Linköping Studies in Science and Technology
Dissertations, No. 1950

Blind Massive MIMO Base Stations

Downlink Transmission and Jamming

Marcus Karlsson

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.

Blind Massive MIMO Base Stations

ISBN 978-91-7685-249-1
ISSN 0345-7524
URL http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-149898

Printed in Sweden by LiU-Tryck, Linköping 2018

Abstract

Massive MIMO (Multiple-Input–Multiple-Output) is a wireless technology which aims to serve several different devices simultaneously in the same frequency band through spatial multiplexing, made possible by using a large number of antennas at the base station. The many antennas facilitates efficient beamforming, based on channel estimates acquired from uplink reference signals, which allows the base station to transmit signals exactly where they are needed. The multiplexing together with the array gain from the beamforming can increase the spectral efficiency over contemporary systems.

One challenge of practical importance is how to transmit data in the downlink when no channel state information is available. When a device initially joins the network, prior to transmitting uplink reference signals that enable beamforming, it needs system information—instructions on how to properly function within the network. It is transmission of system information that is the main focus of this thesis. In particular, the thesis analyzes how the reliability of the transmission of system information depends on the available amount of diversity. It is shown how downlink reference signals, space-time block codes, and power allocation can be used to improve the reliability of this transmission.

In order to estimate the uplink and downlink channels from uplink reference signals, which is imperative to ensure scalability in the number of base station antennas, massive MIMO relies on channel reciprocity. This thesis shows that the principles of channel reciprocity can also be exploited by a jammer, a malicious transmitter, aiming to disrupt legitimate communication between two devices. A heuristic scheme is proposed in which the jammer estimates the channel to a target device blindly, without any knowledge of the transmitted legitimate signals, and subsequently beamforms noise towards the target. Under the same power constraint, the proposed jammer can disrupt the legitimate link more effectively than a conventional omnidirectional jammer in many cases.

Populärvetenskaplig Sammanfattning

Massiv MIMO (eng: Multiple-Input–Multiple-Output) är en teknologi inom cellulär kommunikation som förutspås ha en betydande roll i framtida kommunikationssystem på grund av de många fördelar som denna teknologi medför. Massiv MIMO innebär att basstationen har ett stort antal antenner där varje antenn kan styras individuellt. De många antennerna gör att basstationen kan rikta de elektromagnetiska signalerna på ett sådant sätt att de förstärks på positioner där användarna befinner sig men släcks ut i övrigt. Detta i sin tur innebär att flera användare kan betjänas samtidigt, på samma frekvensband utan att de stör varandra. Detta medför att massiv MIMO kan erbjuda en högre datatakt än nutida cellulära kommunikationssystem.

För att kunna rikta signalerna på ett effektivt sätt måste basstationen känna till kanalen, eller utbredningsmiljön, mellan sig själv och de användare som betjänas. När en användare precis ansluter till systemet vet basstationen inte var användaren befinner sig, men måste likväl tillgodose användaren med information om hur systemet fungerar. Nu måste alltså basstationen kommunicera med användaren, utan möjligheten att kunna rikta signalen på ett effektivt sätt. Det är detta problem som vi i huvudsak studerar i denna avhandling: hur man kan utnyttja de många antennerna på basstationen för att skicka information till användarna utan någon kanalkännedom.

Vi studerar även hur en gruppantenn med många antenner, baserad på samma teknologi som massiv MIMO, kan användas som en störsändare. Störsändarens mål är att hindra kommunikationen mellan två enheter på ett effektivt sätt. En störsändare med ett stort antal antenner kan, utan någon kännedom av vad de två enheterna skickar, i många fall prestera bättre än en konventionell störsändare på grund av att störsignalen kan riktas mot en specifik enhet.

Contents

Acknowledgements

I am immensely grateful to all the people in my life that have helped me, in one way or another, not only during the last five years but my whole life. This thesis could not have been written without you.

I would like to extend my sincerest gratitude to my supervisors Professor Erik G. Larsson and Associate Professor Emil Björnson. You have have taught me the importance of intuitive reasoning and what it means to be a researcher and a teacher. I have very much enjoyed the discussions we have had over the years, especially when the way forward is unclear. To witness two people with your wisdom and knowledge discuss the road ahead has been amazing and educational.

I am amazed and humbled by my colleagues and friends at Communication Systems in Linköping. It is so much interesting discussing possible solutions to a new problem, or minuscule details regarding some fundamental concept. It is also very fun to be able to take breaks and discuss other things, like frozen strawberries or perfect asymmetry. I'd like to say a special thank you to Ema, who helped with many, many details concerning this thesis.

To Christopher—my colleague, office mate, and friend. It has been comforting to know that I can ask you anything—no matter how stupid—and you will always give a sincere and thought out response. Apart from our many research discussions, it has also been wonderful to discuss the review process, the housing crisis, the economy, or anything else that baffled one of us during these five years.

To my friends who keep me grounded and remind me that there are other things in life than matrices, expected values, and random distributions. It has been a true privilege to have known you for a third of my life. It really doesn't matter what we do—spending time in a pool on a cruise ship, preparing a gimlet, or just watching an unfortunate kid get blasted into space—you guys always deliver.

To my family, who always let me go my own way and made me feel like I can do whatever I want, without any pressure.

