CHALMERS

Massive MIMO Systems with Hardware Imperfections

NIKOLAOS KOLOMVAKIS

Signal Processing Group
Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2019

Thesis for the degree of Doctor of Philosophy

Massive MIMO Systems with Hardware Imperfections

Nikolaos Kolomvakis

CHALMERS

Signal Processing Group
Department of Electrical Engineering
Chalmers University of Technology

Gothenburg, Sweden 2019

Kolomvakis, Nikolaos
Massive MIMO Systems with Hardware Imperfections.

ISBN 978-91-7905-123-5
Series No. 4590
ISSN 0346-718X
Signal Processing Group
Department of Electrical Engineering
Göteborg, Sweden, 2019

This thesis has been prepared using LaTeX.

Printed by Chalmers Reproservice,
Göteborg, Sweden, May 2019.

To my beloved son Alexandros and my lovely wife Sophie

Abstract

Recent years have witnessed an unprecedented explosion in mobile data traffic, due to the expansion of numerous types of wireless devices. Moreover, each device needs a high throughput to support demanding applications such as real-time video, movie streaming and games. Thus, future wireless systems have to satisfy three main requirements: 1) having a high throughput; 2) simultaneously serving many users; and 3) less energy consumption. Massive multiple-input multiple-output (MIMO) systems meet the aforementioned requirements , and is nowadays a well-established technology which forms the backbone of the fifth-generation (5G) cellular communication systems. However, massive MIMO systems, i.e. employing hundreds or even thousands of antennas, will be a viable solution in the future only if low-cost and energy-efficient hardware is deployed. Unfortunately, low-cost, low-quality hardware is prone to hardware impairments such as in-phase and quadrature imbalance (IQI) and phase noise.

Moreover, one of the dominant sources of power consumption in massive MIMO systems are the data converters at the BS. The baseband signal at each radio-frequency (RF) chain is generated by a pair of analog-to-digital converters (ADCs). The power consumption of these ADCs increases exponentially with the resolution (in bits). For massive MIMO systems this would lead to prohibitively high power consumption due to the large number of required ADCs. Hence, the ADC resolution must be limited to keep the power budget within tolerable levels.

In this thesis, we investigate the performance of massive MIMO systems in non-ideal hardware. We begin with by studying the impact of IQI on massive MIMO systems. Important insights are gained through the analysis of system performance indicators, such as channel estimation and achievable rates by deriving tractable approximations of the ergodic spectral efficiency. First, a novel pilot-based joint estimator of the uplink augmented MIMO channel matrix and receiver IQI coefficients is described and then, a low-complexity IQI compensation scheme is proposed which is based on the receiver IQI coefficients' estimation and it is independent of the channel gain.

Second, we investigate the impact of the transceiver IQI in massive MIMO considering a time division duplexing (TDD) system where we assume uplink/downlink channel reciprocity in the downlink precoding design. The uplink channel estimation accuracy and the achievable downlink rate of the regularized zero-forcing (RZF) and maximum ratio transmission (MRT) is studied when there is mismatch between the uplink and downlink channels.

Finally, we analyse the quantization distortion in limited-precision ADCs in uplink

massive MIMO systems whose channel state information (CSI) is not known *a priori* to transmitter and receiver. We show that even a small percentage of clipped samples at the receiver can downgrade considerably the systems performance and propose near-optimal low-complexity solutions to reconstruct the clipped signal.

Keywords: Massive MIMO, in-phase and quadrature imbalance, analogue-to-digital converter, achievable rate, channel estimation, random matrix theory.

List of Publications

The thesis is based on part of the following papers:

[A] N. Kolomvakis, M. Matthaiou, and M. Coldrey, "IQ Imbalance in Multiuser Systems: Channel Estimation and Compensation," in *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 3039–3051, Jul. 2016.

[B] N. Kolomvakis, M. Coldrey, T. Eriksson and M. Viberg, "Massive MIMO Systems With IQ Imbalance: Channel Estimation and Sum Rate Limits," in *IEEE Transactions on Communications,* vol. 65, no. 6, pp. 2382 - 2396, Mar. 2017.

[C] N. Kolomvakis, T. Eriksson, M. Coldrey and M. Viberg, "Reconstruction of Clipped Signals in Massive MIMO Systems" submitted to *IEEE Transactions on Communications,* May 2019.

Acknowledgements

First of all, I would like to express my appreciation and thanks to my supervisors and friends, Prof. Mats Viberg, Prof. Thomas Eriksson and Dr. Mikael Coldrey, who have been mentors for me over the last three years. Their endless patience and unconditional help constantly encourage me to mature as a researcher.

Special thanks goes to Prof. Michail Matthaiou, for his creative inspirations and fruitful discussions during our collaboration and for being my supervisor for the first years as a Ph.D student.

Many thanks also goes to my colleagues at E2 and Ericsson, thank you all for creating a pleasant working environment. In particular, I would like to thank Prof. Tomas McKelvey for all the effort he has dedicated in making the group better and better.

Last but not least, I would like to thank my family Sophie and Alexandros as well my parents, and Aphrodite, for their constant support, motivation and love.

This work has been supported in part by the Swedish Governmental Agency for Innovation Systems (VINNOVA) within the VINN Excellence Center Chase.

Nikolaos Kolomvakis
Göteborg, February 2019

Acronyms

3GPP:	3rd Generation Partnership Project
AWGN:	Additive White Gaussian Noise
CDF:	Cumulative Distribution Function
CSI:	Channel State Information
DCR:	Direct-Conversion Radio
LO:	Local Oscillator
LTE:	Long-term Evolution
MIMO:	Multiple-Input Multiple-Output
MMSE:	Minimum-Mean-Squared-Error
MRC:	Maximum-Ratio-Combining
PDF:	Probability Density Function
RF:	Radio Frequency
SINR:	Signal-to-Noise-plus-Interference Ratio
SNR:	Signal-to-Noise Ratio
SVD:	Singular Value Decomposition
ZF:	Zero-Forcing

Notation

$\mathbf{X}$:	Matrix $\mathbf{X}$
$\mathbf{x}$:	Vector $\mathbf{x}$
$\mathbf{X}^T$:	Transpose of matrix $\mathbf{X}$
$\mathbf{X}^H$:	Conjugate transpose of matrix $\mathbf{X}$
$\mathbf{X}^*$:	Conjugate of matrix $\mathbf{X}$
$\mathbf{X}^{-1}$:	Inverse of matrix $\mathbf{X}$
$\|\|\mathbf{X}\|\|$:	L_2 norm of matrix $\mathbf{X}$
$\det\{\mathbf{X}\}$:	Determinant of matrix $\mathbf{X}$
$\mathrm{tr}\{\mathbf{X}\}$:	Trace of matrix $\mathbf{X}$
$\mathbf{I}_N$:	$N \times N$ identity matrix
x^*:	complex conjugate of x
$\mathbb{E}\{x\}$:	expected value of x
$\mathbb{E}\{x\|y\}$:	conditional expected value of x given y
$\Re\{x\}$:	Real part of x
$\Im\{x\}$:	Imaginary part of x
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:	Normal distribution with mean of $\boldsymbol{\mu}$ and co-variance matrix $\boldsymbol{\Sigma}$
$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:	Complex normal distribution with mean of $\boldsymbol{\mu}$ and co-variance matrix $\boldsymbol{\Sigma}$

Contents

Part I

Overview

Chapter 1

Introduction

Data transmission over wireless networks has increased rapidly during the last years and it is predicted that this trend will continue also in the coming years [1]. However, physical resources will remain the same (e.g. frequencies, number of time slots). Therefore, new technologies have to be developed in order to enable this growth in the future. One of these technologies is massive multiple-input multiple-output (MIMO) systems. Massive MIMO (a.k.a. large-scale MIMO, very large MIMO) systems use antenna arrays with a few hundred antennas, simultaneously serving many tens of terminals in the same time-frequency resource. The basic premise behind massive MIMO is to reap all the benefits of conventional MIMO, but on a much greater scale.

Massive MIMO systems has several attractive features [2], [3]. Extra antennas help to focus energy into small regions of space to bring huge improvements in throughput compared to conventional MIMO systems, and simultaneously improve the energy efficiency [4]. Other benefits of massive MIMO include: extensive use of inexpensive low-power components, simplest linear receivers e.g, MRC, become nearly optimal [5], [6], while thermal noise, inter-cell interference and channel estimation errors vanish [7]. However, the features described above can be reaped under favorable propagation conditions, i.e., the channel vectors between different users should become pairwisely orthogonal as the number of antennas grow [8] and assuming that perfect hardware is deployed.

