Linköping Studies in Science and Technology
Dissertations, No. 1896

High-End Performance with Low-End Hardware

Analysis of Massive MIMO Base Station Transceivers

Christopher Mollén

Division of Communication Systems
Department of Electrical Engineering (ISY)
Linköping University, 581 83 Linköping, Sweden
www.commsys.isy.liu.se

Linköping 2018

This is a Swedish Doctor of Philosophy thesis.
The Doctor of Philosophy degree comprises 240 ECTS credits of postgraduate studies.

Cover: A heat map showing the slow variations of the received power in the right adjacent band in an area around a user, to whom a massive MIMO base station is beamforming a signal using nonlinear power amplifiers. The scale shows the length of half a wavelength, which is $\lambda/2 \approx 7.5$ cm at a carrier frequency of 2 GHz. The setup is explained in Paper F, where an excerpt of the cover is shown in color in Figure 4 on page 257.

High-End Performance with Low-End Hardware

© 2017 Christopher Mollén, unless otherwise stated.

ISBN 978-91-7685-388-7
ISSN 0345-7524
Printed in Sweden by LiU-Tryck, Linköping 2017

Abstract

Massive MIMO (multiple-input–multiple-output) is a multi-antenna technology for cellular wireless communication, where the base station uses a large number of individually controllable antennas to multiplex users spatially. This technology can provide a high spectral efficiency. One of its main challenges is the immense hardware complexity and cost of all the radio chains in the base station. To make massive MIMO commercially viable, inexpensive, low-complexity hardware with low linearity has to be used, which inherently leads to more signal distortion. This thesis investigates how the degenerated linearity of some of the main components—power amplifiers, analog-to-digital converters (ADCs) and low-noise amplifiers—affects the performance of the system, with respect to data rate, power consumption and out-of-band radiation. The main results are: Spatial processing can reduce PAR (peak-to-average ratio) of the transmit signals in the downlink to as low as 0 dB; this, however, does not necessarily reduce power consumption. In environments with isotropic fading, one-bit ADCs lead to a reduction in effective signal-to-interference-and-noise ratio (SINR) of 4 dB in the uplink and four-bit ADCs give a performance close to that of an unquantized system. An analytical expression for the radiation pattern of the distortion from nonlinear power amplifiers is derived. It shows how the distortion is beamformed to some extent, that its gain never is greater than that of the desired signal, and that the gain of the distortion is reduced with a higher number of served users and a higher number of channel taps. Nonlinear low-noise amplifiers give rise to distortion that partly combines coherently and limits the possible SINR. It is concluded that spatial processing with a large number of antennas reduces the impact of hardware distortion in most cases. As long as proper attention is paid to the few sources of coherent distortion, the hardware complexity can be reduced in massive MIMO base stations to overcome the hardware challenge and make massive MIMO commercial reality.

Sammanfattning

Massiv MIMO (eng: multiple-input–multiple-output) är en flerantennsteknologi för cellulär trådlös kommunikation, där basstationen använder ett stort antal individuellt styrbara antenner för att multiplexa användare i rummet. Denna teknologi kan tillhandahålla en hög spektral effektivitet. En av dess främsta utmaningar är den enorma hårdvarukomplexiteten och kostnaden hos basstationens alla radiokedjor. För att massiv MIMO skall bli kommersiellt attraktivt, måste billiga, enkla hårdvarukomponenter med låg linjäritet användas, vilket oundvikligen leder till mer signaldistorsion. Denna avhandling undersöker hur den försämrade linjäriteten hos några av huvudkomponenterna – effektförstärkare, analog-digital-omvandlare (AD-omvandlare) och lågbrusförstärkare – påverkar systemets prestanda, i termer av datatakt, effektförbrukning och utombandsstrålning. Huvudresultaten är: Rumslig signalbehandling kan reducera sändsignalernas toppvärde i nerlänken ända ner till 0 dB, vilket dock inte nödvändigtvis minskar effektförbrukningen. I miljöer med isotrop fädning leder enbits-AD-omvandlare till 4 dB lägre signal-till-interferens-och-brus-förhållande i upplänken, och fyrabits-AD-omvandlare ger en prestanda nära den ett system utan kvantisering kan uppnå. Ett analytiskt uttryck för strålningsmönstret för distorsionen från icke-linjära effektförstärkare härleds. Det visar hur distorsionen till viss del lobformas, att dess förstärkning aldrig är starkare än förstärkningen för den önskade signalen och att distorsionens förstärkning minskar med ett högre antal betjänade användare och ett högre antal kanaltappar. Icke-linjära lågbrusförstärkare ger upphov distorsion som delvis kombinerar koherent och begränsar det möjliga signal-till-brus-och-interferens-förhållandet. Slutsatsen är att rumslig signalbehandling med ett stort antal antenner reducerar hårdvarudistorsionens inverkan i de flesta fall. Så länge som de få källorna till koherent distorsion ges tillbörlig uppmärksamhet, kan hårdvarukomplexiteten minskas i basstationer för massiv MIMO för att övervinna hårdvaruutmaningen och göra massiv MIMO kommersiell verklighet.

摘要

蜂窩無線通訊領域中的大規模多天線技術以多個單獨可控的天線通過空間複用的方式服務多個用戶。如是可以大幅提高頻譜效率。實現此技術的主要難題在於基站所用射頻單元的極大複雜度及成本。爲使大規模多天線技術適用在商業系統中，需使用導致失真的低複雜度低成本的非線性硬件。本文探討若將一些主要部件——功放、模數轉換器、低噪聲放大器——的線性程度降低，系統性能是如何受到影響的，即系統的速率、功耗、帶外泄露等指標。主要的結果爲：空間信號處理可以降低下行信號的峯均比，直至0分貝；然而低峯均比不一定能夠降低功耗。用一比特模數轉換器使上行的信干噪比減少4分貝；用四比特模數轉換器可在各向同性衰落的環境裏實現接近無量化系統的性能。本文推導出非線性功放失真輻射方向的解析公式。該公式展示失真在某種程度上會被波束成形的；具體而言，失真的波束成形增益不大於有效信號的增益，波束成形增益會根據服務用戶數量和信道階數的增長而降低。非線性低噪聲放大器引起的失真，一部分會相干地合併，因此會限制信干噪比的增長。結論爲多天線的空間信號處理可以減少硬件失真的影響。只要適當地處理少數相干失真的來源，大規模多天線基站可以降低硬件複雜度，解決硬件難題，使大規模多天線技術成功地應用在商業系統中成爲現實。

Till Mor och Far.

&

致瑜杰，吃煞儂!

Contents

Acknowledgments

Just like water moulds to the shape that is submerged in it, good supervision is skilfully adjusted to the nature of the student. I am very grateful to Prof. Erik Larsson for his knowledgeable guidance that has enabled me to progress and develop as a researcher and as a person. The research findings herein are the fruits of his teaching. Facts and hard knowledge apart, I have also learnt from his principled attitude to research and structured way of work. Without moving, water lets one sink. With some effort, however, being taught how to swim has been a pleasure.

An inspiration, especially during the visit to his group in the summer 2015, has been my co-supervisor Prof. Thomas Eriksson, who is a wellspring of ideas. Dr. Ulf Gustavsson, commonly known as the man with the magnificent beard, has always readily picked up my phone calls to answer my many questions on complicated hardware matters. Time-varying, nonlinear and with multidimensional memory kernels, reality would have been hopelessly recondite without Prof. Eriksson and Dr. Gustavsson.

I thank Prof. Robert Heath, Jr., for expanding my research perspectives, for making research fun and for having me as his visiting student the academic year 2015–16, and Prof. Choi Junil for his careful supervision during this visit. The research visit to Prof. Heath's group at the University of Texas at Austin, USA, was made possible by the generous scholarships from the Fulbright Commission, Ericsson's Research Foundation, Stiftelsen Blanceflor and Ingenjörsvetenskapsakademien's Hans Werthén Fond.

Another interesting research visit that I have been fortunate enough to have made was that to the sixth information coding laboratory at South West Jiao-Tong University in Chengdu, China, during the summer 2017, where the last parts of this thesis came together.

A majority of my time—from the very first day of our doctoral studies—I have spent together with my colleague and friend Marcus Karlsson, with whom I share office. To have somebody wise that is always available to discuss matters with—be it commas, shadow fading or incomprehensible

social situations—is a rare source of great help and comfort.

The steady support from my family is what makes it possible to always persist in my endeavors. When support is not enough, some insistent pushing is indeed an effective means to reach higher. Yu-Jie, you make me who I am.

These extraordinary years at the university, one third of my life so far, have been filled with memorable experiences: fascinating lectures, faraway exchanges, bustling conferences, late-night deadlines, dancing lions, and so much more. During these years, I have had the good fortune to come into contact with many colleagues—motivating, interesting, smart. It is in their presence I have been able to produce this piece of work.

I recognize the privilege that, in our country, where everybody is encouraged to pursue learning and where anybody truly is allowed access to higher education, the path to knowledge and research, albeit strenuous, is straight. I therefore bow to my family, teachers, every member of our research group Kommunikationssystem and the people of Sweden—it is to you I owe this thesis.

Christopher Mollén

Linköping, January 2018

Populärvetenskaplig sammanfattning

Massiv MIMO (eng: Multiple-Input–Multiple-Output) är en trådlös transmissionsteknik för mobil kommunikation där basstationen använder ett hundratal eller fler samarbetande antenner. Genom att synkront kontrollera signalerna vid varje antenn, kan man: (i) förbättra den mottagna signalstyrkan och leverera en högre datatakt, samt (ii) skicka och ta emot ett stort antal parallella dataströmmar från flera mobilenheter samtidigt över samma frekvensband och på så sätt utnyttja vårt begränsade frekvensspektrum mer effektivt.

Enär massiv MIMO ställer låga krav på hårdvaran i mobilenheterna, är basstationens hårdvara med de många samarbetande antennerna mycket komplex. För att massiv MIMO skall kunna användas i uppgraderingen av vårt allt viktigare samhällstäckande mobila nätverk, så att framtidens många nya tekniska lösningar kan stödjas utan att öka kostnaderna för användarna, måste basstationens hårdvara förenklas.

Denna avhandling undersöker möjligheten att bygga basstationer till massiv MIMO av enkel och billig hårdvara. Genom teoretisk och numerisk analys av signalöverföringen visas att prestandan hos massiv MIMO är robust mot den hårdvarudistorsion som enkel hårdvara ger upphov till och att många specialiserade hårdvarufunktioner i basstationen inte bara kan förenklas, utan rentav elimineras eller ersättas av simpla, grundläggande komponenter. Ett exempel är omvandlingen från analoga till digitala signaler, där det visar sig att en enkel komparator kan göra jobbet.

Att enkla komponenter kan användas, trots att de medför svår signal distorsion, innebär att det är möjligt att bygga praktiska basstationer med hundratals antenner för massiv MIMO och att tekniken kan göras billig nog att tjäna kommersiella syften. Med massiv MIMO kommer allt fler och allt enklare apparater att kunna betjänas trådlöst med höghastighetsuppkopplingar till rimliga priser, vilket är nödvändigt för att driva vidare den pågående utvecklingen av nya allmännyttiga trådlösa applikationer.

List of Abbreviations

ACLR	adjacent-band leakage ratio
ADC	analog-to-digital converter
AGC	automatic gain control
AWGN	additive white Gaussian noise
bpcu	bits per channel use
CTCE	continuous-time constant-envelope precoding
DL	downlink
DPD	digital predistortion
DAC	digital-to-analog converter
FDD	frequency-division duplex
i.i.d.	independent and identically distributed
LMMSE	linear minimum mean-square error
LNA	low-noise amplifier
MIMO	multiple-input–multiple-output
MRD	maximum-ratio decoding
MRP	maximum-ratio precoding
MSE	mean-square error
NMSE	normalized mean-square error
OFDM	orthogonal frequency-division multiplexing
OOB	out-of-band
PAR	peak-to-average ratio
QAM	quadrature-amplitude modulation
RZFD	regularized zero-forcing decoding
RZFP	regularized zero-forcing precoding
SINR	signal-to-interference-and-noise ratio
SISO	single-input–single-output
SNR	signal-to-noise ratio
TDD	time-division duplex
UL	uplink
ZFD	zero-forcing decoding
ZFP	zero-forcing precoding

Chapter 1

Introduction

A teenager heads to a distant part of the globe to live and study in a foreign culture that is largely different from the one she grew up in. Such a venture is no longer seen as a great expedition, but rather as a commonplace trip. Why is that? – The parents can talk to and see their child over their smartphone using a messenger application at any time. The teenager has beforehand got a good grasp of her future living circumstances from the Internet. She can navigate and obtain information about her new neighborhood in her own language on the go with the GPS and the cellular network.

The possibility to communicate electronically has obviously changed modern life and become an integral part of it. With the small example above, I want to point out one of its greatest benefits: uncensored global connectivity brings people closer, promotes intercultural understanding and enables borderless exchange of ideas, which in turn has the potential to reduce the risk of conflicts—both global and personal—and speed up scientific and cultural development.

With this as motivation, the research in this thesis aims at giving a better understanding of the practical implementation of *massive MIMO*—the technology that has the potential to replace today's base stations and enhance our wireless communication systems to enable the increasing data traffic load, allow for higher user densities and open up for new functionality requirements that the surging use of wireless technology is expected to demand in a near future. A better understanding of how to implement the theoretical idea of massive MIMO in practical hardware is essential for its adoption in the continued development of our *networked society*.

1.1 General Background

Massive MIMO, as was envisioned in [1], is a communication technology, where a base station is equipped with hundreds of antennas and concurrently communicates with multiple single-antenna users over the same time and frequency resource. By spatial multiplexing of many users, massive MIMO can increase the data rate that the users are served with by orders of magnitude compared to conventional systems, without using more frequency spectrum, and possibly also without using more power. Furthermore, massive MIMO can provide uniformly good service to all users in a large area, both to users on the cell edge and to users near the base station. These qualities make massive MIMO a good choice of technology for the evolution of today's wireless communication systems to meet the new and greater demands of the future [2, 3].

The main qualities of massive MIMO are:

array gain that grows with the number of antennas, which improves the signal quality and lowers the amount of power that has to be radiated.

spatial multiplexing that makes it possible to concurrently serve multiple users at the same time over the same frequencies, which enables high sum rates.

simple handsets that only have a single antenna and do not perform any complicated channel equalization, which can allow for the integration of small, low-power, low-cost devices into the cellular system.

linear signal processing that makes the baseband processing of the base station feasible in terms of computational complexity and, in many cases, gives a performance that is close to the optimal, highly complex dirty paper coding [4].

There are other technologies also considered for the development of today's wireless communication systems. Many of them are complements to massive MIMO and can serve to further increase the data throughput and coverage in a given area.

For example, by densifying the base station deployment to obtain smaller cells, one can increase the per-area throughput [5, 6], but decreasing the cell size too much increases interference and deployment costs.

By using new spectrum at high frequencies, so called millimeter-wave communication, one can access larger bandwidths and thus increase data rates [7]. However, radio waves are easily blocked by obstacles at higher

frequencies, which can make millimeter wave communication difficult in non-line-of-sight scenarios.

Cell-free massive MIMO [8], though, is not a complement to massive MIMO, but rather a different way of deploying the transmitters of the base stations. By spreading out individually controllable antennas over a large area, these antennas can cooperate to serve all users in that area by coherent transmission and reception. Since all antennas serve all users, the concept of a "cell" is not defined and hence the name.

Prototypes of massive MIMO base stations have already been built. Some of the earlier testbeds developed at academic research institutes are the ones at: Rice University (Argos) [9], Lund University [10] and Bristol University [11]. Among industrial research institutes are the ones at: Samsung, Nutaq and Facebook. These testbeds have shown that the theoretical benefits of massive MIMO are real. However, their implementation has been expensive.

To bring cost down and make massive MIMO commercially viable, its base stations have to be built from inexpensive, low-end hardware. The possibility and consequences of using low-end hardware—power amplifiers, analog-to-digital converters (ADCs) and low-noise amplifiers—on the system are studied in this thesis. It is found that massive MIMO is robust against the imperfections of low-end hardware.

The topics covered in this thesis can be divided into the following three parts.

waveforms How different modulation formats of the transmit signals in the downlink compare when the effects of low-end hardware are taken into account.

analog-to-digital converters How the use of ADCs with extremely low resolutions affects the system performance.

amplifiers How the distortion from nonlinear amplifiers at the base station, both power amplifiers for the downlink and low-noise amplifiers in the uplink, is beamformed in-band and out-of-band and how it affects the decoding.

The thesis shows that the spatial processing from the many base station antennas reduces the impact of the distortion from different low-end components. Some terms of the distortion do combine coherently however, which can cause performance degradation to the own system, and to other wireless systems operating in the vicinity in the downlink. Even though these coherent terms are not suppressed by the spatial processing, in many cases with

a finite number of base station antennas, however, the coherent terms are negligible and the distortion can be said to combine noncoherently.

Hence, not all sources of hardware distortion can be suppressed by an increased number of antennas. Nevertheless, the use of more antennas does not make the impact of hardware distortion worse—the array gain of the distortion is never greater than that of the desired signal. Since the number of antennas is large, this thesis supports the claim that low-end hardware can be used in massive MIMO, if only proper care is taken to the cases, where coherent combining might arise.

1.2 Related Work

Previous work has analyzed hardware imperfections by collectively modeling them as a signal distortion that can be described by a simple parametric function. This approach is used in [12] and [13] for example. In [12], both the downlink and uplink are studied and the impact of the hardware is treated as additive uncorrelated noise. In [13], the uplink is investigated and a refined model treats the hardware imperfections of the base station as multiplicative phase drifts, additive distortion noise and noise amplification, which should model the effects of the ADCs, LNAs and the oscillators. A similar approach to [12] is used to model hardware imperfections in [14]. In [14], more sophisticated measurement-based models are used to show that the effect of in-band distortion on system performance under certain circumstances are in-line with what is predicted from simple models similar to the one in [12]. These studies argued that the hardware quality can be degraded in massive MIMO and that the more antennas the base station has the less accurate its hardware can be allowed to be, since the uncorrelated distortion combines noncoherently and vanishes when more base station antennas are used.

An ostensible weakness of an additive, uncorrelated noise model for hardware impairments is that many types of signal distortion are deterministic functions of the input signal. Even if the input signal is modeled as stochastic, the distortion will depend on the input signal and the distribution of the distortion will not be accurately described as an independent Gaussian. Indeed, this thesis shows that uncorrelated-distortion models are too simplistic and fail to accurately describe some fundamental phenomena that arise due to hardware imperfections. Distortion from nonlinear base station hardware only partially combines noncoherently. Some terms, in fact, do combine coherently. In accordance with the results in [14], however, it is shown that there are cases, where the overall impact of distortion on the effective data

rate is comparable to what the uncorrelated-distortion model predicts. Even if the coherent terms will prevent the distortion from vanishing completely as the number of base station antennas is increased, it is important to note that the performance, notwithstanding, is improved by employing a larger number of antennas. Massive MIMO is resilient to hardware imperfections, its many antennas are just not able to completely remove all distortion.

A unique feature of massive MIMO is the possibility to use spatial degrees-of-freedom to do crest-factor reduction without causing any distortion in the receive signal. This crest-factor reduction comes at the cost of an increased transmit power however. When crest-factor reduction is done together with the symbol precoding, so called *low-PAR (peak-to-average ratio) precoding*, the amplitude distribution of the transmit signals can be made more hardware friendly. The first low-PAR precoding method for massive MIMO was presented in [15], where a single-user precoder was proposed. This method was extended, first to a multiuser precoder in [16] and then to a multiuser precoder for frequency-selective channels in [17], which is called the *discrete-time constant-envelope precoder* in this thesis. Two other low-PAR precoders are presented in [18, 19], which can control the trade-off between crest-factor reduction and increased transmit power. Low-PAR precoders are especially interesting for massive MIMO, because conventional precoders result in transmit signals with high PAR that are heavily affected by nonlinear low-end hardware. In this thesis, low-PAR precoding is evaluated against conventional precoding methods in terms of spectral efficiency and the power consumption of the amplifiers of the base station. A new low-PAR precoder is also proposed—the *continuous-time constant-envelope precoder*—which produces continuous-time transmit signals with 0 dB PAR, i.e. passband signals with constant envelope but time-varying phase. Such signals can be amplified with high power efficiency in highly nonlinear inexpensive amplifiers without causing distortion and spectral regrowth.

The use of one-bit ADCs in massive MIMO, was initially studied in [20]. The use of one-bit ADCs in other MIMO systems had previously been studied in, e.g., [21, 22] and for millimeter wave MIMO systems in, e.g., [23]. The feasibility of one-bit ADCs, in terms of achievable rate, was studied in [22], where it was shown that one-bit ADCs only lead to a small capacity reduction in a MIMO system at high noise levels, and in [24], where it was shown that the capacity at low noise levels is surprisingly high and scales linearly with the number of receive antennas. A detector with near maximum-likelihood performance for one-bit ADC systems was proposed in [25]. However, the computational complexity of the detector becomes prohibitive when the channel is frequency selective. Therefore, this thesis investigates the use

of low-complexity linear receivers for massive MIMO systems with one-bit ADCs and other low-resolution ADCs. It is observed that simple receivers work well when there is *stochastic resonance*, i.e. when the desired part of the receive signal is small in comparison to the interference and noise, such as in a frequency-selective channel, in a multiuser setup or in high noise levels. Achievable rates are derived for such systems and shows that, if the receive filter can be implemented as an analog filter and if there is stochastic resonance, the effective SINR loss caused by the one-bit ADCs typically is close to 4 dB compared to unquantized systems. Other papers [26, 27] have come up with similar achievable rates without noticing the necessity of stochastic resonance. The work in [28] also arrived at similar conclusions. Furthermore, this thesis shows that ADC resolutions greater than or equal to 4 bit give a performance that is close to unquantized systems. This conclusion was also drawn in [29–31]. Whereas, in [32] an additive uncorrelated-distortion model and a parametric power-consumption model were used to show that slightly higher resolutions of 4–8 bit are optimal from a power consumption point of view.

The beamforming of intermodulation products from nonlinear amplifiers in certain arrays have been studied for line-of-sight propagation in [33, 34]. These studies show how two sinusoids with different frequencies, each beamformed in a given direction, create intermodulation products that are beamformed in different directions. From these findings, it is not immediately obvious how the distortion from nonlinear amplifiers in a massive MIMO system behaves spatially, especially in non-line-of-sight transmission. Some symbol-sampled models of the effect of nonlinear amplifiers have been established and analyzed, e.g. [35, 36], which result in contradicting conclusions. While [36] claims that the distortion combines noncoherently, which is inline with the uncorrelated-distortion models, [35] claims the contrary—that the distortion from nonlinear amplifiers can combine coherently. In this thesis, a continuous-time signal model is established and used together with a behavioral model to accurately describe the transfer characteristics of a massive MIMO base station with nonlinear amplifiers. Both the downlink with nonlinear power amplifiers and the uplink with nonlinear low-noise amplifiers are studied, both in-band and out-of-band. It is shown that the distortion indeed combines coherently and is beamformed, partially. In the downlink, the coherent part diminishes when the input signal is beamformed in more directions, i.e. when the system serves more users or when the channel is more frequency selective. This behavior of the out-of-band radiation is also observed in [37], where another source of distortion, low-resolution digital-to-analog converters, are studied. In the uplink, however, increasing

the number of received signals does not make the distortion less coherent, since more signals also increases the input power to the low-noise amplifier, which drives it further into saturation. In the uplink, the presence of a blocker—an unwanted strong signal—creates distortion whose strongest term combines noncoherently. There are, however, smaller terms that combine coherently and therefore become significant in massive MIMO.

1.3 Contributions of the Thesis

This thesis explains and evaluates some of the practical aspects of massive MIMO that are not captured by the conventional idealistic linear system model. It is found that massive MIMO is robust to many nonlinear imperfections, and that problems, such as amplifier distortion and coarse quantization, are naturally reduced as the number of base station antennas is increased.

An introduction to the topic—massive MIMO base stations with nonlinear hardware—is given in the first part of the thesis. The second part is a collection of papers, in which the main results are derived and detailed. The papers can be divided into three groups: waveforms, analog-to-digital conversion and amplifiers. In the following subsections, a more detailed presentation of the particular contributions of each of the included papers is given.

All papers are written by the first author himself based on ideas that have sprung from discussions with the co-authors of each paper. The theoretical and empirical results in the papers are derived and implemented by the first author himself. The first author recognizes the great contribution of all the co-authors, who have spent a significant amount of time on supervising his work, revising his texts and have shared their deep professional expertise with him.

1.3.1 Waveforms

The PAR of a signal affects at what power efficiency a power amplifier can be operated and how much distortion that is created. From the perspective of power efficiency and distortion, a low PAR is to prefer. In Papers A and B, different waveforms with low PAR are investigated and compared to conventional waveforms, which tend to have high PAR in massive MIMO.

Paper A: Waveforms for the Massive MIMO Downlink: Amplifier Efficiency, Distortion and Performance

Authored by: Christopher Mollén, Erik G. Larsson and Thomas Eriksson

Published in: IEEE Transactions on Communications, vol. 64, no. 12, pp. 5050–5063, December 2016.

