol{H}}^H \left( \hat{\boldsymbol{H}} \hat{\boldsymbol{H}}^H \right)^{-1}, \tag{3.6}$$

where $\tilde{\boldsymbol{x}}_t = [\tilde{x}_{t,1}, \ \ldots, \ \tilde{x}_{t,K}]$ is the soft output for all users at time $t \in \{P+1, \ \ldots, \ T\}$. The ZF receiver exploits the entire channel matrix to suppress interference from other users. The MRC receiver instead relies on channels associated with different users

(rows of the channel matrix) becoming orthogonal as the size of the antenna array grows. ZF typically performs better [7], but using MRC is more appealing, not only because the computational complexity is lower, but also as it is performed in a distributed fashion, i.e., messages are extracted independently for each user. Note, that in the single user case, MRC and ZF are equivalent as both techniques reduces to (2.42).

To get some insight into the performance of such schemes, consider the MRC and ZF receiver outputs in Figure 3.3 for $\rho = 0$ dB. The figure shows the receiver output for one of the 20 users that communicate with the same BS at the same time. The BS is equipped with $N = 400$ antennas and $P/K = 20$ pilots per user is used to estimate the channel. The different 16QAM symbols are illustrated by different colors so that they can be distinguished more easily. The figure shows that the 16QAM symbols can in fact be detected, though in comparison with the single-user case for the same parameters in Figure 2.13c there is a much larger spread in the constellation in the multiuser case. It also seems that the ZF receiver yields a slightly more distinguishable constellation than the MRC receiver.

Also shown in the figure is the receiver output in the infinite precision case. For the ZF receiver, there is hardly any difference to the single user case in Figure 2.14a, implying that the channels associated to the different users are almost orthogonal.

## 3.3 Massive MIMO uplink throughput

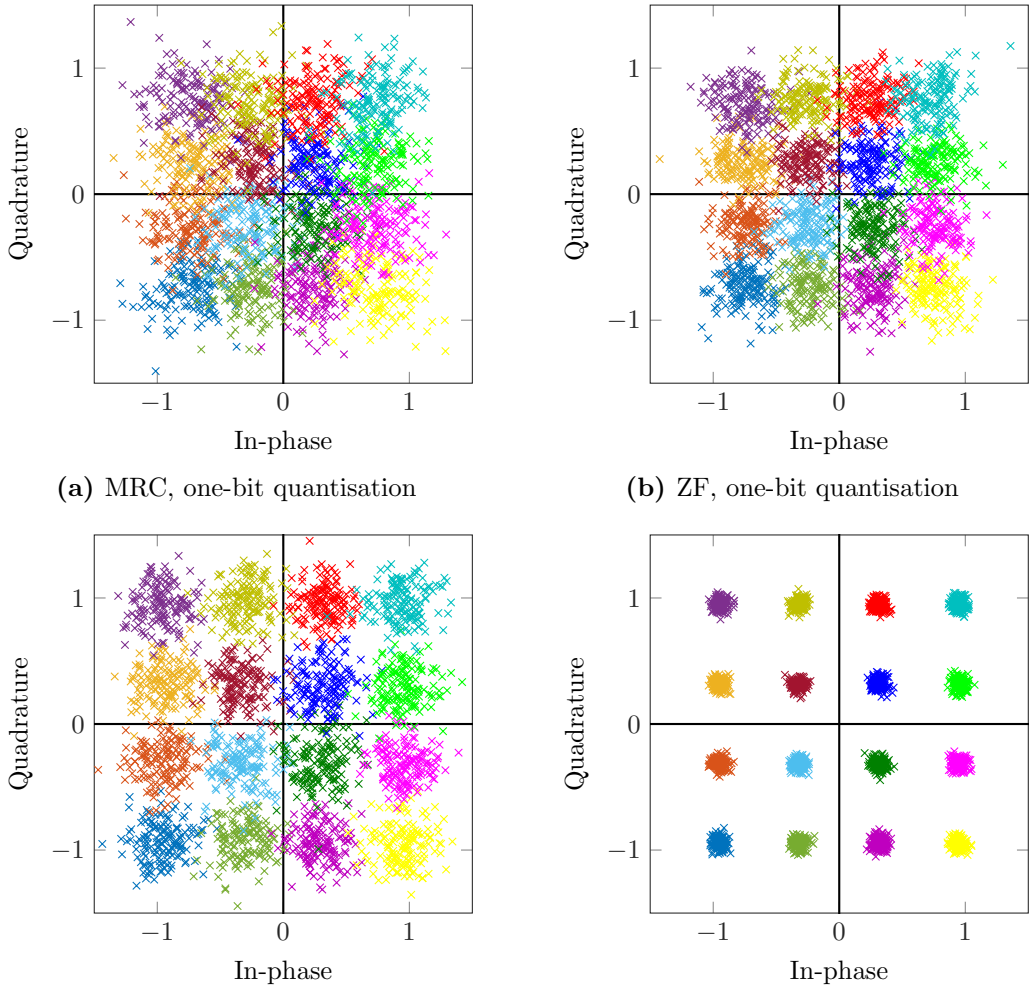
The achievable rate per user in the multiuser case is

$$\frac{T-P}{T}\,\mathcal{I}(x_{t,k};\tilde{x}_{t,k}\,|\,\hat{\boldsymbol{H}}), \tag{3.7}$$

where the mutual information between the channel input $x_{t,k}$ and receiver output $\tilde{x}_{t,k}$ for user $k$ at time $t$ is

$$\mathcal{I}(x_{t,k};\tilde{x}_{t,k}|\hat{\boldsymbol{H}}) = \mathbb{E}_{\hat{\boldsymbol{H}}}\left[\sum_{x_{t,k},\tilde{x}_{t,k}} P_{x_{t,k},\tilde{x}_{t,k}|\hat{\boldsymbol{H}}}(x_{t,k},\tilde{x}_{t,k}|\hat{\boldsymbol{H}})\log_2\frac{P_{\tilde{x}_{t,k}|x_{t,k},\hat{\boldsymbol{H}}}(\tilde{x}_{t,k}|x_{t,k},\hat{\boldsymbol{H}})}{P_{\tilde{x}_{t,k}|\hat{\boldsymbol{H}}}(\tilde{x}_{t,k}|\hat{\boldsymbol{H}})}\right]. \tag{3.8}$$

The mutual information in (3.8) has been approximated using the same approach as in the single user case, i.e., by discretising the output domain. The mutual information is averaged over 200 random channel matrix realisations, and for each channel realisation, the required probabilities are obtained by averaging with respect to 200 random noise and user interference realisations.




**(a)** MRC, one-bit quantisation

**(b)** ZF, one-bit quantisation

**(c)** MRC, infinite precision

**(d)** ZF, infinite precision

**Figure 3.3:** Received 16QAM symbols for one of the users after the linear receiver, with and without one-bit ADC. The BS is equipped with $N = 400$ antennas and there are in total $K = 20$ users being served simultaneously. The fading coefficients has been estimated using $P/K = 20$ pilots per user, and the SNR is $\rho = 0$ dB.