The use of low quality hardware is desirable in order to make massive MIMO an economically sustainable technological shift, or its total deployment cost will scale to a high level with the number of radio-frequency (RF) front-ends and components. Unfortunately, these low-quality RF components are more prone to hardware imperfections, such as phase noise [9] and in-phase and quadrature-phase imbalance (IQI), which refers to the mismatch between the I and Q branches, i.e., the mismatch between the real and imaginary parts of the complex signal. The latter imperfection occurs due to the limited accuracy of analogue hardware, such as finite tolerance of capacitors and transistors [10]. This leads to a degradation in the overall performance and, therefore, to a deteriorated user experience.

Moreover, in a conventional multiple antenna BS, each radio-frequency (RF) port is connected to a pair of high-resolution ADCs (typically, the in-phase and quadrature signal

components are quantized with resolutions exceeding 10 bits). Scaling such architectures to massive MIMO with hundreds or thousands active antenna elements would result in prohibitively high power consumption and hardware costs. In particular, the hardware complexity and power consumption of ADCs scales roughly exponentially in the number of quantization bits [11]. Thus, an effective solution to keep the power consumption and system costs within desirable limits is to reduce the precision of ADCs (e.g. up to 8 bits). An additional motivation for reducing the resolution of the employed ADCs is to limit the amount of data that has to be transferred over the link that connects the RF components and the baseband-processing unit.

1.1 Aim of the Thesis

Motivated by the above discussion, in this dissertation we study the realizable potential of massive MIMO systems in the aforementioned hardware imperfections, i.e., IQ imbalance and limited-precision ADCs. The specific objectives can be summarized as follows:

- analyse the impact of IQ imbalance on the achievable rate and channel estimation in massive MIMO systems;
- develop a joint estimator of the propagation MIMO channel and IQI parameters at the base station;
- propose a low-complexity IQI compensation scheme in order to mitigate the impact of IQ imbalance;
- analyse the performance of uplink massive MIMO under limited-precision ADCs at the base station.
- propose a near-optimal low-complexity scheme in order to reconstruct the clipped signals at the base station.

1.2 Thesis Outline

The thesis is organized as follows: In Chapter 2, we introduce the multiuser MIMO systems, which is the basis of our theoretical analysis. In Chapter 3, we provide the basics of random matrix theory and its applications in massive MIMO systems. Furthermore, in Chapter 4, characterize the hardware imperfections that are investigated in this dissertation. More precisely, we discuss the system models that capture the effect of the IQ imbalance in MIMO systems and introduce the ADCs used in communication receivers. Finally, we summarize our papers and main contributions in Chapter 5.

Chapter 2

Multiuser MIMO Cellular Systems

Massive MIMO is a MU-MIMO cellular system where the number of BS antennas and the number users are large. In this chapter, we provide the basic background of MU-MIMO systems in terms of communication schemes, channel estimation and signal detection. For the sake of simplicity, we limit our discussions to the single-cell systems.

2.1 Uplink Transmission

We consider the uplink of a single-cell MU-SIMO system, which includes a BS equipped with N antennas communicating with K single-antenna mobile stations (MSs) as shown in Fig. 2.1. Since, K users share the same time-frequency resource, the $N \times 1$ received vector at the BS is the combinations of all signals transmitted from the K users, i.e.,

$$\mathbf{r}_{\mathtt{ul}} = \sqrt{\rho_u} \sum_{k=1}^{K} \mathbf{h}_k x_k + \mathbf{w} \tag{2.1}$$

$$= \sqrt{\rho_u} \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{2.2}$$

where $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ is the channel vector between the kth user and the BS, x_k is the unit-power data symbols transmitted from the k-th UEs, and ρ_u is the average signal-to-noise ratio (SNR). Furthermore, for notational convenience we define the channel matrix $\mathbf{H} \triangleq [\mathbf{h}_1, ...\mathbf{h}_K] \in \mathbb{C}^{N\times K}$ and the transmit vector $\mathbf{x} \triangleq [x_1, ...x_K]^T \in \mathbb{C}^{K\times 1}$. In general, the propagation channel $\mathbf{H}$ is modeled via large-scale fading and small-scale fading. However, in this chapter, we ignore large-scale fading and furthermore we assume that the elements of $\mathbf{H}$ are i.i.d. Gaussian distributed with zero mean and unit variance. Finally, $\mathbf{w}$ is the additive white Gaussian noise (AWGN) and, without loss of generality, we assume that its elements are i.i.d Gaussian distributed with zero mean and unit variance, and independent of $\mathbf{H}$.

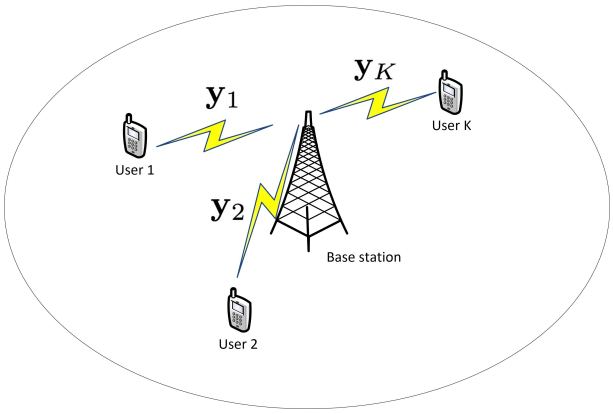

Figure 2.1: A multiuser MIMO cellular system

The BS will coherently detect the signals transmitted from K users by using the received signal vector $\mathbf{r}_{\mathtt{ul}}$ together with the channel state information (CSI). The channel model (2.1) is the multiple-access channel which has the sum-capacity [12]

$$C_{\mathtt{up}} = log_2 det\left(\mathbf{I}_K + p_u \mathbf{H}^H \mathbf{H}\right) \tag{2.3}$$

The aforementioned sum-capacity can be achieved by using the successive interference cancellation (SIC) technique [13]. The SIC scheme is achieved by the receiver decoding the stronger signal first, subtracting it from the combined signal and then decoding the difference as the weaker signal.

2.1.1 Linear Receivers

For the ease of exposition, in this section we assume that the BS has perfect knowledge of the channel.[1]. The BS wants to detect the transmitted signal from K users. To obtain optimal performance, the maximum-likelihood (ML) detect can be used as follows:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{s} \in \mathcal{X}^K} ||\mathbf{r}_{\mathtt{ul}} - \sqrt{\rho_u}\mathbf{H}\mathbf{s}||^2$$

where $\mathcal{X}$ is the finite alphabet of x_k $(k = 1, 2, ..., K)$. The BS has to search over $|\mathcal{X}|^K$ vectors, where $|\mathcal{X}|$ denotes the cardinality of the set $\mathcal{X}$. It has a complexity which is exponential in the number of antennas and modulation size. On the other hand, the BS can use linear detectors in order to reduce the detection complexity. However, these schemes have worse performance compared with ML. Nevertheless, when the BS antennas is large, linear detectors are nearly-optimal.

[1] In the case of imperfect CSI, any linear receiver scheme will utilize the channel estimate to recover the transmitted signal

2.1.1.1 Maximum-Ratio Combining

Maximum-Ratio Combining (MRC) maximizes the received signal-to-noise ratio (SNR) of each stream, ignoring the multiuser interference. In order to detect the transmitted symbol from the kth user, the received signal $\mathbf{r}_{\mathtt{ul}}$ is multiplied by the conjugate-transpose of the channel vector $\mathbf{h}_k$, i.e.,

$$\tilde{\mathbf{r}}_{\mathtt{ul},k} \triangleq \mathbf{h}_k^H \mathbf{r}_{\mathtt{ul}} = \sqrt{\rho_u}||\mathbf{h}_k||^2 x_k + \sqrt{\rho_u} \sum_{i=1, i\neq k}^{K} \mathbf{h}_k^H \mathbf{h}_i x_i + \mathbf{h}_k^H \mathbf{w}. \tag{2.4}$$

The received signal-to-interference-plus-noise ratio (SINR) of the kth stream for MRC is given by

$$\mathtt{SINR}_k^{\mathtt{mrc}} \triangleq \frac{\rho_u ||\mathbf{h}_k||^4}{\rho_u \sum_{i=1, i\neq k}^{K} |\mathbf{h}_k^H \mathbf{h}_i|^2 + ||\mathbf{h}_k||^2}. \tag{2.5}$$

The estimated symbol for the k-th user is given by $\hat{x}_k = \frac{\mathbf{h}_k^H \mathbf{r}_{\mathtt{ul},k}}{\rho_u ||\mathbf{h}_k||^2}$. The implementation of MRC is very simple since the BS multiplies the received vector with the conjugate-transpose of the channel matrix, and this can be realized in a distributed manner. Moreover, notice that for small ρ_u, $\mathrm{SINR}_k \approx \rho_u ||\mathbf{h}_k||^2$. This implies that for low SNR, MRC can achieve the same array gain as in the case of a single-user system. However, the disadvantage of MRC is that it performs poorly in interference-limited scenarios.