The massive MIMO downlink relies on precoded transmission to spatially multiplex individual data streams to different users. Data can be transmitted either over the whole spectrum with single-carrier transmission or over separate subcarriers with OFDM transmission. Furthermore, precoding can be done in many different ways: by conventional methods that do not consider the effect of the nonlinear power amplifier and by hardware-aware precoders that produce hardware-friendly signals, which allow the amplifier to be operated with higher power efficiency. In this paper, different transmission and precoding methods are evaluated in terms of spectral efficiency and power consumption of the amplifiers. It is found, that conventional and hardware-friendly precoders result in approximately the same amplifier power consumption when operated at the same spectral efficiency, even if the hardware-friendly precoders require a higher transmit power to achieve the same performance as conventional precoders. It is also observed that single-carrier and OFDM transmission have the same performance in massive MIMO in terms of achievable data rates and that both transmission methods result in signals with similar PAR, which is not the case in conventional communication systems, where usually only OFDM suffers from high PAR. Consequently, the two transmission techniques therefore result in the same amplifier power consumption.

Paper B: Multiuser MIMO Precoding with Per-Antenna Continuous-Time Constant-Envelope Constraints

Authored by: Christopher Mollén and Erik G. Larsson

Published in: The Proceedings of the International Workshop on Signal Processing Advances in Wireless Communications, Stockholm, Sweden, June 2015, pp. 261–265.

From a signal generation point of view, continuous-time signals with constant envelopes are the most preferable. Such signals allow for highly power efficient and inexpensive radio chain designs. For example, class C or switched mode amplifiers could be used without causing prohibitive signal distortion. This paper presents a precoder for the massive MIMO downlink that results in transmit signals with continuous-time constant envelopes. It is shown that there is a trade-off between excess bandwidth and performance. In

one example system, at low data rates and an excess bandwidth of 40 %, the proposed precoder needs 3 dB more radiated power to achieve the same data rate as conventional zero-forcing precoding. It is argued that this extra radiated power might be compensated for by the increased power efficiency and the decreased complexity of the hardware.

1.3.2 Analog-to-Digital Conversion

The base station in a massive MIMO system requires a large number of radio chains—in order to be individually controllable, each antenna needs one. This can make the hardware cost and the power consumption prohibitively large, especially if the same hardware is used as in conventional base stations with only one radio chain. One way to reduce hardware complexity and power consumption of the receiver part of the radio chain is to use analog-to-digital converters (ADCs) with low resolutions. In Papers C and D, the effect of using low-resolution ADCs is investigated and an achievable rate is derived, with which system performance easily can be evaluated.

Paper C: Uplink Performance of Wideband Massive MIMO with One-Bit ADCs

Authored by: Christopher Mollén, Junil Choi, Erik G. Larsson and Robert W. Heath, Jr.

Published in: IEEE Transactions on Wireless Communications, vol. 16, no. 1, pp. 87–100, January 2017.

This paper investigates the feasibility, in terms of achievable rate, of letting the base station use ADCs with the lowest possible resolution—one-bit ADCs operated at the baudrate. Such ADCs are very easy to implement, consume negligible amounts of power and do not require any advanced automatic gain control. It is shown that, also with one-bit ADCs, channel estimation and symbol detection can be done with linear signal processing, as long as the receive filter can be implemented as an analog filter. Furthermore, it is shown that the use of one-bit ADCs leads to an SINR loss of approximately 4 dB at low spectral efficiencies, which could be overcome by using a factor 2.5 more base station antennas.

Paper D: Achievable Uplink Rates for Massive MIMO with Coarse Quantization

Authored by: Christopher Mollén, Junil Choi, Erik G. Larsson and Robert W. Heath, Jr.

Published in: The Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, March 2017, pp. 6488–6492.

This paper extends the results from Paper C to ADCs with arbitrary resolution and quantization thresholds. Specifically, an achievable rate is given for the massive MIMO uplink, where the received signals are quantized. The impact of system imperfections, such as imperfect automatic-gain-control and imperfect power control, is investigated. It is found that using four-bit ADCs in the massive MIMO base station results in a performance that is close to the performance of an unquantized system, also in the presence of certain system imperfections.

1.3.3 Amplifiers

In practice, amplifiers—power amplifiers for the transmitter chain and low-noise amplifiers for the receiver chain—are nonlinear devices that cause in-band distortion and spectral regrowth. Both phenomena can be mitigated by setting higher linearity requirements on the amplifiers. Increased linearity, however, increases hardware complexity, cost and power consumption. Since the cost and hardware complexity of the power amplifiers and the power consumption of the low-noise amplifiers become significant in a massive MIMO base station, it is desirable to set as low linearity requirements as possible. The papers in this section investigate the effects of using nonlinear amplifiers in the massive MIMO base station.

Paper E: The Hermite-Polynomial Approach to the Analysis of Nonlinearities in Signal Processing Systems

Authored by: Christopher Mollén and Erik G. Larsson

Previously unpublished

The Hermite expansion of a Gaussian random variable with finite variance is explained in this paper. It is a useful tool when analyzing the statistical properties of a nonlinear system, whose output can be given as a Hermite expansion in terms of the input signal. Further, the Hermite expansion technique is generalized to describe nonlinearities in complex baseband, such as a nonlinear amplifier. Using a Mehler-like orthogonality property

of the complex expansion of the output signal, its autocorrelation is easily computed. It is also shown how the same technique can be used to compute the cross-correlation between nonlinearly processed signals in terms of the cross-correlation of the signals prior to the processing, and to partition the signals into a linear part and a distortion part, whose cross-correlation with the linear part is zero. This is useful, for example, when studying the radiation pattern of the distortion from an array, where each antenna is equipped with a nonlinear amplifier.

Paper F: Out-of-Band Radiation from Large Antenna Arrays

Authored by: Christopher Mollén, Erik G. Larsson, Ulf Gustavsson, Thomas Eriksson and Robert W. Heath, Jr.

Accepted to: IEEE Communications Magazine

Nonlinear hardware in the transmitter causes the base station to radiate power outside the allocated band, so called out-of-band radiation. When the signal is transmitted over a shared wireless medium, the out-of-band radiation can disturb other victim systems operating in adjacent frequency bands. In a MIMO system, there is also the risk that the radiation is beamformed and builds up coherently at the victim, which would amplify the disturbance. In this paper, we study the spatial behavior of the out-of-band radiation. It is found that the out-of-band radiation is beamformed to some extent and that this beamforming becomes more prominent the less frequency selective the channel is and the fewer users that are served by the system. In a frequency-selective multiuser channel however, the out-of-band radiation is close to isotropic. Further, it is observed that the array gain of the out-of-band radiation is smaller than the array gain of the desired signal. Since the array gain of the desired signal allows for reduced radiated power, the total effective out-of-band radiation from a MIMO array is never greater than than from a conventional single-antenna transmitter when they are operated at the same spectral efficiencies and with the same linearity requirements. In fact, the radiation is either significantly smaller or is beamformed such that it is confined spatially to a small area. Large MIMO arrays thus allow for less linear hardware, which increases power efficiency and reduces the cost of the array.

Paper G: Spatial Characteristics of Distortion Radiated from Antenna Arrays with Transceiver Nonlinearities

Authored by: Christopher Mollén, Ulf Gustavsson, Thomas Eriksson and Erik G. Larsson

Submitted to: IEEE Transactions on Wireless Communications

To analyze the radiation pattern of the distortion created by nonlinear power amplifiers in a massive MIMO base station, the tools from Paper E are used. It is shown that the distortion is beamformed, in what directions it is beamformed and with what power. The number of beamforming directions that the distortion is spread across is shown to grow cubically with the number of beamforming directions of the intended signal and quadratically in the number of channel taps. In many cases, the distortion therefore is beamformed in more directions than the resolution of the array, which equals the number of antennas. In such cases, the distortion behaves isotropically if all the beamforming directions have the same power, which is the case if all the beamforming directions of the intended signal have the same power. In other cases, the distortion is beamformed, among other directions, in the same directions as the intended signal, which effectively will limit the SNR of the received signals, which will saturate with an increased number of antennas.

Paper H: Impact of Spatial Filtering on Distortion from Low-Noise Amplifiers in Massive MIMO Base Stations

Authored by: Christopher Mollén, Ulf Gustavsson, Thomas Eriksson and Erik G. Larsson

Submitted to: IEEE Transactions on Communications

The use of nonlinear low-noise amplifiers in a massive MIMO base station in order to improve their power efficiency is investigated in this paper. This is motivated by the fact that the power consumption of the low-noise amplifiers grows linearly in the number of antennas, and becomes significant when the number of antennas is large. A scenario, where an out-of-band blocker is contaminating the received signal from the served users is considered. Because of the nonlinearity, the interference from the blocker leaks into the band of the served users, which creates an additional error in the symbol estimates. The tools from Paper E are used to derive the autocorrelation of the estimation error due to the nonlinear amplification. It is shown that, in many cases, spatial processing can filter out the main term of the leaked interference from the blocker, the part whose power scales with the cube of the power of the blocker. However, the term that scales with the square of the power of the blocker cannot be filtered out, because it has the same

spatial pattern as the signal from the served user.

1.4 Excluded Papers

The papers in Table 1 are co-authored by me, but have been excluded from the thesis because either they are superseded by newer papers that are included in the thesis or they are not within the main focus of the thesis.

Table 1: Excluded papers

C. Mollén, E. G. Larsson and T. Eriksson, "On the impact of PA-induced in-band distortion in massive MIMO", in *The Proceedings of the European Wireless Conference*, Barcelona, Spain, May 2014, pp. 201–206.

C. Mollén, J. Choi, E. G. Larsson and R. W. Heath, Jr., "Performance of linear receivers for wideband massive MIMO with one-bit ADCs", in *The Proceedings of the International ITG Workshop on Smart Antennas*, Munich, Germany, Mar. 2016, pp. 509–515.

C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, Jr., "One-bit ADCs in wideband massive MIMO systems with OFDM transmission", in *The Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Shanghai, China, Mar. 2016, pp. 3386–3390.

C. Mollén, U. Gustavsson, T. Eriksson and E. G. Larsson, "Out-of-Band Radiation Measure for MIMO Arrays with Beamformed Transmission", in *The Proceedings of the IEEE International Conference on Communications*, Kuala Lumpur, Malaysia, May 2016.

S. Kashyap, C. Mollén, E. Björnson and E. G. Larsson, "Frequency-domain interpolation of the zero-forcing matrix in massive MIMO-OFDM", in *The Proceedings of the IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Edinburgh, UK, Jul. 2016.

S. Kashyap, C. Mollén, E. Björnson and E. G. Larsson, "Performance analysis of (TDD) massive MIMO with Kalman channel prediction", in *The Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, USA, Mar. 2017, pp. 3554–3558.

C. Mollén, U. Gustavsson, T. Eriksson and E. G. Larsson, "Analysis of nonlinear low-noise amplifiers in massive MIMO base stations", in *The Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, USA, Nov. 2017.

Chapter 2

Communication Theory

This chapter introduces the fundamentals of the communication theory that is used throughout this thesis. The well-versed reader may skip through this chapter.

2.1 Signal Representations

The exact signal that is transmitted in a communication system can take many different forms. In the study of a general communication system however, the exact nature of a specific signal is seldom interesting. Instead, only properties of the signal, such as its power, its spectrum and the statistics of its amplitude, are relevant. For this reason, the signals in a communication system are usually modeled as stochastic processes [38]. The power of such a process $x(t)$ is defined by

$$P(x(t)) \triangleq \lim_{t_0 \to \infty} \frac{1}{2t_0} \mathsf{E}\left[\int_{-t_0}^{t_0} |x(t)|^2 \mathrm{d}t\right]. \tag{1}$$

The spectral properties of the signal are captured by its *power spectral density* if it exists, which is a function $S(f)$ such that

$$\int_{-\infty}^{\infty} S(f)|u(f)|^2 \mathrm{d}f = P\left(\int_{-\infty}^{\infty} u(\tau)x(t-\tau)\mathrm{d}\tau\right), \tag{2}$$

for all integrable functions $u(\tau) : \mathbb{R} \to \mathbb{C}$ with Fourier transform $u(f)$.

If, as a special case, $u(\tau)$ is taken to be a narrow bandpass filter around f_0, it can be seen that the power spectral density $S(f_0)$ can be interpreted as the power of its stochastic process at this frequency, normalized by the

bandwidth of the filter. If $x(t)$ is a wide-sense stationary process, i.e. its *autocorrelation*

$$R(\tau) \triangleq \mathsf{E}\left[x^*(t)x(t+\tau)\right] \tag{3}$$

and mean $\mathsf{E}[x(t)]$ do not depend on t, then the power spectral density is given as the function $S(f)$ whose inverse Fourier transform equals the autocorrelation:

$$R(\tau) = \int_{-\infty}^{\infty} S(f)e^{j2\pi f\tau}\mathrm{d}f, \quad \forall\tau \in \mathbb{R}. \tag{4}$$

It can be shown that, if it exists, the power spectral density is unique and a real-valued non-negative function (except, possibly, on a set of Lebesgue measure zero). In the included papers, signals are often treated as ordinary deterministic functions to simplify the exposition and no notational difference is made between signals that are stochastic and signals that are deterministic—context has to distinguish the two.

A wireless communication medium, such as air or empty space, is commonly shared between many systems. To avoid interference, each system is usually allocated a frequency band $[\pm f_\mathrm{c} - B/2, \pm f_\mathrm{c} + B/2]$ of its own of some width B around a *carrier frequency* f_c. The typical transmit signal used for wireless communication is thus a real-valued signal $x_\mathrm{pb}(t)$ whose energy is zero outside this band, i.e. a signal whose power spectral density $S_\mathrm{pb}(f) = 0$ when $|f| \notin [f_\mathrm{c} - B/2, f_\mathrm{c} + B/2]$. Such a signal is called a *passband signal*.

The wireless communication medium can be modeled as a linear system whose properties change slowly in relation to the time duration of the transmit signal. Therefore it is assumed that the signal $y_\mathrm{pb}(t)$ that is received during the transmission is given by the transmit signal and the impulse response $g_\mathrm{pb}(\tau)$ of the channel in the following way

$$y_\mathrm{pb}(t) = \sqrt{P}\int_{-\infty}^{\infty} g_\mathrm{pb}(\tau)x_\mathrm{pb}(t-\tau)\mathrm{d}\tau + z_\mathrm{pb}(t), \tag{5}$$

where we let P denote the transmit power by requiring that

$$P\left(x_\mathrm{pb}(t)\right) = 1, \tag{6}$$

and where $z_\mathrm{pb}(t)$ is a noise term that models the thermal noise of the receiving hardware. The noise is modeled as a Gaussian stochastic process that is white over the allocated band, i.e. its power spectral density is constant for $|f| \in [f_\mathrm{c} - B/2, f_\mathrm{c} + B/2]$. The constant spectral height is denoted N_0. The general setting is depicted in Figure 1.

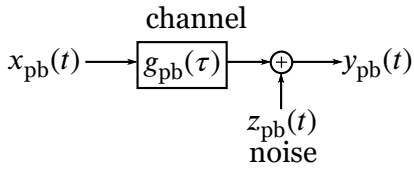


Figure 1: General communication system

The channel introduces two distortion effects: *large-scale fading* and *small-scale fading*. The large-scale fading is the signal attenuation due to both the distance the signal has traversed and the materials that the signal has penetrated on its way to the receiver. The small-scale fading is the aggregate amplitude and phase distortion that stems from multipath propagation, where the received signal is the superposition of many copies of the same signal with different time delays. By denoting the attenuation due to large-scale fading by $\beta \in [0, 1]$ and the effects of small scale fading by $h_{\mathrm{pb}}(\tau)$, the impulse response of the channel can be factorized as follows:

$$g_{\mathrm{pb}}(\tau) = \sqrt{\beta} h_{\mathrm{pb}}(\tau). \tag{7}$$

The large-scale fading changes very little over the course of the transmission. It is therefore relatively easy to estimate and is assumed to be known to both transmitter and receiver. The small-scale fading, on the other hand, changes over the course of the transmission and has to be estimated. The factorization in (7) is thus helpful to distinguish what is known and what has to be estimated in our models.

The general range, in which the carrier frequency lies, determines certain propagation characteristics of the wireless medium, e.g., the amount of path loss, penetration loss and molecular absorption that can be expected [39]. Other than that, the carrier frequency is of little importance for the theoretical study of the general communication system in Figure 1. It is therefore common practice to represent the physical real-valued passband signal by its complex *baseband* equivalent:

$$x(t) = \mathrm{LP}_{B/2}\left(x_{\mathrm{pb}}(t) e^{-j2\pi f_{\mathrm{c}} t}\right), \tag{8}$$

where $\mathrm{LP}_{B/2}$ denotes an ideal lowpass filter with cutoff frequency $B/2$, see [40] for a thorough introduction to the baseband model. In the baseband notation, the passband signal $x_{\mathrm{pb}}(t)$ at any time instant t is represented by a complex number that naturally represents its phase and envelope by $\arg(x(t))$ and $|x(t)|$ respectively. Note that the transform in (8) is invertible:

$$x_{\mathrm{pb}}(t) = 2\,\mathfrak{Re}\,(x(t))\cos(2\pi f_{\mathrm{c}} t) - 2\,\mathfrak{Im}\,(x(t))\sin(2\pi f_{\mathrm{c}} t) \tag{9}$$

and that the power spectral density of the baseband signal $S(f) = S_{\mathrm{pb}}(f + f_{\mathrm{c}})$ in the band $f \in [-B/2, B/2]$ and $S(f) = 0$ outside that band. Furthermore, the received signal in (5) is given by

$$y(t) = \sqrt{\beta P} \int_{-\infty}^{\infty} h(\tau) x(t - \tau) \mathrm{d}\tau + z(t) \tag{10}$$

in the equivalent baseband representation, where the baseband signals $y(t)$, $z(t)$ and $h(\tau)$ are defined in the same way as $x(t)$ in (8). Note that the thermal noise $z(t)$ then becomes a realization of a complex circularly symmetric Gaussian stochastic process, whose power spectral density is equal to the constant N_0 in the band $f \in [-B/2, B/2]$.

A common way of encoding information onto the transmit signal is to map it onto a sequence of complex values $\{x[n]\}$ first. These values are then *pulse-amplitude modulated* by a *transmit filter* $p'(\tau)$ into the baseband transmit signal:

$$x(t) = \sum_{n=-\infty}^{\infty} p'(t - nT) x[n], \tag{11}$$

where T is called the *symbol duration*. To make the transmit signal fit within its allocated band, the filter $p(\tau)$ has to be bandlimited to within $[-B/2, B/2]$. Further, if $\{x[n]\}$ is a series of i.i.d. random variables such that $\mathsf{E}[|x[n]|^2] = 1$, then the energy of the filter has to be

$$\int_{-\infty}^{\infty} \left| p'(\tau) \right|^2 \mathrm{d}\tau = T \tag{12}$$

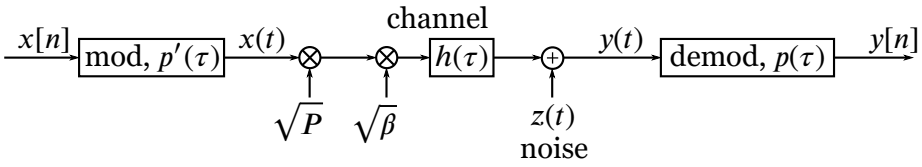
to make the transmit signal fulfill its power constraint (6).

To obtain a sequence of complex values again, the reverse operation, *demodulation*, is performed on the receive signal

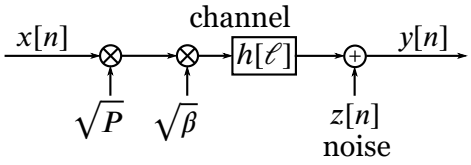
$$y^{(\kappa)}[n] = \int_{-\infty}^{\infty} p(\tau) y(nT/\kappa - \tau) \mathrm{d}\tau, \tag{13}$$

where $p(\tau)$ is the *receive filter* and κ the *oversampling factor*. When $\kappa = 1$, the signal is given in *symbol-sampled* time and the superscript is omitted: $y[n] \triangleq y^{(1)}[n]$. When pulse-amplitude modulation and demodulation are used, the communication channel given in (10) and shown in Figure 2(a) can equivalently be given in symbol-sampled time as in Figure 2(b), where the received signal is given by:

$$y[n] = \sqrt{\beta P} \sum_{\ell=-\infty}^{\infty} h[\ell] x[n - \ell] + z[n], \tag{14}$$



(a) continuous time



(b) symbol-sampled discrete time

Figure 2: Equivalent baseband model of a general communication system

the discrete-time channel impulse response is given by

$$h[\ell] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p'(\tau) h(\tau' - \tau) p(\ell T - \tau') \mathrm{d}\tau \mathrm{d}\tau' \tag{15}$$

and the discrete-time noise is

$$z[n] = \int_{-\infty}^{\infty} p(\tau) z(nT - \tau) \mathrm{d}\tau. \tag{16}$$

Since the communication model in (14) is equivalent to (1) when modulation and demodulation are done according to (11) and (13), communication systems are usually studied in symbol-sampled time for simplicity. However, when nonlinear systems are studied, such as in the included papers, oversampled signals have to be considered because nonlinearities can cause spectral regrowth that results in undesired aliasing in the sampling.

A *Nyquist pulse of parameter* T is a pulse whose Fourier transform $\Gamma(f)$ satisfies

$$\sum_{n=-\infty}^{\infty} \Gamma\left(f - \frac{n}{T}\right) = \text{constant} \tag{17}$$

for all f, except possibly for f in some set of Lebesgue measure zero. To make the discrete-time noise $z[n]$ white, the receive filter is usually chosen as a root-Nyquist pulse of parameter T, i.e. a pulse whose self-similarity function

$\int p(\tau)p^*(\tau-t)\mathrm{d}\tau$ is a Nyquist pulse of parameter T. When the receive filter is a root-Nyquist pulse and has the energy

$$\int_{-\infty}^{\infty} |p(\tau)|^2 \,\mathrm{d}\tau = \frac{1}{T}, \tag{18}$$

then $z[n] \sim \mathcal{CN}(0, N_0/T)$ and independently and identically distributed across n. To maximize the signal-to-noise ratio (SNR)[1], i.e. the power of the desired signal in relation to the power of the noise

$$\frac{\mathsf{E}\left[\left|y[n]-z[n]\right|^2\right]}{\mathsf{E}\left[\left|z[n]\right|^2\right]}, \tag{19}$$

the transmit filter is *matched* to the receive filter [41], i.e. $p'(\tau) = Tp^*(-\tau)$, where the transmit filter is scaled by T so that the power constraint (12) holds.

A common choice of filters in practical communication systems are the root-raised cosine filters, which have a good trade-off between narrow bandwidth and short delays. The root-raised cosine filters are a family of filters parameterized by their *excess bandwidth* $\alpha \triangleq TB - 1$, which is a measure of how much wider the bandwidth B is compared to the *baudrate* $1/T$. Note that the bandwidth of a Nyquist pulse never can be smaller than the baudrate, but that the bandwidth of the filter usually is chosen as close to the baudrate as possible to use the spectrum efficiently—a small excess bandwidth is desired. For the root-raised cosines, α can vary between 0 and 1. The common choice $\alpha = 0.22$, which means that the bandwidth of the pulse is 22 % wider than the baudrate, is used to evaluate the theoretical results in the included papers.

The symbol-sampled description in (14) of the transmission can be simplified by introducing a *cyclic prefix* and studying it in the frequency domain. By observing the received signals for a block of N symbol durations and adding a cyclic prefix to the transmit signal in the following way:

$$x[n] = x[n+N], \quad n = -L+2, \ldots, -1, \tag{20}$$

the time indices of the signals in the convolution in (14) can be taken modulo N and the received signals $\{y_k^{\mathrm{DL}}[n], n = 0, \ldots, N-1\}$ can be rewritten

[1]It is desirable to leave the channel equalization to the digital part of the radio chain. For this reason, "maximize the SNR " is taken to mean to maximize the average SNR over all channel realizations at this point, i.e. to ensure the highest possible SNR if the channel were flat.

as a cyclic convolution, i.e. as the convolution of periodic signals. Since the discrete Fourier transform of a cyclic convolution is a multiplication in the frequency domain, the received signals are given by their Fourier transform as:

$$y[\nu] = \sqrt{\beta P} h[\nu] x[\nu] + z[\nu], \tag{21}$$

where the transmit signal and channel are defined in the frequency domain as:

$$x[\nu] \triangleq \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] e^{-j2\pi n\nu/N} \tag{22}$$

$$h[\nu] \triangleq \sum_{\ell=0}^{L-1} h[\ell] e^{-j2\pi \ell \nu/N}. \tag{23}$$

The Fourier transforms $y[\nu]$ and $z[\nu]$ of $y[n]$ and $z[n]$ are defined in analogy with $x[\nu]$ in (22). It is noted that the frequency response of the channel $h[\nu]$ is not scaled by $1/\sqrt{N}$, so that the block length N will not show up in the relation in (21). The time-domain receive signal is given by the inverse Fourier transform:

$$y[n] = \frac{1}{\sqrt{N}} \sum_{\nu=0}^{N-1} y[\nu] e^{j2\pi n\nu/N}. \tag{24}$$

Because this frequency-domain description of the symbol-sampled discrete-time channel abstracts the frequency-selectivity of the channel—the channel in (21) does not involve a convolution as the one in (14)—it is many times used in the study of communication systems to simplify the exposition. The papers that are included in this thesis will assume block transmission with cyclic prefix and often switch between the two, time-domain and the frequency-domain, descriptions.