### 3.3.1 Throughput versus the signal-to-noise ratio

The achievable rate as a function of the SNR for QPSK and 16QAM inputs is plotted in Figure 3.4 for $N = 400$, $K = 20$ and $T = 1000$. Also shown is the achievable rate for the same setup in the single user case. First, note that with QPSK there is not a significant loss in the rate per user compared to the no interference case. However, the QPSK rate settles slightly below 2 bits per channel use for high SNR. This holds because more pilots have to be transmitted in the multiuser case, to estimate the

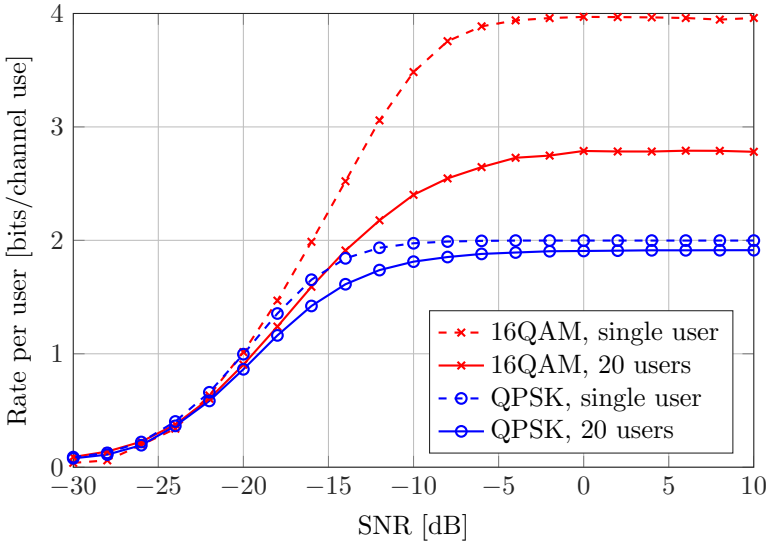


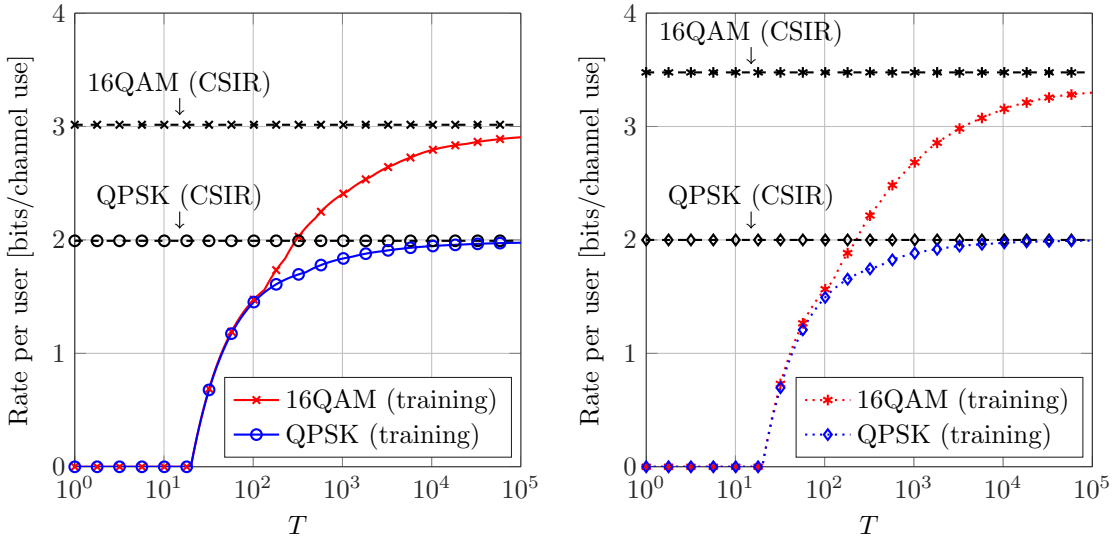

**Figure 3.4:** Achievable rate per user as a function of the SNR in a training based scheme with LS estimation and MRC. Evaluated for $N = 400$ antennas, $K = 20$ users and $T = 1000$. The number of pilots is optimised for every SNR.

channels for all users. For 16QAM signalling, the rate loss per user is more significant with a gap of approximately 1.2 bits per channel use for high SNR, which indicates that the system is in fact interference limited so that the channels for the different users can not be considered orthogonal with MRC due to the one-bit quantisation. In any case, the system still performs well in the multiuser setup and can support high-order constellations, despite the one-bit quantisers. A rate per user of 2.4 bits per channel use per user for $\rho = -10$ dB and 20 users, adds up to a sum-rate for 48 bits per channel use.

When the SNR increases, the AWGN can be neglected so that all disturbances on the received signal per antenna can be attributed to interference. For each user, there are 19 interfering users so the signal-to-interference-plus-noise ratio (SINR) at each antenna is approximately

$$\text{SINR} \approx 10 \log_{10}\left(\frac{1}{19}\right) \approx -13 \text{ dB}. \tag{3.9}$$

The achievable rate for $\rho = -13$ dB in the single user case is approximately 2.8 bits per channel use for 16QAM, and 1.9 bits per channel use for QPSK. These rates correspond well to the achievable rate per user with QPSK and 16QAM in the multiuser case for high SNR values. Hence, it seems that interference and noise with the same power has a similar affect on the system performance.




**(a)** MRC receiver. **(b)** ZF receiver.

**Figure 3.5:** Achievable rate per user as a function of $T$ in a training based scheme with LS estimation. Evaluated for $N = 400$ antennas, $K = 20$ users and $\rho = -10$ dB. The number of pilots is optimised for every $T$.

### 3.3.2 Throughput versus the coherence interval

An additional problem that has to be investigated is how the system performance varies with user mobility, i.e., how the achievable rate depends on the coherence interval. Recall that with more users, more training symbols have to be transmitted in order to estimate the fading channels.

Figure 3.5 shows the achievable rate as a function of the coherence interval for the MRC and ZF receivers respectively. The SNR is again set to $\rho = -10$ dB per user and antenna. Each user has to send pilots in separate time slots so that the achievable rate is zero for coherence intervals less than $T = 20$, as at least 20 pilots have to be transmitted. In other words, the duration of the coherence interval directly limits the number of active users, as stated in Section 1.1.1.

Also shown in the figure is the upper bound on the rate when perfect CSIR is available. For QPSK signalling, the rate per user approaches the perfect CSIR rate fairly quickly, with only an acceptable loss at $T = 1000$. For 16QAM symbols, a longer coherence interval is required to approach the perfect CSIR rate, and there is a quite significant rate loss in the high mobility case.

With 16QAM, the achievable rate eventually converges to the perfect CSIR rate, both for the MRC and ZF receiver. However, the required length of the coherence interval is substantially increased.

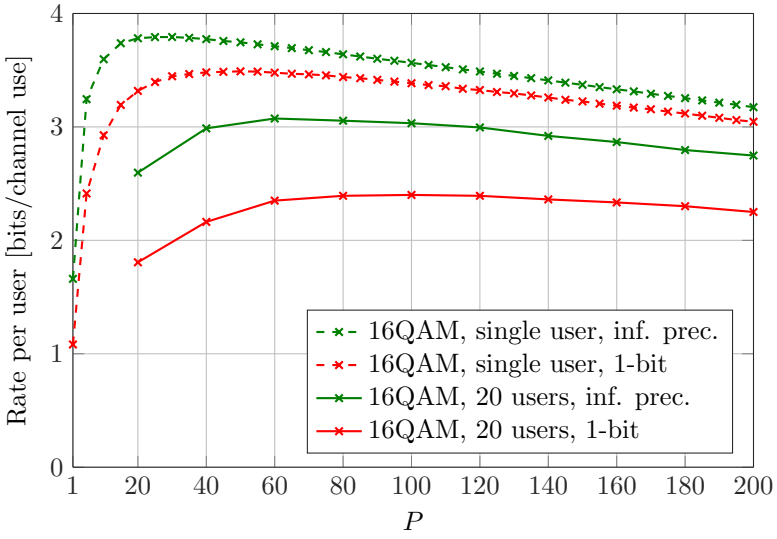



**Figure 3.6:** Achievable rate per user as a function of $P$ in a training based scheme with LS estimation and MRC. Evaluated for $N = 400$ antennas and $\rho = -10$ dB.