2.1.1.2 Zero-Forcing Receiver

In contrast with MRC, Zero-forcing (ZF) receivers cancel out the multiuser interference but neglecting the effect of noise. Particularly, the received vector is multiplied by the pseudo-inverse of the channel matrix $\mathbf{H}$ as

$$\tilde{\mathbf{r}} \triangleq \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{r}_{\mathtt{up}} = \sqrt{\rho_u}\mathbf{x} + \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{w}. \tag{2.6}$$

We see that the post-processing signal in (2.6) is free of multiuser interference. It is worth mentioning that this scheme requires that the number of antennas is greater than the simultaneously served users ($N \geq K$). The received SINR of the kth stream is given by

$$\mathtt{SINR}_k^{\mathtt{zf}} \triangleq \frac{\rho_u}{\left[\left(\mathbf{H}^H \mathbf{H}\right)^{-1}\right]_{kk}}. \tag{2.7}$$

The advantage of ZF receivers is that they can completely null out multiuser interference and their signal processing is relatively simple. However, they perform poorly under noise-limited scenarios since they boost the noise variance (a.k.a. the noise coloring effect). Moreover, ZF receivers have higher implementation complexity than MRC receivers due to the computation of the pseudo-inverse of the channel matrix.

2.1.1.3 Minimum Mean-Square Error Receiver

The linear minimum mean-square error (MMSE) receiver aims to minimize the mean-square error between the estimate and the transmitted signal. More precisely, let $\mathbf{A}$ be an linear detection matrix, then the MMSE receiver is given by

$$\begin{aligned}\mathbf{A}_{\texttt{mmse}} &= \underset{\mathbf{A}\in\mathcal{C}^{N\times K}}{\arg\min}\,\mathbb{E}\left\{||\mathbf{A}\mathbf{r}_{ul}-\mathbf{x}||^2\right\}\\ &= \underset{\mathbf{A}\in\mathcal{C}^{N\times K}}{\arg\min}\sum_{k=1}^{K}\mathbb{E}\left\{|\mathbf{a}_k^H\mathbf{r}_{ul}-x_k|^2\right\}\end{aligned}$$

where $\mathbf{a}_k$ is the kth column of $\mathbf{A}$. Therefore, the kth column of the MMSE detection matrix is

$$\begin{aligned}\mathbf{A}_{\texttt{mmse}} &= \underset{\mathbf{a}_k\in\mathcal{C}^{N\times K}}{\arg\min}\sum_{k=1}^{K}\mathbb{E}\left\{|\mathbf{a}_k^H\mathbf{r}_{ul}-x_k|^2\right\}\\ &= \left(p_u\sum_{i\neq k}^{K}\mathbf{h}_i\mathbf{h}_i^H+\mathbf{I}_N\right)^{-1}\mathbf{h}_k\\ &= c_k\left(p_u\mathbf{H}\mathbf{H}^H+\frac{1}{p_u}\mathbf{I}_N\right)^{-1}\mathbf{h}_k\end{aligned}$$

where

$$c_k=\frac{1}{p_u-p_u\mathbf{h}_k^H\left(\mathbf{H}\mathbf{H}^H+\frac{1}{p_u}\mathbf{I}_N\right)^{-1}\mathbf{h}_k}$$

It is known that the MMSE receiver maximizes the received SINR. Therefore, among the MMSE, ZF, and MRC receivers, MMSE is the best. Furthermore, the received SINR for the MMSE receiver is given by

$$\texttt{SINR}_k^{\texttt{mmse}}\triangleq p_u\mathbf{h}_k^H\left(\sum_{i\neq k}^{K}\mathbf{h}_i\mathbf{h}_i^H+\mathbf{I}_N\right)^{-1}\mathbf{h}_k$$

In the next section, we will briefly overview the principles behind the channel estimation. We assume that the channel stays constant over T symbol durations. During each coherent frame, there are two phases. In the first phase, a part τ of the coherence frame is used for uplink training to estimate the channel of each user. In the second phase, all K users simultaneously send their data to the BS. The BS then detects the transmitted symbols using the channel estimates acquired in the first phase.

2.2 Uplink Training Phase

A part of the coherence frame is used for the uplink training. We assume that each user is assigned an orthogonal pilot sequence of length τ. The pilot sequence used by the K

users can be represented by a $K \times \tau$ matrix $\sqrt{\rho_p}\mathbf{S}$, which satisfies $\mathbf{S}\mathbf{S}^H = \mathbf{I}_K$, where ρ_p is the power of each pilot symbol. Then, the equivalent MIMO signal model for pilot symbol transmission at the BS is given by

$$\mathbf{R}_{\mathtt{p}} = \sqrt{\rho_p}\mathbf{H}\mathbf{S} + \mathbf{W} \tag{2.8}$$

where $\mathbf{R}_{\mathtt{p}}$ represents the $N \times \tau$ received signal matrix during pilot transmission, $\mathbf{W}$ refers to the $N \times \tau$ additive noise matrix at the BS and we set the power of each pilot symbol $\rho_p \triangleq \tau\rho_u$. Assuming that $\tau \geq K$, the estimate of the channel $\mathbf{H}$ can be obtained as

$$\widetilde{\mathbf{R}}_{\mathtt{p}} \triangleq \mathbf{R}_{\mathtt{p}}\mathbf{S}^H = \sqrt{\rho_p}\mathbf{H} + \widetilde{\mathbf{W}} \tag{2.9}$$

where $\widetilde{\mathbf{W}} \triangleq \mathbf{W}\mathbf{S}^H$ is an $N \times K$ complex complex Gaussian matrix whose elements are i.i.d. Gaussian distributed with zero mean and unit variance. Since $\mathbf{H}$ has independent columns, we can estimate each column of $\mathbf{H}$ independently. Let $\tilde{\mathbf{r}}_{p,k}$ and $\widetilde{\mathbf{w}}_k$ be the kth columns of $\widetilde{\mathbf{R}}_{\mathtt{p}}$ and $\widetilde{\mathbf{W}}$, respectively. Then

$$\tilde{\mathbf{r}}_{\mathtt{p},\mathtt{k}} = \sqrt{\rho_p}\mathbf{h}_k + \widetilde{\mathbf{w}}_k. \tag{2.10}$$

With the MMSE, the BS estimates the channel which minimizes the mean-square error. Mathematically speaking, we have

$$\hat{\mathbf{h}}_k^* = \arg\min_{\hat{\mathbf{h}}_k \in \mathbb{C}^N} \mathbb{E}\left\{||\hat{\mathbf{h}}_k - \mathbf{h}_k||^2\right\} \tag{2.11}$$

$$= \mathbb{E}\left\{\mathbf{h}_k|\tilde{\mathbf{r}}_{p,k}\right\} = \frac{\sqrt{\rho_p}}{\rho_p + 1}\tilde{\mathbf{r}}_{p,k}. \tag{2.12}$$

The right hand of the expression (2.12) can be found by rewritting (2.10) as [14]

$$\sqrt{\rho_p}\mathbf{h}_k = \nu\tilde{\mathbf{r}}_{p,k} + \boldsymbol{\theta}_k \tag{2.13}$$

where $\nu = \frac{\rho_p}{\rho_p+1}$ and each element of $\boldsymbol{\theta}_k$ is complex Gaussian distributed variable, with zero mean and variance $\sigma^2_{\theta_k} = \frac{\rho_p}{\rho_p+1}$, uncorrelated with the elements of $\mathbf{h}_k$. Dividing the expression (2.13) by $\sqrt{\rho_p}$ and taking the conditional expectation we get (2.12).