The transmission that has been explained above is sometimes termed *single-carrier transmission* in order to distinguish it from its generalization *multi-carrier transmission*, where the transmit signals are given as the sum of N pulse-amplitude modulated discrete-time signals $x_\nu[n]$:

$$x(t) = \sum_{\nu=0}^{N-1} \sum_{n=-\infty}^{\infty} p'_\nu(t - nT) x_\nu[n]. \tag{25}$$

In full generality, the pulses $p'_\nu(\tau)$ can be chosen arbitrarily, but are usually chosen such that each pulse is orthogonal to all other pulses and their time shifts in order to avoid inter-carrier interference in the demodulation. One common multi-carrier format is *OFDM* (orthogonal frequency-division multiplexing) with rectangular pulses:

$$p'_\nu(\tau) = \text{rect}\left(\frac{t}{T} - n\right) e^{j2\pi t\nu/T}, \tag{26}$$

where $\text{rect}(t) = 1$ when $-0.5 \leq t < 0.5$ and equal to zero otherwise. The specific structure of the OFDM signal allows for an easier modulation method that is based on the Fourier transform, for details see [42, Ch. 4.6 & 4.9], which might be a more common way to present the OFDM signal. While the two modulation methods—the pulse-amplitude modulation in (26) and the one based on the Fourier transform—produce the same transmit signal, the method based on the Fourier transform is easier to realize in hardware. Except for Paper G, the modulation format in (11) with one pulse is used. An overview of different multi-carrier transmission techniques can be found in [43].

2.2 The Wireless Communication Channel

A signal that is transmitted will be reflected, refracted and diffracted by the different materials in the surrounding environment and, as a result, the transmitted signal will travel many different paths to the receiver. The received signal is therefore a superposition of many differently attenuated and delayed copies of the transmitted signal, as can be seen in Figure 3. In wireless communication over the air, the two phenomena reflection and diffraction are the most pronounced sources of multipath propagation. In comparison, refraction, which happens when a signal enters a medium of a different refractive index, is less common in most earth-bound wireless channels over the air. Refraction can, however, bend the signal path due to gradual change of pressure, temperature and chemical composition of the air, which all affect the refractive index, for example in the communication with satellites.

If the number of paths is N_c, then the receive signal can be written:

$$y_\text{pb}(t) = \sqrt{\beta} \sum_{i=1}^{N_\text{c}} \alpha_i x_\text{pb}(t - \tau_i), \tag{27}$$

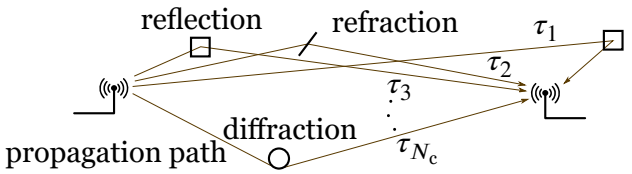


Figure 3: Multipath propagation

where β is the large-scale fading and α_i the real-valued amplitude scaling and τ_i the delay of the i-th path [40]. Using the baseband notation, the relation in (27) becomes

$$y(t) = \sqrt{\beta} \sum_{i=1}^{N_c} \alpha_i e^{-j2\pi f_c \tau_i} x(t - \tau_i). \tag{28}$$

Hence, the channel impulse response is

$$h(\tau) = \sum_{i=1}^{N_c} \alpha_i e^{-j2\pi f_c \tau_i} \delta(\tau - \tau_i) \tag{29}$$

and, according to (15), the discrete-time channel becomes

$$h[\ell] = \sum_{i=1}^{N_c} \gamma(\ell T - \tau_i) \alpha_i e^{-j2\pi f_c \tau_i}, \tag{30}$$

where $\gamma(\tau) \triangleq \int p'(t)p(\tau - t)\mathrm{d}t$ is the aggregate transmit–receive filter.

We will assume that $\{\gamma(nT - \tau_i)\}_{n=-\infty}^{\infty}$ is a sequence with only a few non-zero taps around $n_i = \mathrm{round}(\tau_i/T)$, since the transmit pulse $p(\tau)$ has been designed to have short delays. The support of the discrete-time impulse response of the channel is thus approximated by the difference

$$\sigma_\tau \triangleq \max_{i,i'} \left(\tau_i - \tau_{i'} \right), \tag{31}$$

which is one measure of the *delay spread* of the channel. The number L of non-zero taps in the discrete-time channel is then roughly

$$L \approx \frac{\sigma_\tau}{T}. \tag{32}$$

A channel, where the discrete-time channel impulse response has $L = 1$, is called a *frequency-flat* channel, because the discrete-time Fourier transform of $\{h[\ell]\}$ is a constant. If $L > 1$, the channel is *frequency selective*, because the discrete-time Fourier transform varies with frequency.

The delay of a path i is a function of the length d_i of that path: $\tau_i = d_i/c$, where c is the speed of the signal (usually the speed of light). For example, in an outdoor environment, where the wireless communication system is supposed to cover an area with diameter 1000 m, it would be reasonable to assume that the maximum difference in path lengths $\max(d_i - d_{i'}) \approx 1000\,\text{m}$. In such an environment, the delay spread is approximately

$$\sigma_\tau \approx \frac{1}{c} \max_{i,i'} \left(d_i - d_{i'} \right) \approx 3.3\,\mu\text{s}, \tag{33}$$

where $c = 300\,\text{Mm/s}$, the speed of light, was assumed. The discrete-time channel is thus approximately frequency flat ($L = 1$) in a system with baudrate $1/T = 300\,\text{kHz}$, and frequency selective with $L = 67$ taps with baudrate 20 MHz.

Movements in the propagation environment will change the time delays, amplitude scalings and the number of paths of the channel. We have previously claimed that the channel impulse response is well approximated as static during the course of the transmission. This claim is equal to saying that all the objects that influence the signal paths, as well as the transmitter and the receiver, all are still during the course of the transmission. This is obviously not the case for a general channel. If the channel changes in a way that can be accurately tracked over time, however, a static channel model and a dynamic one are practically equivalent.

The time during which the channel accurately can be tracked is called the *coherence time*. The coherence time can be derived by letting the length of path i at time t be $d_i(t)$. Then the delay of that path is given by $\tau_i = d_i(t)/c$. If v_i is the constant rate, at which the length $d_i(t)$ shrinks or elongates over time due to movements in the environment, then the *Doppler shift* of path i is given by $f_i \triangleq v_i/\lambda$, where the wavelength of the signal is given by $\lambda \triangleq c/f_\text{c}$. By defining $\alpha_i'[\ell] \triangleq \gamma(\ell T - \tau_i)\alpha_i$, the ℓ-th tap of the discrete-time channel in (30) can be written as

$$h[\ell] = \sum_{i=0}^{N_\text{c}} \alpha_i'[\ell] e^{-j2\pi(d_i(0)/\lambda + f_i t)}, \tag{34}$$

where $d_i(t) = d_i(0) + v_i t$ is used to write

$$f_\text{c}\tau_i = \left(d_i(0) + v_i t \right) f_\text{c}/c \tag{35}$$

$$= \left(d_i(0) + v_i t \right)/\lambda \tag{36}$$

$$= d_i(0)/\lambda + f_i t. \tag{37}$$

To gain some intuition, first assume that only one path $i = 1$ contributes to tap ℓ. Then

$$h[\ell] = \underbrace{\alpha_1'[\ell] e^{-j2\pi d_1(0)/\lambda}}_{\text{constant}} e^{-j2\pi f_1 t}. \tag{38}$$

Even though the channel coefficient is not constant over time, it can easily be tracked because the evolution is a phase shift with constant rate f_1, as long as the value at some time $t = t_0$ is known.

Now assume that only two of the paths $i = 1, 2$ significantly contribute to tap ℓ. The following reasoning can be extended to arbitrary many taps, but it would only obfuscate the exposition and lead to little additional insight. Channel tap ℓ is thus

$$h[\ell] = \alpha_1'[\ell] e^{-j2\pi(d_1(0)/\lambda + f_1 t)} + \alpha_2'[\ell] e^{-j2\pi(d_2(0)/\lambda + f_2 t)} \tag{39}$$

$$= \Big(\underbrace{\alpha_1'[\ell] e^{-j2\pi d_1(0)/\lambda}}_{\text{constant}} + \underbrace{\alpha_2'[\ell] e^{-j2\pi d_2(0)/\lambda}}_{\text{constant}} \underbrace{e^{-j2\pi(f_2 - f_1)t}}_{\text{not constant}} \Big) e^{-j2\pi f_1 t}. \tag{40}$$

Now the evolution over time is no longer a constant-rate phase shift. Knowledge of the value of $h[\ell]$ at some time $t = t_0$ is no longer enough to track the channel coefficient over time. With knowledge of $\alpha_1'[\ell]$ and $\alpha_2'[\ell]$, it would be possible but they are difficult to estimate.

The time period, during which it is possible to keep track of the channel coefficient is therefore approximated by how long the term

$$e^{-j2\pi(f_2 - f_1)t} \tag{41}$$

can be considered constant. It therefore follows that the coherence time is on the order $\sim 1/(f_2 - f_1)$. By generalizing this two-path model, the coherence time can be approximately determined by $1/\sigma_v$, where

$$\sigma_v \triangleq \max_{i,i'} \left(f_i - f_{i'} \right), \tag{42}$$

which is one measure of the *Doppler spread* of the channel. In symbol-sampled time, the coherence time is thus approximately given by

$$N_{\text{coherent}} \approx \frac{1}{T\sigma_v}. \tag{43}$$

For example, in an outdoor environment, where objects can be assumed to move around at a maximum speed of $|v_i| < 30\,\text{m/s}$, the maximum difference

in speeds is then $\max(v_i - v_{i'}) \approx 60\,\text{m/s}$. In such an environment,

$$\sigma_v \approx \frac{1}{\lambda} \max_{i,i'} \left(v_1 - v_{i'} \right) \approx 400\,\text{Hz}, \tag{44}$$

where it was assumed that $\lambda = 0.15\,\text{m}$—the wavelength of an electromagnetic wave that propagates in air with the frequency 2 GHz. It is thus possible to track the channel for $N_{\text{coherent}} = 750$ symbol durations if the baudrate is 300 kHz and for 50 000 symbol durations if the baudrate is 20 MHz.

To simplify the exposition, it is many times assumed that the channel is static and does not change. As long as the transmission duration is shorter than the coherence time, such a model can be adopted since the actual time-varying channel anyway can be obtained at any time during that coherence time through tracking. In the study of a communication system, it is sometimes assumed that the channel is *block fading*, i.e. that it is described accurately by the impulse response $h[\ell]$ during one coherence time—for N_{coherent} symbols—and by another, different, impulse response during the next coherence time.

This block-fading model is used in all the included papers. A multipath channel model similar to (30) was used in Paper B. In the other papers, it is assumed that the channel is frequency selective and that the taps are *Rayleigh fading*, i.e. that in each coherence time the channel is equal to a realization of the random variable

$$h[\ell] \sim \mathcal{CN}\left(0, \sigma^2[\ell]\right), \tag{45}$$

where $\sigma^2[\ell]$ is the *power delay profile* of the channel. Since the large-scale fading is given by β, the small-scale fading is normalized such that:

$$\sum_{\ell=-\infty}^{\infty} \sigma^2[\ell] = 1. \tag{46}$$

Because each path mainly contributes to only one channel tap in (30), it is also assumed that $\{h[\ell]\}$ are independent and identically distributed over ℓ. Note that for $h[\ell]$ to be Gaussian, α_i^2 should be Rayleigh distributed and $(f_c \tau_i \mod 1)$ uniformly distributed over $[0, 1)$. If the number of paths is large however, the taps become Gaussian without any assumptions on the distributions of the individual paths because of the central limit theorem.

Another channel model used in the papers is the *line-of-sight channel*, in which it is assumed that there is only one path $N_c = 1$. This model can be a good approximation of the kind of channel that appears when the

sight between transmitter and receiver is unobstructed and the signal is not reflected in the surrounding, for example, because of heavy signal attenuation, low reflectivity or because of the plain topology of the environment.

If the attenuation α_1 of the line-of-sigh path is absorbed by the large-scale fading term β, the channel is then given by

$$h(\tau) = e^{-j2\pi f_c \tau_1} \delta(\tau - \tau_1) \tag{47}$$

and the discrete-time baseband channel by

$$h[\ell] = \gamma(\ell/T - \tau_1) e^{-j2\pi f_c \tau_1}. \tag{48}$$

When the bandwidth of the pulse $\gamma(\tau)$ is small relative to $1/\tau_1$, the impulse response $\gamma(\ell/T - \tau_1) \approx \gamma(\ell/T)$. If the pulse is a Nyquist pulse, $\gamma(\ell/T) = \delta[\ell]$, the channel is well approximated by

$$h[\ell] = \begin{cases} e^{-j2\pi f_c \tau_1}, & \text{if } \ell = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{49}$$

Since coherence time is a function of the maximum difference between Doppler shifts and there is only one Doppler shift f_1 in a line-of-sight channel, it is not defined. In fact, the notion of coherence time as defined in (43) makes little sense in a line-of-sight channel, where a first order approximation of the evolution of the channel phase $f_c \tau_1 = d_1(0)/\lambda + f_1 t$ many times makes it possible to accurately track the channel for a relatively long time.

In most environments, it is more realistic to believe that there is one or a few strong paths and many weaker paths, which is a situation in between Rayleigh fading and line-of-sight. In such a situation, the channel can be modeled as a weighted sum of a line-of-sight channel and a Rayleigh fading channel. The taps then follow a Rice distribution and the channel is said to be *Ricean fading*. In the included papers, only the Rayleigh fading and line-of-sight channels are studied. It is argued, as in [44], that if the results hold in both of these cases, it is reasonable to believe that the results also hold approximately for a Ricean fading channel.

Chapter 3

Basics of Massive MIMO

Here the discrete-time system model established in the previous chapter will be used to explain how massive MIMO works in its canonical embodiment, i.e. a time-division duplex (TDD) system that relies on reciprocity-based uplink channel estimation to perform fully digital multiuser beamforming to serve its users. The chapter will describe a single-cell system, which is sufficient for the purpose of the thesis. Some consequences of having multiple cells in the system, which is the common way to operate a cellular communication system, will be discussed at the end of the chapter.

3.1 Multiuser MIMO

Massive MIMO is a *multiuser MIMO* system, which is a communication system where a base station that is equipped with M antennas concurrently communicates with K users. "Concurrently" here means that the signals to or from each user are sent at the same time and over the same frequency. The setup is shown in Figure 4. The transmission from the base station to the users is called the *downlink* and the reverse transmission, i.e. from the users to the base station, the *uplink*.

In Figure 4, each user has one antenna. Potentially, users can be equipped with multiple antennas to improve the data rate of the communication. One way to extend the analysis from single-antenna users to multi-antenna users is to treat each of the antennas of one user as if it were a user on its own. The actual data rate of a user would then be the total rate of all its antennas. However, one should have in mind that the more antennas the power limited mobile user is equipped with, the bulkier and more power consuming its hardware becomes. For this reason, this thesis only studies users with single

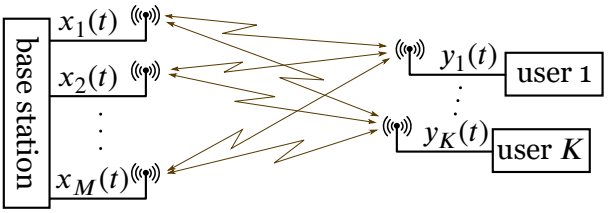


Figure 4: Multiuser MIMO with single-antenna users

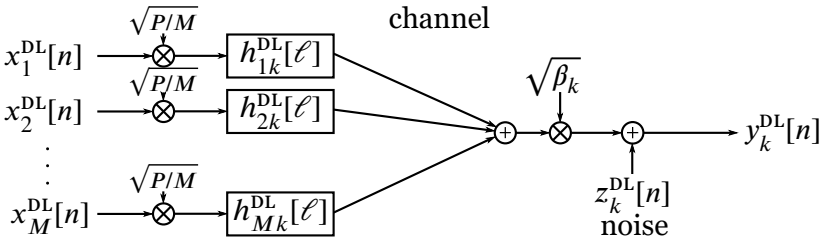


Figure 5: Discrete-time model of the downlink in a multiuser MIMO system

antennas. This is also a common assumption in the massive MIMO literature [45].

Each individual channel between one base station antenna and one user can be analyzed in the same way as the fundamental communication system discussed in Section 2.1. The effect of having multiple signals concurrently transmitted in the downlink is that each user receives a sum of signals as shown in Figure 5 instead of only one signal as in Figure 2(b) on page 19. If the channel between antenna m and user k is denoted $h_{mk}^{\mathrm{DL}}[\ell]$, then the signal received by user k is

$$y_k^{\mathrm{DL}}[n] = \sqrt{\beta_k P} \sum_{m=1}^{M} \sum_{\ell=-\infty}^{\infty} h_{mk}^{\mathrm{DL}}[\ell] x_m^{\mathrm{DL}}[n-\ell] + z_k^{\mathrm{DL}}[n], \tag{50}$$

where β_k is the large-scale fading attenuation to the user, $x_m^{\mathrm{DL}}[n]$ the transmit signal from antenna m and $z_k^{\mathrm{DL}}[n]$ the thermal noise at user k. It is assumed that the base station antennas are co-located, so that all transmit signals are attenuated by the same large-scale fading coefficient β_k.

In the same way, the signal received by base station antenna m in the uplink can be given in terms of the impulse response $h_{mk}^{\mathrm{UL}}[\ell]$ of the channel

between user k and the antenna in the following way:

$$y_m^{\mathrm{UL}}[n] = \sum_{k=1}^{K} \sqrt{\beta_k P_k} \sum_{\ell=-\infty}^{\infty} h_{mk}^{\mathrm{UL}}[\ell] x_k^{\mathrm{UL}}[n-\ell] + z_m^{\mathrm{UL}}[n], \tag{51}$$

where $x_k^{\mathrm{UL}}[n]$ and P_k are the transmit signal and transmit power from user k respectively and $z_m^{\mathrm{UL}}[n]$ is the thermal noise of the antenna.

The description of both the downlink and uplink transmissions can be simplified in the frequency domain by observing the transmission for a block of N symbol durations and introducing a cyclic prefix to the transmit signals, in the same way as was shown in (21) on page 21. Then the received signals are given in the frequency domain as:

$$y_k^{\mathrm{DL}}[\nu] = \sqrt{P\beta_k} \sum_{m=1}^{M} h_{mk}^{\mathrm{DL}}[\nu] x_m^{\mathrm{DL}}[\nu] + z_k^{\mathrm{DL}}[\nu], \tag{52}$$

$$y_m^{\mathrm{UL}}[\nu] = \sum_{k=1}^{K} \sqrt{P_k\beta_k} h_{mk}^{\mathrm{UL}}[\nu] x_k^{\mathrm{UL}}[\nu] + z_m^{\mathrm{UL}}[\nu]. \tag{53}$$

The Fourier transforms of the signals $y_m^{\mathrm{DL}}[\nu]$, $x_m^{\mathrm{DL}}[\nu]$, $z_k^{\mathrm{DL}}[\nu]$, $y_m^{\mathrm{UL}}[\nu]$, $x_k^{\mathrm{UL}}[\nu]$ and $z_m^{\mathrm{UL}}[\nu]$ are defined as in (22) on page 21; and the frequency responses $h_{mk}^{\mathrm{UL}}[\nu]$ and $h_{mk}^{\mathrm{DL}}[\nu]$ as in (23).

3.2 Channel Model

In the thesis, two types of channel models are considered: the block fading, Rayleigh fading model and the far-field, free-space, line-of-sight model. To simplify presentation, the uplink channel will be described. The argumentation for the downlink is practically identical however—instead of using the term angle of incidence, angle of departure should be used instead.

3.2.1 Line-of-Sight

In free space, there is only one distinct propagation path—the straight line from the transmitter to the receiver. If it is further assumed that the length of this propagation path is long, all incoming waves at the array are approximately parallel, or *planar*, as is shown in Figure 6. It can be shown through simple geometric arguments that this is a good approximation if the distance d to the user is farther than the Fraunhofer distance:

$$D_{\mathrm{f}} \triangleq \frac{D^2}{2\lambda}, \tag{54}$$

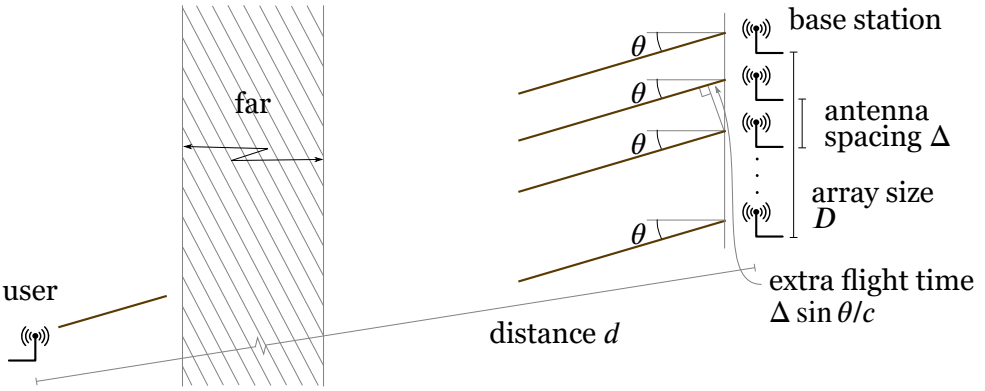


Figure 6: A planar wave in the far-field of a uniform linear array.

where D is the largest cross section of the array and λ the wavelength of the carrier frequency. When $d > D_\mathrm{f}$, the user is said to be in the *far-field*.

For a uniform linear array, like the one in Figure 6, with an antenna spacing Δ, the largest cross section would be $D = (M-1)\Delta$. If the antenna spacing is assumed to be half a wavelength, $\Delta = \lambda/2$, where the wavelength $\lambda = c/f$ is the speed of light c divided by the frequency f, the Fraunhofer distance is

$$D_\mathrm{f} = (M-1)^2 \frac{c}{8f}. \tag{55}$$

With $M = 100$ antennas and $f = 2\,\mathrm{GHz}$ carrier frequency, the Fraunhofer distance is $D_\mathrm{f} \approx 190\,\mathrm{m}$. For arrays with more compact geometries, however, the Fraunhofer distance is much smaller. For a uniform circular array, for example,

$$D_\mathrm{f} = M^2 \frac{c}{8f\pi^2} \approx 19\,\mathrm{m} \tag{56}$$

with the same number of antennas and carrier frequency.

In the far-field, the relative propagation delays between the antennas do not depend on the exact location of the user, only on the direction, in which the user is located: its angle in azimuth θ and angle in elevation φ. If the propagation delay to some reference point, e.g. one of the antennas in the array, is τ, the propagation delay to any other antenna m is given by

$$\tau_m(\theta, \varphi) + \tau, \tag{57}$$

where $\tau_m(\theta, \varphi)$ is a function of only the angles. Following the arguments for (48) on page 27, the channel of the m-th antenna is given by

$$h^{(m)}_{\theta,\varphi,\tau}[\ell] \triangleq \gamma(\ell/T - \tau_m(\theta,\varphi) - \tau)e^{-j2\pi f_c(\tau_m(\theta,\varphi)+\tau)}. \tag{58}$$

When studying line-of-sight channels, it is common to assume that the bandwidth of the pulse $\gamma(t)$ is small relative to all $1/(\tau_m(\theta,\varphi)+\tau)$, as in (49) on page 27, so that the channel can be approximated as frequency-flat:

$$h^{(m)}_{\theta,\varphi,\tau}[\ell] \approx \delta[\ell]e^{-j2\pi f_c(\tau_m(\theta,\varphi)+\tau)}. \tag{59}$$

If these channels are stacked in a vector

$$\mathbf{h}_{\theta,\varphi,\tau}[\ell] \triangleq (h^{(1)}_{\theta,\varphi,\tau}[\ell], \ldots, h^{(M)}_{\theta,\varphi,\tau}[\ell])^{\mathsf{T}}, \tag{60}$$

the so-called *steering vector* of the array is obtained. Using the previously established notation, a user located in a direction given by the angles θ_k and φ_k has the channel

$$h^{\mathrm{UL}}_{mk}[\ell] = h^{(m)}_{\theta_k,\varphi_k,\tau_k}[\ell], \quad \forall m, \tag{61}$$

for some delay τ_k.

In the special case of the horizontal uniform linear array that was depicted in Figure 6, where all antennas were located on a line with uniform spacing Δ, the delay function is a function of only the azimuth angle and given by

$$\tau_m(\theta,\varphi) = m\frac{\Delta}{c}\sin(\theta). \tag{62}$$

Similar expressions can be derived for any given array geometry.