To understand how the transmitted number of pilots affects the achievable rate, consider Figure 3.6. Clearly, there is a trade-off between sending more pilots to estimate the channel better, and increasing the number of data symbols that can be transmitted during the coherence interval. Hence, for a given coherence interval, there is always an optimal number of pilots. In Figure 3.6 the optimal number of pilots in the multiuser case is found to be $P = 80$, so that every user transmits $P/K = 4$ pilots each. The optimal number of pilots in the single user case on the other hand is about $P = 50$. Hence, with more users, more pilots are required and each user is typically dedicated a smaller amount of pilots. Note that the achievable rate is fairly insensitive to large variations in the number of pilots. There is for example not a significant difference in doubling the number of pilots from 60 to 120 in the multiuser case. This again reflects the trade-off between having less time slots available for data transmission and having more channel knowledge.

The figure also shows the achievable rate when infinite precision is assumed. It can also be concluded that the required amount of pilots is reduced when infinite precision is assumed, both in the single-user and multiuser case.

### 3.3.3 Throughput versus the number of antennas

A final question that has to be answered is how many antennas are required in order for one-bit massive MIMO to work. Figure 3.7 compares the achievable rate for QPSK and 16QAM as the number of antennas is increased. It is again shown that 16QAM outperforms QPSK, also for smaller arrays. The fact that 16QAM symbols

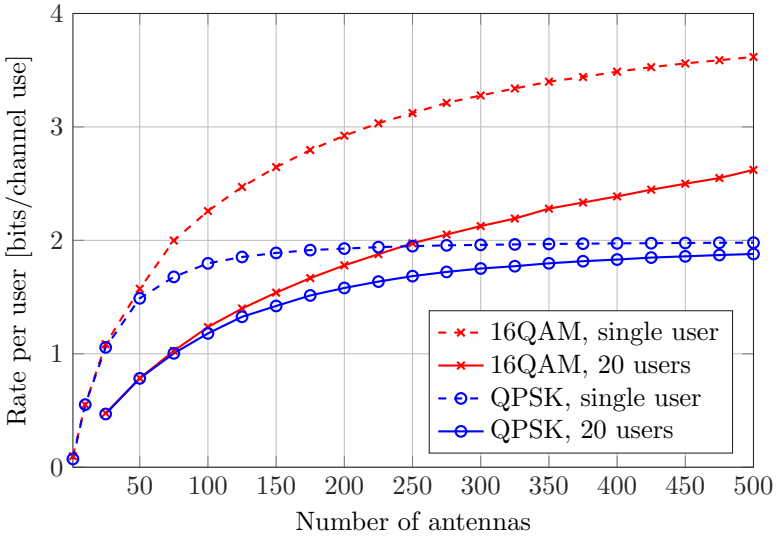



**Figure 3.7:** Achievable rate per user as a function of the numer of antennas in a training based scheme with LS estimation and MRC. Evaluated for $T = 1000$ antennas and $\rho = -10$ dB. The number of pilots is optimised for every $N$.

can be detected also with a smaller number of antennas could be guessed from Figure 2.13b, where a 16QAM constellation can be seen for only 40 elements, albeit a noisy one. A considerably smaller array than 400 antenna elements is also enough to support multiuser operation. However, as the number of antennas is reduced, the rate decreases accordingly.

In Figure 3.8, the performance of the MRC and ZF receivers is compared to the infinite precision case. With the interference cancellation of the ZF receiver, a higher throughput can be achieved. However, for smaller antenna arrays, the opposite is true. When the effect of the quantiser is ignored, a considerably smaller array delivers the same throughput. For example, a rate of 3 bits per channel use per user is achievable with only 200 elements in the infinite precision case. When the one-bit quantisers are taken into account, an array size in excess of 500 elements is required.

Finally, as a proof of concept, in Figure 3.9 it is shown that also higher-order modulation schemes are supported with one-bit massive MIMO. However, a much larger array is required in order to extract useful information. For example, with 4000 antenna elements and ZF, a 64QAM constellation can be detected.

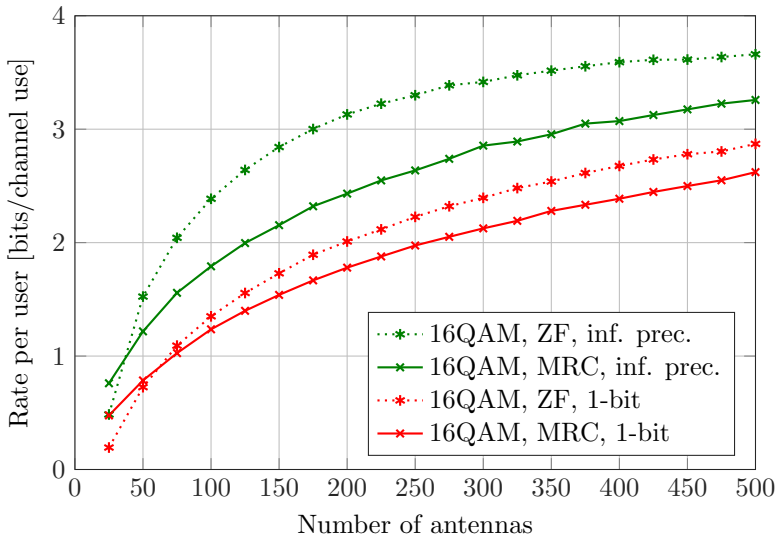
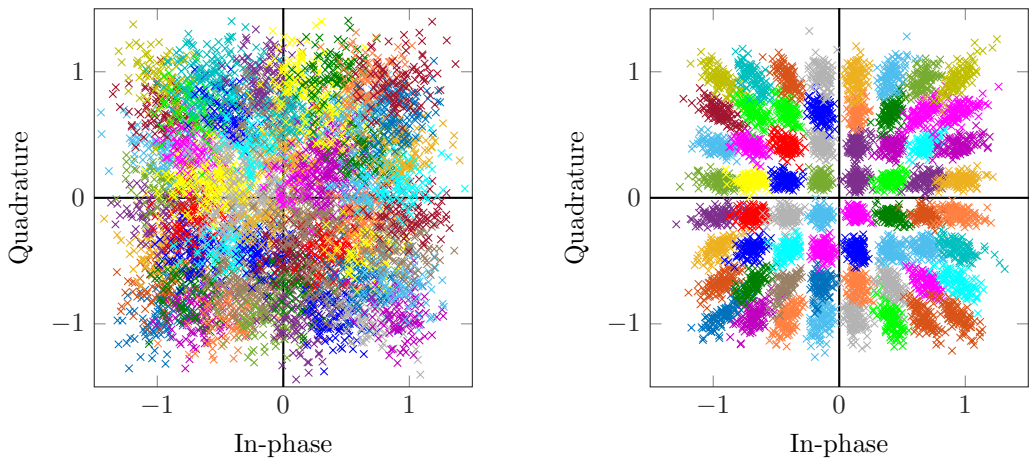



**Figure 3.8:** Achievable rate per user as a function of the number of antennas in a training based scheme with LS estimation and MRC/ZF. Evaluated for $T = 1000$ antennas and $\rho = -10$ dB. The number of pilots is optimised for every $N$.