Marcus Karlsson
Linköping, Summer 2018

List of Abbreviations

AP	access point (a small base station)
AWGN	additive white Gaussian noise
bpcu	bits per channel use
BS	base station
CDF	cumulative distribution function
i.i.d.	independent and identically distributed
LS	least squares
LTE	Long-Term Evolution
MAP	maximum a posteriori
MIMO	multiple-input–multiple-output
MISO	multiple-input–single-output
ML	maximum likelihood
MMSE	minimum mean-square error
MVU	minimum-variance (and) unbiased
NR	New Radio
PDF	probability density function
RF	radio frequency
SIMO	single-input–multiple-output
SINR	signal-to-interference-and-noise ratio
SISO	single-input–single-output
SNR	signal-to-noise ratio
TDD	time-division duplex

Chapter 1

Introduction and Motivation

Wireless communication today is an integral part of everyday life for millions of people. To not have your phone or laptop to check schedules or answer emails, is perceived as a nightmare. To not have access to the Internet is paralyzing. The demand for accessing and spreading information, in particular over a wireless connection will surely continue to rise for years to come. Somehow we need to meet this demand, without spending more bandwidth, time, or energy as these wireless resources are already limited. The frequency spectrum is finite, time is scarce, and in an energy-starving world, increasing energy consumption to meet increased demand is not a sustainable option. Wireless research is all about trying to simultaneously fulfill the conflicting goals of faster and more reliable communication without spending more resources.

1.1 Massive MIMO

The fifth generation of cellular network technology (5G), also known as New Radio (NR), will offer higher data rates, improved coverage and reliability, and reduced latency compared to current technology [1, 2]. One of the most promising physical-layer access technologies for NR is massive multiple-input multiple-output (MIMO) [3, 4]. Massive MIMO is an incarnation of the multi-user MIMO concept, where the base station (BS) is equipped with a large number of antennas and serves many users[1] in the same time-frequency resource by spatial multiplexing. Each antenna has its own radio-frequency (RF) chain, which facilitates fully digital beamforming that works in any propagation environment. Massive MIMO can increase the data rate (bits per second) for a fixed bandwidth and power, com-

[1]We will use the term terminal/user/device interchangeably throughout the thesis. Examples of terminals include smartphones, tablets, laptops, sensors, or anything else that needs service.

pared to contemporary systems [5]. Specifically, it is known that this advantage grows with the number of antennas since more antennas yield both a higher array gain and improved orthogonality between the user channels [6]. There is no need for intricate transmission, reception or detection methods as linear processing performs well [7]. Massive MIMO is also robust to many of the impairments caused by the use of inexpensive, low-end hardware [8].

All these things are made possible at the same time by the use of a large number of BS antennas. The multitude of antennas makes it possible to *beamform* different signals to different users, so the signals add up constructively at the desired user and destructively everywhere else. This enables the BS to *multiplex spatially*, serving different users in parallel, using the same time-frequency resource. The beamforming also provides an *array gain*, as the transmitted energy is not wasted transmitting in all directions, but focused towards the terminals. Linear processing performs very well because of the phenomenon known as *favorable propagation* [7, 9]. The terminals do not need to estimate the channel, as long as the statistics of the channel are known, because the channel behaves almost deterministically—a phenomenon known as *channel hardening*.

Since the initial paper [3], a sea of papers have been published, analyzing, for example, spectral efficiency [10, 11] and non-conventional ways to benefit from the many antennas [12–16]. Focus has been on showing that the theory behind massive MIMO is solid, and that the gains are impressive even at a finite number of antennas and with non-ideal hardware [17, 18], making it viable in practice.

There exist several testbeds, both from the industry and academia [19–21], showing that the beautiful closed-form rate expressions obtained in theory are actually achievable in reality. Recently, several companies, such as Facebook, Samsung, and Nokia, have demonstrated the power of massive MIMO by breaking world records in spectral efficiency and by deploying massive MIMO to increase the network capacity of particularly highly loaded systems, such as the 2018 World Cup in Russia. However, this is just the beginning and more research is needed for massive MIMO to reach its true potential.

System Information

Much of the research involving massive MIMO has been done on the physical layer, analyzing how to acquire channel state information (CSI), transmit data, and deal with interference. However, before the users can transmit anything, they must first receive and decode the system information—instructions on how to operate within the network and contact the BS [22]. These instructions are continuously broadcasted from the BS without any CSI. Not having CSI can severely limit the coverage, especially for users at the cell edge, since the array gain from the coher-

ent beamforming is nonexistent without CSI [23]; thus, efficient techniques for downlink transmission without CSI at the BS are needed in massive MIMO.

Jamming

Using wireless technology to a greater extent means that more sensitive and private information is transmitted over the air, for anyone to receive [24]. It is thus necessary for massive MIMO to be resilient towards jamming and eavesdropping in order to prevent unauthorized parties to have access to this information. The massive MIMO operation seems to be inherently immune to some types attacks, while it is more vulnerable to others [25]. In order to ensure the security and privacy of future wireless communication, understanding the strengths and weaknesses of massive MIMO is imperative.

1.2 Contributions of the Thesis

This thesis is divided into two parts: an introduction and a collection of papers. In the introductory part, basic concepts regarding wireless communication and massive MIMO are covered. These chapters are meant to give the reader a concise introduction to massive MIMO and at the same time put the particular research topics covered in the subsequent papers into context. The papers in the second part of the thesis cover two different topics: transmission of system information and jamming. The transmission of system information is considered for both conventional (collocated) and cell-free massive MIMO and we analyze to what extent transmit diversity in the form of space-time block codes can improve coverage and outage. For jamming, we show that massive MIMO technology can blindly—without any prior knowledge of the channels or transmitted signals—outperform a conventional omnidirectional jammer.

All papers below, including the code generating the numerical results, are written by the first author. The co-authors (supervisors) have with an abundance of comments, ideas, and proofreading made the papers more understandable, more rigorous and, better in every way. None of the papers below would have the same quality without the help of both supervisors.