3.2.2 Rayleigh Fading

In its most general form, the Rayleigh fading model assumes that the channel coefficients are correlated across the antennas according to the correlation matrix:

$$\mathbf{R}_k[\ell] \triangleq \mathsf{E}\left[\mathbf{h}_k[\ell]\mathbf{h}^{\mathsf{H}}_k[\ell]\right], \tag{63}$$

where the vector $\mathbf{h}_k[\ell] \triangleq (h^{\mathrm{UL}}_{1k}[\ell], \ldots, h^{\mathrm{UL}}_{Mk}[\ell])^{\mathsf{T}}$. The coefficients $h^{\mathrm{UL}}_{mk}[\ell]$ and $h^{\mathrm{UL}}_{m'k'}[\ell']$ are assumed to be jointly Gaussian, circularly symmetric random

variables. It is also assumed that channels of different users are uncorrelated, just as different taps are:

$$\mathsf{E}\left[\mathbf{h}_k[\ell]\mathbf{h}_{k'}^{\mathsf{H}}[\ell']\right] = \mathbf{0}, \quad (k,\ell) \neq (k',\ell'). \tag{64}$$

In accordance with (46) on page 26, the energy of the small-scale fading coefficients is normalized such that:

$$\sum_{\ell=0}^{L-1} \mathrm{tr}(\mathbf{R}_k[\ell]) = M, \quad \forall k. \tag{65}$$

A common way to model the correlation matrix is to use the steering vector of the array and to assume that the signal from user k is incident on the array with a certain probability $p_k(\theta,\varphi,\tau)$. This probability function is the normalized version of what is called the *instantaneous power azimuth-delay spectrum* in [46]. For consistency, however, the probability function will be referred to as the *power angle-delay profile*. The correlation matrix in (63) then becomes:

$$\mathbf{R}_k[\ell] = \int_0^{\sigma_\tau}\int_0^{2\pi}\int_0^{2\pi} p_k(\theta,\varphi,\tau)\mathbf{h}_{\theta,\varphi,\tau}[\ell]\mathbf{h}_{\theta,\varphi,\tau}^{\mathsf{H}}[\ell]\,\mathrm{d}\theta\mathrm{d}\varphi\mathrm{d}\tau. \tag{66}$$

A simplified channel model is the *i.i.d. (independent and identically distributed) Rayleigh fading model*, where the correlation matrix is a scaled identity matrix,

$$\mathbf{R}_k[\ell] = \sigma_k^2[\ell]\mathbf{I}_M, \tag{67}$$

i.e. the channel coefficients are uncorrelated also across antennas and all have the same variance $\sigma_k^2[\ell]$.

In fact, there are measurements [47,48] of actual channels that show that the channel seldom is i.i.d. Rayleigh fading. However, in many situations, the performance that can be achieved over the measured channels closely follows the performance that is predicted by the theoretical analysis based on the i.i.d. Rayleigh fading assumption [49]. From the perspective of channel estimation, the i.i.d. Rayleigh fading model can also be seen as a worst-case model—correlation between channel coefficients can improve the estimation quality. For that reason and for tractability, the i.i.d. Rayleigh fading model is used in many of the included papers.

3.3 Channel Estimation

The massive MIMO base station needs to estimate the uplink and downlink channels in order to do coherent spatial processing. One way to estimate the channel is to dedicate part of the transmission to training signals, so called *pilots*. If the pilots are known to the receiver, they allow the receiver to estimate the channel over which the pilots have been transmitted.

In massive MIMO, the uplink channel is estimated by letting the users transmit known N_{p}-symbol long pilots $x_k^{\mathrm{UL}}[n] = \phi_k[n]$, $n = 0, \dots, N_{\mathrm{p}}-1$. For notational simplicity, a dedicated transmission block for channel estimation is considered whose length is $N = N_{\mathrm{p}}$. The channel is also assumed to have support L, i.e. $h_{mk}^{\mathrm{UL}}[\ell] = 0$ when $\ell \notin [0, \dots, L-1]$, $\forall m, k$. If the pilots are chosen such that:

$$\sum_{n=0}^{N_{\mathrm{p}}-1} \phi_k[n]\phi_{k'}^*[n+\ell] = \begin{cases} N_{\mathrm{p}}, & \text{if } k = k', \ell = 0, \\ 0, & \text{if } k = k', \ell = 1, \dots, L-1, \\ 0, & \text{if } k \neq k', \ell = 0, \dots, L-1, \end{cases} \tag{68}$$

where the indices are taken modulo N_{p}, the base station can estimate the channel by correlating the received signal with the known pilot sequences:

$$r_{mk}[\ell] = \sum_{n=0}^{N_{\mathrm{p}}-1} y_m^{\mathrm{UL}}[n]\phi_k^*[n+\ell]. \tag{69}$$

By using (51) and (68), it can be seen that

$$r_{mk}[\ell] = \sqrt{\beta_k P_k} N_{\mathrm{p}} h_{mk}^{\mathrm{UL}}[\ell] + z'_{mk}[\ell], \tag{70}$$

where

$$z'_{mk}[\ell] \triangleq \sum_{n=0}^{N_{\mathrm{p}}-1} z_m^{\mathrm{UL}}[n]\phi_k^*[n+\ell] \sim \mathcal{CN}\left(0, \frac{N_{\mathrm{p}} N_0}{T}\right) \tag{71}$$

is a white Gaussian noise sequence that is uncorrelated across ℓ. The correlation output $r_{mk}[\ell]$ is thus a noisy observation of the channel coefficient $h_{mk}^{\mathrm{UL}}[\ell]$. Note that the length of the pilot has to be $N_{\mathrm{p}} \geq KL$ in order to fulfill (68).

When the base station knows the correlation matrices of the channels $\{\mathbf{R}_k[\ell]\}$, it can compute the linear minimum mean-square error (LMMSE)

estimates of the channel from the observations. If the observations are stacked into a vector:

$$\mathbf{r}_k[\ell] \triangleq (r_{1k}[\ell], \ldots, r_{Mk}[\ell])^{\mathsf{T}} \tag{72}$$

$$= \sqrt{\beta_k P_k} N_{\mathrm{p}} \mathbf{h}_k^{\mathrm{UL}}[\ell] + \mathbf{z}'_k[\ell], \tag{73}$$

where $\mathbf{h}_k^{\mathrm{UL}}[\ell]$ and $\mathbf{z}'_k[\ell]$ are defined analogously to $\mathbf{r}_k[\ell]$, the LMMSE estimate of the channel is given by:

$$\hat{\mathbf{h}}_k^{\mathrm{UL}}[\ell] = \mathsf{E}\left[\mathbf{h}_k^{\mathrm{UL}}[\ell]\mathbf{r}_k^{\mathsf{H}}[\ell]\right] \mathsf{E}\left[\mathbf{r}_k[\ell]\mathbf{r}_k^{\mathsf{H}}[\ell]\right]^{-1} \mathbf{r}_k[\ell] \tag{74}$$

$$= \sqrt{\beta_k P_k} N_{\mathrm{p}} \mathbf{R}_k[\ell] \left(\beta_k P_k N_{\mathrm{p}}^2 \mathbf{R}_k[\ell] + \frac{N_{\mathrm{p}} N_0}{T} \mathbf{I}_M\right)^{-1} \mathbf{r}_k[\ell]. \tag{75}$$

The correlation matrix of the error is:

$$\mathbf{E}_k[\ell] \triangleq \mathsf{E}\left[(\mathbf{h}_k^{\mathrm{UL}}[\ell] - \hat{\mathbf{h}}_k^{\mathrm{UL}}[\ell])(\mathbf{h}_k^{\mathrm{UL}}[\ell] - \hat{\mathbf{h}}_k^{\mathrm{UL}}[\ell])^{\mathsf{H}}\right] \tag{76}$$

$$= \mathbf{R}_k[\ell] - \mathsf{E}\left[\mathbf{h}_k^{\mathrm{UL}}[\ell]\mathbf{r}_k^{\mathsf{H}}[\ell]\right] \mathsf{E}\left[\mathbf{r}_k[\ell]\mathbf{r}_k^{\mathsf{H}}[\ell]\right]^{-1} \mathsf{E}\left[\mathbf{h}_k^{\mathrm{UL}}[\ell]\mathbf{r}_k^{\mathsf{H}}[\ell]\right]^{\mathsf{H}} \tag{77}$$

$$= \mathbf{R}_k[\ell](\mathbf{I}_M - \mathbf{C}_k[\ell]), \tag{78}$$

where the matrix

$$\mathbf{C}_k[\ell] \triangleq \beta_k P_k N_{\mathrm{p}} \left(\beta_k P_k N_{\mathrm{p}} \mathbf{R}_k[\ell] + \frac{N_0}{T} \mathbf{I}_M\right)^{-1} \mathbf{R}_k[\ell]. \tag{79}$$

In the frequency domain, the channel estimate is:

$$\hat{\mathbf{h}}_k^{\mathrm{UL}}[\nu] = \sum_{\ell=0}^{L-1} \hat{\mathbf{h}}_k^{\mathrm{UL}}[\ell] e^{-j2\pi\ell\nu/N} \tag{80}$$

and the correlation matrix of the error:

$$\mathbf{E}[\nu] \triangleq \mathsf{E}\left[(\mathbf{h}_k^{\mathrm{UL}}[\ell] - \hat{\mathbf{h}}_k^{\mathrm{UL}}[\ell])(\mathbf{h}_k^{\mathrm{UL}}[\ell] - \hat{\mathbf{h}}_k^{\mathrm{UL}}[\ell])^{\mathsf{H}}\right] \tag{81}$$

$$= \sum_{\ell=0}^{L-1} \mathbf{E}_k[\ell]. \tag{82}$$

In the i.i.d. Rayleigh fading case, where $\mathbf{R}_k[\ell] = \sigma_k^2[\ell]\mathbf{I}_M$ for all taps ℓ, all matrices in (78) are diagonal and the LMMSE estimates can be computed

per antenna as:

$$\hat{h}_{mk}^{\mathrm{UL}}[\ell] = \frac{\mathsf{E}\left[h_{mk}^{\mathrm{UL}}[\ell] r_{mk}^{*}[\ell]\right]}{\mathsf{E}\left[|r_{mk}[\ell]|^2\right]} r_{mk}[\ell] \tag{83}$$

$$= \frac{\sqrt{\beta_k P_k}\sigma_k^2[\ell]}{\beta_k P_k N_{\mathrm{p}} \sigma_k^2[\ell] + N_0/T} r_{mk}[\ell], \tag{84}$$

whose mean square-error is given by

$$\mathsf{E}\left[|\hat{h}_{mk}^{\mathrm{UL}}[\ell] - h_{mk}^{\mathrm{UL}}[\ell]|^2\right] = \sigma_k^2[\ell] - \frac{\left|\mathsf{E}\left[h_{mk}^{\mathrm{UL}}[\ell] r_{mk}^{*}[\ell]\right]\right|^2}{\mathsf{E}\left[|r_{mk}[\ell]|^2\right]} \tag{85}$$

$$= \sigma_k^2[\ell]\left(1 - c_k[\ell]\right), \tag{86}$$

where

$$c_k[\ell] \triangleq \frac{\beta_k P_k N_{\mathrm{p}} \sigma_k^2[\ell]}{\beta_k P_k N_{\mathrm{p}} \sigma_k^2[\ell] + N_0/T}. \tag{87}$$

The LMMSE estimate of the frequency response of the channel is given by

$$\hat{h}_{mk}^{\mathrm{UL}}[\nu] = \sum_{\ell=0}^{L-1} \hat{h}_{mk}^{\mathrm{UL}}[\ell] e^{-j2\pi\ell\nu/N}. \tag{88}$$

The mean-square error of the estimated frequency response is

$$\mathsf{E}\left[|\hat{h}_{mk}^{\mathrm{UL}}[\nu] - h_{mk}^{\mathrm{UL}}[\nu]|^2\right] = 1 - c_k, \tag{89}$$

where the *channel estimation quality* is defined as

$$c_k \triangleq \sum_{\ell=0}^{L-1} \sigma_k^2[\ell] c_{mk}[\ell] = \sum_{\ell=0}^{L-1} \frac{\sigma_{mk}^4[\ell] \beta_k P_k N_{\mathrm{p}}}{\sigma_k^2[\ell] \beta_k P_k N_{\mathrm{p}} + N_0/T}. \tag{90}$$

Remember that the large-scale fading β_k of the downlink and uplink channels were the same in (50) and (51). This is a very natural assumption: the uplink and downlink signals travel the same paths, just in opposite directions, and should experience the same attenuation. If the small-scale fading is *reciprocal* in the same way, i.e. that $h_{mk}^{\mathrm{DL}}[\ell] = h_{mk}^{\mathrm{UL}}[\ell]$, then it is enough to estimate the uplink channel to also get an estimate of the downlink channel.

If the transmission is done in so-called *time-division duplex* (TDD), i.e. the uplink and downlink use the same frequency band and are separated in time,

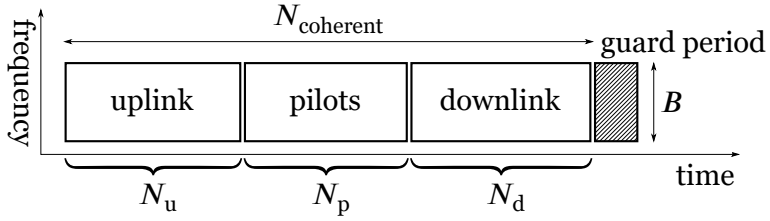


Figure 7: Massive MIMO frame structure

it makes sense to assume that the channel is reciprocal as long as the time separation is short. In practice, the hardware is not reciprocal and requires calibration to avoid mismatches between the uplink and downlink when exploiting channel reciprocity. Since this hardware mismatch changes slowly over time, it can be handled by relatively simple calibration methods [10].

Massive MIMO is usually assumed to be implemented in TDD, as illustrated in Figure 7, in order to use the reciprocity argument and to avoid estimating the downlink channel directly. In TDD mode, the whole allocated bandwidth is first used for the uplink, where the users transmit a block of $N = N_{\text{u}}$ symbols, and then for the training period[1], where the users transmit a block of $N = N_{\text{p}}$ pilots. Thereafter, the whole bandwidth is used for the downlink and the base station transmits a block of $N = N_{\text{d}}$ symbols to each user.

Potentially, a guard period is required when the system switches back to the uplink after the downlink, to make sure that the weak uplink signals do not drown in any residuals of the strong downlink signals at the base station. This kind of guard period is used in legacy systems, such as the TTD mode of LTE [50]. There is no guard period in between the uplink and the downlink however.

In this thesis, TDD transmission and perfect reciprocity is assumed. The superscript in the notation is therefore dropped

$$h_{mk}[\ell] = h_{mk}^{\text{UL}}[\ell] = h_{mk}^{\text{DL}}[\ell]. \tag{91}$$

Similarly, the channel estimate is written:

$$\hat{h}_{mk}[\ell] \triangleq \hat{h}_{mk}^{\text{UL}}[\ell], \tag{92}$$

[1]It is not strictly required that the uplink data transmission and the training period are separate in time. The users could send pilots interleaved with the uplink data, as long as the data and the pilots are orthogonal. Treating the uplink and the training period as separate blocks does not compromise the generality of the results however, and is a way to simplify the mathematical exposition.

which also is the estimate of the downlink channel. In analogy to the time-domain notation, the frequency response of the channel is denoted $h_{mk}[\nu]$ and its estimate $\hat{h}_{mk}[\nu]$. Note that these Fourier transforms are implicit functions of the block size N, which is not necessarily the same in the uplink and downlink.

In reality, the uplink and downlink will not be perfectly reciprocal due to channel aging—the fact that the channel changes over time—and due to difference in transmitter hardware in the uplink and downlink. The latter issue can be dealt with through *reciprocity calibration* [51, 52]. The former issue of channel aging, however, is more difficult to deal with since the evolution of the channel often depends on too many unknown factors in the surroundings.

Some methods to deal with channel aging and to make the downlink longer are proposed in, e.g., [53–55], where channel aging is investigated. If a guard period is needed between the downlink and uplink, longer downlink transmissions would require less frequent switching and less resources wasted on guard periods. Since the performance degradation due to channel aging increases with the time since the last uplink pilots were sent, it will limit the maximum possible duration of the downlink channel.

To estimate the actual downlink channel directly, each base station antenna would have to transmit pilots that fulfill a constraint similar to (68) (the indices would run from 1 to M instead of to K). These pilots therefore would have to be at least ML symbols long. Since it would be the users who estimate the channel, they would have to feed the estimate back to the base station. Both the ML-symbol long downlink pilots and the feedback would consume valuable resources that otherwise could be used for data transmission. This is one reason for wanting to avoid direct estimation of the downlink channel. Another reason to avoid ML-symbol long downlink pilots is to make the system fully scalable in the number of base station antennas, i.e. the system design should be such that the number of antennas M can be increased without limit, which is not possible if ML-symbol long pilots shall fit in a finite coherence time.

The number of symbol durations that have to be used for channel estimation is thus $N_{\mathrm{p}} = \mu KL$, $\mu \geq 1$, where a *pilot excess factor* $\mu = 1$ gives the shortest possible pilots that still fulfill (68). Since the channel has to be estimated once every coherence time, i.e. once every N_{coherent} symbol durations, one has to consider if the length of the pilots fits in this interval and how much of the coherence time that has to be used for channel estimation and not for data transmission. Using the relations in (32) on page 23 and (43) on page 25, the fraction of the coherence time that is used for channel

estimation is

$$\frac{N_{\mathrm{p}}}{N_{\mathrm{coherent}}} \approx \frac{\mu K \sigma_\tau / T}{1/(T\sigma_\upsilon)} \tag{93}$$

$$= \mu K \sigma_\tau \sigma_\upsilon. \tag{94}$$

For this fraction to be less than one, the number of users has to be limited to below

$$K < \frac{1}{\mu \sigma_\tau \sigma_\upsilon}. \tag{95}$$

The coherence time will therefore limit the number of user that can be served, but not the number of antennas the base station can be equipped with. Using the example numbers derived for the outdoor channel in Section 2.2 on page 22 ($\sigma_\tau = 3.3\,\mu\mathrm{s}$ and $\sigma_\upsilon = 400\,\mathrm{Hz}$), we get

$$\frac{N_{\mathrm{p}}}{N_{\mathrm{coherent}}} \approx 0.00132 \mu K, \tag{96}$$

$$K < \frac{757}{\mu}. \tag{97}$$

Since the number of concurrently active users in one communication system seldom is as large as 757, this indicates that this kind of pilot-based channel estimation is feasible with regards to the coherence time as long the pilot excess factor is small.

If channel reciprocity does not hold, for example if *frequency-division duplex* (FDD) is employed and the uplink and downlink use different carrier frequencies, techniques such as beam training, channel quantization, compressed sensing etc. can possibly be used to reduce the training overhead.

If the downlink channel can be described in terms of a small set of parameters, it is said to be sparse. In this case, the downlink channel can be fed back to the base station by transmitting only these, or a subset of these channel parameters, which reduces the feedback overhead [56]. It might also be possible to estimate a sparse channel based on pilots shorter than ML symbols using methods such as beam training [57] or compressed sensing [58].

These techniques, however, rely on the assumption of a sparse channel or, at least, that the channel consists of only a few significant components. Measurements show that many of these methods lead to significant degradation in performance in comparison to reciprocity based full channel estimation just because the channel is not sparse in many common situations [59]. In these cases, TDD operation is to prefer over FDD operation.

In order to do pilot based channel estimation, each user needs to know which pilot sequence to transmit and when it is supposed to be transmitted. In this thesis, this kind of *system information* is assumed to be known throughout the system. In practice, however, system information has to be transmitted to the users before the channel is estimated. In massive MIMO, the transmission of system information can be done at least as efficient as from any legacy single-antenna transmitter, but without channel state information it is difficult to obtain a useful array gain as in the coherent transmission of the data in the uplink and the downlink. In fact, by using orthogonal space-time block codes or knowledge about the correlation of the channel, the transmission of system information can be done more reliable and with less latency than in many legacy systems [60, 61].

Collisions can occur when more than one user try to access the same channel by sending the same pilot. Such events can be resolved by traditional random access techniques—e.g., scheduling or ALOHA—or by more efficient techniques that rely on channel hardening that allows for the simple detection of collisions, see [62] for example. Throughout the thesis, however, it is assumed that the pilot sequences already are allocated among the users and that no collision occurs.

3.4 Precoding and Decoding

To send information, the base station *precodes* the signals in such a way that each user receives the signal that is intended for it. To receive information that each user has transmitted, the base station *decodes* the received signals in such a way that the information from each user is separated. Sometimes precoding is also referred to as *beamforming* and decoding as *receive combining*. Collectively, precoding and decoding are referred to as *spatial processing*.

In the downlink, the *symbols* that are to be transmitted to the user k are denoted $s_k^{\mathrm{DL}}[\nu]$, $\nu = 0, \ldots, N_{\mathrm{d}} - 1$. The general precoder is a function of these symbols and the channel estimates that outputs the transmit signals:

$$(\hat{h}_{mk}[\nu], s_k^{\mathrm{DL}}[\nu], \forall k, m) \mapsto (x_m^{\mathrm{DL}}[\nu], \forall m), \quad \forall \nu. \tag{98}$$

Nonlinear precoders, which map the symbols onto the transmit signals in a nonlinear fashion, include the discrete-time constant-envelope precoder that is studied in Paper A and the continuous-time constant-envelope precoder that is proposed in Paper B. Both of these precoders aim at producing precoded signals that are easy to generate with low-end hardware.

A *linear precoder* weights the symbols linearly by the precoding weights $w_{mk}[\nu]$ to produce the transmit signal:

$$x_m^{\mathrm{DL}}[\nu] = \sum_{k=1}^{K} w_{mk}[\nu]\sqrt{\xi_k}s_k^{\mathrm{DL}}[\nu], \tag{99}$$

where the power allocation factors ξ_k make it possible to allocate the power between the users. To ensure that the power of the transmit signal meets the power constraint:

$$\sum_{m=1}^{M} \mathsf{E}\left[|x_m^{\mathrm{DL}}[\nu]|^2\right] = 1, \quad \forall \nu, \tag{100}$$

the power allocation factors sum to unity: $\sum_{k=1}^{K} \xi_k = 1$, the symbols are normalized to unit power: $\mathsf{E}\left[|s_k^{\mathrm{DL}}[\nu]|^2\right] = 1$, and the precoding weights are scaled such that

$$\sum_{m=1}^{M}\sum_{k=1}^{K} \xi_k \mathsf{E}\left[|w_{mk}[\nu]|^2\right] = 1, \quad \forall \nu. \tag{101}$$

For mathematical tractability, the expectations in (100) and (101) are taken with respect also to the small-scale fading of the channel throughout the thesis. This is called a *long-term power constraint*, because it ensures that the radiated power meets the constraint when averaged over many coherence times. In a practical system, it might be desirable to fulfill the power constraint in each coherence time and take the expectation in (100) with respect to the random symbols and remove the expectation operator in (101). This constraint is called a *short-term power constraint*. Due to the law of large numbers, however, long-term and short-term power constraints are very similar in massive MIMO when the channel to different users are uncorrelated. Therefore the choice of power constraint makes little difference to the system performance.

To see the effect of the precoding on the received signal, the channel is written as the sum of the estimated channel and the estimation error: $h_{mk}[\nu] = \hat{h}_{mk}[\nu] + e_{mk}[\nu]$. Then the received signal of user k in (52) on

page 31 is given by:

$$y_k^{\mathrm{DL}}[\nu] = \sqrt{\beta_k P \xi_k} s_k^{\mathrm{DL}}[\nu] \sum_{m=1}^{M} \hat{h}_{mk}[\nu] w_{mk}[\nu] + \sqrt{\beta_k P} \sum_{k' \neq k} \sqrt{\xi_{k'}} s_{k'}^{\mathrm{DL}}[\nu] \sum_{m=1}^{M} \hat{h}_{mk}[\nu] w_{mk'}[\nu] + \sqrt{\beta_k P} \sum_{k'=1}^{K} \sqrt{\xi_{k'}} s_{k'}^{\mathrm{DL}}[\nu] \sum_{m=1}^{M} e_{mk}[\nu] w_{mk'}[\nu] + z_{k'}^{\mathrm{DL}}[\nu]. \quad (102)$$

The received signal is thus the sum of four terms: one desired signal that contains the symbol $s_k^{\mathrm{DL}}[\nu]$, one interuser interference that contains the symbols to other users $s_{k'}^{\mathrm{DL}}[\nu]$, one term that contains the channel estimation error and one noise term.