**(a)** 400 antennas. **(b)** 4000 antennas.

**Figure 3.9:** Received 64QAM symbols for one of the users after the ZF receiver with one-bit ADC. There are in total $K = 20$ users being served simultaneously. The fading coefficients has been estimated using $P/K = 20$ pilots per user at $\rho = 0$ dB.

# Chapter 4

# Discussion

The single-antenna case was initially investigated to understand the limitations incurred by the one-bit ADC. QPSK is the capacity-achieving distribution. Hence, the throughput in the SISO case can never exceed 2 bits per channel use. It was found that for the Rayleigh block-fading channel under the assumption of high SNR, the perfect CSIR rate can be approached in a training based scheme with LS estimation. This is a very interesting result, given that the estimates in Figure 2.10 is far from perfect. The assumption of perfect CSIR is very optimistic as the information that can be inferred from the quantised pilots is very limited. However, it seems that knowing only the phase of the fading is in some sense sufficient.

It is shown in Figure 2.11 that when operating in the low SNR regime, using LS estimation incurs a signifiant rate loss. A data-aided version of LS estimation was proposed, which improved the achievable rate significantly, by transmitting only one pilot. Note that data-aided LS estimation might not be a very appealing channel estimation method, as it requires a rather complicated joint decoding and estimation scheme, where different symbols in the codeword are encoded at different rates. Also, in any practical scenario there will be delay constraints, so that the code length has to be limited, which implies that the received data symbols, which are treated as pilots, may be prone to errors.

In the case of multiple antennas at the receiver, MRC was used to map the received signal over all antennas to a scalar. Because data-aided LS is a complicated estimation method, it might not be suitable to a massive MIMO setting. With a large antenna array, the required signal processing scales accordingly, so that simple strategies are preferred. Therefore, standard LS estimation is considered.

The MRC receiver relies heavily on adequate channel knowledge. In particular, the received signals per antenna have to be co-phased in order for the combining to be constructive. Luckily, this is exactly the information that the LS estimate provides. In Figure 2.13, it is shown that high-order constellations are supported with large antenna arrays. The fact that different symbol amplitudes can be detected is due to the AWGN. Symbols of different phase can be detected due to the independent fading at each antenna element.

It was shown that with the LS estimator and MRC, high throughputs can be achieved with large antenna arrays. For example, the full 16QAM rate of 4 bits per channel use can be approached as SNR increases.

Next, multiple users are considered so that the massive MIMO scenario can be studied. It was shown that the channels to different users lose their orthogonality when subject to one-bit quantisation. In spite of this, high throughputs per user is still possible, e.g., a sum-rate of almost 50 bits per channel use with 400 antennas, 20 users and $\rho = -10$ dB. This rate is over 500 times higher than the capacity in the single-antenna case. Comparing the sum-rate in a massive MIMO system to the rate in a SISO system is partly misleading, indeed, though no additional power is radiated at the transmitter, the power consumption for the BS will increase drastically. However, it serves to portray the potential benefits of adding multiple antennas. The BS is also typically subject to more stringent power constraints than the UEs.

An important conclusion from Figure 3.4 is that 16QAM outperforms QPSK also at low SNR. Hence, also when operating in the noise-limited regime, high-order constellations should not be discarded if the antenna array is sufficiently large. This point was missed in [24] where it is shown that the full QPSK rate is achieved rapidly as SNR increases, both in the case of one-bit ADC and with infinite precision ADC. When the system operates in the saturated QPSK regime, higher-order constellations should be used. Alternatively, the number of active antenna elements in the array should be decreased, to save power and complexity.

When the effect of the quantisers is disregarded completely (see Figure 3.8), a much smaller antenna array suffices to reap the massive MIMO gains. A more complete study, would entail comparing a smaller array equipped high-precision converters, to a large array with one-bit converters. The two systems would have to be compared not only in achievable throughput but also in total power consumption, including the power consumed by the RF chains.

In Figure 3.5, it is shown that when $T$ is sufficiently large, the achievable rate with LS estimation in a one-bit massive MIMO system approaches the perfect CSIR rate. However, very long coherence intervals are required to achieve this rate. Furthermore, for high mobility, the rate reduction is significant. Thus, LS estimation may not be the best approach. Other estimation methods should be analysed to establish whether higher throughput can be achieved for high mobility scenarios.

The fact that one-bit quantisers are used in the uplink does not exclude the use of high-resolution DACs in the downlink. However, in a TDD scheme, the uplink channel estimates will need to be used for beamforming also in the downlink. Whether the estimates obtained from the quantised channel are adequately good for efficient beamforming in the downlink is an interesting problem to investigate. As shown in Figure 2.10, different pilot sequences result in different channel estimates. Perhaps custom pilot sequences could be used to extract required/desired information for different constellations.

The operating point of $\rho = -10$ dB, considered in most figures, may at first glance seem a bit restrictive. However, recall that low-resolution ADCs are mainly

considered for mmWave systems. With higher carrier frequency, the received power per antenna, is reduced significantly due to the increased attenuation and path loss.

It should be mentioned that for mmWave systems there will often be a LOS channel between the transmitter and receiver. Furthermore, the non-line-of-sight (NLOS) components are typically significantly attenuated due to propagation issues at higher frequencies. Hence, the assumption of IID Rayleigh fading channels may not be valid. Therefore, it could be interesting to study one-bit massive MIMO for other channel models, e.g., Rician fading, with correlated antennas.

Another assumption made in this thesis is that the SNR for all UEs is the same. As the distance between UEs and BS may vary a lot for different users, so will the SNR. If the received power from one of the users is much higher than the rest, the quantiser output might depend only on the dominant signal. Hence, the achievable rate for remaining users could decrease significantly. For the same reason, systems with one-bit quantisers may be particularly sensitive to radio jamming attempts. Thus, an interesting problem is whether multiuser support is still possible with a predominant user in the cell.

In many systems, the assumption of a flat-fading channel can not be justified. Multi-carrier modulations, such as OFDM, are typically used to compensate for the frequency-selectivity of channels. Whether OFDM can be supported with one-bit massive MIMO is unclear.

Finally, it should be mentioned that few if any systems will most likely be built with one-bit quantisers. However, due to the high power consumption associated with high-speed ADCs, low-precision solutions using 1-6 bits, are required to keep the energy consumption low for devices in next generation cellular networks. The results presented in this thesis for the extreme case of one-bit converters, shows that equipping large antenna arrays with a large number of low-precision ADCs is a very promising technology.

# Chapter 5

# Summary

This thesis has analysed the achievable uplink throughput in a massive MIMO system with one-bit quantisers. Information was conveyed over a Rayleigh block-fading channel under an average power constraint. The no *a priori* CSI setup was studied, so that the channel fading coefficients are unknown to the transmitter and receiver.

It was shown that with only a single antenna at the receiver, QPSK is the capacity-acheiving signalling scheme. The channel was estimated using the least squares approach. Using this technique, the phase of the fading coefficients could with some precision be estimated. Despite this limited channel knowledge, training with least squares estimation was shown to be capacity-achieving for a coherence interval of two symbols. Furthermore, a data-aided version of least squares was proven to be capacity-achieving for coherence intervals of arbitrary length.