Paper A: Performance of In-band Transmission of System Information in Massive MIMO Systems

Authored by: Marcus Karlsson, Emil Björnson, and Erik G. Larsson

Published in IEEE Transactions on Wireless Communications vol. 17, no. 3, pp. 1700–1712, March 2018.

The transmission of system information in massive MIMO is analyzed. In particular the use of orthogonal space-time block codes to facilitate reliable communication in the downlink without channel state information at the base station. The orthogonal space-time block codes are precoded to reduce the pilot overhead and we discuss the effects of this precoding when the channels to the different base station antennas are correlated. We further analyze the performance of four orthogonal space-time block codes in settings with different number of time/frequency diversity branches, and compare the performance of massive MIMO base station to that of a single-antenna base station.

Paper B: Techniques for System Information Broadcast in Cell-Free Massive MIMO

Authored by: Marcus Karlsson, Emil Björnson, and Erik G. Larsson

Submitted to: IEEE Transactions on Communications

We investigate how to transmit system information in a cell-free system with many geographically distributed, single-antenna access points. We use space-time block codes to achieve spatial diversity without channel state information at the access points. From this setup, a new problem of how to group the access points, in order to jointly transmit the space-time block codes, appears. We investigate how to group the access points, and also introduce a heuristic power allocation that can further improve coverage.

Paper C: Jamming a TDD Point-to-Point Link Using Reciprocity-Based MIMO

Authored by: Marcus Karlsson, Emil Björnson, and Erik G. Larsson

Published in IEEE Transactions on Information Forensics and Security vol. 12, no. 12, pp. 2957–2970, December 2017.

In this paper, we consider a massive MIMO jammer, a malicious transmitter that aims to stop communication of a legitimate point-to-point link consisting of two single-antenna users operating in time-division duplex. We present a jamming algorithm where the jammer has very limited knowledge of the legitimate link—no knowledge of the legitimate transmit signals or any channel state information. To

estimate the frame timing, the jammer analyzes the structure of the sample covariance matrix. After this, the jammer exploits the channel reciprocity to estimate the channel to one of the users in order beamform noise to reduce the rate of the legitimate link.

1.3 Conclusions

Paper A shows the importance of spatial diversity when transmitting system information in massive MIMO, especially in cases with very limited time and frequency diversity. Even though extremely large space-time block codes could technically be used for downlink transmission, this is not practically viable. Each additional spatial diversity branch requires additional pilot overhead which in turn is limited by the finite coherence interval. As a consequence, larger codes, with diversity order much greater than 10 are probably not too useful in practice, unless the coherence interval is extremely long and there is a stringent constraint on outage probability. Paper A also shows the importance of choosing an appropriate precoder for transmitting the small space-time block code with the large antenna array. In particular, if the channels are correlated, special care has to be taken when precoding the space-time block code. In addition, the paper shows the benefit of allocating different powers to the downlink pilots and the data. This power allocation can be done in the absence of channel state information.

Paper B is similar to Paper A in many ways, as it studies a similar problem, albeit in a vastly different setting. To be able to utilize spatial diversity in the form of a space-time block code, the single-antenna access points need to cooperate. Specifically, the access points need to group up, in order to transmit different parts of the space-time block code. The paper shows that the grouping of access points may not need any complicated algorithms—randomly grouping the access points works fairly well. The performance can be enhanced slightly by more sophisticated grouping methods, which takes the locations of the access points into account. What is important to note is that neither of these methods require channel state information, and can be done offline. What proves more important, when considering the outage performance, is the power allocation between pilots and data. Again, a heuristic power allocation based on the principles of the one in Paper A is employed and is shown to improve the performance.

Paper C shows how potent a jammer armed with the massive MIMO concept can be. By leveraging fundamental properties of the legitimate transmission, the jammer can disrupt the legitimate link without almost any prior information. It is moreover much more difficult to locate than a traditional omnidirectional jammer because of the employed beamforming. Paper C sheds some light on what

Table 1: Excluded papers

M. Karlsson, E. Björnson, and E. G. Larsson, "Broadcasting in massive MIMO using OSTBC with reduced dimension," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, 2015, pp. 386–390.
M. Karlsson and E. G. Larsson, "On the operation of massive MIMO with and without transmitter CSI," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2014, pp. 1–5.
M. Karlsson and E. G. Larsson, "Massive MIMO as a cyber-weapon," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 661–665.

may happen in the future, when massive MIMO is used to destroy communication, rather than to enable it. The paper further discusses possible defensive countermeasures, of which transmit and receive beamforming at the legitimate link seem the most promising.

1.4 Excluded Papers

The papers in Table 1 are not included in the thesis because they are preliminary (conference) versions of the included papers, and are therefore deemed superfluous.