By choosing the weights properly, it is possible to make the desired term large and to suppress the interuser interference term; without knowledge of the channel estimation error or the noise, little can be done about these two terms in the precoding however. As will be seen later, fortunately, these terms combine noncoherently and their power becomes small in comparison to the power of the desired signal. With proper precoding, the desired signal will thus dominate the received signal and the users can treat the other terms as noise. The users never have to do any advanced equalization and therefore do not need to estimate the channel. The received signal can directly be used as the symbol estimate:

$$\hat{s}_k^{\mathrm{DL}}[\nu] \triangleq y_k^{\mathrm{DL}}[\nu]. \quad (103)$$

In the uplink, the base station performs decoding instead of precoding. Again, the users do not need any knowledge of the channel nor do they have to do any special signal processing. They simply transmit their symbols $\{s_k^{\mathrm{UL}}[\nu]\}$ in the frequency domain, i.e. let

$$x_k^{\mathrm{UL}}[\nu] = s_k^{\mathrm{UL}}[\nu]. \quad (104)$$

To meet the power constraint, the symbols are normalized to have unit power: $\mathsf{E}\left[|s_k^{\mathrm{UL}}[\nu]|^2\right] = 1$. The base station then applies a *linear decoder* to the received signals and the output is the estimate of the transmitted symbol:

$$\hat{s}_k^{\mathrm{UL}}[\nu] \triangleq \sum_{m=1}^{M} w_{mk}[\nu] y_m^{\mathrm{UL}}[\nu]. \quad (105)$$

Note that the decoder weights and the precoding weights have been chosen to be the same. This can partly be motivated by the channel reciprocity assumption—they are equalizing the same channel. For a more thorough discussion about using the same weights for decoding and precoding, see [63, Ch. 4.3.2], where an argument about uplink–downlink duality is used to motivate using the same spatial processing weights: It is possible to achieve the same data rates in the downlink as in the uplink by choosing the precoder in the same way as the decoder if the power allocation factors $\{\xi_k\}$ are chosen in a certain way.

The transmission method described above, where the symbols were transmitted on different discrete frequencies ν, is many times referred to as *single-carrier transmission with frequency domain equalization* (SC-FDE). To reduce the PAR of the transmit signal in the uplink, the symbols can just as well be transmitted in the time domain, i.e. instead of choosing the transmit signals as in (104), they are directly given in the time-domain as

$$x_k^{\mathrm{UL}}[n] = s_k^{\mathrm{UL}}[n]. \tag{106}$$

The decoding of these signals in (105) would then instead produce estimates of the discrete-time Fourier transform of the symbols:

$$\hat{s}_k^{\mathrm{UL}}[\nu] \triangleq \sum_{m=1}^{M} w_{mk}[\nu] y_m^{\mathrm{UL}}[\nu]. \tag{107}$$

By an inverse transform, however, the estimates of the time-domain symbols are recovered:

$$\hat{s}_k^{\mathrm{UL}}[n] = \frac{1}{\sqrt{N_{\mathrm{u}}}} \sum_{\nu=0}^{N_{\mathrm{u}}-1} \hat{s}_k^{\mathrm{UL}}[\nu] e^{j2\pi n\nu/N_{\mathrm{u}}}. \tag{108}$$

Time-domain transmission can also be done in the downlink, but in many cases these signals would not have a significantly lower PAR than the frequency-domain transmission described already. The reason for the high PAR is the precoding—the transmit signals in the downlink is the sum of a large number of symbols, which increases PAR. This is observed in Paper A, where the difference between transmission of symbols in the time- and the frequency-domain is studied.

In Paper A, it is also seen that the precoding (and subsequently also the decoding) can be done in the time domain with relatively short FIR filters, i.e. the multiplication in (99) can be done in the time domain by a filter with

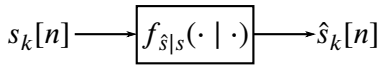


Figure 8: A generic channel is a description of the probability of the output given the input: $f_{\hat{s}|s}(\hat{s}_k[n] \mid s_k[n])$.

few taps. The number of taps in the filter only has to be a few more than the number of channel taps. In latency sensitive applications, a time-domain implementation could be interesting for that reason, because the delay of the decoding would equal the number of taps in the filter and not the length of the Fourier transform, which is significantly longer in general.

Even though the model presented here is for SC-FDE signals, many results drawn from it can indicate that similar results also hold for OFDM signals, which were described in (25) on page 21. The properties of OFDM signals and SC-FDE signals are similar. For example, they have similar PAR values and nonlinearities affect the in-band part of the signal similarly. In fact, the term OFDM is often used also for SC-FDE, which potentially is confusing. The simplicity of the symbol-sampled discrete-time model of SC-FDE systems that is described above is many times used to draw conclusions about OFDM systems too when the time-domain signal has to be studied, for example when nonlinearities are present. This is done in Papers A and C as well as in [19, 28, 64] for example.

3.5 The Use-and-Forget Bound

The performance of a communication system is often measured in the information theoretic data rate that it can convey reliably. If we treat the whole massive MIMO system—precoder or decoder and physical channel—as an information theoretic channel $f_{\hat{s}|s}(\cdot \mid \cdot)$, as in Figure 8, a limit on the possible data rate is given by the capacity of the channel, which will be defined shortly. To allow for a compact exposition, both the uplink and downlink symbols $s_k^{\mathrm{UL}}[n]$, $s_k^{\mathrm{DL}}[n]$ are denoted by $s_k[n]$, and their estimates $\hat{s}_k^{\mathrm{UL}}[n]$, $\hat{s}_k^{\mathrm{DL}}[n]$ by $\hat{s}_k[n]$.

By modeling the transmitted symbol $s_k[n]$ as a stochastic variable, the *mutual information* between it and its estimate is denoted

$$I(s_k[n]; \hat{s}_k[n]) \triangleq \int f_{s,\hat{s}}(s, \hat{s}) \log \left(\frac{f_{s,\hat{s}}(s, \hat{s})}{f_s(s) f_{\hat{s}}(\hat{s})} \right) \mathrm{d}s\mathrm{d}\hat{s}. \tag{109}$$

Despite the notation, the mutual information is a function of the probability density $f_s(s_k[n])$ of the symbol $s_k[n]$ and the channel $f_{\hat{s}|s}(\cdot \mid \cdot)$—the

conditional probability of the estimate $\hat{s}_k[n]$ given the symbol.

By maximizing the mutual information with respect to the distribution of the symbol under a power constraint $\mathsf{E}[|s_k[n]|^2] \leq 1$, we obtain the *capacity* [65, eq. (7.3.1)] of the channel $f_{\hat{s}|s}(\cdot \mid \cdot)$:

$$C_k \triangleq \sup_{f_s : \mathsf{E}[|s_k[n]|^2] \leq 1} I(s_k[n]; \hat{s}_k[n]), \tag{110}$$

which is the limit on the data rate of reliable communication over the channel—for any rate $R_k < C_k$ and error probability $\epsilon > 0$, there exists a communication method with this data rate R_k and with a probability of error smaller than ϵ. A rate $R_k < C_k$ is called an *achievable rate*.

The base of the logarithm in (109) is arbitrary, but decides the unit of the data rate. The base 2 will be used henceforth, so that the information is measured in bits-per-channel-use (bpcu). In the context of the massive MIMO system the rate 1 bpcu shall be interpreted as "1 bit of information can be conveyed on average in one symbol duration T over the bandwidth B".

It should be noted here that what we have defined as "capacity" is the capacity of the single channel between one user and the base station under the assumption that the parameters of the transmission to and from the other users are fixed. The capacity of one user could increase for example if we reduce the data rate of another user. For a complete characterization of the capacity of a multiuser channel, one should really consider a *capacity region*, i.e. the space of all tuples of individual rates $(R_1, \ldots, R_K)$ that are jointly achievable. The capacity region is explained in detail in [66, Ch. 4]. The concept of a capacity, where the parameters of other users are fixed, is adequate for explaining the information theoretic tools used in this thesis however.

An achievable rate for any communication system is given by the *use-and-forget bound*:

$$R_k = \log_2 \left(1 + \frac{\left| \mathsf{E}\left[s_k^*[n] \hat{s}_k[n] \right] \right|^2}{\mathsf{E}\left[|\hat{s}_k[n]|^2 \right] - \left| \mathsf{E}\left[s_k^*[n] \hat{s}_k[n] \right] \right|^2} \right) \quad \text{[bpcu]}. \tag{111}$$

By using Gaussian distributed symbols, the expectations can be evaluated and an achievable rate is obtained. The rate can be derived based on results in [67] as in (68) on page 179 in Paper C. This achievable rate can be used to evaluate the performance of any precoder and decoder, also nonlinear ones such as the low-PAR precoders.

The expectations in (111) are evaluated with respect to all sources of randomness, also with respect to the small-scale fading. As such, it is a

so-called *ergodic rate*, i.e. coding has to be done over many realizations of the small-scale fading—over many coherence intervals—as well as over many realizations of the thermal noise.

To let the expectations be taken with respect to the small-scale fading, implies, in this case, that the detector does not use any information about the small-scale fading other than its statistics. In the downlink, this makes sense, since the users only know the large-scale fading. In the uplink, however, the base station and the detector has access to estimates of the small-scale fading, and could use those to improve the detection. Indeed, a larger achievable rate can be obtained by conditioning the expectations in (111) on the channel estimates and applying another expectation over the whole logarithm. The thus obtained bound is called the *side information bound*; for details, see [68, Ch. 2.3.5]. To derive an expression that does not depend on the small-scale fading, however, it will be assumed that the base station "forgets" its channel state information after the decoding and the use-and-forget bound is evaluated without any conditioning also for the uplink. Hence, its name: the estimates are *used* in the decoding but *forgot* in the detection.

In many cases, the use-and-forget bound is a tight bound on the capacity [69]. This happens when *channel hardening* occurs, i.e. when the coherent channel gain

$$\mathsf{E}\left[s_k^*[n]\hat{s}_k[n] \mid \{\hat{h}_{mk}[\ell]\}\right], \tag{112}$$

conditioned on the estimates of the small-scale fading $\{\hat{h}_{mk}[\ell]\}$, is well approximated by its mean value

$$\mathsf{E}\left[s_k^*[n]\hat{s}_k[n]\right]. \tag{113}$$

In cases, where channel hardening does not occur, the use-and-forget bound might give an overly pessimistic estimation of the actual data rates that are achievable in the system. For example, channel hardening might not occur when the number of antennas is small or when there is a basis in which the channel correlation matrix is sparse, such as when there is keyhole propagation or when there is little multipath propagation. In these cases, other bounds might be needed to evaluate the performance of the system; see [69, 70].

For linear precoders and decoders, the use-and-forget bound in (111) can be further analyzed in the case of i.i.d. Rayleigh fading. It can be seen from (102) and (103) that the numerator in (111) equals $c_k\xi_k\beta_k P G_k$ in the

downlink, where the *array gain* is defined as

$$G_k \triangleq \frac{1}{c_k} \left| \sum_{m=1}^{M} \mathsf{E}\left[w_{mk}[\nu] \hat{h}_{mk}[\nu] \right] \right|^2 . \tag{114}$$

It is noted that just by choosing the weights $w_{mk}[\nu]$ such that they are correlated to the conjugate of the channel estimates $\hat{h}^*_{mk}[\nu]$, the array gain will scale with the number of antennas M. This will become clearer when some common choices of weights are introduced later. Similarly to the downlink, it can be shown that the numerator in (111) for the uplink equals $c_k \beta_k P_k G_k$.

The downlink channel in (102) and (103) can be rewritten as a SISO link with a deterministic effective channel gain:

$$\begin{aligned} \hat{s}^{\mathrm{DL}}_k[\nu] = {} & \sqrt{\beta_k P \xi_k} s^{\mathrm{DL}}_k[\nu] \, \mathsf{E}\Big[\sum_{m=1}^{M} w_{mk}[\nu] \hat{h}_{mk}[\nu] \Big] \\ & + \sqrt{\beta_k P} \sum_{k'=1}^{K} \sqrt{\xi_{k'}} s^{\mathrm{DL}}_{k'}[\nu] \Big(\sum_{m=1}^{M} w_{mk'}[\nu] \hat{h}_{mk}[\nu] - \delta[k-k'] \, \mathsf{E}\Big[\sum_{m=1}^{M} w_{mk'}[\nu] \hat{h}_{mk}[\nu] \Big] \Big) \\ & + \sqrt{\beta_k P} \sum_{k'=1}^{K} \sqrt{\xi_{k'}} s^{\mathrm{DL}}_{k'}[\nu] \sum_{m=1}^{M} e_{mk}[\nu] w_{mk'}[\nu] + z^{\mathrm{DL}}_{k'}[\nu], \end{aligned} \tag{115}$$

where $\delta[n] = 1$ when $n = 0$ and $\delta[n] = 0$ otherwise. The second term is the *interuser interference* from user k' onto user k; its normalized variance is denoted

$$I_{kk'} \triangleq \frac{1}{c_k} \mathsf{E}\left[\left| \sum_{m=1}^{M} w_{mk'}[\nu] \hat{h}_{mk}[\nu] - \delta[k-k'] \, \mathsf{E}\Big[\sum_{m=1}^{M} w_{mk'}[\nu] \hat{h}_{mk}[\nu] \Big] \right|^2 \right] \tag{116}$$

in the downlink and $I_{k'k}$ in the uplink.

It is observed that the self-interference I_{kk} generally is non-zero. In the downlink, this is natural because the user, who only has statistical knowledge about the channel, does not know the actual effective channel gain of the transmission $\sum_{m=1}^{M} w_{mk}[\nu] \hat{h}_{mk}[\nu]$. By using its expected value as the estimate of the effective channel gain, as in (115), an estimation error arises, whose impact on the rate is captured by its variance, the self-interference I_{kk}.

In many massive MIMO channels, this simplistic estimation of the effective channel gain works well, but more sophisticated estimation methods of the effective channel gain have been proposed [70]. Arguably, the self-interference is zero in the uplink, because the base station has channel state information and can compute the effective channel gain. Using $I_{kk} = 0$ in

the uplink would yield a higher rate than what is obtained from using the interference in (116). The improvement is small however and the definition in (116) is used in this thesis also for $k' = k$ for simplicity.

Since the power of the third term, the error due to imperfect channel state information, is $K\beta_k P(1-c_k)$ in the downlink and $\sum_{k'=1}^{K} \beta_{k'} P_{k'}(1-c_{k'})$ in the uplink, the use-and-forget bound in (111) becomes

$$R_k^{\text{DL}} = \log_2\left(1 + \frac{c_k \xi_k \beta_k P G_k}{\beta_k P \sum_{k'=1}^{K} \left(1 - c_k(1-\xi_{k'} I_{kk'})\right) + N_0/T}\right) \quad \text{[bpcu]} \tag{117}$$

for linear precoders in the downlink and

$$R_k^{\text{UL}} = \log_2\left(1 + \frac{c_k \beta_k P_k G_k}{\sum_{k'=1}^{K} \beta_{k'} P_{k'} \left(1 - c_{k'}(1-I_{k'k})\right) + N_0/T}\right) \quad \text{[bpcu]} \tag{118}$$

for linear decoders in the uplink. Because the array gain and interference suppression properties of a precoder and decoder are captured by the two parameters G_k and $I_{k'k}$, they are referred to as the *characteristic parameters* of the spatial processing method. The big fraction inside the argument of the logarithms in (117) and (118) is referred to as the SINR (signal-to-interference-and-noise ratio) of the use-and-forget bound.

Since these rates do not depend on the small-scale fading, they simplify the system design. For example, when doing power control between users to balance all data rates in the system, this only has to be done once for every realization of the large-scale fading $\{\beta_k\}$, which is assumed to change slowly compared to the small-scale fading. Methods for power control are detailed in [68, Ch. 5], where an approach that guarantees that all users within a cell get the same data rate is advocated.

It should be noted that the power of the pilot transmission in the uplink affects both the data rates of the uplink and the downlink, since the channel quality c_k is a function of this power. How to balance the power of the pilots and the power of the data transmission is studied in, for example, [71].

3.6 Common Linear Decoders and Precoders

It was shown in Section 3.5 how the choice of decoding and precoding weights affects the array gain G_k and interference suppression $I_{kk'}$ of the transmission. One way to achieve a good balance between the two, is to choose the

spatial processing weights according to the LMMSE estimates of the uplink symbols, which maximizes the SINR of the side information bound [63, Ch. 4]. Since the spatial processing weights $\{w_{mk}[\nu]\}$ are the same for the decoder and the precoder, the uplink will be used below to motivate the different choices of spatial processing weights.

The decoder will be defined in terms of the estimated channel matrix:

$$\hat{\mathbf{H}}[\nu] \triangleq \begin{pmatrix} \hat{h}_{11}[\nu] & \cdots & \hat{h}_{1K}[\nu] \\ \vdots & \ddots & \vdots \\ \hat{h}_{M1}[\nu] & \cdots & \hat{h}_{MK}[\nu] \end{pmatrix} \mathbf{D}_{\mathbf{c}}^{-1/2}, \tag{119}$$

where the columns are normalized by the positive-definite square root of the diagonal matrix $\mathbf{D}_{\mathbf{c}} = \operatorname{diag}(c_1, \ldots, c_K)$ in order to make each element of the matrix distributed as $\mathcal{CN}(0,1)$ when the channel is i.i.d. Rayleigh fading, which will simplify the derivation of the array gain G_k and the interference suppression $I_{kk'}$ later.

An amplitude scaled version of the LMMSE estimate of the transmit signals is given by using the decoding weights:

$$\mathbf{W}^{\mathsf{H}}[\nu] = \alpha \left(\mathsf{E}\left[\mathbf{x}[\nu]\mathbf{y}^{\mathsf{H}}[\nu]\right] \mathsf{E}\left[\mathbf{y}[\nu]\mathbf{y}^{\mathsf{H}}[\nu]\right]^{-1} \right)^{\mathsf{H}} \mathbf{D}_{\boldsymbol{\beta}}^{1/2} \mathbf{D}_{\mathbf{P}}^{1/2} \mathbf{D}_{\mathbf{c}}^{-1/2}, \tag{120}$$

where the vectors $\mathbf{x}[\nu] \triangleq (x_1[\nu], \ldots, x_M[\nu])^{\mathsf{T}}$, $\mathbf{y}[\nu] \triangleq (y_1[\nu], \ldots, y_M[\nu])^{\mathsf{T}}$, and the diagonal matrices $\mathbf{D}_{\boldsymbol{\beta}} = \operatorname{diag}(\beta_1, \ldots, \beta_K)$, $\mathbf{D}_{\mathbf{P}} \triangleq \operatorname{diag}(P_1, \ldots, P_K)$. The scaling factor $\alpha \in \mathbb{R}^+$ is chosen such that the power constraint for the precoder (101) on page 42 is fulfilled. In the uplink, the scaling factor does not affect the decoding, since it scales both the desired signal, interference and noise by the same amount. For the same reason, the multiplication with the diagonal matrices from the right does not affect the decoding. The introduction of these diagonal matrices will, however, simplify the derivation of other decoders in what follows. The conjugate transpose in (120) is introduced to simplify notation.

Computing the expectations in (120) results in the LMMSE decoder:

$$\mathbf{W}^{\mathsf{H}}[\nu] = \alpha \left(\hat{\mathbf{H}}[\nu] \mathbf{D}_{\boldsymbol{\beta}} \mathbf{D}_{\mathbf{P}} \mathbf{D}_{\mathbf{c}}^{-1} \hat{\mathbf{H}}[\nu] + \sum_{k=1}^{K} P_k \beta_k \mathbf{E}_k + \frac{N_0}{T} \mathbf{I}_M \right)^{-1} \hat{\mathbf{H}}[\nu] \mathbf{D}_{\boldsymbol{\beta}} \mathbf{D}_{\mathbf{P}} \mathbf{D}_{\mathbf{c}}^{-1}. \tag{121}$$

If the sum of the correlation matrices of the channel estimation error is a

scaled identity matrix, which is the case when the fading is i.i.d. Rayleigh,

$$\sum_{k=1}^{K} \beta_k P_k \mathbf{E}_k = \mathbf{I}_M \underbrace{\sum_{k=1}^{K} \beta_k P_k (1 - c_k)}_{\triangleq \lambda'}, \tag{122}$$

then the decoder in (121) can be simplified. Using the matrix identity:

$$(\mathbf{BC} + \lambda \mathbf{I})^{-1} \mathbf{B} = \mathbf{B}(\mathbf{CB} + \lambda \mathbf{I})^{-1}, \tag{123}$$

which is stated in [63, Lem. B.5], the expression in (121) becomes:

$$\mathbf{W}^{\mathsf{H}}[\nu] = \alpha \hat{\mathbf{H}}[\nu] \left(\hat{\mathbf{H}}^{\mathsf{H}}[\nu] \hat{\mathbf{H}}[\nu] + \left(\lambda' + \frac{N_0}{T} \right) \mathbf{D}_{\boldsymbol{\beta}}^{-1/2} \mathbf{D}_{\mathbf{P}}^{-1/2} \mathbf{D}_{\mathbf{c}}^{1/2} \right)^{-1}. \tag{124}$$

If the matrix product

$$\mathbf{D}_{\boldsymbol{\beta}}^{-1/2} \mathbf{D}_{\mathbf{P}}^{-1/2} \mathbf{D}_{\mathbf{c}}^{1/2} = \lambda'' \mathbf{I}_K \tag{125}$$

is a scaled identity matrix, the *regularized zero-forcing* decoder is obtained:

$$\mathbf{W}^{\mathsf{H}}[\nu] = \alpha \hat{\mathbf{H}}[\nu] \left(\hat{\mathbf{H}}^{\mathsf{H}}[\nu] \hat{\mathbf{H}}[\nu] + \lambda \mathbf{I}_K \right)^{-1}, \tag{126}$$

where $\lambda \triangleq \lambda''(\lambda' + N_0/T)$.

It is noted that, when power control is applied, the equality in (125) can be made true, which would make the received power and channel estimation quality the same for all users. Then the regularized zero-forcer is optimal in the sense of the side information bound, just like the LMMSE decoder. In [72], it was shown that the *precoding* weights that are obtained from a scaled identity regularization, as in (126), are optimal for the downlink in massive MIMO, if the aim is to give all users the same rate.

Also when (125) is not true, the regularized zero-forcer can be used. It still gives an overall good performance and balances between the array gain and interference suppression ability. From a practical point of view, the regularized zero-forcer is also easier to implement than the optimal LMMSE decoder. The LMMSE decoder requires the inversion of an M-dimensional matrix, whereas the regularized zero-forcer only requires the inversion of a K-dimensional matrix. Furthermore, the regularized zero-forcer does not depend on the individual powers $\{P_k\}$ of the users, which conceptually simplifies power control among users, since the choice of powers decouples from the choice of decoding weights.

If the scaled identity matrix in (126) is small compared to the diagonal elements in the Gramian matrix $\hat{\mathbf{H}}[\nu]\hat{\mathbf{H}}^{\mathsf{H}}[\nu]$, the inverse can be approximated by

$$\left(\hat{\mathbf{H}}^{\mathsf{H}}[\nu]\hat{\mathbf{H}}[\nu] + \lambda \mathbf{I}_K\right)^{-1} \approx \left(\hat{\mathbf{H}}^{\mathsf{H}}[\nu]\hat{\mathbf{H}}[\nu]\right)^{-1}. \tag{127}$$

This happens in high SNR, when $\lambda \to 0$, or when the number of antennas is large, because the diagonal elements of the Gramian matrix grow with the number of antennas. The decoder that is obtained under this assumption, i.e. the decoder

$$\mathbf{W}[\nu] = \alpha \left(\hat{\mathbf{H}}[\nu]\hat{\mathbf{H}}^{\mathsf{H}}[\nu]\right)^{-1} \hat{\mathbf{H}}^{\mathsf{H}}[\nu], \tag{128}$$

is referred to as the *zero-forcing* decoder. It is the decoder that minimizes the interference $I_{kk'}$ at the cost of slightly smaller array gain G_k than otherwise would be possible. For an i.i.d. Rayleigh fading channel, zero-forcing processing results in the characteristic parameters:

$$G_k = M - K, \quad I_{kk'} = 0. \tag{129}$$

When the regularization factor is large $\lambda \to \infty$, we note that

$$\mathbf{W}[\nu] \to \alpha \hat{\mathbf{H}}^{\mathsf{H}}[\nu]. \tag{130}$$

This happens when the received SNR is low, i.e. when $\beta_k P_k T/N_0$ is small for all users k. The precoder obtained by

$$\mathbf{W}[\nu] = \alpha \hat{\mathbf{H}}^{\mathsf{H}}[\nu] \tag{131}$$

is referred to as the *maximum-ratio* decoder. The maximum-ratio decoder maximizes the array gain. The interference is not suppressed however. For an i.i.d. Rayleigh fading channel, the characteristic parameters of maximum-ratio processing are:

$$G_k = M, \qquad I_{kk'} = 1. \tag{132}$$

The maximum-ratio decoder was the first decoder to be investigated in the original paper [1]. Because of its simplicity, it is still an interesting choice for massive MIMO. Its definition does not require any centralized computations for matrix inversion and its structure lends itself for parallel computations that can be implemented in a distributed manner—intermediate estimates can be computed per antenna and the final estimate is obtained just by summing these intermediate estimates up.

Using (117), (118) on page 49 and the values of G_k and $I_{k'k}$, closed-form achievable rates for maximum-ratio and zero-forcing precoding are obtained. These expressions were first derived for a flat-fading channel in [73] and then extended to frequency-selective channels for maximum-ratio precoding in [74]. These previous data rates are special cases of the more general rates in (117) and (118) that are derived in Paper A for the frequency-selective downlink with nonlinear power amplifiers and in Paper C for the frequency-selective uplink with one-bit ADCs. The achievable rate in (111) on page 46 is also used in Paper B.