The massive MIMO system performance was studied by assuming a receiver structure and comparing achievable rates for different signalling schemes. The channel estimates were used in a linear receiver to produce a soft estimate of the transmitted symbol. In contrast to the single-antenna case, high-order constellations are supported with a large antenna array. Furthermore, multiuser operation with high data-rates is possible, in spite of the one-bit quantisers.

# Appendix A

# Achievable rate with QPSK for SISO

The input is drawn from a QPSK constellation so that $p_{\boldsymbol{x}}(\boldsymbol{x}) = 4^{-T}$ for all $\boldsymbol{x} \in \mathcal{X}^{T\times 1}$. Consider the SISO case in (2.25) so that the channel law in (2.18) is reduced to

$$P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{r}\,|\,\boldsymbol{x}) = \mathbb{E}_h\left[\prod_{t=1}^{T} \Phi\Big(\sqrt{2}\Re\{r_t\}\Re\{x_t h\}\Big)\,\Phi\Big(\sqrt{2}\Im\{r_t\}\Im\{x_t h\}\Big)\right]. \tag{A.1}$$

The channel law is invariant to a rotation of the input constellation. To show this, let $u$ be any complex number of unit length. The rotated input $\boldsymbol{x}u$, results in the quantised channel output

$$\mathcal{Q}(\boldsymbol{x}u\boldsymbol{h} + \boldsymbol{W}). \tag{A.2}$$

Since $h$ is circularly-symmetric, $uh \sim h$ so that $\boldsymbol{x}uh + \boldsymbol{w} \sim xh + \boldsymbol{w}$, which means that

$$\boldsymbol{r} = \mathcal{Q}(\boldsymbol{x}h + \boldsymbol{W}) \sim \mathcal{Q}(\boldsymbol{x}uh + \boldsymbol{W}) \tag{A.3}$$

Hence, rotating the input constellation does not change the distribution of the quantised output. Thus, the channel law in (2.18) is the same for any rotated QPSK input. Therefore, consider the QPSK constellation given by $\mathcal{X} = \{\sqrt{\rho},\, j\sqrt{\rho}, -\sqrt{\rho}, -j\sqrt{\rho}\}$. In [19] it is shown that

$$P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{r}\,|\,\boldsymbol{U}\boldsymbol{x}) = P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{U}\boldsymbol{r}\,|\,\boldsymbol{x}), \tag{A.4}$$

where $\boldsymbol{U}$ is a complex permutation matrix with diagonal entries $u_{i,i} \in \{1, j, -1, -j\}$. It follows from (A.4) that all outputs are equiprobable since

$$P_{\boldsymbol{r}}(\boldsymbol{r}) = \mathbb{E}_{\boldsymbol{x}}\big[P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{r}\,|\,\boldsymbol{x})\big] = \mathbb{E}_{\boldsymbol{x}}\big[P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{r}\,|\,\boldsymbol{U}\boldsymbol{x})\big] = \mathbb{E}_{\boldsymbol{x}}\big[P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{U}\boldsymbol{r}\,|\,\boldsymbol{x})\big] = P_{\boldsymbol{r}}(\boldsymbol{U}\boldsymbol{r}). \tag{A.5}$$

Thus, all $4^T$ possible outputs are equiprobable with $P_{\boldsymbol{r}}(\boldsymbol{r}) = 4^{-T}$. Furthermore, it also follows from (A.4) that when computing the mutual information in (2.20), it is enough to only consider one input sequence, as they all generate the same set of probabilities. Thus,

$$\begin{aligned}\mathcal{I}(\boldsymbol{x};\boldsymbol{r}) &= \frac{1}{4^T}\sum_{\boldsymbol{x}\in\mathcal{X}^T}\sum_{\boldsymbol{r}\in\mathcal{R}^T} P_{\boldsymbol{r}|\boldsymbol{x}}(\boldsymbol{r}|\boldsymbol{x}) \log_2\left(4^T P_{\boldsymbol{r}|\boldsymbol{x}}(\boldsymbol{r}|\boldsymbol{x})\right) \\ &= \sum_{\boldsymbol{r}\in\mathcal{R}^T} P_{\boldsymbol{r}|\boldsymbol{x}}(\boldsymbol{r}|\boldsymbol{x}) \log_2\left(4^T P_{\boldsymbol{r}|\boldsymbol{x}}(\boldsymbol{r}|\boldsymbol{x})\right),\end{aligned} \tag{A.6}$$

for an arbitrary $\boldsymbol{x} \in \mathcal{X}^{T\times 1}$. For simplicity, select the input $\boldsymbol{x}_0 = [\sqrt{\rho}, \ldots, \sqrt{\rho}]$, so that (A.1) reduces to

$$\begin{aligned} P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{r}\,|\,\boldsymbol{x}_0) &= \mathbb{E}_h\left[\prod_{t=1}^{T} \Phi\Big(\sqrt{2\rho}\Re\{r_t\}\Re\{h\}\sqrt{\rho}\Big)\,\Phi\Big(\sqrt{2\rho}\Im\{r_t\}\Im\{h\}\Big)\right] \\ &= \mathbb{E}_h\left[\prod_{t=1}^{T} \Phi\Big(\sqrt{2\rho}\Re\{r_t\}\Re\{h\}\Big)\right]\mathbb{E}_h\left[\prod_{t=1}^{T} \Phi\Big(\sqrt{2\rho}\Im\{r_t\}\Im\{h\}\Big)\right]. \end{aligned} \tag{A.7}$$

Hence, the real and imaginary part of the quantised output are independent so that the channel law can be written as

$$P_{\boldsymbol{r}\,|\,\boldsymbol{x}}(\boldsymbol{r}\,|\,\boldsymbol{x}_0) = P_{\Re\{\boldsymbol{r}\}\,|\,\boldsymbol{x}_0}(\Re\{\boldsymbol{r}\}\,|\,\boldsymbol{x}_0)P_{\Im\{\boldsymbol{r}\}\,|\,\boldsymbol{x}_0}(\Im\{\boldsymbol{r}\}\,|\,\boldsymbol{x}_0). \tag{A.8}$$

The complex channel can therefore be decomposed into two real channels. The real and imaginary part of the quantised signal are identically distributed so that the mutual information of the complex channel is twice the mutual information of one real channel. Let $\bar{\boldsymbol{x}} \in \bar{\mathcal{X}}^T = \{\sqrt{\rho}, -\sqrt{\rho}\}^T$ and $\bar{\boldsymbol{r}} \in \bar{\mathcal{R}}^T = \{1, -1\}^T$ denote the input and quantised output of the real channel, i.e.,

$$\bar{\boldsymbol{r}} = \mathcal{Q}(\bar{\boldsymbol{x}}g + \boldsymbol{v}) \tag{A.9}$$

where $g \sim \mathcal{N}(0,1)$ is the real channel gain and $\boldsymbol{v} = [v_1,\ \ldots,\ v_T]^T$ is the AWGN so that $v_t \sim \mathcal{N}(0,1)$. The corresponding law for the real channel is

$$P_{\bar{\boldsymbol{r}}\,|\,\bar{\boldsymbol{x}}}(\bar{\boldsymbol{r}}\,|\,\bar{\boldsymbol{x}}) = \mathbb{E}_g\left[\prod_{t=1}^{T} \Phi(\bar{r}_t g \bar{x}_t)\right]. \tag{A.10}$$