1.5 Notation

The mathematical notation of the thesis is as follows. Scalars, column vectors, and matrices are denoted by lower-case letters, boldfaced lower-case letters, and boldfaced upper-case letters—such as x, $\mathbf{x}$ and $\mathbf{X}$—respectively. The transpose of a matrix is written as $\mathbf{X}^{\mathsf{T}}$, the Hermitian (conjugate) transpose is written as $\mathbf{X}^{\mathsf{H}}$, and the Euclidean norm (the 2-norm) of a vector is written as $\|\mathbf{x}\|$. The determinant and trace of a (square) matrix are denoted by $\det(\mathbf{X})$ and $\mathrm{tr}\,(\mathbf{X})$, respectively. The identity matrix of dimension x is denoted by $\mathbf{I}_x$. The zero matrix is denoted by $\mathbf{0}$, whose dimensions are clear from the context. The natural exponential function is denoted by $\exp(\cdot)$. The distribution of a circularly-symmetric Gaussian (random) vector with mean $\mathbf{x}$ and covariance $\mathbf{X}$ is denoted by $\mathcal{CN}(\mathbf{x}, \mathbf{X})$. The normal (Gaussian) distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. The probability density function is denoted by $p\,(\cdot)$ for continuous variables and $P\,(\cdot)$

for discrete variables. The probability of an event is denoted by $\mathbb{P}[\cdot]$. The Fourier transform is denoted by $\mathcal{F}\{\cdot\}$. The sum $\sum_{a \neq b}$ is a sum over all permissible values of a (which should be clear from the context) except b. The real and imaginary parts are denoted by $\mathfrak{R}\{\cdot\}$ and $\mathfrak{I}\{\cdot\}$, respectively, while the imaginary unit is denoted by i. The estimate of an unknown variable x is denoted by $\widehat{x}$. The expectation of a random variable is denoted by $\mathbb{E}[\cdot]$ and the covariance matrix of a random vector is denoted by $\operatorname{cov}(\cdot)$. Given two functions $f(x)$ and $g(x)$, the notation $f(x) = \mathcal{O}(g(x)), x \to \infty$ means that $\limsup_{x \to \infty} |f(x)/g(x)| < \infty$.

Chapter 2

Basic Concepts

This chapter introduces basic concepts regarding communications and wireless communications. The chapter is primarily aimed towards engineers without a background in communications and should be seen as a short cut into the theory of wireless communications needed to grasp the subsequent introduction to massive MIMO in Chapter 3 and the included papers. The goal is to motivate the use of basic models and to introduce some essential and useful tools. For a more thorough and rigorous introduction to these concepts, consider [26–30].

2.1 The Communication Problem

It is difficult get a more concise description of the communication problem than what was given by Claude Shannon in [31]: "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point". A schematic overview of a communication system is shown in Figure 1; it consists of a transmitter, a channel, and a receiver.

The transmitter transforms the original message m to a signal x that is transmitted over the channel. The message m belongs to a finite set of possible messages, $\mathcal{M}$, known to both the transmitter and the receiver. This finite set can be, for example, the letters of the alphabet or ASCII characters. For transmission over

Figure 1: The fundamental problem of communication is to transfer a message from one point to another.

the channel, the message is *modulated*, meaning that it is mapped onto a sinusoid with a particular phase, frequency, and amplitude:

$$x(t) = x_{\mathrm{E}}(t) \cos\left(2\pi f_{\mathrm{C}} t + \phi_x(t)\right). \tag{1}$$

The information about the message m is now embedded in the transmitted signal $x(t)$, more precisely in its amplitude (or envelope), $x_{\mathrm{E}}(t)$ and its phase $\phi_x(t)$. The modulation makes it possible to transmit the message over the air in a particular frequency band by choosing the carrier frequency f_{C}. Signals occupying a specific frequency band are called *passband signals*.

As the transmitted signal travels through the channel towards the receiver, it is reflected, diffracted, and scattered by objects in its path. As a consequence, the received signal only resembles the transmitted one approximately. Moreover, noise and interference may distort the received signal even further.

At the receiver, the received signal is *demodulated*, sampled, and further processed in order negate the distortions caused by the channel, to finally recover the original message m. This can prove difficult, as the distortions caused by the channel may be severe and partially or completely unknown. The receiver tries to find a strategy that minimizes the probability of choosing the wrong message, in hope that the decoded message $\hat{m}$ equals the transmitted message m.

2.2 System Model

In this section we will discuss how to model a communication system. We want to obtain a useful, tractable model that captures the effects of modulation, channel propagation, demodulation, and sampling. The end product will be the time-discrete complex baseband model [26, 27].

Consider a signal $x(t)$ transmitted over a wireless channel. On its way from the transmitter to the receiver, the signal is reflected, scattered and diffracted many times as it bounces from object to object. When the receiver then measures the signal, it sees a linear combination of all these time-delayed and attenuated versions of $x(t)$. This phenomenon is called *multipath propagation* due to the signal traveling a number of different paths to reach the receiver.

In principle, if all parameters of the channel were known, one could start with Maxwell's equations and derive the received signal from the transmitted one. This is, however, impractical. As an approximation, a technique known as ray tracing may be used to model the channel. Ray tracing methods are useful when the number of multipath components is small and the propagation environment is known. In practice, wireless channels often change over time, and thus an accurate ray tracing model is not possible. Hence, we resort to a statistical channel model, which allows the channel to change over time [27, Ch. 2].

Since the world around us is moving, the channel will change: new paths from the transmitter to the receiver will arise while others will be blocked, or changed. This makes the wireless channel inevitably time varying. However, the channel may be considered almost constant if observed over a short enough time period. The time when the channel can be considered constant is called the *coherence time* and is denoted by T_{C}. This means that signals transmitted less than T_{C} seconds apart will experience the same channel. The coherence time depends on how quickly the transmitter and receiver are moving, but also how fast the world around them changes. The channel is said to change significantly if objects move a fraction of a wavelength, typically a few centimeters for carrier frequencies considered in this thesis, giving a coherence time in the order of milliseconds [26].

Assuming the duration of the transmitted signal is short enough for the channel to be considered time invariant, i.e., shorter than T_{C}, the (noise-free) received signal can be written as

$$y(t) = \sum_i a_i x(t - \tau_i), \tag{2}$$

where a_i and τ_i are the attenuation and the propagation delay of path i, respectively. We do not explicitly give an upper limit on the summation in (2), but it can be seen as "practically infinite", in a rich scattering environment.