2.3 Downlink Transmission

Downlink is the scenario where the BS transmits signals to all K users. We consider that the BS uses linear precoding techniques to process the signal before transmitting to all users. This requires knowledge of CSI at the BS. By assuming time-division duplexing (TDD) operation, the estimates of CSI are obtained from uplink training. Let $\mathbf{W} \in \mathbb{C}^{N\times K}$ be the linear precoding matrix, then the $N \times 1$ transmit signal vector is given by

$$\mathbf{x} = \sqrt{\rho_d}\mathbf{W}\boldsymbol{d} \tag{2.14}$$

where $\boldsymbol{d}$ is a zero-mean circularly symmetric complex Gaussian $K\times 1$ vector (i.e. $\mathbb{E}\{\boldsymbol{d}\boldsymbol{d}^H\} = \mathbf{I}_K$ and $\mathbb{E}\{\boldsymbol{d}\boldsymbol{d}^T\} = \mathbf{0}_K$) of independent, unit-power data symbols transmitted to the K

UEs and ρ_d is the average transmit power at the BS. To satisfy the power constraint at the BS, $\mathbf{W}$ is chosen such that $\mathbb{E}\{||\mathbf{x}||^2\} = \rho_d$ or equivalently $\mathbb{E}\left\{\mathtt{tr}\left(\mathbf{W}\mathbf{W}^H\right)\right\} = 1$.

The $K \times 1$ received vector at the K UEs is given by

$$\mathbf{r}_{\mathtt{dl}} = \mathbf{H}^T\mathbf{x} + \mathbf{n} = \sqrt{\rho_d}\mathbf{H}^T\mathbf{W}\boldsymbol{d} + \mathbf{n} \tag{2.15}$$

where $\mathbf{H}$ is the $N \times K$ channel matrix that characterizes the propagation environment, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma_w^2\mathbf{I})$ is the additive white Gaussian noise (AWGN).

The channel model (2) is the broadcast channel whose sum-capacity is known to be

$$C_{\mathtt{sum}} = \max_{\{q_k\},\ q_k \geq 0,\ \sum_{k=1}^{K} q_k \leq 1} \log_2 det\left(\mathbf{I}_N + p_d\mathbf{H}^*\mathbf{Q}\mathbf{H}^T\right)$$

where $\mathbf{Q}$ is the diagonal matrix whose kth diagonal element is q_k. The sum-capacity can be achieved by using the dirty-paper coding (DPC).

2.3.1 Linear Precoders

In the downlink, with linear precoding techniques, the signal transmitted from N antennas, x, is a linear combination of the symbols intended for the K users. Let d_k, be the symbol intended for the kth user. Then, the linearly precoded signal vector x is

$$\mathbf{x} = \sqrt{\alpha}\mathbf{W}\mathbf{d}$$

where α is a normalization constant chosen to satisfy the power constraint $\mathbb{E}\{||\mathbf{x}||\} = 1$. Thus,

$$\alpha = \frac{1}{\mathbb{E}\left\{\mathtt{tr}\left(\mathbf{W}\mathbf{W}^H\right)\right\}}$$

The received signal at the kth user is given by

$$\begin{aligned} r_{\mathtt{dl,k}} &= \rho_d\mathbf{h}_k\mathbf{W}\mathbf{d} + n_k \\ &= \rho_d\mathbf{h}_k\mathbf{w}_k d_k + \rho_d\sum_{i\neq k}^{K}\mathbf{h}_k^T\mathbf{w}_i d_i + n_k. \end{aligned}$$

Therefore, the SINR of the transmission from the BS to the kth user is

$$\mathtt{SINR}_k = \frac{\rho_d|\mathbf{h}_k^T\mathbf{w}_k|^2}{\rho_d\sum_{i\neq k}^{K}|\mathbf{h}_k^T\mathbf{w}_i| + 1}$$

Three conventional linear precoders are maximum-ratio transmission (MRT) (also called conjugate beamforming), ZF, and MMSE precoders. These precoders have similar operational meanings and properties as MRC, ZF, MMSE receivers, respectively. Thus, here we just provide the final formulas for these precoders, i.e.,

$$\mathbf{W} = \begin{cases} \mathbf{H}^*, & MRT \\ \mathbf{H}^* \left(\mathbf{H}^T \mathbf{H}^*\right)^{-1}, & ZF \\ \mathbf{H}^* \left(\mathbf{H}^T \mathbf{H}^* + \frac{K}{\rho_d} \mathbf{I}\right)^{-1}, & MMSE \end{cases}$$

Chapter 3

Random Matrix Theory and Wireless Communications

Random matrix theory is widely applied to problems in physics, statistics, data analysis and engineering. In the last few years, a large body of work has emerged in the field of wireless communications and information theory that have not only using random matrix theory results, but also have made fundamental contributions to the field [15]. Tools from random matrix theory have been particularly attractive to researchers for analyzing the performance of massive MIMO systems, where typically the analysis involves random matrices of large dimensions.

3.1 The Role of the Singular Values

Assuming that the channel matrix $\mathbf{H}$ is known at the receiver, the capacity of (2.1) depends on the distribution of the singular values of $\mathbf{H}$ under input power constraints.

The empirical cumulative distribution function of the eigenvalues (also referred to as the spectrum or empirical distribution) of an $N \times N$ Hermitian matrix $\mathbf{A}$ is defined as [15]

$$F_{\mathbf{A}}^{N}(x) \triangleq \frac{1}{N} \sum_{i=1}^{N} 1\{\lambda_i(\mathbf{A}) < x\}$$

where $\lambda_1(\mathbf{A}), ..., \lambda_N(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$ and $1\{\cdot\}$ is the indicator function.

The normalized uplink channel capacity in (2.3) conditioned on $\mathbf{H}$ is [16]

$$\frac{1}{N}C_{\mathtt{up}} = \frac{1}{N}\log det(\mathbf{I} + \mathtt{SNR}\mathbf{H}\mathbf{H}^H) \tag{3.1}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\log\left(1 + \mathtt{SNR}\lambda_i(\mathbf{H}\mathbf{H}^H)\right)\right) \tag{3.2}$$

$$= \int_0^\infty \log(1+\mathtt{SNR}x)dF^N_{\mathbf{H}\mathbf{H}^H}(x) \tag{3.3}$$

with the transmitted signal-to-noise (SNR)

$$\mathtt{SNR} = \frac{N\mathbb{E}\left[||\mathbf{x}||^2\right]}{K\mathbb{E}\left[||\mathbf{n}||^2\right]} \tag{3.4}$$

and with $\lambda_i(\mathbf{H}\mathbf{H}^H)$ equal to the i-th squared singular value of $\mathbf{H}$.

If the channel is known at the receiver and its variation over time is stationary and ergodic, then the expectation of (3.1) over the distribution of $\mathbf{H}$ is the channel capacity (normalized to the number of receive antennas or the number of degrees of freedom per symbol in the CDMA channel).

Another important performance measure is the MMSE achieved by a linear receiver, which determines the maximum achievable output SINR. For an i.i.d. input, the arithmetic mean over the users (or transmit antennas) of the MMSE is given, as function of $\mathbf{H}$ [17]

$$\frac{1}{K}\arg\min_{\mathbf{A}\in\mathcal{C}^{N\times K}}\mathbb{E}\left\{||\mathbf{A}\mathbf{r}_{\mathtt{ul}} - \mathbf{x}||^2\right\} = \frac{1}{K}\mathtt{tr}\left\{\left(\mathbf{I} + \mathtt{SNR}\mathbf{H}^H\mathbf{H}\right)^{-1}\right\} \tag{3.5}$$

$$= \frac{1}{K}\sum_{i=k}^{K}\frac{1}{1+\mathtt{SNR}\lambda_i\left(\mathbf{H}^H\mathbf{H}\right)} \tag{3.6}$$

$$= \int_0^\infty \frac{1}{1+\mathtt{SNR}x}dF^N_{\mathbf{H}^H\mathbf{H}}(x) \tag{3.7}$$

where the expectation in (3.5) is over the transmitted symbols $\mathbf{x}$ and additive noise $\mathbf{n}$.

As we see in (3.3) and (3.7), both fundamental performance measures, namely, capacity and MMSE are with respect to the distribution of the empirical (squared) singular value distribution of the random channel matrix. In the simplest case of $\mathbf{H}$ having i.i.d. Gaussian entries, the density function corresponding to the expected value of $F^N_{\mathbf{H}^H\mathbf{H}}(x)$ can be expressed explicitly in terms of the Laguerre polynomials [15]. Although the integrals in (3.3) and (3.7) lead to explicit solutions, limited insight can be drawn from either the solutions or their numerical evaluation. Deeper insights can be obtained using the tools provided by asymptotic random matrix theory. A rich body of results exists analyzing the asymptotic spectrum of $\mathbf{H}$ as the number of columns and rows goes to infinity while the aspect ratio of the matrix is kept constant.