Obviously, the outputs are uniformly distributed also for the real channel. Thus, the mutual information in (A.6) can be written as

$$\mathcal{I}(\boldsymbol{x};\boldsymbol{r}) = 2\,\mathcal{I}(\bar{\boldsymbol{x}};\bar{\boldsymbol{r}}) = 2\sum_{\boldsymbol{r}\in\bar{\mathcal{R}}^T} P_{\bar{\boldsymbol{r}}\,|\,\bar{\boldsymbol{x}}}(\bar{\boldsymbol{r}}\,|\,\bar{\boldsymbol{x}})\log_2\big(2^T P_{\bar{\boldsymbol{r}}\,|\,\bar{\boldsymbol{x}}}(\bar{\boldsymbol{r}}\,|\,\bar{\boldsymbol{x}})\big), \tag{A.11}$$

Independently of which BPSK symbol was transmitted at time $t$, $\Pr\{\text{sign}\{\bar{r}_t\} \neq \text{sign}\{\bar{x}_t g\}\} = \Phi(-g\sqrt{\rho})$ and $\Pr\{\text{sign}\{\bar{r}_t\} = \text{sign}\{\bar{x}_t g\}\} = \Phi(g\sqrt{\rho})$. Hereafter, the event $\text{sign}\{\bar{r}_t\} \neq \text{sign}\{\bar{x}_t g\}$ is referred to as a sign mismatch. Define $k$ as the number of sign mismatches that occurs in a coherence interval, i.e.,

$$k \triangleq \sum_{t=1}^{T} \mathbb{1}_{-1}(\text{sign}\,\{\bar{r}_t g \bar{x}_t\}) \tag{A.12}$$

where $\mathbb{1}_{\mathcal{A}}(s) : \mathcal{S} \to \{0, 1\}$ for $\mathcal{A} \subseteq \mathcal{S}$ denotes the indicator function which is defined as

$$\mathbb{1}_{\mathcal{A}}(s) \triangleq \begin{cases} 1, & \text{if } s \in \mathcal{A} \\ 0, & \text{if } s \notin \mathcal{A} \end{cases} \quad \forall s \in \mathcal{S}. \tag{A.13}$$

For each input vector, there are in total $\binom{T}{k}$ possible output vectors with exactly $k$ sign mismatches that all occurs with the same probability. Denote such a sequence by $\bar{\boldsymbol{r}}_k$, then for $k = 0, \ldots, T$,

$$P_{\bar{\boldsymbol{r}} \,|\, \bar{\boldsymbol{x}}}(\bar{\boldsymbol{r}}_k \,|\, \bar{\boldsymbol{x}}) = \mathbb{E}_g\Big[\Phi(-g\sqrt{\rho})^k \, \Phi(g\sqrt{\rho})^{T-k}\Big] = \gamma(k, T-k), \tag{A.14}$$

where $\gamma(\alpha,\, \beta)$ is defined as

$$\gamma(\alpha,\, \beta) \triangleq \mathbb{E}_g\Big[\Phi(-g\sqrt{\rho})^{\alpha} \, \Phi(g\sqrt{\rho})^{\beta}\Big]. \tag{A.15}$$

The mutual information in (A.11) can hence be written as

$$\mathcal{I}(\boldsymbol{x}; \boldsymbol{r}) = 2 \sum_{k=0}^{T} \binom{T}{k} \gamma(k,\, T-k) \log_2 \Big(2^T \gamma(k,\, T-k)\Big). \tag{A.16}$$

Thus, it follows that the the achievable rate with QPSK is

$$\begin{aligned} R_{\text{QPSK}}(\rho) &= \frac{1}{T} \, \mathcal{I}(\boldsymbol{x}; \boldsymbol{r}) \\ &= \frac{2}{T} \sum_{k=0}^{T} \binom{T}{k} \gamma(k,\, T-k) \log_2 \Big(2^T \gamma(k,\, T-k)\Big) \\ &= 2 + \frac{2}{T} \sum_{k=0}^{T} \binom{T}{k} \gamma(k,\, T-k) \log_2 \Big(\gamma(k,\, T-k)\Big). \end{aligned} \tag{A.17}$$

# Appendix B

# Achievable rate with LS estimation for SISO

The achievable rate with LS estimation in (2.39) is a lower bound on the achievable rate, as

$$\begin{aligned}\frac{1}{T}\,\mathcal{I}(\boldsymbol{x};\boldsymbol{r}) &\geq \frac{1}{T}\,\mathcal{I}(\boldsymbol{x}^{(d)};\boldsymbol{r}^{(d)}\,|\,\boldsymbol{x}^{(p)},\ \boldsymbol{r}^{(p)}) \\ &\geq \frac{1}{T}\,\mathcal{I}(\boldsymbol{x}^{(d)};\boldsymbol{r}^{(d)}\,|\,\hat{h}) \\ &= \frac{T-P}{T}\mathcal{I}(x_t;r_t\,|\,\hat{h}).\end{aligned} \tag{B.1}$$

The first inequality in (B.1) holds as pilot and data transmission is not necessarily the optimal strategy (it excludes for example the use of blind estimation techniques which could perform better). The second inequality holds as the LS estimation in (2.37) might not be optimal. Assume that the channel input at each time is chosen from a QPSK constellation, so that the achievable rate with QPSK signalling and LS estimation is

$$R^{\mathrm{LS}}_{\mathrm{QPSK}}(\rho) = \frac{T-P}{T}\mathcal{I}(x;r\,|\,\hat{h}). \tag{B.2}$$

Recall from Appendix A, that under the assumption of QPSK, the complex channel can be decomposed into two IID channels. Thus, the achievable rate with QPSK signalling and LS estimation can be written as

$$R^{\mathrm{LS}}_{\mathrm{QPSK}}(\rho) = 2\left(\frac{T-P}{T}\right)\mathcal{I}(\bar{x};\bar{r}\,|\,\hat{g}), \tag{B.3}$$

where $\hat{g}$ is the LS estimate of the real channel, $\bar{x} \in \{\sqrt{\rho}, -\sqrt{\rho}\}$ are BPSK inputs, and $\bar{r}$ is the quantised output of the real channel $\bar{x}g + v$. The mutual information between a BPSK input and the quantised output given the LS estimate of the real channel is

$$\mathcal{I}(\bar{x};\bar{r}\,|\,\hat{g}) = \frac{1}{2}\sum_{\bar{x},\bar{r},\hat{g}} P_{\hat{g}}(\hat{g})P_{\bar{r}\,|\,\bar{x},\hat{g}}(\bar{r}\,|\,\bar{x},\hat{g})\log_2\frac{P_{\bar{r}\,|\,\bar{x},\hat{g}}(\bar{r}\,|\,\bar{x},\hat{g})}{P_{\bar{r}\,|\,\hat{g}}(\bar{r}\,|\,\hat{g})}. \tag{B.4}$$

Recall that for QPSK inputs, the LS channel estimate is given by (2.38). Similarly for BPSK inputs, the estimate of the real channel is

$$\hat{g} = \frac{1}{P\sqrt{\rho}} \sum_{i=1}^{P} \bar{x}_i^{(p)} \bar{r}_i^{(p)}. \tag{B.5}$$