Another key parameter of the wireless channel is the *coherence bandwidth*, denoted by B_{C}. Similar to T_{C} in the time domain, the coherence bandwidth denotes the frequency range over which the channel can be considered constant; hence, frequency components differing less than B_{C} Hz will experience the same channel. The frequency-domain measure of coherence bandwidth is closely related to the time-domain measure *delay spread*; in fact, as we will see, they are two sides of the same coin [32, Ch. 5].

The delay spread—denoted by T_{D}—is the difference in delay between the shortest path and the longest path, i.e., the effective duration of the impulse response. In order to get some insight about the relation between the coherence bandwidth and the delay spread we consider the channel impulse response corresponding to (2):

$$h(t) = \sum_i a_i \delta(t - \tau_i),$$

where $\delta(t)$ denotes the Dirac delta. The channel frequency response is obtained by Fourier transformation of the channel impulse response:

$$H(f) = \mathcal{F}\{h(t)\} = \sum_i a_i e^{-\mathrm{i}2\pi f \tau_i}.$$

Assuming that $\tau_0 \leq \tau_1 \leq \tau_2 \leq \ldots$, we can write

$$H(f) = e^{-\mathrm{i}2\pi f\tau_0}\left(1 + e^{-\mathrm{i}2\pi f(\tau_1-\tau_0)} + \ldots + e^{-\mathrm{i}2\pi fT_{\mathrm{D}}}\right),$$

from which we can deduce that the rate of change of $|H(f)|$ is governed by the delay spread T_{D}. We also see that if $B_{\mathrm{C}}T_{\mathrm{D}} \ll 1$, then $H(f \pm \Delta) \approx H(f)$ for $\Delta < B_{\mathrm{C}}/2$. In general, the coherence bandwidth is inversely proportional to the delay spread, but an exact relationship requires specific information about the multipath structure [32, Ch. 5.4].

2.2.1 Complex Signals

All transmitted and received signals are real valued, however, to simplify the analysis the signals are often viewed as complex entities. Let us rewrite the passband signal in (1) as

$$x_{\mathrm{PB}}(t) = x_{\mathrm{E}}(t)\cos(2\pi f_{\mathrm{C}}t + \phi_x(t)) = x_{\mathrm{I}}(t)\cos(2\pi f_{\mathrm{C}}t) - x_{\mathrm{Q}}(t)\sin(2\pi f_{\mathrm{C}}t),$$

where

$$x_{\mathrm{I}}(t) = x_{\mathrm{E}}(t)\cos(\phi_x(t)),\ x_{\mathrm{Q}}(t) = x_{\mathrm{E}}(t)\sin(\phi_x(t)).$$

Equivalently we can express the passband signal in terms of the baseband signal as

$$x_{\mathrm{PB}}(t) = \Re\left\{x_{\mathrm{BB}}(t)e^{\mathrm{i}2\pi f_{\mathrm{C}}t}\right\},$$

where

$$x_{\mathrm{BB}}(t) = x_{\mathrm{I}}(t) + \mathrm{i}x_{\mathrm{Q}}(t)$$

is called the complex baseband equivalent to $x_{\mathrm{PB}}(t)$ and $\Re\{\cdot\}$ denotes the real part. Note that we can obtain the in-phase and quadrature component, $x_{\mathrm{I}}(t)$ and $x_{\mathrm{Q}}(t)$, respectively, by demodulating the passband signal:

$$x_{\mathrm{I}}(t) = \mathrm{LP}\left\{2x_{\mathrm{PB}}(t)\cos(2\pi f_{\mathrm{C}}t)\right\}$$

and

$$x_{\mathrm{Q}}(t) = -\mathrm{LP}\left\{2x_{\mathrm{PB}}(t)\sin(2\pi f_{\mathrm{C}}t)\right\},$$

where $\mathrm{LP}\{\cdot\}$ denotes ideal lowpass filtering with an appropriate cut-off frequency.

One way of thinking about how to go from a real signal in the passband, to a complex baseband signal is to look at the spectrum of the real passband signal $x(t)$. As $x(t)$ is real, its spectrum is symmetric around zero, as seen in Figure 2a. This means we can find out everything we need to know about $x(t)$ by only considering the spectrum for frequencies larger than zero. Moving this spectrum down to the

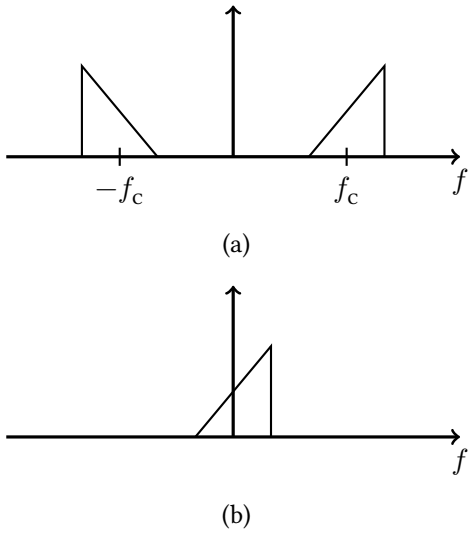


Figure 2: The complex baseband representation of a signal can be thought of as shifting the positive part of the spectrum of the real-valued passband signal down to the origin. Both signals are equivalent as they carry the same information, but the latter allows for a more compact description: (a) the spectrum of the real-valued passband signal is symmetric around $f = 0$; (b) the positive part of the spectrum centered, but not necessarily symmetric, around $f = 0$.

baseband means that the baseband signal, in general, is complex, as the spectrum is not necessarily symmetric, see Figure 2b.