3.2 Asymptotic Results

Assume a random matrix denoted by $\mathbf{H}$ of size $N \times K$, whose entries are Gaussian i.i.d. random variables (RVs). The element in the i-th row and the j-th column in $\mathbf{H}$ is denoted as $[\mathbf{H}]_{i,j} \sim \mathcal{CN}(0, \frac{1}{N})$. In a massive MIMO system, $\mathbf{H}$ can represent the small-scale Rayleigh fading channel matrix between K user terminals and N BS antennas. As the number of rows and columns in $\mathbf{H}$ grows without bound, i.e., $N, K \to \infty$, while $N/K = \beta$, the empirical cumulative distribution function of the eigenvalues (also called the spectrum) of $\mathbf{H}$ shows interesting convergence properties. Specifically, the spectrum of $\mathbf{H}$ and its functionals become deterministic in the asymptotic limit. This observation leads to the central notion in asymptotic random matrix theory that the empirical distribution of the moments of the eigenvalues of $\mathbf{H}$ and its functionals become deterministic, and this is independent of the distribution of the matrix entries. Specifically, the spectrum of $\mathbf{H}\mathbf{H}^H$ converges almost surely to a non-random distribution function called the Marchenko-Pastur law whose density function is

$$f_\beta(x) = \left(1 - \frac{1}{\beta}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\beta x} \tag{3.8}$$

where $(z)^+ = \max(0, z)$ and

$$a = (1 - \sqrt{\beta})^2, \qquad b = (1 + \sqrt{\beta})^2. \tag{3.9}$$

Analogously, the empirical distribution of the eigenvalues of $\mathbf{H}^H\mathbf{H}$ converges almost surely to a non-random limit whose density function is

$$\tilde{f}_\beta(x) = (1-\beta)\,\delta(x) + \beta f_\beta(x) \tag{3.10}$$

$$= (1-\beta)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\beta x}. \tag{3.11}$$

These results are particularly useful given that eigenvalues of random matrices are used to characterize the performance of communication links (for e.g., MIMO links).

Using the asymptotic spectrum, the following closed-form expression for the limits of (3.1) and (3.5) can be obtained [15]

$$\frac{1}{N}\log det\left(\mathbf{I} + \mathtt{SNR}\mathbf{H}\mathbf{H}^H\right) \to \beta \int_a^b \log\left(\mathbf{I} + \mathtt{SNR}x f_\beta(x)dx\right) \tag{3.12}$$

$$= \beta \log\left(1 + \mathtt{SNR} - \frac{1}{4}\mathcal{F}(\mathtt{SNR}, \beta)\right) \tag{3.13}$$

$$+ \log\left(1 + \mathtt{SNR}\beta - \frac{1}{4}\mathcal{F}(\mathtt{SNR}, \beta)\right) \tag{3.14}$$

$$- \frac{loge}{4\mathtt{SNR}}\mathcal{F}(\mathtt{SNR}, \beta) \tag{3.15}$$

and

$$\frac{1}{K}\mathtt{tr}\left\{\left(\mathbf{I}+\mathtt{SNR}\mathbf{H}^H\mathbf{H}\right)^{-1}\right\} \to \int_\alpha^\beta \frac{1}{1+\mathtt{SNR}x} f_\beta(x)dx$$
$$= 1 - \frac{\mathcal{F}(\mathtt{SNR},\beta)}{4\beta\mathtt{SNR}}.$$

respectively, where $\mathcal{F}(x,y) = \left(\sqrt{x(1+\sqrt{y})^2+1} - \sqrt{x(1-\sqrt{y})^2+1}\right)^2$.

The convergence of the singular values of $\mathbf{H}$ yields several interesting features with engineering significance. The asymptotic analysis is especially useful when the convergence is so fast that, even for small values of the parameters, the asymptotic results come close to the finite-size results. Recent works have shown that the convergence rate is of the order of the reciprocal of the number of entries in the random matrix [18].

Another feature is the insensitivity of the asymptotic eigenvalue distribution to the shape of the p.d.f. of the random matrix entries. This property implies that the results obtained asymptotically hold for any type of fading statistics.

On more example is the ergodic behaviour of the capacity. That is, it suffices to observe a single matrix realization in order to obtain convergence to a deterministic limit. In other words, the eigenvalue histogram of any matrix realization converges almost surely to the average asymptotic eigenvalue distribution. This hardening of the singular values lends operational significance to the capacity formulas even in cases where the random channel parameters do not vary ergodically within the span of a codeword.

Note that the m-th moment of the eigenvalues of $\mathbf{H}$ is calculated as

$$\frac{1}{N}\sum_{n=1}^{N}\lambda_n^m = \frac{1}{N}\mathrm{tr}\{\left(\mathbf{H}\mathbf{H}^H\right)^m\} \tag{3.16}$$

where λ_n denotes an eigenvalue of $\mathbf{H}\mathbf{H}^H$. This implies that the normalized trace of the functionals of $\mathbf{H}$, $\frac{1}{N}\mathrm{tr}\left\{\left(\mathbf{H}\mathbf{H}^H\right)^m\right\}$, becomes deterministic in the asymptotic limit. Even though the convergence of the spectrum is based on the assumption that both N and K become asymptotically large, this result is a good approximation even for small dimensions of $\mathbf{H}$ [19].

We now review some useful limit results about very long random vectors which will be used for the analysis in the rest of the thesis.

Let $\mathbf{x} \triangleq [x_1 ... x_N]^T$ and $\mathbf{y} \triangleq [y_1 ... y_N]^T$ be $N \times 1$ vectors whose elements are independent identically distributed (i.i.d.) random variables (RVs) with $\mathbb{E}\{x_i\} = \mathbb{E}\{y_i\} = 0$, $\mathbb{E}\{|x_i|^2\} = \sigma_x^2$ and $\mathbb{E}\{|y_i|^2\} = \sigma_y^2$. Using the law of large number, we obtain

$$\frac{1}{N}\mathbf{x}^H\mathbf{x} \xrightarrow{a.s.} \sigma_x^2, \ \ as\ N \to \infty$$
$$\frac{1}{N}\mathbf{y}^H\mathbf{y} \xrightarrow{a.s.} \sigma_y^2, \ \ as\ N \to \infty$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence.

Applying the Lindeberg-Lévy central limit theorem, we obtain

$$\frac{1}{\sqrt{N}}\mathbf{x}^H\mathbf{y} \xrightarrow{d} \mathcal{CN}\left(0, \sigma_x^2\sigma_y^2\right), \ \textit{as } N \to \infty$$

where $\xrightarrow{d}$ denotes convergence in distribution.

3.3 Stieltjes Transform

For a wide class of random matrices, the asymptotic eigenvalue distributions are either explicitly known or can be calculated numerically. However, the problem of determining an unknown probability distribution given its moments is addressed using the Stieltjes Transform [15].

DEFINITION 1: Let X be a real-valued RV with distribution F. Then the Stieltjes transform $m(z)$ of F, for $z \in \mathbb{C}$ such that $\Im\{z\} > 0$, is defined as

$$m(z) \triangleq \mathbb{E}\left[\frac{1}{X-z}\right] = \int_{-\infty}^{\infty} \frac{1}{x-z} dF(x). \tag{3.17}$$

The Stieltjes transform of the Marcenko-Pastur law $f_\beta(\cdot)$ in (3.8) is

$$m(z) = \int_a^b \frac{1}{x-z} f_\beta(x)dx \tag{3.18}$$

$$= \frac{1-\beta-z \pm \sqrt{z^2-2(\beta+1)z+(\beta-1^2)}}{2\beta z}. \tag{3.19}$$

The Stieltjes transform of $\tilde{f}_\beta(\cdot)$ in (3.10) is

$$m(z) = \int_a^b \frac{1}{x-z} \tilde{f}_\beta(x)dx \tag{3.20}$$

$$= \frac{-1+\beta-z \pm \sqrt{z^2-2(\beta+1)z+(\beta-1^2)}}{2z}. \tag{3.21}$$

Given $m(z)$, The pdf of X, $p(x)$, can be obtained by applying the Stieltjes inversion formula, which is given as [20]

$$p(x) = \lim_{\omega \to 0^+} \frac{1}{\pi} m(x + j\omega) \tag{3.22}$$

Assuming $F(x)$ has compact support, we can expand $m(z)$ in a Laurent series involving the moments of X. Expanding $\frac{1}{x-z}$ with respect to z, exchanging summation and integration and using analytical extension, (3.17) can be written as

$$m(z) = -\frac{1}{z}\sum_{n=1}^{\infty} \frac{\mathbb{E}\left[X^n\right]}{z^n}. \tag{3.23}$$

If the distribution of X is the averaged empirical eigenvalue distribution of an $N \times N$ random matrix $\mathbf{A}$, then $\mathbb{E}\left[X^k\right]$ can be regarded as the k-th moment $\mathbb{E}\left[\frac{1}{N}\mathrm{tr}\left\{\mathbf{A}^k\right\}\right]$. As a consequence, $m(\cdot)$ can be regarded as a generating function for the moments of the random matrix whose averaged empirical eigenvalue distribution is F_X.