Regularized zero-forcing can be used to adjust the performance continuously between the array gain maximization of maximum-ratio processing and interference suppression of zero-forcing processing, by varying the regularization parameter λ. This can be useful in systems with unknown noise and distortion sources that are not captured by the ideal linear model shown here. For example, in a system with nonlinear amplifiers that produce an unknown amount of distortion, the parameter λ can be chosen adaptively to maximize performance as is done in Paper A.

In multicell systems, more advanced precoding methods can be used to also suppress the intercell interference. To do that, the base station has to estimate the channel to all users in the system, including the users belonging to other cells. This estimation can be done by listening to the pilots sent by users in other cells in the same way as estimating the channel to the users in the own cell, if the timing of the training period of the other cells is known. In [75], it was shown that this *multicell precoding* outperforms precoding methods that do not consider the interference from other cells. The LMMSE processing in (121) can also be extended into *multicell LMMSE* processing that takes intercell interference into account, see [76, 77], [63, Cor. 4.2] for details.

3.7 Pilot Contamination

In a single-cell system, where one base station serves a set of users without being disturbed by neighboring transmission, a pilot excess factor of one is often enough to obtain good channel estimates. The reason is that the total energy of the received observations from the pilots in (70) on page 35 increases with the pilot length, which is large in a multiuser system also with $\mu = 1$.

A pilot excess factor greater than one might be needed in a multicell system, however, where many base stations concurrently serve disjoint sets

of users. If transmission from other cells is present on the same frequency and at the same time that the pilots are sent, then *pilot contamination* can arise [1], i.e. the channel estimates obtained can be contaminated by the nonorthogonal interfering signals from other cells. Apart from decreasing the channel estimation quality, pilot contamination also results in a channel estimate that is the sum, not only, of the desired channel but also of the channel of the contaminating transmitter.

When decoding is done based on the sum of two channels, the symbol estimate will be the sum of the data that is transmitted on those two channels, i.e. the symbol estimates in case of pilot contamination will contain the data from the contaminator as well, which will constitute additional coherent interference that is not apparent in the single-cell setup that has been described in the previous chapters. The coherent interference from the contaminating user will result in an additional term in the denominator of the SINR inside the logarithms of the rate expressions in (117) and (118) on page 49, which scales with G_k—just like the numerator. When the number of antennas grows, the array gain G_k grows with it and the SINR will approach a finite limit. This stands in contrast to the single-cell system, where the SINR grows without limit in the number of antennas. Consequently, pilot contamination can potentially degrade the performance significantly in a massive MIMO system.

One way to avoid pilot contamination would be to synchronize the transmission of pilots in all cells using the same frequency and to make each pilot orthogonal to all other pilots in all cells. Apart from the additional system complexity that stems from having to coordinate all cells, this could also require a pilot excess factor larger than one [78].

Other methods include: pilot decontamination of the received signals [79], to design the pilots in a clever way that mitigates the [80], filter out the contaminating channel from the estimate [81, 82], or try to estimate the channel of the contaminating transmitter separately [83, 84].

The authors in [76, 85], however, argue that the pilot contamination issue is an artifact of the i.i.d. Rayleigh fading model. For channel models, where the eigenvalues and eigenspaces of the correlation matrices are different across users, proper channel estimation, decoding and precoding can get rid of the pilot contamination problem at the cost of a smaller array gain G_k.

Still, pilot contamination has to be considered in the design of a multicell massive MIMO system if all benefits of massive MIMO shall be exploited. This thesis is about the impact of hardware impairments and is therefore limited to the study of single-cell systems without pilot contamination not to make the exposition unnecessarily convolved.

Chapter 4

Hardware

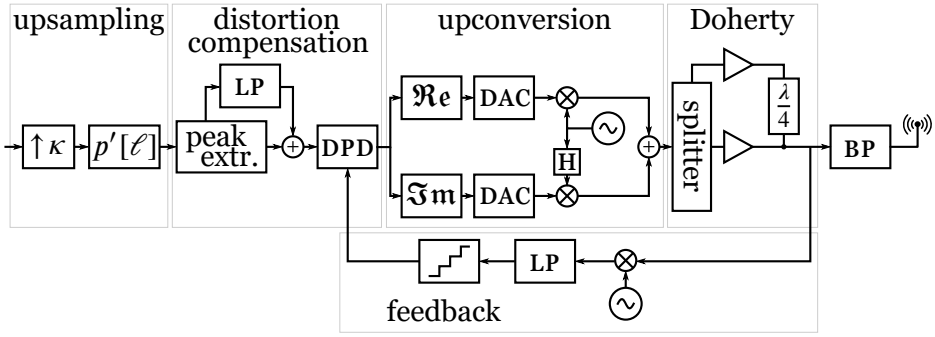
The hardware that generates a transmit signal from the discrete-time complex baseband signal is called a *transmitter*. The corresponding hardware that converts the received passband signal into a discrete-time complex baseband signal is called a *receiver*. Below follows a brief introduction to the hardware of the transmitters and receivers of a massive MIMO base station.

4.1 Transmitter Design

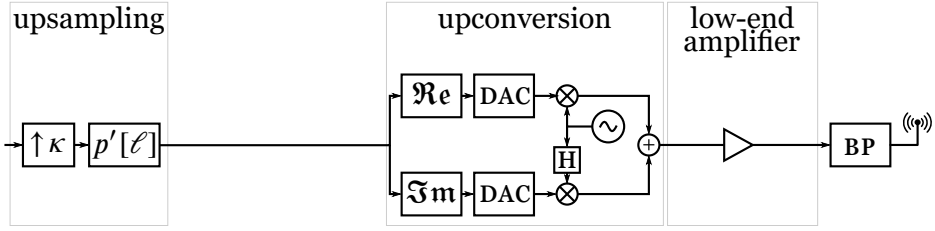
As a reference design, the conventional transmitter in Figure 9(a) is studied. The complexity, power consumption and cost of this conventional transmitter, however, makes it difficult to use in a massive MIMO base station, which requires hundreds of transmitters—one for each antenna. For massive MIMO, instead it is more relevant to study a low-end transmitter that employs inexpensive hardware and excludes much of the more complex hardware, such as distortion compensation that is not absolutely necessary for the signal generation. An example of such a low-end design is shown in Figure 9(b), which is the design studied in the included Papers A, B, F and G.

4.1.1 Upsampling

The first step in the transmitter is to upsample the symbol-sampled baseband signal $x[n]$ to produce the κ-times oversampled signal $x^{(\kappa)}[n]$. This is accomplished by first inserting $\kappa - 1$ zero-valued samples between each pair



(a) Conventional design



(b) Low-end design

Figure 9: A general transmitter design

of consecutive original samples to produce

$$x_{\text{ZS}}^{(\kappa)}[n] \triangleq \begin{cases} x[n/\kappa], & n \equiv 0 \mod \kappa, \\ 0, & \text{otherwise.} \end{cases} \tag{133}$$

This signal is then pulse-shape filtered with the discrete-time pulse $p'[\ell] \triangleq p'(\ell T/\kappa)$. The resulting upsampled signal is

$$x^{(\kappa)}[n] = \sum_{\ell=-\infty}^{\infty} p'[\ell] x_{\text{ZS}}^{(\kappa)}[n-\ell]. \tag{134}$$

The number κ is referred to as the *oversampling factor*. The discrete-time pulse $p'[\ell]$ should be the equivalent of the continuous-time filter $p'(\tau)$ discussed in Section 2.1 on page 15. In practice, the ideal, continuous-time pulse $p'(\tau)$ might have to be windowed before the sampling to avoid a discrete-time pulse with infinitely many taps.

In case of OFDM signals with N rectangular pulse shapes (subcarriers) as in (26) on page 22, the oversampling can be more efficiently implemented with a fast Fourier transform:

$$x^{(\kappa)}[n] \triangleq x\left(n\frac{T}{\kappa N}\right) \tag{135}$$

$$= \sum_{\ell=-\infty}^{\infty} \operatorname{rect}\left(\frac{n}{\kappa N} - \ell\right) x_{\ell}^{(\kappa)}[n], \tag{136}$$

where $\operatorname{rect}(\tau) = 1$ when $-0.5 \leq \tau < 0.5$ and $\operatorname{rect}(\tau) = 0$ otherwise, and the signal

$$x_{\ell}^{(\kappa)}[n] \triangleq \sum_{\nu=0}^{\kappa N-1} x_{\nu}[n] e^{j2\pi\nu n/(N\kappa)} \tag{137}$$

is the κN-point inverse Fourier transform of the symbol-sampled signals $\{x_{\nu}[n]\}$ with respect to the subcarrier index ν, which can be efficiently computed by a fast Fourier transform algorithm. The symbol-sampled signals are zero-padded, i.e. it is assumed that $x_{\nu}[n] = 0$ for all subcarrier indices $\nu \geq N$.

4.1.2 Distortion Compensation

The upsampled baseband signal is then preprocessed to compensate for the imperfections of the analog components, such as the oscillators, mixers and

power amplifiers, that come later on in the transmitter chain. Two compensation techniques commonly used to handle imperfections of the power amplifiers are: *crest-factor reduction* that lowers the PAR of the signal to enable amplification with higher power efficiency and less distortion, and *digital predistortion* (DPD) that tries to linearize the combined transfer characteristics of the system from the DPD to after the nonlinear power amplifier to avoid nonlinear signal distortion.

Crest-factor reduction can be done, e.g., by repeated clipping-and-filtering, tone reservation and peak cancellation, see [86, 87] for overviews of crest-factor reduction techniques. Some of the techniques are specific for OFDM, such as tone reservation, others can be used independently of the transmission method, such as clipping-and-filtering. Sometimes several techniques are used together to reduce the PAR further.

Different crest-factor reduction techniques, however, have different drawbacks. Some cause distortion to the in-band signal, i.e. distortion that has frequency components inside the frequency band of the original signal, e.g., clipping-and-filtering. Some cause delays to the signal, e.g., clipping-and-filtering and peak cancellation. And some have to use a fraction of the degrees-of-freedom of the signal that otherwise could have been used for data transmission, which reduces the effective data rate of the system, e.g., tone reservation and peak-cancellation that use reserved frequencies.

To accomplish this, the DPD requires knowledge of the transfer characteristics of the power amplifier, which varies over time. This is the purpose of the feedback loop in Figure 9(a). Through it, the DPD regularly observes the amplifier output in order to compute the transfer characteristics. Since the amplifier output is an analog passband signal, it has to be downconverted and digitized in an ADC before being fed back to the DPD, which adds to the hardware complexity.

To build a massive MIMO base station, one would need hundreds of transmitters. Functionality, such as the DPD, that consumes a fixed amount of power independently of the output power of the transmitter does not allow for scalability in the number of base station antennas and becomes impractical in a massive MIMO base station. Other components, such as crest-factor reduction, similarly have a complexity that is affordable in systems with a few transmitters. In massive MIMO however, their complexity becomes high and might not be feasible. For this reason, the low-end transmitter in Figure 9(b) that is studied in this thesis has no distortion compensation nor a feedback loop. Crest-factor reduction can still be done on the symbol-sampled baseband signal by applying a low-PAR precoding method, as the ones studied in Papers A and B.

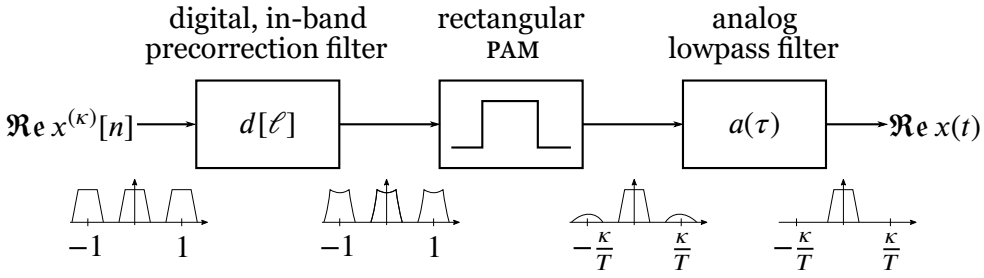


Figure 10: Digital-to-analog conversion in the I-branch (the real part of the complex baseband signal). To illustrate the process, the spectra of the different signals are shown below each step in the conversion. The pulse-amplitude modulation is assumed to be perfect, i.e. no quantization of the amplitudes is done.

4.1.3 Digital-to-Analog Conversion

To obtain a passband signal as in (9) on page 17, the complex baseband signal is split up in its real and imaginary parts—the I and Q components—which are converted into an analog signal by a digital-to-analog converter (DAC), as is shown in Figure 10.

The simplest DAC consists of a circuit that outputs an analog signal that is constant between each transition, where the transition period is equal to the sampling period T/κ. In each transition, the amplitude can only change to one of a finite number of amplitudes, a number which is given by the resolution of the DAC—a b bit DAC has 2^b output amplitudes. Hence, the signal undergoes a quantization, which causes nonlinear distortion, both in the form of out-of-band radiation and in-band distortion [88]. On the one hand, increasing the resolution of the DAC can mitigate the distortion. On the other hand, a higher resolution makes the DAC more complex and power hungry. For this reason, some work has investigated the use of low-resolution DACs in massive MIMO base stations [89, 90].

The simple DAC that outputs constant amplitudes can be modeled as a pulse-amplitude modulator that employs a rectangular pulse shape:

$$x(t) = \sum_{n=-\infty}^{\infty} Q(x^{(\kappa)}[n]) \operatorname{rect}\left(\frac{t\kappa}{T} - n\right), \tag{138}$$

where $Q(x) \in \mathcal{Q}$ is a rounding function that outputs values from the set $\mathcal{Q}$ of possible amplitudes and the pulse shape $\operatorname{rect}(t) = 1$ for $-0.5 \leq t < 0.5$ and $\operatorname{rect}(t) = 0$ otherwise. Such a simple DAC, however, would distort the in-band

signal also without the quantization (when $Q(x) = x$), because the spectrum of a rectangular pulse shape is a cardinal sine, which (i) is not constant over the bandwidth of the discrete-time signal, (ii) is not bandlimited and would leave traces of the cyclic repetitions of the spectrum of the discrete-time signal in the analog signal, as can be seen in Figure 10.

The distortion of the in-band signal can be handled by a digital precorrection filter, which pre-compensates for the non-constant amplitude of the spectrum of the pulse-amplitude modulation. The digital precorrection filter is marked by its impulse response $d[\ell]$ in Figure 10 and has a frequency response that is the inverse of the cardinal-sine spectrum of the pulse-amplitude modulator inside the frequency band of the original signal. In practice, the pre-correction filter is combined with the filter $p'[\ell]$ used in the upsampling (134) on page 57, and has little influence on the overall system complexity.

The cyclic repetitions around the frequencies $\pm\kappa/T$ are then removed after the pulse-amplitude modulation by an analog lowpass filter. If the oversampling factor κ is large enough, these high-frequency components are far away in frequency from the desired signal. This simplifies the implementation of the analog filter because the transition between the stopband and passband can be more gradual. Increased oversampling, however, increases the power consumption of the DAC.

4.1.4 Upconversion

The two real, analog signals from the baseband DACs are then upconverted to passband by mixing them with a carrier wave as in (9) on page 17. The carrier wave for the real part is obtained directly from an oscillator that operates at the carrier frequency f_c. The carrier wave for the imaginary part is a copy of the oscillator output that is delayed to obtain a 90° phase shift from the first carrier wave. The delay is denoted by H in Figure 9, where H denotes the *Hilbert transform*. The passband signal is obtained by summing up the two signals. This process of producing a real passband signal from a complex baseband signal is referred to as *IQ modulation* from the terms *in-phase* and *quadrature phase* for the phases of the two carrier waves.

4.1.5 Amplification

After upconversion, the passband signal is amplified in a power amplifier to transmit power. Often in base stations with transmit signals with high PAR, the amplification is done by multi-transistor amplifiers such as the

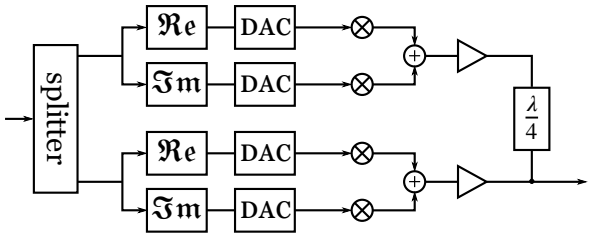


Figure 11: A digital Doherty amplifier. Compare with the analog design in Figure 9(a).

Doherty amplifier [91], where the signal is split up and amplified by separate amplifiers and then recombined in order to increase power efficiency. For a detailed description of the Doherty amplifier, see [92, Ch. 9].

The splitting and the combining of analog signals are difficult to implement perfectly and usually result in distortion. This is traditionally handled by the DPD. Another method to avoid some of the distortion is depicted in Figure 11, where the splitting is done in the digital domain. To do that a different transmitter architecture has to be considered: The splitter would have to be moved to the front of the upconversion stage and the two output signals from the splitter would need one upconversion stage each. This is called a *digital Doherty amplifier*.

In massive MIMO, where each antenna needs its own amplifier, simple amplifier designs, such as class A, B or AB amplifiers, are necessary to keep the total hardware complexity down. If the splitting and combining of the Doherty amplifier can be done in the analog domain without a DPD (or, possibly, with a static DPD without feedback), simple analog single-input Doherty amplifiers could also be feasible in massive MIMO base stations. The drawback of these simple amplifier designs is their nonlinear transfer characteristics, which inherently causes undesired distortion.

The memoryless nonlinearity of an amplifier can be illustrated by its transfer characteristics, an example of which is shown in Figure 12. The transfer characteristics show what amplitude the output signal will have as a function of the amplitude of the input signal. A linear amplifier would have a transfer characteristic that is linear with slope one in the logarithmic scale of Figure 12. The actual transfer characteristics are only linear for small amplitudes, however. For large amplitudes, the transfer characteristics *saturate*, i.e. the output amplitude asymptotically approaches a constant level as the input amplitude grows large. Saturation is said to occur close to the *1-dB compression point*, which is where the curve of the actual transfer

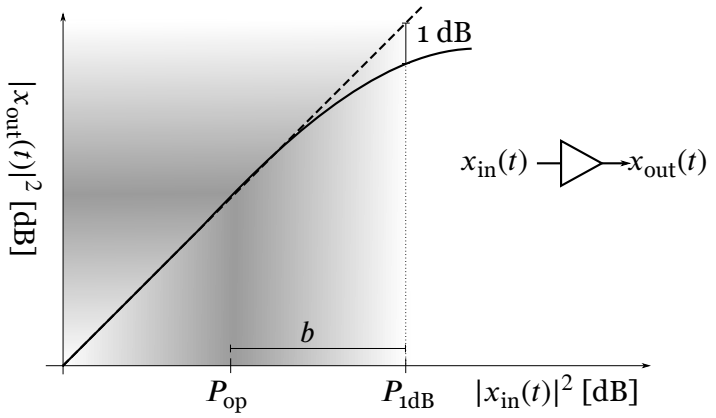


Figure 12: The transfer characteristics of a general power amplifier (solid) compared to a perfectly linear characteristics (dashed).

characteristics is 1 dB below the ideal linear one; it is denoted by P_{1dB} in Figure 12.

When an input signal varies inside the nonlinear region, the region of large amplitudes that cannot be said to be linearly amplified, two types of distortion arise: *in-band distortion* that disturbs the own system and *out-of-band radiation* that disturbs systems that use adjacent frequency bands. The in-band distortion will cause distortion to the received demodulated signal; the effect can be seen in Figure 5 in Paper A on page 125. The out-of-band radiation will cause the spectrum of the nonlinearly amplified signal to grow and some power to leak outside the desired frequency band, as can be seen in Figure 1 in Paper F on page 248.

To avoid distortion in a system without a DPD, the input signal is usually *backed off* prior to amplification, i.e. its power P_{op} is chosen such that the amplitude of the input signal most of the time varies within the linear region of the transfer characteristics. The color of the shade in Figure 12 illustrates the amount of time the signal assumes a certain amplitude. The light shade in the nonlinear region indicates that it seldom assumes values that are nonlinearly amplified. Through back-off, the shaded area can be moved left and right and the amount of distortion can thus be controlled. The drawback of back-off is the reduced power efficiency, which is illustrated in Figure 13 for a class B and a Doherty amplifier. The back-off is defined as the fraction between the 1-dB compression point and the power of the input signal, $b \triangleq P_{\text{1dB}}/P_{\text{op}}$ in Figure 12.

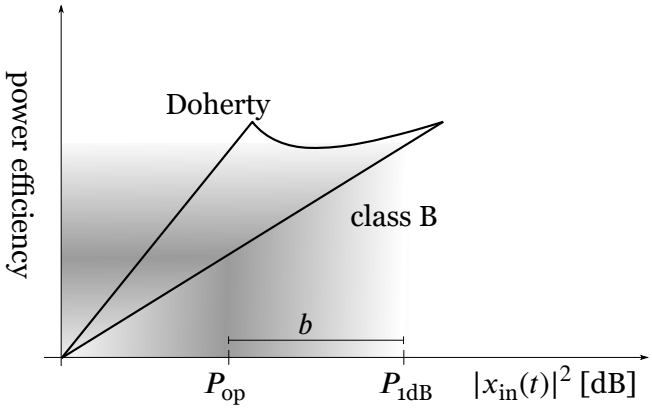


Figure 13: Comparison of the power efficiencies of a class B and a Doherty amplifier as a function of the input power.

As an initial estimate of how much back-off that is needed, the *peak-to-average ratio* (PAR) is often used. Because many signals are stochastic in their nature and can assume very high amplitudes with small probabilities, the PAR is many times defined as the ϵ-percentile of the complimentary cumulative distribution of the amplitude of the input signal, i.e.

$$\text{PAR} \triangleq \sup\left\{ A \,:\, \Pr\left(\frac{|x_{\text{in}}(t)|}{\sqrt{\mathsf{E}[|x_{\text{in}}(t)|^2]}} > A \right) > \epsilon \right\}. \tag{139}$$

The percentile can be chosen differently depending on the sensitivity of the system to nonlinear amplification. A common choice is $\epsilon = 10^{-4}$. The amplitude distribution for some massive MIMO signals is shown in Figure 1 in Paper A on page 102. There it can be seen, on the one hand, that the PAR is around 10 dB with linear precoding, independently of modulation order and whether single-carrier or OFDM transmission is used. With the low-PAR precoding method, discrete-time constant-envelope precoding, on the other hand, the PAR is just over 4 dB, which is significantly lower.

If a nonlinearity is fed with a passband signal, its output will consist of many frequency components, each centered around multiples of the carrier frequency, as is shown in Figure 14. The components around other frequencies $f \neq f_{\text{c}}$ than the carrier frequency are referred to as *higher-order harmonics*. Since the higher-order harmonics are far away from the frequency band in which the transmit signal lies, however, they can be filtered out relatively easy by the low-order bandpass filter preceding the antenna in Figure 9. In fact, the antenna itself can act as a bandpass filter and filter out

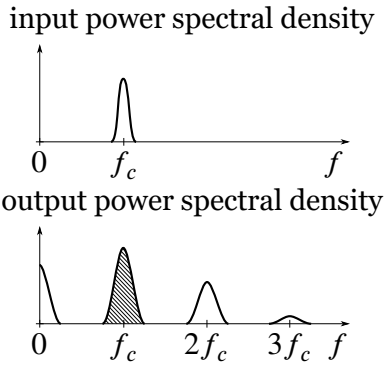


Figure 14: Typical power spectral density of the input and output of a memoryless amplifier. The figure is a sketch of the illustration [93, Fig. 5.3]. The shaded part of the output power spectral density is the baseband part of the output signal. A baseband model of the amplifier only describes this part of the output signal.

some of the harmonics. For simplicity, the remaining component around the carrier frequency will be referred to as the complex baseband equivalent of the output signal. It is also referred to as the *fundamental*.

The transfer characteristics of an amplifier can be modeled in many different ways [94]. In most models, the complex baseband equivalent of the output signal $x_{\text{out}}(t)$ is given in terms of the baseband equivalent of the input signal $x_{\text{in}}(t)$. As long as the higher-order harmonics are filtered out, a baseband model that neglects the effects of the higher-order harmonics still accurately describes the system.

A common way to characterize nonlinear amplifiers is through their *AM-AM conversion*, which describes how the amplitude of the output signal is affected by the amplitude of the input signal, and their *AM-PM conversion*, which describes how the phase distortion is affected by the input amplitude. In this framework, the amplified signal is given by:

$$x_{\text{out}}(t) = a\left(|x_{\text{in}}(t)|\right) e^{j2\pi(\arg x_{\text{in}}(t) + \Phi(|x_{\text{in}}(t)|))}, \tag{140}$$

where the function $a(x) : \mathbb{R} \to \mathbb{R}$ describes the AM-AM conversion and $\Phi(x) : \mathbb{R} \to \mathbb{R}$ the AM-PM conversion. There are many ways to model $a(x)$ and $\Phi(x)$, see [95] for a survey. In Paper A, the *Rapp model* is used, which is

given by

$$a(x) = A_{\max} \frac{x/x_{\max}}{\left(1 + (x/x_{\max})^{2d}\right)^{\frac{1}{2d}}} \tag{141}$$

$$\Phi(x) = 0. \tag{142}$$

The parameter value $d = 2$ was used to model a typical moderate-cost solid-state amplifier [96]. The factors $A_{\max}$ and $x_{\max}$ are parameters that determine the amount of amplification and the point where saturation occurs.