Moving the positive spectrum down to the baseband also halves the bandwidth of the signal, making the bandwidth of the real passband signal twice that of its complex baseband equivalent. This might give the impression that one can break the sampling theorem, which states that a sample frequency of at least $2B$ Hz is necessary to fully represent a signal with bandwidth B Hz [33, Th. 1]. Sure, one can get away with sampling the complex baseband equivalent with B Hz, but now these are complex samples (two real samples each sampling time), effectively giving the same dimensionality.

2.2.2 Frequency Flat Fading

At first glance, it may look like there is not much to gain from introducing this complex notation in Section 2.2.1. However, consider now a passband signal $x_{\text{PB}}(t)$ with bandwidth B; that is, its spectrum zero outside the frequency bands $|\pm f_c \pm B/2|$. If the transmitted signal's bandwidth B is smaller than the coherence

bandwidth, B_{C},

$$\mathcal{F}\{x_{\text{PB}}(t-\tau)\} = X(f)e^{-\text{i}2\pi f\tau} \approx X(f) = \mathcal{F}\{x_{\text{PB}}(t)\},$$

as long as $\tau < T_{\text{D}}$, which means that $x_{\text{PB}}(t-\tau) \approx x_{\text{PB}}(t)$. We say that, paths whose delays differ less than T_{D} are *unresolvable* [27]. The received passband signal can then be written as

$$\begin{aligned} y_{\text{PB}}(t) &= \sum_i a_i x_{\text{PB}}(t-\tau_i) \\ &= \sum_i a_i \Re\left\{x_{\text{BB}}(t-\tau_i)e^{\text{i}2\pi f(t-\tau_i)}\right\} \\ &= \Re\left\{\sum_i a_i e^{-\text{i}2\pi f\tau_i} x_{\text{BB}}(t-\tau_i)e^{\text{i}2\pi ft}\right\} \\ &\approx \Re\left\{x_{\text{BB}}(t)\left(\sum_i a_i e^{-\text{i}2\pi f\tau_i}\right)e^{\text{i}2\pi ft}\right\}. \end{aligned} \tag{3}$$

Now, with $y_{\text{BB}}(t)$ analogously defined as $x_{\text{BB}}(t)$ and

$$y_{\text{PB}}(t) = \Re\left\{y_{\text{BB}}(t)e^{\text{i}2\pi ft}\right\},$$

we can write

$$y_{\text{BB}}(t) = hx_{\text{BB}}(t), \tag{4}$$

where

$$h = \sum_i a_i e^{-\text{i}2\pi f\tau_i} \tag{5}$$

is called the *channel coefficient*, or simply the channel. The randomness in the channel (stemming from the many different paths) is known as *fading*; in particular, the channel in (4) is called *frequency flat*, since the fading affects all frequency components of the signal, $x_{\text{BB}}(t)$, the same way.

Sampling (4) ideally with the Nyquist frequency $f_{\text{S}} = B = 1/T_{\text{S}}$ gives

$$y_{\text{BB}}[n] = hx_{\text{BB}}[n], \tag{6}$$

where $y_{\text{BB}}[n] \triangleq y_{\text{BB}}(nT_{\text{S}})$, $x_{\text{BB}}[n] \triangleq x_{\text{BB}}(nT_{\text{S}})$, and T_{S} is the sampling time. Quite often, the sample index is suppressed for notational simplicity and as all subsequent signals will be represented in the baseband notation, the subscript is superfluous. With these simplifications, (6) becomes

$$y = hx. \tag{7}$$

That is, with the help of the complex-baseband notation, we can relate the received sample to the transmitted sample by a single complex scalar h.

Let us now have a closer look at the channel coefficient, h, in (5) and how to model it statistically. Arguably the most common channel model in wireless communication is the Rayleigh-fading model, where the attenuation a_i is assumed to be Rayleigh distributed and the phase $2\pi f\tau_i$ modulo 2π is assumed to be uniformly distributed on $[0, 2\pi[$. Moreover, the parameters of different paths are independent and identically distributed (i.i.d.). Under these assumptions, the channel is a zero-mean, circularly-symmetric complex Gaussian random variable, i.e.,

$$h \sim \mathcal{CN}(0, \beta), \tag{8}$$

for some variance β.

By invoking the central limit theorem, h is a circularly-symmetric complex Gaussian random variable irrespective of the distributions of the amplitude and the delays, if the i.i.d. assumption remains and the number of paths is large [27, Ch. 3]. If, in addition, the terms in the sum constituting h, (5), have zero mean, then so does h and (8) holds.

It is common to split the random (fading) channel into two separate parts

$$h = \sqrt{\beta}g,$$

where β is the *large-scale fading* and $g \sim \mathcal{CN}(0, 1)$ is the *small-scale fading*. The large-scale fading models how the average channel magnitude changes on a macroscopic level, when the transmitter or receiver moves tens or hundreds of meters. It usually comprises distant-dependent path loss and shadow fading, caused by large objects. The small-scale fading models constructive and destructive interference of the multipath propagation and how the channel magnitude varies when moving short distances, in the order of a few wavelengths.