Chapter 4

Hardware Imperfections

4.1 In-phase and Quadrature Components

Massive MIMO systems are built from an excessive number of antenna elements and show great promise for mobile wireless technologies. However, by increasing the number of antennas and associated radio frequency (RF) chains, the size and cost-efficiency of individual RF chains becomes more and more critical. Furthermore, a growing number of wireless standards forces for flexible solutions, which can support several communications applications.

The concept of direct-conversion radio (DCR) [21] for frequency translation is a good candidate for the massive MIMO transceiver structure. First, it is flexible and thus able to operate with several different air interfaces, frequency bands and waveforms [22]. Moreover, it does not need external intermediate frequency (IF) filters and image rejection filters [23]. Instead, the image rejection is provided by the signal processing in the in-phase (I) and quadrature (Q) arm. Therefore, this architecture opens the door to monolithic integration of the analog front-end and, thus, low-cost implementations [24].

The DCR architecture, also referred to as homodyne or zero-IF architecture, however, has some disadvantages compared to more conventionally used heterodyne architectures. These disadvantages include DC offset through self-mixing, $1/f$-noise and severe IQ mismatch [24]. This chapter will focus on the latter impairment which is caused by mutual differences in the components used for frequency translation. These differences result in a phase and/or amplitude imbalance between the I and Q signals, an effect which we will refer to as IQ imbalance.

There exist two models of IQI in a RF transceiver. One is the static amplitude and phase mismatch between the local oscillator (LO) signals used for down- and up-conversion of the I and Q signals, which remain constant over the whole bandwidth and is known as static or frequency-independent IQI. The other is the time impulse response mismatch between the low pass filters on the I and Q branches, which is known as dynamic or frequency-dependent IQI. In this thesis, the effects of dynamic IQI are not considered and we focus on the static IQI model. The latter is a precise description of IQI in narrowband systems, while it contributes the most to IQI originating in the RF stage. On the other

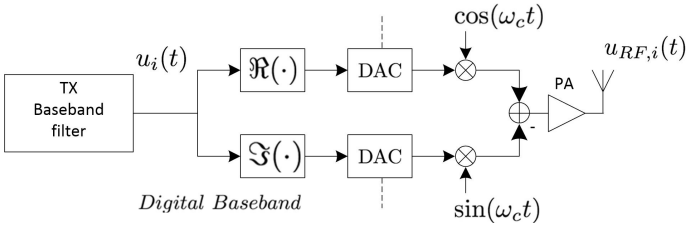

Figure 4.1: Block diagram of a homodyne transmitter [24].

hand, the dynamic IQI is estimated and corrected either in time or frequency domain by adaptive techniques or by using subsample resolution basis functions.

In this chapter, we will consider the influence of IQ mismatch in both the transmitter (TX) and receiver (RX) front-ends. First, Section 4.1.1 introduces the homodyne transceiver structure and Section 4.1.2 shows the influence of IQ mismatch on the transmitted and received signals.

4.1.1 Transmit/Receive front-end architecture

We first consider the up-conversion of the baseband signal $u_i(t)$ in the ith TX branch ($1 \leq i \leq N_T$, where N_T is the total number of transmit antennas), as illustrated in Fig. 3.1. The real and imaginary part of the digital baseband signal are passed through the digital-to-analog converters (DACs). The signal is then up-converted to radio frequency (RF) with carrier frequency f_c, using the quadrature mixing structure as illustrated in the figure. The RF signal pass through the power amplifier (PA), which, we will assume to be perfect with unity gain.

In case of ideal matching between the I and Q branches, the LO signals that multiplies the I and Q branches differ by a 90° phase shift. Thus, they can be expressed as

$$a_Q(t) = \sin(\omega_c t), \tag{4.1}$$
$$a_I(t) = \cos(\omega_c t) \tag{4.2}$$

Using these expressions, the RF TX signal for the i-th branch can be written as

where $\omega = 2\pi f_c$ and where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ give the real and imaginary part of their arguments. The factor 2 is added for notational convenience.

At the rth RX branch ($1 \leq r \leq N_R$, where N_R is the total number of transmit antennas), as illustrated in Fig. 3.2., the received RF signal $y_{RF,r}(t)$ is first amplified by a low-noise amplifier (LNA), which we will assume to be ideal with unity gain. Down-conversion is done again by two 90° phase shifted LO signals at RF f_c. Low-pass filtering is applied in both branches to remove higher order modulation products. Both signals are then passed through the analog-to-digital convertors (ADCs) and combined to form the baseband signal $y_r(t)$, which is input to the baseband RX filter. In the case of ideal matching between the I and Q branch, the LO signals multiplying the I and Q signal

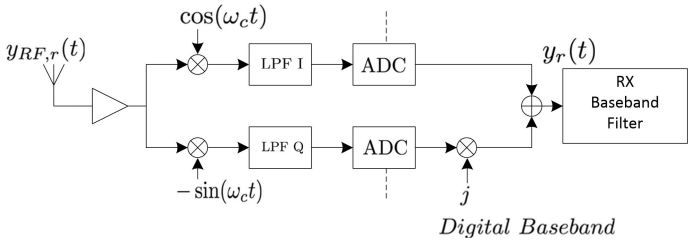

Figure 4.2: Block diagram of a homodyne receiver [24].

again differ by a 90° phase shift. They can be written as

$$b_Q(t) = -\sin(\omega_c t), \tag{4.3}$$
$$b_I(t) = \cos(\omega_c t). \tag{4.4}$$

From (3.4) it can be concluded that the received RF signal on the rth RX branch is given by

$$y_{RF,r}(t) = y_r(t)e^{j\omega_c t} + y_r^*(t)e^{-j\omega_c t}. \tag{4.5}$$

Using (3.5), (3.6) and (3.7) the baseband RX signals are given by

$$y_r(t) = y_{I,r}(t) + jy_{Q,r}(t), \tag{4.6}$$

where, as we define LPF$\{\cdot\}$ to be the low-pass filtering operation, while

$$y_{I,r}(t) = \text{LPF}\{b_I(t)y_{RF,r}(t)\} = \text{LPF}\{\cos(\omega_c)y_{RF,r}(t)\} \tag{4.7}$$
$$= \frac{1}{2}\text{LPF}\{y_r(t)(1 + e^{j2\omega_c t}) + y_r^*(t)(1 + e^{-j2\omega_c t})\} \tag{4.8}$$
$$= \Re\{y_r(t)\}, \tag{4.9}$$

and

$$y_{Q,r}(t) = \text{LPF}\{b_Q(t)y_{RF,r}(t)\} = \text{LPF}\{-\sin(\omega_c)y_{RF,r}(t)\} \tag{4.10}$$
$$= \frac{1}{2}\text{LPF}\{y_r(t)(e^{j2\omega_c t} - 1) + y_r^*(t)(1 - e^{-j2\omega_c t})\} \tag{4.11}$$
$$= \Im\{y_r(t)\}. \tag{4.12}$$