In Paper E, it is shown how the *polynomial model* is derived from the general Volterra series description of a nonlinearity. The Volterra series [97] is an expansion of the output passband signal in terms of the input signal:

$$x_{\text{out}}^{\text{pb}}(t) = b_0 + \sum_{\varpi=1}^{\Pi} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} b'_{\varpi}(\tau_1, \ldots, \tau_{\varpi}) \prod_{\varpi'=1}^{\varpi} x_{\text{in}}^{\text{pb}}(t - \tau_{\varpi'}) \, \mathrm{d}\tau_{\varpi'}, \tag{143}$$

where Π is the order of the model and $b'_{\varpi}(\tau_1, \ldots, \tau_{\varpi})$ is the kernel of the ϖ-dimensional convolution. It has, however, been observed that often an amplifier can be accurately described using only the diagonal kernels, i.e. by making the ϖ-dimensional convolutions one dimensional:

$$b'_{\varpi}(\tau_1, \ldots, \tau_{\varpi}) = A b_{\varpi}^{\text{pb}}(\tau_1) \prod_{\varpi'=2}^{\varpi} \delta(\tau_{\varpi'} - \tau_1), \tag{144}$$

where the impulse response $b_{\varpi}^{\text{pb}}(\tau)$ defines the one-dimensional convolutions and A is a scaling factor.

As is shown in Paper E, the even-order impulse responses $\{b_{\varpi}^{\text{pb}}(\tau), \varpi \text{ even}\}$ only affect the spectrum of the output signal around even multiples of the carrier frequency, i.e. they describe the higher-order harmonics. When the distortion around these frequencies are filtered away by the bandpass filter at the antenna, only the odd-order convolutions need to be considered. The *memory polynomial model* is obtained by using (144) with

$$A = \left(\frac{\varpi - 1}{2}\right)! \left(\frac{\varpi + 1}{2}\right)! \,/\, \varpi! \tag{145}$$

and only the odd-order kernels to rewrite the Volterra series in (143) to complex baseband as:

$$x_{\text{out}}(t) = \sum_{\varpi=1,3,5,\ldots,\Pi} \int_{-\infty}^{\infty} b_{\varpi}(t - \tau) x_{\text{in}}(\tau) \, |x_{\text{in}}(\tau)|^{\frac{\varpi-1}{2}} \, \mathrm{d}\tau, \tag{146}$$

where the baseband kernel $b_{\varpi}(\tau)$ is related to the passband kernel $b_{\varpi}^{\text{pb}}(\tau)$ as in (9) on page 9. The kernels $b_{\varpi}(\tau)$ and order Π are chosen to fit measured data.

The polynomial model is a simplification of the memory polynomial model, where the kernels are assumed to be memoryless:

$$b_{\varpi}(\tau) = b_{\varpi}\delta(\tau), \tag{147}$$

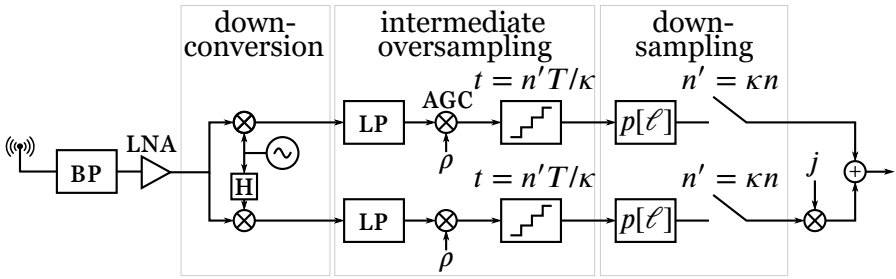
for some constants $b_{\varpi} \in \mathbb{C}$. This simplification is accurate in systems, where the kernels $b_{\varpi}(\tau)$ are constant over the frequency band of interest. Since the constants can be complex valued, the polynomial model takes short memory (shorter than one over the studied bandwidth) into account. Nonlinearities with such short memory are said to be *quasi-nonlinear*. A generalization of the memory polynomial is the *generalized memory polynomial model*, where the kernels of the Volterra series are zero outside a two-dimensional slice of the memory space, instead of a one-dimensional slice as in the case of the memory polynomial model.

These models are summarized in Table 4 in Paper E on page 233. Another explanation of the polynomial model and other specializations of the Volterra series can be found in [98]. It is shown, in Paper E, how the polynomial model can be rewritten in terms of the Itô-Hermite polynomials to analytically obtain expressions for the frequency-dependent covariance matrix of the amplified transmit signals in an antenna array. This technique is used in Papers F and G to analyze the frequency dependent radiation pattern of the distortion from the power amplifiers, as well as in Paper H, which analyzes the error caused in the decoding by nonlinear low-noise amplifiers.

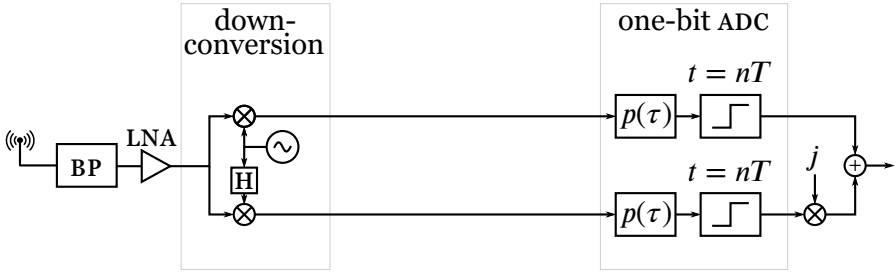
4.2 Receiver Design

Just like the transmitter, one receiver is required per antenna. It is therefore desirable to also simplify the receivers in the massive MIMO base station. There are many ways to simplify the conventional receiver that is illustrated in Figure 15(a). The design with one-bit ADCs that is studied in Paper C is shown in Figure 15(b). There the intermediate, oversampled stage has been removed, which requires the receive filter to be implemented in the analog domain, and the analog-to-digital conversion is done by a simple comparator (a so-called one-bit ADC).

Other possible simplifications include *direct downconversion*, where the mixers are removed and the aliasing that occurs in the sampling is used to downconvert the signal. Other low-resolution ADC designs, such as the



(a) Conventional homodyne design



(b) Studied one-bit design

Figure 15: The general receiver design

ones studied in Paper D, and the use of simple low-noise amplifiers (LNAs), which is studied in Paper H, are some other possible ways to reduce hardware complexity.

4.2.1 Low-Noise Amplification

Directly after the antenna, the weak receive signal is bandpass filtered and amplified by a low-noise amplifier (LNA). The LNA is a nonlinear component whose transfer characteristics are of the same type as the power amplifier of the transmitter, which was shown in Figure 12.

The power amplifier of the transmitter and the LNA of the receiver might seem to have the same function: to increase the power of a given signal. The design of the LNA is more challenging, however, since an LNA has to amplify an unknown signal with much noise, whereas a power amplifier deals with a known signal with little noise. To avoid degrading the SNR, the LNA should have a small noise figure. It should also be linear and have a large enough dynamic range to handle the presence of signals with much higher power than

the signal of interest, so called *blockers*, without creating in-band distortion.

To obtain a small noise figure, the LNA cannot use the same power-efficient technologies as in the power amplifiers. To handle blockers, the LNA also has to be operated with a high back-off, which further reduces the power efficiency. As a consequence, the power efficiency of the LNA is worse than the power efficiency of the power amplifier. This is the reason the power consumption of a single LNA can be around 1 W, despite only outputting a much weaker signal than the power amplifier of the transmitter. In a massive MIMO base station, where each antenna needs an LNA, a significant part of the total power consumption can stem from the LNAs. It is therefore desirable to increase their power efficiency by relaxing the requirements on noise figure and linearity as much as possible.

Because thermal noise combines noncoherently in massive MIMO and can be handled by increasing the number of antennas, it is possible that more power efficient LNAs with higher noise figures can be used in massive MIMO. The question studied in Paper H is whether more power efficient nonlinear operation close to saturation and the resulting distortion also can be handled by the spatial processing in massive MIMO. It is shown that the major part of the distortion from the blocker combines noncoherently. There are, however, smaller terms that combine coherently, which spatial processing with a large number of antennas cannot suppress.

4.2.2 Downconversion

The receiver design shown in Figure 15(a) is a so called *homodyne* receiver, where the analog passband signal is downconverted directly to baseband. The alternative would be a *heterodyne* receiver with an intermediate stage, where the analog signal is downconverted to an intermediate low frequency and the baseband signal is recovered in the digital domain through a second downconversion.

To obtain the real and imaginary (I and Q) parts of the baseband signal, the passband signal is split into two copies and downconverted by mixing each copy with their carrier waves, one being a 90° phase shifted copy of the other. Since this is done in the analog domain in a homodyne receiver, the phase shift might be slightly different from 90° and the two carrier waves might have slightly different amplitudes. Such imperfections can cause a nonlinear distortion to the complex baseband signal, so called *IQ imbalance*. In contrast, IQ imbalance can be avoided in the heterodyne receiver at the cost of more complex hardware and higher sampling rate. Since a homodyne receiver is the less complex of the two in theory, the homodyne structure is

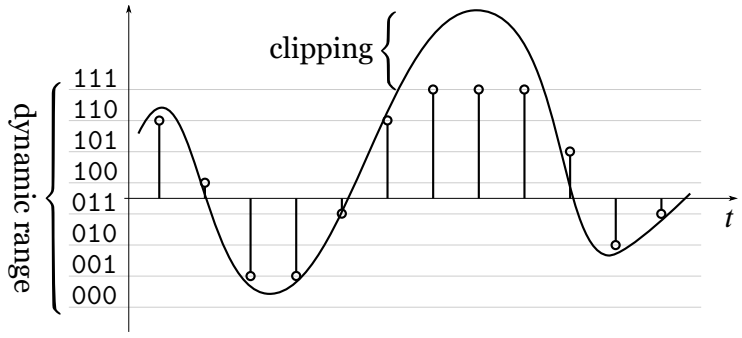


Figure 16: Quantization in a 3-bit ADC to a uniform grid of $2^3 = 8$ quantization levels.

assumed in the Papers C and D about ADCs.

The mixing can also give rise to phase noise, because the frequency of the oscillator tends to drift over time. How phase noise affects massive MIMO systems is investigated in, e.g., [99].

4.2.3 Analog-to-Digital Conversion

To avoid aliasing in the sampling, the two downconverted continuous-time signals are first lowpass filtered. The signals are then sampled by an ADC (analog-to-digital converter) with an oversampling factor κ. The ADC rounds the sample values off to a grid of predetermined quantization levels as shown in Figure 16. If $y(t)$ is the real baseband signal after lowpass filtering, the discrete-time output signal of the ADC is given by:

$$y^{(\kappa)}[n] = Q\left(y\left(\frac{nT}{\kappa}\right)\right), \tag{148}$$

where the function $Q(y) \in \mathcal{Q}$ describes the quantization to the set $\mathcal{Q}$ of quantization levels. The *resolution* of the ADC is often given in the number of bits required to represent all the quantization levels. An ADC with $|\mathcal{Q}| = 8$ levels thus has $\log_2 |\mathcal{Q}| = 3$-bit resolution and is referred to as a three-bit ADC. The complex baseband signal, that is fed to the baseband processing unit that performs decoding and symbol estimation, is then obtained by interpreting the two discrete-time signals as the real and imaginary parts of a complex signal.

Ideally, the thresholds of the function $Q(\cdot)$ are in the middle of each pair of quantization level in $\mathcal{Q}$, so that values are rounded off to the nearest level. In practice, however, the thresholds can have drifted away from their ideal

value, which creates additional distortion [100]. If this distortion is severe, it has to be handled by a digital post-correction, one method for which can be found in [101].

The *dynamic range* of the ADC is the range of amplitudes between the lowest quantization level and the highest, i.e. the interval $[\max \mathcal{Q}, \min \mathcal{Q}]$. If the amplitude of the input signal assumes values outside the dynamic range, the signal is *clipped*, i.e. its samples are rounded off to the highest or the lowest quantization level depending on which is closer. If, on the contrary, the amplitudes of the input signal are much smaller than the dynamic range, not all quantization levels are used and some of the resolution of the ADC is wasted.

To avoid that clipping occurs too often, and to efficiently use the whole resolution, the power of the input signal is controlled by an *automatic gain control* (AGC), which adjusts the power of the signal such that its amplitude most of the time stays just within the dynamic range. Since the power level of the received signal changes over time, the AGC has to be built as a dynamic control loop with both variable attenuators and amplifiers. Because the AGC is a relatively complex technology, a typical base station employs high-resolution ADCs (10–14 bits) to accommodate for a low-end AGC with imperfect power control.

The high ADC resolution is also used to handle interference from blockers, undesired high-power signals, which otherwise can cause the desired signal to drown and disappear in the quantization. Such signals can come from a malicious jammer [102], or from a transmission in an adjacent frequency band that is not filtered out by the lowpass filter. The latter case is described in Paper H in the context of low-noise amplifiers.

It is critical that the receive filter $p(\tau)$ in (13) is implemented accurately to maximize the signal-to-noise ratio in the final symbol-sampled baseband signal and to avoid that interference from blockers leaks into the demodulated signal. Since it is easier to implement the receive filter as a digital filter, the received signal is usually first sampled by an ADC into an intermediate stage with κ-times oversampling. After that, it is filtered by the digital receive filter $p[\ell] \triangleq p(\ell T/\kappa)$ and downsampled to the baudrate.

In massive MIMO receivers, much of the complexity and power consumption stem from the ADCs, whose total complexity scales linearly in the number of antennas. For example, the power consumption and hardware complexity grows exponentially with the number of resolution bits in *flash ADCs*, which is a type of high-speed ADCs that are used in ultra-wideband systems [103].

Another common ADC design for base stations is the *successive approximation* ADC, which has lower hardware complexity than flash ADCs and

thus can be used to obtain a higher resolution. Successive approximation ADCs, however, are slow and are therefore difficult to use in wideband systems. Especially, at sampling rates above 100 MHz, which are necessary if $\kappa/T \geq 100\,\text{MHz}$, the power consumption of the ADCs becomes critical, independently of their hardware architecture [104].

Both power consumption and hardware complexity would be reduced by using ADCs with lower resolution. In Paper C, ADCs with the lowest possible resolution (1 bit) and lowest possible sampling rate ($\kappa = 1$) are therefore investigated. For the same reason, ADCs with other low-resolutions are evaluated in Paper D.

To operate the ADC without oversampling, $\kappa = 1$, i.e. at the baudrate, the analog-to-digital conversion has to be done together with the downsampling without the intermediate stage, as in Figure 15(b). Such a design requires the receive filter to be implemented in the analog domain. Because the receive filter, in general, is a steep filter (often a root-raised cosine filter), it would have to be implemented as a high-order active filter, which might be challenging in practice. The architecture without an intermediate oversampled step is nevertheless considered in order to derive tractable expressions for the performance.

To motivate the study of the symbol-sampled system, it is argued that introducing an oversampled stage would increase the performance—a proposition that is supported by the results about low-resolution DACs in [88]. The achievable rates derived in Paper C and D should thus be seen as lower bounds on the performance of massive MIMO in a more practical oversampled system with low-resolution ADCs, and give insight into the feasibility of coarse quantization in massive MIMO.

It is, for example, observed that it will be difficult for a weak user, i.e. a user whose received signal is much weaker than the signals from the other users, to obtain good channel estimates, which can degrade its performance compared to an unquantized system significantly. Therefore, low-resolution ADCs, and one-bit ADCs in particular, have to be accompanied by user scheduling and proper power allocation to avoid that users are received with widely different powers.

The use of one-bit ADCs, apart from low power consumption and low hardware complexity, have at least two more advantages. If the one-bit ADCs use a zero threshold, they are insensitive to the absolute magnitude of the amplitudes, only the sign of the signal matters. Therefore the ADC does not require any AGC, which reduces the hardware complexity further. As discussed in Paper C, if the analog receive filter is an active filter, it hypothetically might still need a mechanism to control its input voltage to avoid

producing nonlinear distortion. Since only the sign of the signal is kept in the quantization in the end, nonlinear distortion might not necessarily have to be avoided. In any case, such a control mechanism ought to be simpler to implement than the complex AGC required for conventional ADCs.

The second advantage is that the resulting complex baseband signals will be represented with only two bits. It is possible that the large amount of arithmetic operations that have to be performed on the signals in the baseband processing unit can be simplified by this small number of bits. The low-resolution is especially advantageous in a wideband massive MIMO system, where the amount of baseband data that has to be sent between the receiver and the baseband processing unit is huge—it scales linearly with the numbers of antennas, the baudrate and the number of resolution bits. For example, with 100 antennas that each has two conventional 10-bit ADCs, a baudrate of 50 MHz and an oversampling factor of $\kappa = 5$, there will be 500 Gbit/s of data to transfer from all antennas. With an energy consumption of 28 pJ/bit [105], the data transfer will consume 14 W of power, which might be a significant part of the total hardware power consumption. Results in [106] indicate that using one-bit ADCs is optimal when the data transfer rate between the remote radio heads and the baseband unit is fixed.

Chapter 5

Future Work

From a theoretical point of view, massive MIMO is a fairly mature technology and a good understanding for what it can do and its limitations has been established through research in recent years. The remaining challenges are foremost implementational: how to parallelize the signal processing, how to design the precoding, decoding and channel estimation algorithms in power efficient ASICs (application-specific integrated circuits) and how to handle the data shuffling at high baudrates.

One way to make the data shuffling easier is to represent the digital signals with few bits. As the results on low-resolution ADCs in, e.g., Papers C and D indicate, some signals in the base station can be coarsely quantized, i.e. represented by few bits, without hampering the functionality of massive MIMO. If also the number of bits of the digital filters and other intermediate digital signals can be reduced to a few bits, the data rate over the internal buses and cables could be made smaller.

The possibility to design massive MIMO base stations with continuous-time transmit signals with constant envelope, 0 dB PAR, without limiting the transmitted symbols to have constant envelope was demonstrated in Paper B. Such transmit signals would enable radically simpler analog designs of the base station transmitters and make then cheap and power efficient. The continued research on this precoder should focus on the possibility to reduce its computational complexity and improve upon its performance in terms of increasing the data rate of the transmission and reducing the excess bandwidth of the transmit signal.

As is demonstrated in this thesis, not all types of hardware distortion combine noncoherently and vanish when the number of base stations antennas is increased. Future research should focus on finding predistortion and

other compensation techniques that can mitigate the coherent distortion, in order to further reduce the hardware complexity of the massive MIMO base station and to increase the effective SINR of the transmission.

A promising wireless communication technology that builds on massive MIMO is what has been called “cell-free massive MIMO”, where a massive amount of simple access points together transmits information coherently to all users in the system. The development and evolution of that concept would make for interesting future research in wireless communications.

Chapter 6

Swedish Terminology

According to the Swedish Research Council [107, Ch. 5.1], ”Researchers in Sweden [...] have a responsibility to help maintain and develop an adequate scientific vocabulary in Swedish.“ This chapter is dedicated to fulfill this responsibility.

När forskning inom nya områden bedrivs, stöter man ofta på nya begrepp och fenomen som ännu saknar etablerad nomenklatur. För att underlätta eventuell diskurs inom massiv MIMO, presenterar jag nedan en tabell över de svenska massiv-MIMO-termer som jag använder. Till varje term ges en definition och dess engelska och kinesiska motsvarighet. Notera att många av dessa termer (ännu) inte är etablerade och bör, vid användande, möjligen åtföljas av en förklaring.

svensk term, engelsk term, kinesisk term
AM-AM-karaktäristik, AM-AM conversion, 振幅轉換 beskrivning av hur amplituden hos en förstärkares utsignal påverkas av insignalens amplitud
AM-PM-karaktäristik, AM-PM conversion, 振幅相位轉換 beskrivning av hur fasen hos en förstärkares utsignal påverkas av insignalens amplitud
analog-till-digital-omvandlare, analog-to-digital converter, 模擬數字轉換器 en krets som omvandlar en analog signal till en digital
anpassat filter, matched filter, 匹配濾波器 ett mottagarfilter vars impulssvar är den tidsomvända och komplexkonjugerade versionen av sändarfiltret

antennförstärkning, antenna gain, 天線增益
förhållandet mellan den effekt som tas emot från en riktantenn och den effekt som tas emot från en rundstrålande antenn när den utstrålade effekten i båda fallen är densamma

autokorrelation, autocorrelation, 自相關函數
korrelationen mellan olika sampel i en stokastisk process

automatisk förstärkningsreglering, automatic gain control, 自動增益控制
att adaptivt reglera en signals effekt

avkoda, decode, 解碼
att utföra en avkodning

avkodning, decoding, 解碼
tekniken att, med en antennuppställning, urskilja en signal från en eller flera punkter i rummet baserat på kännedom om utbredningsvägen från punkten

backa av, back-off, 功率回退
att välja en lägre arbetspunkt i en förstärkare för att minska mängden icke-linjära distorsion som den ger upphov till

basbandsignal, baseband signal, 基帶信號
en komplex representation av en passbandsignal, vars frekvensinnehåll är centrerat kring frekvens noll

blanda upp/ner, up-/downconvert, 上/下變頻
att multiplicera en signal med en bärvåg för att flytta den upp eller ner i frekvens

blockfädning, block fading, 分組衰落
en fädningsmodell i vilken kanalen antas vara konstant inom ett block av symboler och förändras oberoende mellan blocken

bärfrekvens, carrier frequency, 載波頻率
frekvensen kring vilken en passbandsignal är centrerad

cykliskt prefix, cyclic prefix, 循環前綴
upprepningen av den sista delen av en skur av symboler, vilken föregår transmissionen av skuren

digital-till-analog-omvandlare, digital-to-analog converter, 數字模擬轉換器
en krets som omvandlar en digital signal till en analog

dohertyförstärkare, Doherty amplifier, 多爾蒂功放
förstärkararkitektur som använder flera förstärkare för att öka verkningsgraden

dopplerskift, Doppler shift, 多普勒頻移
skillnaden i frekvens mellan utsänd och mottagen signal som uppstår på grund av utbredningsvägens elongering

dopplerspridning, Doppler spread, 多普勒擴展
den största skillnaden mellan olika dopplerskift i en kanal med flervägsutbredning

dynamiskt omfång, dynamic range, 動態範圍
det amplitudintervall hos en krets inom vilket ingen klippning sker

effektfördröjningsprofil, power delay profile, 功率延遲分布
en statistisk beskrivning av variansen hos sampel av kanalens impulssvar

encellssystem, single-cell system, 單小區系統
ett kommunikationssystem med en basstation

enkelbärvågsmodulation, single-carrier transmission, 單載波傳輸
transmission, där data sänds på en bärvåg

enkelbärvågsmodulation med frekvensdomänsutjämning, single-carrier transmission with frequency domain equalization, 頻域均衡單載波傳輸
modulationsmetod där symbolernas diskreta fouriertransform sänds i block för att utjämnas i frekvensdomänen

ergodisk datatakt, ergodic rate, 遍歷速率
en datatakt som är uppnåelig om kodning sker över många av kanalens koherensintervall

fjärrfält, far-field, 遠場
området där antagandet om en planär vågfront ger en god approximation av kanalens frekvenssvar i frisikt

fleranvändarförkodning, multiuser precoding, 多用戶預編碼
koordineringen av signalerna som sänds från basstationsantennerna med avsikt att flera användare samtidigt skall ta emot för dem ämnade individuella meddelanden

fleranvändar-MIMO, multiuser MIMO, 多用戶多輸入多輸出系統
ett system med en flerantenns basstation som betjänar flera användare över samma tids-frekvensresurs

flerbärvågsmodulation, multi-carrier transmission, 多載波傳輸
transmission, där data sänds på flera bärvågor

flercellssystem, multicell system, 多小區系統
ett kommunikationssystem med flera basstationer som samtidigt betjänar olika grupper av användare

frekvensdelad duplex, frequency-division duplex, 頻分雙工
tvåvägstransmission där upp- och nerlänkstransmissionen sker samtidigt men i olika frekvensband

frekvensflat, frequency flat, 頻率平頻性
att ha en fouriertransform vars belopp är konstant över de studerade frekvenserna och vars fas är en linjär funktion av frekvens

frekvensselektiv, frequency selective, 頻率選頻性
inte frekvensflat

fri sikt, line-of-sight, 視距
en oblockerad utbredningsväg mellan sändare och mottagare

fördistorsion, predistortion, 預矯正
signalbehandling som ämnar förvränga signalen på ett sådant sätt att hela överföringen, inklusive en senare icke-linjär förvrängning, sammantaget blir linjär

förkoda, precode, 預編碼
att utföra en förkodning

förkodning, precoding, 預編碼
tekniken att styra en signal från en gruppantenn till en eller flera punkter i rummet baserat på kännedom om utbredningsvägen till punkten

grundton, fundamental, 基波
frekvenskomponenten hos en utsignal till en icke-linjäritet som ligger kring bärfrekvensen när insignalen är en passbandsignal

gruppantenn, antenna array, 天線陣列
en samling samarbetande, individuellt styrbara antenner

gruppantennförstärkning, array gain, 陣列增益
förhållandet mellan den effekt som tas emot från en gruppantenn och den effekt som tas emot från en enda antenn när den utstrålade effekten i båda fallen är densamma

guldfiskgränsen, use-and-forget bound, 先用後忘下界
en nedre gräns på kanalkapaciteten, se (111) på sida 46

gynnsam utbredning, favorable propagation, 適宜傳播
en utbredningsmiljö som tillåter hög rumsmultiplexering

heterodyn mottagare, heterodyne receiver, 外差接收機
en radiomottagare som först blandar ner den mottagna signalen med en lokal oscillator till en bestämd mellanfrekvens för vidare behandling

homodyn mottagare, homodyne receiver, 零差接收機
: en radiomottagare som blandar ner den mottagna signalen med lokal oscillator vars frekvens är identisk, eller mycket lik, bärfrekvensen hos den mottagna signalen

inombandsdistorsion, in-band distortion, 帶內失真
: den del av distorsionen som ligger inom det använda frekvensbandet och överlappar med den önskade signalen

interanvändarinterferens, interuser interference, 用戶間干擾
: interferens orsakad av parallell transmission från eller till en annan användare

inversförkodning, zero-forcing precoding, 迫零預編碼
: en förkodare som ämnar minimera interanvändarinterferensen