It is interesting to note that the outcome of the LS estimator, with BPSK symbols, does not depend on the transmitted pilot sequence, but rather on the number of sign mismatches between the transmitted and received pilot sequences,

$$l \triangleq \sum_{i=1}^{P} \mathbb{1}_{-1}(\text{sign}\,\{\bar{r}_i^{(p)} h \bar{x}_i^{(p)}\}), \tag{B.6}$$

where an outcome with $l$ sign mismatches, which is denoted by $\hat{g}_l$, takes on the value

$$\hat{g}_l = \frac{1}{P\sqrt{\rho}} \left( \sum_{i=1}^{P-l} \sqrt{\rho} - \sum_{i=1}^{l} \sqrt{\rho} \right) = 1 - \frac{2l}{P}, \tag{B.7}$$

for $l = 0, \ldots, P$. This means that there are $P+1$ possible outcomes, distributed in the interval $[-1, +1]$ with equidistant spacing. Hence, with QPSK symbols for the complex channel, there are $(P+1)^2$ possible LS estimation outcomes distributed on a rectangular grid. Note that the QPSK constellation in this appendix is actually shifted by 45°. Hence, the LS estimation outcomes of the complex channel, for QPSK pilots, are distributed as in Figure 2.9. Since for each pilot sequence there are $\binom{P}{l}$ possible outcomes with $l$ sign mismatches, such an outcome is associated with the conditional probability

$$P_{\hat{g}\,|\,g}(\hat{g}_l \,|\, g) = \binom{P}{l} \Phi\left(-g\sqrt{\rho}\right)^l \Phi\left(g\sqrt{\rho}\right)^{P-l}. \tag{B.8}$$

Furthermore, by taking the expectation with respect to $g$, the output probability of the LS estimate can be written as

$$P_{\hat{g}}(\hat{g}_l) = \mathbb{E}_g \left[ \binom{P}{l} \Phi\left(-g\sqrt{\rho}\right)^l \Phi\left(g\sqrt{\rho}\right)^{P-l} \right]. \tag{B.9}$$

By the law of total probability and the chain rule for probabilities, the required probability in (B.4) can be rewritten as

$$\begin{aligned} P_{\bar{r}\,|\,\bar{x},\hat{g}}(\bar{r} \,|\, \bar{x}, \hat{g}) &= \frac{1}{P_{\hat{g}}(\hat{g}) P_x(x)} \int_{-\infty}^{\infty} P_{r,x,\hat{g},g}(r, x, \hat{g}_P, g) dg \\ &= \frac{1}{P_{\hat{g}}(\hat{g}) P_x(x)} \int_{-\infty}^{\infty} p_g(g) P_{\hat{g}\,|\,g}(\hat{g} \,|\, g) P_x(x) P_{r\,|\,x,g}(r \,|\, x, g) dg \\ &= \frac{1}{P_{\hat{g}}(\hat{g})} \mathbb{E}_g \left[ P_{\hat{g}\,|\,g}(\hat{g} \,|\, g) P_{r\,|\,x,g}(r \,|\, x, g) \right]. \end{aligned} \tag{B.10}$$

Using (B.8) and (B.9), the conditional probability in (B.10) for an estimation outcome caused by $l$ sign mismatches can be written as

$$P_{\bar{r}\,|\,\bar{x},\hat{g}}(\bar{r}\,|\,\bar{x},\hat{g}_l) = \frac{\mathbb{E}_g\big[\Phi(-g\sqrt{\rho})^l\Phi(g\sqrt{\rho})^{P-l}\Phi(\bar{r}g\bar{x})\big]}{\gamma(l,P-l)}. \tag{B.11}$$

Finally, by averaging with respect to all inputs, the probability $P_{\bar{r}\,|\,\hat{g}}(\bar{r}\,|\,\hat{g}_l)$ is

$$P_{\bar{r}\,|\,\hat{g}}(\bar{r}\,|\,\hat{g}_l) = \mathbb{E}_{\bar{x}}\left[\frac{\mathbb{E}_g\big[\Phi(-g\sqrt{\rho})^l\Phi(g\sqrt{\rho})^{P-l}\Phi(\bar{r}g\bar{x})\big]}{\gamma(l,P-l)}\right] = \frac{1}{2}\frac{\gamma(l,P-l)}{\gamma(l,P-l)} = \frac{1}{2}. \tag{B.12}$$

Hence, independently of the LS outcome, the outputs are equally likely. It follows from (B.12) that $P_r(r) = 1/2$. Note that $P_{\bar{r}\,|\,\bar{x},\hat{g}}(\bar{r}\,|\,\bar{x},\hat{g}_l)$ has only two possible outcomes. The probabilities of the outcomes do not depend on the actual input, but only on number of sign mismatches between the input and output. Thus, when computing the mutual information in (B.4), it is enough to consider only one of the inputs. It follows by inserting (B.11) and (B.12), that the mutual information is

$$\begin{aligned}\mathcal{I}(\bar{x};\bar{r}\,|\,\hat{g}) &= 1 + \sum_{l=0}^{P}\sum_{m=0}^{1}\binom{P}{l}\gamma(l+m,P+1-l-m) \\ &\qquad \cdot \log_2\left(\frac{\gamma(l+m,P+1-l-m)}{\gamma(l,P-l)}\right) \\ &= 1 - \sum_{l=0}^{P}\binom{P}{l}\gamma(l,P-l)\ \mathcal{H}\left(\frac{\gamma(l+1,P-l)}{\gamma(l,P-l)}\right).\end{aligned} \tag{B.13}$$

Hence, the achievable rate in (B.3) is

$$R^{\mathrm{LS}}_{\mathrm{QPSK}}(\rho) = 2\left(\frac{T-P}{T}\right)\left(1 - \sum_{l=0}^{P}\binom{P}{l}\gamma(l,P-l)\ \mathcal{H}\left(\frac{\gamma(l+1,P-l)}{\gamma(l,P-l)}\right)\right). \tag{B.14}$$

Recall that the achievable rate with training and LS estimation is a lower bound on the achievable rate with QPSK. However as stated in Lemma 1, repeated below for convenience, the full QPSK rate is achieved for $T = 2$.

**Lemma 1.** *Training and LS channel estimation with one pilot and one data symbol achieves the QPSK rate in (2.33) for $T = 2$.*

*Proof.* First, note that for $P = 1$,

$$\gamma(l,1-l) = \mathbb{E}_g\Big[\Phi\left(-g\sqrt{\rho}\right)^l\Phi\left(g\sqrt{\rho}\right)^{1-l}\Big] = \frac{1}{2}, \tag{B.15}$$

so that the achievable rate with LS estimation in (B.14) for $T = 2$ is

$$\begin{aligned} R_{\text{QPSK}}^{\text{LS}}(\rho) &= 2 + \sum_{l=0}^{1}\sum_{m=0}^{1} \gamma(l+m, 2-l-m) \log_2 \gamma(l+m, 2-l-m) \\ &= 2 + \gamma(0,2)\log_2\gamma(0,2) + 2\gamma(1,1)\log_2\gamma(1,1) + \gamma(2,0)\log_2\gamma(2,0) \qquad \text{(B.16)} \\ &= 2 + \sum_{k=0}^{2} \binom{2}{k} \gamma(k,2-k)\log_2\gamma(k,2-k), \end{aligned}$$

which is exactly the achievable rate with QPSK in (2.33) for $T = 2$. □

Using data-aided LS estimation as described in Section 2.4.2, the achievable rate for a coherence interval of $T$ symbols is

$$R_{\text{QPSK}}^{\text{DA},(T)}(\rho) = \frac{1}{T}\left((T-1)R_{\text{QPSK}}^{\text{DA},(T-1)} + 2I(\bar{x};\bar{r}\,|\,\hat{g}_T)\right) \qquad \text{(B.17)}$$

where $\hat{g}_T$ is the LS estimate of the real channel using $T$ pilots. The achievable rate with data-aided LS estimation is given in Theorem 2, repeated below for convenience.