4.1.2 IQ imbalance

The results in Section 4.1.1 show that for a system with ideal I/Q brances the baseband signals are perfectly up-converted in the TX and that the image signal centered around $-f_c$ is perfectly removed by the low-pass filters in the down-conversion. In practical

systems, however, ideal matching between the I and Q branch of the quadrature TX/RX is not possible due to limited accuracy of RF front-end hardware. This will result in phase and amplitude mismatch between the I and Q branch. Several stages in the transceiver structure can contribute to the IQ mismatch, e.g., errors in the nominal 90° phase shift between the LO signals used for up- and down-conversion of the I and Q signals and the difference in amplitude transfer of the total I and Q arms [24]. These imbalances are generally modelled as phase and/or amplitude errors in the LO signal used for up- and down-conversion. The imbalances can be modeled either symmetrical or asymmetrical. In the symmetrical method, each branch (I and Q) experiences half of the phase and amplitude errors, see e.g. [25]. In the asymmetrical method, the I branch is modeled to be ideal and the errors are modeled in the Q branch, see e.g. [26]. Nevertheless, these two methods are equivalent [24]. We will use the asymmetrical model for our analysis in this chapter. For this model the imbalanced LO signals used for up-conversion are given by

$$a_Q(t) = g_T \sin(\omega_c t + \phi_T), \tag{4.13}$$
$$a_I(t) = \cos(\omega_c t) \tag{4.14}$$

where g_T and ϕ_T model the TX gain and phase mismatch, respectively. We can conclude from (3.1) and (3.2) that for perfect matching, these imbalance parameters are given by $g_T = 1$ and $\phi_T = 0$, respectively. The TX RF signal on the ith branch can then be expressed as

$$u_{RF,i}(t) = 2(\Re\{u_i(t)\}\cos(\omega_c t) - \Im\{u_i(t)\} g_T \sin(\omega_c t + \phi_T)) \tag{4.15}$$
$$= e^{j\omega_c t}\left(\Re\{u_i(t)\} + j g_T e^{j\phi_T}\Im\{u_i(t)\}\right) \tag{4.16}$$
$$+ e^{-j\omega_c t}\left(\Re\{u_i(t)\} - j g_T e^{-j\phi_T}\Im\{u_i(t)\}\right). \tag{4.17}$$

By defining the coefficients G_1 and G_2, as

$$G_1 \triangleq (1 + g_T e^{j\phi_T})/2, \tag{4.18}$$
$$G_2 \triangleq (1 - g_T e^{-j\phi_T})/2, \tag{4.19}$$

respectively, $u_{RF,i}(t)$ can be rewritten as

$$u_{RF,i}(t) = (G_1 u_i(t) + G_2^* u_i^*(t))\, e^{j\omega_c t} + (G_1^* u_i^*(t) + G_2 u_i(t))\, e^{-j\omega_c t}. \tag{4.20}$$

It is noted that for perfect TX matching $G_1 = 1$ and $G_2 = 0$ and that (3.22) reduces to (3.4). When we subsequently consider the imbalance on the received side, the imbalanced LO signals used for down-conversion are given by

$$b_Q(t) = -g_R \sin(\omega_c t + \phi_R), \tag{4.21}$$
$$b_I(t) = \cos(\omega_c t), \tag{4.22}$$

where g_R and ϕ_R model the RX gain and phase mismatch, respectively. Note that we can conclude from (3.5) and (3.6) that when there is ideal matching, these imbalance parameters are given by $g_R = 1$ and $\phi_R = 0$, respectively. Down-conversion of the RF RX signal, as expressed by (3.7), then yields

$$\begin{aligned}\hat{y}_r(t) &= \hat{y}_{I,r}(t) + j\hat{y}_{Q,r}(t) && (4.23)\\ &= \text{LPF}\{\cos(\omega_c t) y_{RF,r}(t)\} + j\text{LPF}\{-g_R \sin(\omega_c t + \phi_R) y_{RF,r}(t)\} && (4.24)\\ &= \Re\{y_r(t)\} + j\Im\{g_R e^{-j\phi_T} y_r(t)\} && (4.25)\\ &= K_1 y_r(t) + K_2 y_r^*(t), && (4.26)\end{aligned}$$

Note that the coefficients K_1 and K_2 are given by

$$K_1 \triangleq (1 + g_R e^{-j\phi_R})/2, \quad (4.27)$$

$$K_2 \triangleq (1 - g_R e^{j\phi_R})/2, \quad (4.28)$$

respectively. Again, for perfect matching we find that $K_1 = 1$ and $K_2 = 0$. For that case (3.28) reduces to (3.8).

Finally, in the context of MIMO systems, for the transmit antenna array, (3.22), can be rewritten as a $N_T \times 1$ vector

$$\mathbf{u}_{RF}(t) = \left(\mathbf{G}_1 \mathbf{u}(t) + \mathbf{G}_2^* \mathbf{u}^*(t)\right) e^{j\omega_c t} + \left(\mathbf{G}_1^* \mathbf{u}^*(t) + \mathbf{G}_2 \mathbf{u}(t)\right) e^{-j\omega_c t}. \quad (4.29)$$

where $\mathbf{u}(t) = [u_1(t), ...u_{N_T}(t)]^T$. Moreover, $\mathbf{G}_1$ and $\mathbf{G}_2$ are diagonal matrices defined as

$$\mathbf{G}_1 = (\mathbf{I} + \boldsymbol{g}_T e^{j\boldsymbol{\phi}_T})/2, \quad (4.30)$$

$$\mathbf{G}_2 = (\mathbf{I} - \boldsymbol{g}_T e^{-j\boldsymbol{\phi}_T})/2, \quad (4.31)$$

where $\mathbf{I}$ denotes the identity matrix and where

$$\boldsymbol{g}_T \triangleq \text{diag}\left\{\boldsymbol{g}_{T,1}, ..., \boldsymbol{g}_{T,N_T}\right\} \quad (4.32)$$

$$\boldsymbol{\phi}_T \triangleq \text{diag}\left\{\boldsymbol{\phi}_{T,1}, ..., \boldsymbol{\phi}_{T,N_T}\right\}, \quad (4.33)$$

are the diagonal matrices contain the TX amplitude and phase mismatches.

Similarly, after down-conversion with the imbalanced quadrature RX, the received baseband signal $N_R \times 1$ vector is given by

$$\hat{\mathbf{y}}(t) = \mathbf{K}_1 \mathbf{y}(t) + \mathbf{K}_2^* \mathbf{y}^*(t) \quad (4.34)$$

where $\mathbf{y}(t) = [y_1(t), ...y_{N_R}(t)]^T$ and

$$\mathbf{K}_1 = (\mathbf{I} + \boldsymbol{g}_R e^{-j\boldsymbol{\phi}_R})/2, \quad (4.35)$$

$$\mathbf{K}_2 = (\mathbf{I} - \boldsymbol{g}_R e^{j\boldsymbol{\phi}_R})/2, \quad (4.36)$$

with

$$\boldsymbol{g}_R \triangleq \text{diag}\left\{\boldsymbol{g}_{R,1}, ..., \boldsymbol{g}_{R,N_T}\right\} \quad (4.37)$$

$$\boldsymbol{\phi}_R \triangleq \text{diag}\left\{\boldsymbol{\phi}_{R,1}, ..., \boldsymbol{\phi}_{R,N_T}\right\}, \quad (4.38)$$

are the diagonal matrices contain the RX amplitude and phase mismatches.

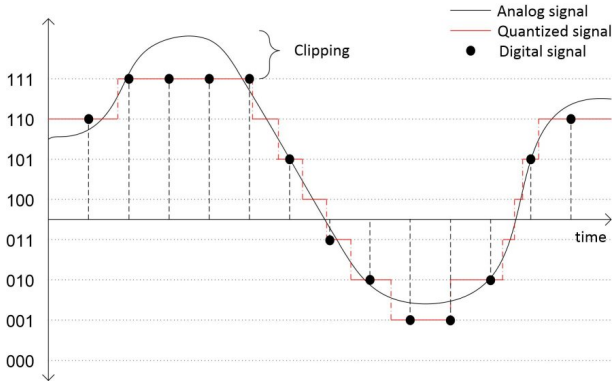

Figure 4.3: Quantization in a 3-bit ADC to a mid-rise uniform grid of $2^3 = 8$ quantization levels.

4.2 Analog-to-digital Converter

Digital signal processing is an integral part of all modern cellular systems [27]. In the uplink, in order to process data digitally, the received signal at the BS has to be converted into the digital domain, which requires conversion in time and amplitude. The device that performs these operations is called an analog-to-digital converter (ADC).

More specifically, an ADC can be modelled as two processes: sampling and quantization. Sampling converts a time-varying voltage signal into a discrete-time signal, a sequence of real numbers. Quantization replaces each real number with an approximation from a finite set of discrete values. As an example consider an ADC with sampling rate f_s Hz and a resolution of b bits. Then, the ADC maps each sample of a continuous-time, continuous-amplitude signal to one out of 2^b possible quantization labels, by operating $f_s \cdot 2^b$ conversion steps per second.