IQ-obalans, IQ imbalance, ＩＱ兩路不平衡
: distorsion som uppstår när I- och Q-grenarna inte är lika

kanalförstelning, channel hardening, 信道固化
: fenomenet att egenskaperna hos en stokastisk kommunikationskanal med god noggrannhet kan approximeras som deterministiska

kanalkapacitet, channel capacity, 信道容量
: en gräns på de datatakter som kan överföras över en kanal med godtyckligt liten felsannolikhet

kanalkännedom, channel state information, 信道狀態信息
: det som är känt om kanalen

kanalskattningskvalitet, channel estimation quality, 信道估計質量
: ett mått på den skattade kanalens kvalitet

kanalreciprocitet, channel reciprocity, 信道互易
: likheten mellan kanalerna i upp- och nerlänken

kapacitetsregion, capacity region, 容量域
: alla datatakter som är simultant uppnåeliga i ett fleranvändarsystem

klippning, clipping, 限幅
: trunkeringen av en signals alla överskjutande amplituder till ett givet tröskelvärde

koherensbandbredd, coherence bandwidth, 相干帶寬
: det största frekvensintervall, med vilket en frekvensselektiv kanals frekvenssvar kan samplas, som tillåter perfekt rekonstruktion av frekvenssvaret utifrån samplen

koherensinterval, coherence interval, 相干時頻區間
: produkten mellan koherenstiden och koherensbandbredden för en kanal

koherenstid, coherence time, 相干時間
den tid över vilken en tidsvarierande kanal kan predikteras med god noggrannhet

(mottagar-)kombinering, (receive) combining, (接收)合併
samma som *avkodning*

kompressionspunkten vid 1 dB, 1-dB compression point, 一分貝壓縮點
den arbetspunkt hos en förstärkare där utsignalen är 1 dB lägre än den skulle ha varit vid perfekt linjär förstärkning

konstantenveloppförkodning, constant-envelope precoding, 恆定包絡預編碼
en förkodare som producerar sändsignaler, vars envelopper är konstanta

kortsiktiga effektbivillkor, short-term power constraint, 短期功率束縛
effektvillkor som sträcker sig över en koherenstid

kvantiseringsupplösning, quantization resolution, 量化分辨率
ett mått på antalet kvantiseringsnivåer

lobformning, beamforming, 波束賦形
tekniken att rikta en signal från en gruppantenn i en eller flera riktningar enbart baserat på gruppantennens geometri

lågbrusförstärkare, low-noise amplifier, 低噪聲放大器
en förstärkare designad för att ge ett högt signal-till-brus-förhållande

lågtoppvärdesförkodning, low-PAR precoding, 低峯均比預編碼
en förkodare som producerar sändsignaler med lågt relativt toppvärde

långsiktiga effektbivillkor, long-term power constraint, 長期功率束縛
effektvillkor som sträcker sig över många koherenstider

massiv MIMO, massive MIMO, 大規模多天線系統
en kommunikationsteknik, där basstationen utrustas med ett stort antal individuellt styrbara antenner, vilka används för att multiplexa flera användare i samma tids-frekvensresurs

maximieffektförkodning, maximum-ratio precoding, 最大合並比預編碼
en förkodare som maximerar den mottagna nyttoeffekten hos varje användare

mottagare, receiver, 接收機
en elektrisk krets som tar emot och demodulerar en signal

mottagarfilter, receive filter, 接收濾波器
det filter genom vilket den mottagna signalen filtreras före nersampling till symboltakt

multiplexvinst, multiplexing gain, 複用增益
antalet dataströmmar som sänds parallellt över samma tid-frekvensresurs

nerlänk, downlink, 下行
transmissionen från basstationen till användaren

nyquistpuls, Nyquist pulse, 尼奎斯特脈衝
en puls vars fouriertransform uppfyller (17), se sida 19

ortogonal frekvensdelningsmultiplex, orthogonal frequency-division multiplexing, 正交頻分復用
en flerbärvågsmodulation där data sänds parallellt på ortogonala bärvågor som överlappar i tid

passbandsignal, passband signal, 通帶信號
en signal vars frekvensinnehåll är begränsat till ett smalt spektrum som inte innehåller frekvensen noll

pilot, pilot, 導頻
kända signaler som sänds för att skatta kanalen

pilotens överskottsfaktor, pilot excess factor, 導頻多餘因子
den faktor fler pilotsymboler som används i jämförelse med den kortaste ortogonala pilotlängden

pilotkontaminering, pilot contamination, 導頻污染
det fel som uppstår i en pilotbaserad kanalskattning på grund av interferens från andra användare

polynommodell, polynomial model, 多項式模型
en förstärkarmodell

pulsamplitudmodulering, pulse-amplitude modulation, 脈衝振幅調制
en avbildning av en tidsdiskret signal på en tidskontinuerlig med hjälp av en puls, se (11) på sida 18

rayleighfädning, Rayleigh fading, 瑞利衰落
en fädningsmodell där kanalkoefficienterna antas ha ett absolutvärde som är rayleighfördelat

reciprocitetskalibrering, reciprocity calibration, 互易校準
justering av kanalskattningarna för att ta hänsyn till icke-reciproka överföringsfenomen

regulariserad inversförkodning, regularized zero-forcing precoding, 調整迫零預編碼
en förkodare som ämnar balansera minimeringen av interanvändarinterferensen och maximeringen av antennförstärkningen

relativt toppvärde, peak-to-average ratio, crest factor, 峯均比
förhållandet mellan högsta toppen i en signal och signalens medelamplitud

ricefädning, Ricean fading, 萊斯衰落
en fädningsmodell vilken kombinerar frisiktsutbredning och rayleighfädning

rotnyquistpuls, root-Nyquist pulse, 根尼奎斯特脈衝
en jämn puls vars fouriertransform kvadrerad är fouriertransformen av en nyquistpuls

samordnad förkodning, coordinated precoding, 協調式預編碼
förkodning, där symbolerna förkodas centralt baserat på kännedom om kanalen från alla antenner till alla användare

skymd sikt, non-line-of-sight, 非視距
en blockerad utbredningsväg mellan sändare och mottagare

småskalig fädning, small-scale fading, 小尺度衰落
amplitudskalningen och fasvridningen som en kanal ger upphov till, vilken ändrar sig i varje koherensintervall

spektraltäthet, power spectral density, 功率譜密度
en statistisk beskrivning över hur effekten hos en stokastisk process är fördelad i frekvensled

spridningspunkt, scatterer, 散射體
en punkt, ur vilken en infallande signal sprids i många riktningar

storskalig fädning, large-scale fading, 大尺度衰落
den dämpning av signalen som orsakas av kanalen och som kan anses vara konstant över den korta tid varöver transmissionen sker

styrvektor, steering vector, 導向向量
vektorn som beskriver en gruppantenns frekvenssvar i termer av infallsvinkeln

symbol, symbol, 符號
informationsbärande signal

symbolperiod, symbol duration, 信號周期
tiden mellan två symboler

symboltakt, baudrate / symbol rate, 符號率
antal symboler per sekund

systeminformation, system information, 系統信息
informationen som ett cellulärt kommunikationssystem måste delge sina användare innan de kan koppla upp sig

sändarfilter, transmit filter, 傳輸濾波器
pulsen med avseende på vilken pulsamplitudmoduleringen sker i sändaren

sändare, transmitter, 發射機
en elektrisk krets som modulerar och sänder ut en signal

sändtagare, transceiver, 收發機
en elektrisk krets som både sänder ut och tar emot signaler

tidsdelad duplex, time-division duplex, 時分雙工
tvåvägstransmission där upp- och nerlänkstransmissionen sker i samma frekvensband men i olika tidsintervall

tidsspridning, delay spread, 時延擴展
tiden mellan mottagandet av den första och sista kopian av en signal som sänds över en kanal med flervägsutbredning

toppvärdesreducering, crest-factor reduction, 峯均比降低
en metod som minskar en signals relativa toppvärde

total takt, sum-rate, 總速率
summan av datatakterna som varje individuell användare betjänas med

uppländk, uplink, 上行
transmissionen från användaren till basstationen

uppnåelig datatakt, achievable rate, 可達信息率
en datatakt som är mindre än kanalens kapacitet

utombandsstrålning, out-of-band radiation, 帶外泄露
energin hos en utsänd signal, vilken kan fångas upp utanför det allokerade frekvensbandet

utspridd förkodning, distributed precoding, 分佈式預編碼
förkodning, där symbolerna förkodas lokalt vid varje antenn baserat på kännedom om kanalen från den lokala antennen till alla användare

översamplingsfaktor, oversampling factor, 過採樣因子
antalet sampel per symbolperiod i en tidsdiskret signal

överskottsbandbredd, excess bandwidth, 剩餘帶寬
förhållandet mellan den överskjutande bandbredden och symboltakten

överton, harmonic, 諧波
en frekvenskomponent hos utsignalen till en icke-linjäritet som ligger kring en heltalsmultipel, skild från ett, av bärfrekvensen när insignalen är en passbandsignal

Bibliography

[1] 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, Oct. 2010.

[2] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

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

[4] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *The Proceedings of the Asilomar Conference on Signals, Systems and Computers*, Nov. 2012, pp. 295–299.

[5] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, Feb. 2014.

[6] E. Björnson, L. Sanguinetti, and M. Kountouris, "Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 832–847, Apr. 2016.

[7] T. S. Rappaport, R. W. Heath, Jr., R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Pearson Education, 2014.

[8] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[9] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *The Proceedings of the Annual International Conference on Mobile Computing and Networking*. ACM, Aug. 2012, pp. 53–64.

[10] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Öwall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive MIMO," in *The Proceedings of the IEEE Global Communications Conference*, Dec. 2014, pp. 287–293.

[11] P. Harris, S. Zang, A. Nix, M. Beach, S. Armour, and A. Doufexi, "A distributed massive MIMO testbed to assess real-world performance and feasibility," in *The Proceedings of the IEEE Vehicular Technology Conference*, May 2015, pp. 1–2.

[12] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[13] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[14] U. Gustavsson, C. Sanchez Perez, T. Eriksson, F. Athley, G. Durisi, P. N. Landin, K. Hausmair, C. Fager, and L. Svensson, "On the impact of hardware impairments on massive MIMO," in *The Proceedings of the IEEE Global Communications Conference*, Dec. 2014.

[15] S. Mohammed and E. G. Larsson, "Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3992–4005, Sep. 2012.

[16] ——, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

[17] ——, "Constant-envelope multi-user precoding for frequency-selective massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 2, no. 5, pp. 547–550, Oct. 2013.

[18] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 303–313, Feb. 2013.

[19] H. Bao, J. Fang, Z. Chen, H. Li, and S. Li, "An efficient bayesian PAPR reduction method for OFDM-based massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4183–4195, Jun. 2016.

[20] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *ArXiv E-Print*, Apr. 2014, arXiv:1404.7736 [cs.IT].

[21] M. T. Ivrlač and J. A. Nossek, "Challenges in coding for quantized MIMO systems," in *The Proceedings of the IEEE International Symposium on Information Theory*, Jul. 2006, pp. 2114–2118.

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

[23] J. Mo and R. W. Heath, Jr., "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Information Theory and Applications Workshop, Conference Proceedings*, Feb. 2014, pp. 1–5.

[24] ——, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[25] J. Choi, J. Mo, and R. W. Heath, Jr., "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2005–2018, Mar. 2016.

[26] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO with low-resolution ADC," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[27] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Communications Letters*, vol. 20, no. 5, pp. 842–845, May 2016.

[28] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Transactions on Signal Processing*, May 2017.

[29] C. Desset and L. Van der Perre, "Validation of low-accuracy quantization in massive MIMO and constellation EVM analysis," in *The Proceedings of the European Conference on Networks and Communication*, Jun. 2015, pp. 21–25.

[30] D. Verenzuela, E. Björnson, and M. Matthaiou, "Hardware design and optimal ADC resolution for uplink massive MIMO systems," in *The Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*. IEEE, Jul. 2016, pp. 1–5.

[31] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Transactions on Wireless Communications*, 2017.

[32] M. Sarajlić, L. Liu, and O. Edfors, "When are low resolution ADCs energy efficient in massive MIMO?" *IEEE Access*, vol. 5, pp. 14 837–14 853, Jul. 2017.

[33] W. Sandrin, "Spatial distribution of intermodulation products in active phased array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 21, no. 6, pp. 864–868, Nov. 1973.

[34] C. Hemmi, "Pattern characteristics of harmonic and intermodulation products in broadband active transmit arrays," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 6, pp. 858–865, Jun. 2002.

[35] Y. Zou, O. Raeesi, L. Antilla, A. Hakkarainen, J. Vieira, F. Tufvesson, Q. Cui, and M. Valkama, "Impact of power amplifier nonlinearities in multi-user massive MIMO downlink," in *IEEE Globecom Workshops*, Dec. 2015, pp. 1–7.

[36] S. Blandino, C. Desset, A. Bourdoux, L. V. der Perre, and S. Pollin, "Analysis of out-of-band interference from saturated power amplifiers in massive MIMO," in *The Proceedings of the European Conference on Networks and Communication*, Jun. 2017, pp. 1–6.

[37] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "On out-of-band emissions of quantized precoding in massive MU-MIMO-OFDM," *ArXiv E-Print*, Dec. 2017, arXiv:1712.01249 [cs.IT].

[38] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge University Press, 2009.

[39] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 2002.

[40] J. G. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill, New York, 2008.

[41] S. Haykin, *Communication Systems*, 4th ed. Wiley Publishing, 2001.

[42] G. L. Stüber, *Principles of Mobile Communication*, 4th ed. Springer, 2017.

[43] R. Nissel, S. Schwarz, and M. Rupp, "Filter bank multicarrier modulation schemes for future mobile communications," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1768–1782, Aug. 2017.

[44] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *The Proceedings of the European Signal Processing Conference*, Sep. 2014, pp. 76–80.

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

[46] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 437–447, Mar. 2000.

[47] O. Edfors and F. Tufvesson, "MaMi channel characteristics: Measurement results," https://mammoet-project.eu/publications-deliverables, Massive MIMO for Efficient Transmission (MAMMOET), Project Deliverable D1.2, Jun. 2015, online: accessed 2016-10-26.

[48] M. Peter, K. Sakaguchi, S. Jaeckel, S. Wu, M. Nekovee, J. Medbo, K. Haneda, S. L. H. Nguyen, R. Naderpour, J. Vehmas, F. Mani,

A. Bamba, R. D'Errico, M. Rybakowski, J.-M. Conrat, A. Goulianos, P. Cain, M. Rumney, M. Dieudonne, H. Wang, and M. Kottkamp, "Measurement campaigns and initial channel models for preferred suitable frequency ranges," https://bscw.5g-mmmagic.eu/pub/bscw.cgi/93650, Millimetre-Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications (mmMagic), Project Deliverable H2020-ICT-671650-mmMAGIC/D2.1, Mar. 2016, online: accesed 2016-10-26.

[49] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.

[50] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation*, 3GPP Std., Rev. V14.4.0, 2017.

[51] R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. F. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1815–1831, Apr. 2014.

[52] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu, and F. Tufvesson, "Reciprocity calibration for massive MIMO: Proposal, modeling, and validation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3042–3056, May 2017.

[53] K. T. Truong and R. W. Heath, Jr., "Effects of channel aging in massive MIMO systems," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[54] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1428–1442, Feb. 2017.

[55] S. Kashyap, C. Mollén, E. Björnson, and E. G. Larsson, "Performance analysis of (TDD) massive MIMO with kalman channel prediction," in *The Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2017, pp. 3554–3558.

[56] D. J. Love, R. W. Heath, Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[57] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[58] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[59] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive MIMO performance–TDD versus FDD: What do measurements say?" *ArXiv E-Print*, Apr. 2017, arXiv:1704.00623 [cs.IT].

[60] M. Karlsson, E. Björnson, and E. G. Larsson, "Broadcasting in massive MIMO using OSTBC with reduced dimension," in *International Symposium on Wireless Communication Systems*. IEEE, Aug. 2015, pp. 386–390.

[61] M. Karlsson, E. Björnson, and E. G. Larsson, "Performance of in-band transmission of system information in massive MIMO systems," *ArXiv E-Print*, Nov. 2017, arXiv:1711.07307 [cs.IT].

[62] E. de Carvalho, E. Björnson, J. H. Sørensen, P. Popovski, and E. G. Larsson, "Random access protocols for massive MIMO," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 216–222, May 2017.

[63] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3–4, pp. 154–655, 2017.

[64] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[65] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., 1968, vol. 2.

[66] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[67] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[68] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[69] G. Caire, "On the ergodic rate lower bounds with applications to massive mimo," *ArXiv E-Print*, Oct. 2017, arXiv:1705.03577 [cs.IT].

[70] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2921–2935, Mar. 2017.

[71] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2363–2378, May 2017.

[72] L. Sanguinetti, E. Björnson, M. Debbah, and A. L. Moustakas, "Optimal linear precoding in multi-user MIMO systems: A large system analysis," in *The Proceedings of the IEEE Global Communications Conference*, Dec. 2014, pp. 3922–3927.

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

[74] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Communications Letters*, vol. 1, no. 4, pp. 276–279, Apr. 2012.

[75] H. H. Yang, G. Geraci, T. Q. S. Quek, and J. G. Andrews, "Cell-edge-aware precoding for downlink massive MIMO cellular networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3344–3358, Jul. 2017.

[76] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Transactions on Wireless Communications*, Nov. 2017.

[77] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *EURASIP Journal on Wireless Communications and Networking*, Dec. 2017.

[78] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[79] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[80] V. C. Trinh, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *ArXiv E-Print*, Jul. 2017, arXiv:1707.03072 [cs.IT].

[81] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[82] J. Vinogradova, E. Björnson, and E. G. Larsson, "On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination," in *The Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2016, pp. 3421–3425.

[83] T. T. Do, E. Björnson, E. G. Larsson, and S. M. Razavizadeh, "Jamming-resistant receivers for the massive MIMO uplink," *IEEE Transactions on Information Forensics and Security*, 2017.

[84] H. Akhlaghpasand, S. M. Razavizadeh, E. Björnson, and T. T. Do, "Jamming detection in massive MIMO systems," *IEEE Wireless Communications Letters*, Nov. 2017.

[85] D. Neumann, T. Wiese, M. Joham, and W. Utschick, "A generalized matched filter framework for massive MIMO systems," *ArXiv E-Print*, Jul. 2017, arXiv:1707.09940 [cs.IT].

[86] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 56–65, Apr. 2005.

[87] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 4, pp. 1567–1592, Mar. 2013.

[88] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear precoding with low-resolution DACs for massive MU-MIMO-OFDM downlink," *ArXiv E-Print*, Sep. 2017, arXiv:1709.04846 [cs.IT].

[89] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "On one-bit quantized ZF precoding for the multiuser massive MIMO downlink," in *The Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, Jul. 2016, pp. 1–5.

[90] Y. Li, C. Tao, A. L. Swindlehurst, A. Mezghani, and L. Liu, "Downlink achievable rate analysis in massive MIMO systems with one-bit DACs," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1669–1672, Jul. 2017.

[91] W. H. Doherty, "A new high efficiency power amplifier for modulated waves," *Proceedings of the Institute of Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, Sep. 1936.

[92] A. Grebennikov, *RF and Microwave Power Amplifier Design*. McGraw-Hill, 2005.

[93] D. Middleton, *An introduction to statistical communication theory*. IEEE Press Classical Reissue, 1996, reprint. Originally published: New York, McGraw-Hill, 1960.

[94] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.

[95] H. Ochiai, "An analysis of band-limited communication systems from amplifier efficiency and distortion perspective," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1460–1472, Feb. 2013.

[96] N. Benvenuto, R. Dinis, D. Falconer, and S. Tomasin, "Single carrier modulation with nonlinear frequency domain equalization: An idea whose time has come—again," *Proceedings of the IEEE*, vol. 98, no. 1, pp. 69–96, Jan. 2010.

[97] W. J. Rugh, *Linear System Theory*. Prentice Hall, 1996, vol. 2.

[98] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Sep. 2006.

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

[100] N. Björsell and P. Händel, "Achievable ADC performance by post-correction utilizing dynamic modeling of the integral nonlinearity," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, Dec. 2007.

[101] H. Lundin, M. Skoglund, and P. Händel, "Optimal index-bit allocation for dynamic post-correction of analog-to-digital converters," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 660–671, Feb. 2005.

[102] M. Karlsson, E. Björnson, and E. G. Larsson, "Jamming a TDD point-to-point link using reciprocity-based MIMO," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 2957–2970, Dec. 2017.

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

[104] B. Murmann, "ADC performance survey 1997–2016," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html, accessed: 2016-07-19.

[105] P. Upadhyaya, J. Savoj, F.-T. An, A. Bekele, A. Jose, B. Xu, D. Wu, D. Furker, H. Aslanzadeh, H. Hedayati *et al.*, "3.3 A 0.5-to-32.75 Gb/s

flexible-reach wireline transceiver in 20 nm CMOS," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, Feb. 2015, pp. 1–3.

[106] K. Senel, E. Björnson, and E. G. Larsson, "Optimal base station design with limited fronthaul: Massive bandwidth or massive MIMO?" *ArXiv E-Print*, Sep. 2017, arXiv:1709.05172 [cs.IT].

[107] G. H. B. Gustavsson and B. Petersson, *Good Research Practice – What is it?* Vetenskapsrådet, 2006.

Included Papers

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-143455

Other Recently Published Theses From

The Division of Communication Systems
Department of Electrical Engineering (ISY)
Linköping University, Sweden

Victor Hei Cheng, *Aspects of Power Allocation in Massive MIMO*, Linköping Studies in Science and Technology. Licentiate Thesis, No. 1767, 2017.

Marcus Karlsson, *Aspects of Massive MIMO*, Linköping Studies in Science and Technology. Licentiate Thesis, No. 1764, 2017.

Christopher Mollén, *On Massive MIMO Base Stations with Low-End Hardware*, Linköping Studies in Science and Technology. Licentiate Thesis, No. 1756, 2016.

Antonios Pitarokoilis, *Phase Noise and Wideband Transmission in Massive MIMO*, Linköping Studies in Science and Technology. Dissertations, No. 1756, 2016.

Anu Kalidas M. Pillai, *Signal Reconstruction Algorithms for Time-Interleaved ADCs*, Linköping Studies in Science and Technology. Dissertations, No. 1672, 2015.

Ngô Quốc Hiển, *Massive MIMO: Fundamentals and System Designs*, Linköping Studies in Science and Technology. Dissertations, No. 1642, 2015.

Mirsad Čirkić, *Efficient MIMO Detection Methods*, Linköping Studies in Science and Technology. Dissertations, No. 1570, 2014.

Reza Moosavi, *Improving the Efficiency of Control Signaling in Wireless Multiple Access Systems*, Linköping Studies in Science and Technology. Dissertations, No. 1556, 2014.

Johannes Lindblom, *The MISO Interference Channel as a Model for Non-Orthogonal Spectrum Sharing*, Linköping Studies in Science and Technology. Dissertations, No. 1555, 2014.

Antonios Pitarokoilis, *On the Performance of Massive MIMO Systems with Single Carrier Transmission and Phase Noise*, Linköping Studies in Science and Technology. Licentiate Thesis, No. 1618, 2013.

Tumula V. K. Chaitanya, *HARQ Systems: Resource Allocation, Feedback Error Protection, and Bits-to-Symbol Mappings*, Linköping Studies in Science and Technology. Dissertations, No. 1555, 2013.