**Theorem 2.** *Training and data-aided LS estimation with only one pilot achieves the QPSK rate in (2.33) for all $T$.*

*Proof.* The proof is by induction. Note that data-aided LS obviously is equivalent to standard LS for $T = 2$. Assume that data-aided LS is optimal for a coherence length $T$, then the rate for a coherence interval of $T + 1$ symbols from (B.17) is

$$\begin{aligned} R_{\text{QPSK}}^{\text{DA},(T+1)}(\rho) = \frac{2}{T+1}\Bigg(& T + \sum_{k=0}^{T}\binom{T}{k}\gamma(k,\,T-k)\log_2\gamma(k,\,T-k) \\ &+ 1 + \sum_{k=0}^{T}\binom{T}{k}\gamma(k,\,T-k+1)\log_2\frac{\gamma(k,\,T-k+1)}{\gamma(k,\,T-k)} \qquad \text{(B.18)} \\ &+ \sum_{k=0}^{T}\binom{T}{k}\gamma(k+1,\,T-k)\log_2\frac{\gamma(k+1,\,T-k)}{\gamma(k,\,T-k)}\Bigg). \end{aligned}$$

Expanding the logarithms yields

$$\begin{aligned} R_{\text{QPSK}}^{\text{DA},(T+1)} = \frac{2}{T+1}\Bigg( & T+1+\sum_{k=0}^{T}\binom{T}{k}\gamma(k,\,T-k)\log_2\gamma(k,\,T-k) \\ & -\sum_{k=0}^{T}\big[\gamma(k,\,T-k+1)+\gamma(k+1,\,T-k)\big]\log_2\gamma(k,\,T-k) \\ & +\sum_{k=0}^{T}\binom{T}{k}\gamma(k,\,T+1-k)\log_2\gamma(k,\,T+1-k) \\ & +\sum_{k=0}^{T}\binom{T}{k}\gamma(k+1,\,T-k)\log_2\gamma(k+1,\,T-k)\Bigg). \end{aligned} \tag{B.19}$$

It is easily verified that for every $k$ and $T$, it holds that

$$\gamma(k,\,T-k+1)+\gamma(k+1,\,T-k)=\gamma(k,\,T-k). \tag{B.20}$$

Hence, the data-aided LS rate in (B.19) reduces to

$$\begin{aligned} R_{\text{QPSK}}^{\text{DA},(T+1)} &= \frac{2}{T+1}\Bigg(T+1+\sum_{k=0}^{T}\binom{T}{k}\gamma(k,\,T+1-k)\log_2\gamma(k,\,T+1-k) \\ &\quad +\sum_{k=0}^{T}\binom{T}{k}\gamma(k+1,\,T-k)\log_2\gamma(k+1,\,T-k)\Bigg) \\ &= \frac{2}{T+1}\Bigg(T+1+\sum_{k=0}^{T}\binom{T}{k}\gamma(k,\,T+1-k)\log_2\gamma(k,\,T+1-k) \\ &\quad +\sum_{k=0}^{T}\binom{T}{k-1}\gamma(k,\,T+1-k)\log_2\gamma(k,\,T+1-k)\Bigg) \\ &= \frac{2}{T+1}\Bigg(T+1+\sum_{k=0}^{T}\binom{T+1}{k}\gamma(k,T+1-k)\log_2\gamma(k,T+1-k)\Bigg). \end{aligned} \tag{B.21}$$

In the last step, the following recursive formula for binomial coefficients was used

$$\binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}. \tag{B.22}$$

The last row in (B.21) is recognised as the expression for the achievable rate with QPSK for a coherence interval of $T+1$ in (2.33).

From Lemma 1, data-aided LS achieves the QPSK rate for $T=2$. Thus, by induction it also holds for $T=3$ and beyond, i.e., data-aided LS estimation achieves the QPSK rate for all $T$. □

# Bibliography


[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] E. Dahlman, G. Mildh, J. Peisa, J. Sachs, Y. Selén, and S. Parkvall, "5G radio access," *Ericsson Review*, pp. 1–7, Jun. 2014.

[3] Ericsson, "More than 50 billion connected devices," *White paper*, Feb. 2011. [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/wp-50-billions.pdf

[4] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, May 2014.

[5] F. Boccardi, R. W. Heath Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five Disruptive Technology Directions for 5G," *IEEE Communications Magazine*, no. February, pp. 74–80, Feb. 2014.

[6] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[7] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Communications Magazine*, pp. 186–195, Feb. 2014.

[8] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," Jan. 2014. [Online]. Available: http://arxiv.org/abs/1307.2584

[9] Z. Pi and F. Khan, "An Introduction to Millimeter-Wave Mobile Broadband Systems," *IEEE Communications Magazine*, pp. 101–107, Jun. 2011.

[10] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, and K. Cheun, "Millimeter-Wave Beamforming as an Enabling Technology for 5G Cellular Communications : Theoretical Feasibility and Prototype Results," *IEEE Communications Magazine*, pp. 106–113, Feb. 2014.

[11] R. H. Walden, "Analog-to-digital converter technology comparison," in *IEEE GaAs IC Symposium Technical Digest*, Oct. 1994, pp. 217–219.

[12] ——, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[13] B. Murmann, "ADC Performance Survey 1997-2014." [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html

[14] O. E. Ayach, S. Rajagopal, S. Abu-surra, Z. Pi, and R. W. Heath Jr., "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[15] Texas Instruments, "ADC Products." [Online]. Available: http://www.ti.com/lsds/ti/data-converters/analog-to-digital-converter-products.page

[16] S. Hoyos, B. M. Sadler, and G. R. Arce, "Monobit Digital Receivers for Ultra-wideband Communications," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1337–1344, Jul. 2005.

[17] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *IEEE International Symposium on Information Theory*, Nice, Jun. 2007, pp. 1286–1289.

[18] J. Mo and R. W. Heath, "Capacity Analysis of One-Bit Quantized MIMO Systems with Transmitter Channel State Information," Oct. 2014. [Online]. Available: http://arxiv.org/abs/1410.7353

[19] A. Mezghani and J. A. Nossek, "Analysis of Rayleigh-fading channels with 1-bit quantized output," in *IEEE International Symposium on Information Theory*, no. 4, Toronto, Jul. 2008, pp. 260–264.

[20] ——, "Analysis of 1-bit output noncoherent fading channels in the low SNR regime," in *IEEE International Symposium on Information Theory*, Seoul, Jul. 2009, pp. 1080–1084.

[21] T. Koch and A. Lapidoth, "One-bit quantizers for fading channels," in *International Zurich Seminar on Communications*, Zurich, 2012, pp. 36–39.

[22] O. Dabeer and U. Madhow, "Channel Estimation with Low-Precision Analog-to-Digital Conversion," in *IEEE International Conference on Communications*, Cape Town, May 2010, pp. 1–6.

[23] G. Zeitler, G. Kramer, and A. C. Singer, "Bayesian parameter estimation using single-bit dithered quantization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.

[24] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," Apr. 2014. [Online]. Available: http://arxiv.org/abs/1404.7736

[25] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE press, 1993.

[26] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2012.

[27] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: Optimal modulation, ergodic capacity and outage probability," in *IEEE Information Theory Workshop*, no. 2. Dublin: Ieee, Aug. 2010, pp. 1–5.

[28] A. Goldsmith, *Wireless communications*. Cambridge, England: Cambridge University Press, 2005.

[29] J. Huang and S. P. Meyn, "Characterization and Computation of Optimal Distributions for Channel Coding," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2336–2351, Jul. 2005.