2.2.3 Frequency Selective Fading

If the signal bandwidth, B, is larger than the coherence bandwidth, B_{C}, all paths are no longer unresolvable and the approximation $x_{\text{PB}}(t-\tau) \approx x_{\text{PB}}(t)$ for $\tau < T_{\text{D}}$ used in (3) is no longer valid. Some paths will have delays that differ significantly (relative to the inverse of the signal bandwidth), so that the paths are *resolvable*. We can group mutually unresolvable paths with similar delays together in different batches, with paths in different batches being mutually resolvable. This gives the *frequency selective channel* modeled as

$$y[n] = \sum_{l=0}^{L-1} h[l]x[n-l]. \tag{9}$$

In (9), $h[l]$ is called the lth channel tap and each channel tap corresponds to a batch of unresolvable paths. In the frequency domain, a frequency selective channel affects different frequency components of the signal differently, hence the name. The number of channel taps, L, depends on the exact relation between the signal bandwidth and the coherence bandwidth. In the case of a single channel tap, (9) reduces to (7).

2.2.4 Additive White Gaussian Noise

So far, we have considered a receiver able to measure the output of the channel perfectly. In practice, the measurements unavoidably contain noise from the receiver circuit and signals originating from other electrical devices in the receiver's vicinity. Consider the following model, obtained by adding a noise term to the received signal in (2):

$$y(t) = \sum_i a_i x(t - \tau_i) + w(t), \tag{10}$$

where $w(t)$ denotes the received noise which is assumed to be white and Gaussian. The Gaussian assumption can be motivated by the fact that the noise often comes from many independent sources, hence, the central limit theorem tells us that the sum is approximately Gaussian. Moreover, Gaussian noise is tractable and pleasant to deal with when doing analytical work. The white assumption means that the noise does not have any specific structure.

After demodulation and ideal sampling, the general model for noisy frequency-selective channels can be written as

$$y[n] = \sum_{l=0}^{L-1} h[l]x[n-l] + w[n]. \tag{11}$$

The noise samples $w[n]$ are jointly Gaussian, circularly-symmetric random variables with variance σ^2, i.e., $w[n] \sim \mathcal{CN}(0, \sigma^2)$. Sampling with the Nyquist frequency, $f_s = B$, results in uncorrelated noise samples. Since the samples are jointly Gaussian and uncorrelated, they are independent.

2.2.5 Multiple-Input Multiple-Output

Up until now we have only considered a single-antenna transmitter communicating with a single-antenna receiver: a single-input single-output (SISO) channel. The model (11) can be generalized to a system involving multiple receiver and transmitter antennas: a MIMO channel. We focus on the case with a frequency flat channel ($L = 1$) to not convolve the analysis, although the frequency selective channel can also be extended to include MIMO.

Let us consider a transmitter with N_{T} antennas and a receiver with N_{R} antennas. If we consider a transmitter-receiver antenna pair, say transmit antenna $m \in \{1, \dots, N_{\text{T}}\}$ and receive antenna $k \in \{1, \dots, N_{\text{R}}\}$, we have the same situation as in the SISO case covered earlier in this section, (11). The received signal at antenna k, if only antenna m transmits (all other antennas are silent), can be written as

$$y_k = h_{km} x_m + w_k,$$

where h_{km} is the channel from transmit antenna m to receive antenna k; w_k is the noise measured at antenna k. When the entire array transmits, the received signal at antenna k will be the sum of all signals transmitted from the N_{T} transmit antennas:

$$y_k = \sum_{m=1}^{N_{\text{T}}} h_{km} x_m + w_k.$$

Using matrix notation, the N_{R} simultaneously received samples can conveniently be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{12}$$

where

$$\mathbf{y} = [y_1, \dots, y_{N_{\text{R}}}]^{\mathsf{T}},$$
$$\mathbf{x} = [x_1, \dots, x_{N_{\text{T}}}]^{\mathsf{T}},$$
$$\mathbf{w} = [w_1, \dots, w_{N_{\text{R}}}]^{\mathsf{T}},$$

and $\mathbf{H}$ is a matrix with element (k, m) equal to h_{km}.

As special cases of the MIMO channel we have $N_{\text{T}} > 1$ and $N_{\text{R}} = 1$, called the multiple-input–single-output (MISO) channel, and $N_{\text{T}} = 1$ and $N_{\text{R}} > 1$, called the single-input–multiple-output (SIMO) channel.

2.2.6 The Block-Fading Model

From the sampling theorem, we know that in order to fully represent a real-valued, continuous signal $x(t)$ with bandwidth B Hz, a sampling rate of at least $2B$ Hz is required. Assuming the signal duration is T seconds, it can thus be represented by $2BT$ real-valued samples. Conversely, in a time-frequency space of $T \times B$ (with T and B large enough), it is possible to fit $2BT$ real-valued samples. Similarly, it can be argued that a time-frequency space of $B \times T$ enables BT complex samples, or *channel uses* [34].

The model in (12) describes the received samples (signal) $\mathbf{y}$ when $\mathbf{x}$ was sent for a single channel use. The channel is considered static in the time-frequency space of $T_{\text{c}} \times B_{\text{c}}$, or $\tau_{\text{c}} = T_{\text{c}} B_{\text{c}}$ channel uses. τ_{c} is commonly referred to as

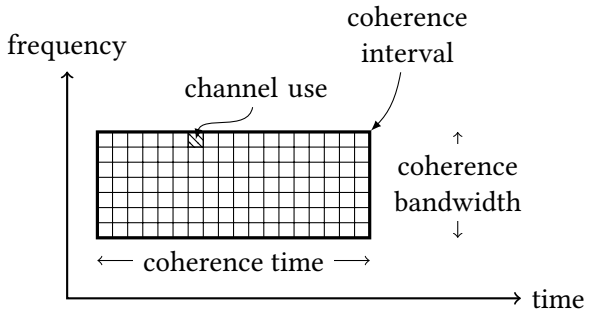


Figure 3: An illustration of the coherence interval in the time-frequency grid.

the *coherence interval* and is the time-frequency space over which a channel can accurately be modeled as a linear and time-invariant system; see Figure 3. The coherence interval can differ vastly between different applications, from only a few samples to practically infinite.