To ensure that the input to the sampling circuit comply (at least approximately) with the sampling theorem, the analog input signal is passed through an anti-aliasing filter (a low-pass filter) prior to the sampling circuit [28]. For the rest of the thesis, It is assumed that the sampling circuit is ideal, i.e., the amplitude of the output of the sampling circuit at any sampling instant is exactly the amplitude of the input signal. Furthermore, it is assumed that the anti-aliasing filter is an ideal low-pass filter with a cutoff frequency that equals the symbol rate, such that any out-of-band noise present in the analog input does not enter into the sampling circuit.

While the sampling operation incurs no loss of information for band-limited signals, the quantization will cause an error between the input and output of the quantizer, which can be made smaller by increasing the resolution of the quantizer. However, increasing the resolution of an ADC also increases the consumed power and complexity of the circuit.

Often the design of a quantizer involves supporting only a limited range of possible output values and performing clipping to limit the output to this range whenever the input exceeds the supported range. The error introduced by this clipping is referred to as overload distortion. Within the extreme limits of the supported range, the amount of

spacing between the selectable output values of a quantizer is referred to as its granularity, and the error introduced by this spacing is referred to as granular distortion. It is common for the design of a quantizer to involve determining the proper balance between granular distortion and overload distortion. For a given supported number of possible output values, reducing the average granular distortion may involve increasing the average overload distortion, and vice versa. A technique for controlling the amplitude of the signal (or, equivalently, the quantization step size Δ) to achieve the appropriate balance is the use of automatic gain control (AGC). However, in some quantizer designs, which are not studied in this thesis, the concepts of granular error and overload error may not apply (e.g., for a quantizer with a limited range of input data or with a countably infinite set of selectable output values).

Chapter 5

Contributions

This thesis consists of three main contributions. The first two investigate the impact of IQI on both conventional and massive MIMO systems. The third elaborates on the effect of clipping distortion in limited-precision quantizers on massive MIMO systems. In Section 5.1, we list the papers that are appended to this thesis and summarize our contributions.

5.1 Included Publications

1. **Paper A: "IQ Imbalance in Multiuser Systems: Channel Estimation and Compensation"**
   In this paper, we consider the uplink of a single-cell multi-user single-input multiple-output (MU-SIMO) system with in-phase and quadrature-phase imbalance (IQI). This scenariois of particular importance, especially, in MU-SIMO systems with large antenna arrays, where the deployment of lower cost, lower-quality components is desirable to maintain their total implementation cost to affordable levels. Particularly, we investigate the effect of RX IQI on the performance of MU-SIMO systems with large antenna arrays employing maximum ratio combining (MRC) receivers. In order to study how IQI affects channel estimation, we derive a new channel estimator for the IQI-impaired model and show that IQI can downgrade the spectral efficiency (SE) of MRC receivers. Moreover, a novel pilot-based joint estimator of the augmented MIMO channel matrix and IQI coefficients is described and then, a low-complexity IQI compensation scheme is proposed which is based on the IQI coefficients' estimation and it is independent of the channel gain. The performance of the proposed compensation scheme is analytically evaluated by deriving a tractable approximation of the ergodic SE assuming transmission over Rayleigh fading channels with large-scale fading. Finally, by deriving asymptotic power scaling laws, and proving that the SE loss due to IQI is asymptotically independent of the number of BS antennas, we show that massive MIMO is resilient to the effect of IQI.

2. **Paper B: "Massive MIMO Systems with IQ Imbalance: Channel Estimation and Sum Rate Limits"**
   This paper studies the impact of in-phase and quadrature imbalance (IQI) on single-cell multiuser multiple-input multiple-output (MU-MIMO) systems with large antenna arrays. Moreover, we consider a time-division duplex (TDD) system where we assume uplink/downlink channel reciprocity in the downlink precoding design. First, we investigate the effect of transceiver IQI on the uplink channel estimation by deriving the linear minimum-mean-square-error (LMMSE) estimator for the IQ-impaired model and prove that only the receiver IQI at the base station (BS) limits the estimation accuracy. Then, we study the impact of uplink/downlink channel mismatch on the downlink rate caused by different IQI at the BS and user equipments (UEs). We prove that the achievable downlink rate of each UE is limited either by the receiver IQI at the UEs or jointly by the transmit and receive IQI at the BS, when there is mismatch between the uplink and downlink channels.

3. **Paper C: "Reconstruction of Clipped Signals in Massive MIMO Systems"**
   This paper considers the uplink of a single-cell multiuser massive multiple-input multiple-output (MIMO) system. Each receiver antenna of the base station is assumed to be equipped with a pair of analog-to-digital converters (ADCs) to quantize the real and imaginary part of the received signal. We first derive the minimum mean-square-error (MMSE) receiver for quantized MIMO systems, which reconstructs the quantized received vector. Motivated by the high computational complexity of the MMSE receiver, we propose a novel clipping-aware MMSE-based receiver (CA-MMSE) with significantly less computational complexity, which reconstructs only the clipped samples of the quantized received signal. On this basis, we present a low-complexity iterative algorithm to implement the proposed receiver along with its convergence analysis. Next, we extend the CA-MMSE receiver to the channel and data estimation for quantized multiuser MIMO systems and additionally propose the quantization-aware weighted zero-forcing (QA-WZF) receiver based on the Bussgang's theorem. Our results demonstrate that the CA-MMSE is almost identical to the MMSE receiver and thus, it is sufficient to reconstruct only the clipped samples. Moreover, when the length of channel training is higher than the number of users, then, the CA-MMSE is shown to outperform existing quantization-aware linear receivers, while at high SNR the mean square error of the QA-WZF is near to the CA-MMSE receiver.

References

[1] Ericsson, "Ericsson mobility report: On the pulse of the networked society," Nov. 2014. [Online]. Available: http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–46, Jan. 2013.

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

[4] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[6] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[7] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[8] J. Chen, "When does asymptotic orthogonality exist for very large arays?" in *Proc. IEEE GLOBECOM*, Dec. 2013.

[9] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 711–723, Feb. 2015.

[10] A. Hakkarainen, J. Werner, K. Dandekar, and M. Valkama, "Widely-linear beamforming and RF impairment suppression in massive antenna arrays," *J. Commun. Netw.*, vol. 15, no. 4, pp. 383–397, Aug. 2013.

[11] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[12] ——, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[13] D. N. C. Tse and P. Viswanath, "Fundamentals of wireless communications," *Cambridge, UK: Cambridge University Press*, 2005.

[14] D. Gu and C. Leung, "Performance analysis of transmit diversity scheme with imperfect channel estimation," *IEE Electron. Lett.*, vol. 39, pp. 402–403, Feb. 2003.

[15] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004. [Online]. Available: http://dx.doi.org/10.1561/0100000001

[16] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[17] S. Verdú, "Multiuser detection," *Cambridge, UK: Cambridge University Press*, 1998. [Online]. Available: http://dx.doi.org/10.1561/0100000001

[18] Z. D. Bai, "Convergence rate of expected spectral distributions of large random matrices. Part i: Wigner matrices," *Annals of Probability*, vol. 21, no. 2, pp. 625–648, 1993.

[19] S.Wagner, R. Couillet, M. Debbah, and D. T. Slock, "Large system analysis of linear precoding in correlated miso broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[20] J. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," *J. of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, 1995.

[21] A. A. Abidi, "Direct-conversion radio transceivers for digital communications," *IEEE Journ. of Solid-State Circuits*, vol. 30, pp. 1399–1410, Dec. 1995.

[22] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell System Techn. Journal*, vol. 45, pp. 1775–1796, 1966.

[23] B. Razavi, "Design considerations for direct-conversion receivers," *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 44, no. 6, pp. 428–435, June 1997.

[24] T. Schenk, *RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation.* Springer Netherlands, 2008.

[25] B. Razavi, *RF Microelectronics.* ser. Prentice-Hall Communications Engineering and Emerging Technologies Series. Prentice Hall, 1998.

[26] M. R. M. Valkama and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," *IEEE Trans. Signal Process.*, vol. 49, pp. 2335–2344, Oct. 2001.

[27] B. Widrow and I. Kollár, "Quantization noise: Roundoff error in digital computation, signal processing, control, and communications," *Cambridge Univ. Press*, 2008.

[28] F. Maloberti, "Data converters," *Springer*, 2007.

Part II

Included papers