In a practical system, the communication may take place over a longer duration than T_{c} or a wider bandwidth than B_{c} so the channel will not be static over the whole transmission. However, as we have modeled it, it may be considered piecewise static, when considering small parts of the time-frequency space occupied by the transmission. This is usually referred to as the *block-fading* model. In each coherence interval (block) of τ_{c} channel uses (samples), the channel is static, but the channel varies between different coherence intervals. Usually, the channels in different coherence intervals are considered to be i.i.d.

2.3 Estimation and Detection

Estimation and detection theory are important subjects in statistical signal processing and are used in many parts of electrical engineering. In this section we present a few popular estimation and detection techniques used in communication theory; specific applications include symbol detection and channel estimation.

Many times when faced with a signal processing problem, there may be several different appropriate estimators or detectors to choose from. The optimal choice depends on the problem at hand, how it is modeled, and the performance metric. In this section we focus on the so-called linear model [28, 29],

$$\mathbf{r} = \mathbf{K}\boldsymbol{\theta} + \mathbf{w}, \tag{13}$$

as it is ubiquitous in communication theory in general and the most relevant model for the remainder of this thesis. In (13), $\mathbf{r} \in \mathbb{C}^N$ is the received (observed) samples,

$\mathbf{K} \in \mathbb{C}^{N \times p}$ is a known matrix with $N > p$ and full rank, $\boldsymbol{\theta} \in \mathbb{C}^p$ is the vector we wish to estimate, and $\mathbf{w} \in \mathbb{C}^N$ is the noise vector. The noise is assumed to be white and circularly symmetric. Under the assumption that the receiver knows the noise statistics it is sufficient to consider $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, since the receiver can subtract the mean and whiten the noise.

2.3.1 Estimation Theory

The general purpose of estimation is to use measurement data to estimate one or several parameters, somehow embedded in this data. In the linear model (13), we wish to estimate the value of $\boldsymbol{\theta}$ from the observation $\mathbf{r}$. There are two different philosophies, stemming from the difference in how the parameter $\boldsymbol{\theta}$ is viewed. In classical estimation theory, the parameter $\boldsymbol{\theta}$ is viewed as an unknown, deterministic constant, while in Bayesian estimation theory, the parameter $\boldsymbol{\theta}$ is viewed as an unknown realization of a random variable. As this affects the premise of the estimation problem, it also affects the solution.

Classical Estimation Theory

What makes a good estimator depends on how the performance of the estimation is measured. Do we want to minimize the absolute error, the squared error, or some other cost function between the parameter $\boldsymbol{\theta}$ and its estimate $\hat{\boldsymbol{\theta}}$? One intuitive criterion is the minimum variance unbiased (MVU) estimator [28]. As the name suggests, it requires the estimate to be unbiased, and among all the unbiased estimators, we choose the one with the lowest variance. Unfortunately, an estimator satisfying these conditions may not always exist, and even if it does, we might not be able to find it.

The Cramer-Rao (lower) bound (CRB) specifies the minimum variance an unbiased estimator can have. If an unbiased estimator attains the CRB it is said to be *efficient*. Thus, the CRB can be used to validate or measure the performance of any unbiased estimator. If the proposed estimator attains the CRB it is known that no other estimator can perform better; hence it must be the MVU estimator.

In practice, maximum likelihood (ML) estimation is often used when no MVU estimator can be found. It is a popular method, because it is straight forward to implement in many cases and it makes intuitive sense. Moreover, the ML estimator is asymptotically efficient, meaning it attains the CRB for large data records, and perhaps more importantly—if an MVU estimator does exists, it is given by the ML estimator [28, Ch. 7.4].

The intuition behind ML estimation is that we should choose the parameter $\boldsymbol{\theta}$ that maximizes the likelihood of observing the data we actually did observe.

Mathematically speaking, the ML estimator is given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\mathbf{r};\boldsymbol{\theta}), \tag{14}$$

where $p(\mathbf{r};\boldsymbol{\theta})$ is the probability density function (PDF) of $\mathbf{r}$ and the notation is used to stress that the PDF of the random vector $\mathbf{r}$ is parameterized by $\boldsymbol{\theta}$. When viewed as a function of the parameter $\boldsymbol{\theta}$, $p(\mathbf{r};\boldsymbol{\theta})$ is often referred to as the *likelihood function*. Hence the name ML for (14).

For the linear model, the observed vector is distributed as

$$\mathbf{r} \sim \mathcal{CN}(\mathbf{K}\boldsymbol{\theta}, \mathbf{I}_N),$$

which implies

$$p(\mathbf{r};\boldsymbol{\theta}) = \pi^{-N} \exp\left(-\|\mathbf{r}-\mathbf{K}\boldsymbol{\theta}\|^2\right),$$

where $\|\cdot\|$ denotes the Euclidean norm. Maximizing the above with respect to $\boldsymbol{\theta}$ is equivalent to minimizing

$$\begin{aligned}\|\mathbf{r}-\mathbf{K}\boldsymbol{\theta}\|^2 &= \|\mathbf{K}(\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}}(\mathbf{r}-\mathbf{K}\boldsymbol{\theta}) + (\mathbf{I}_N - \mathbf{K}(\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}})(\mathbf{r}-\mathbf{K}\boldsymbol{\theta})\|^2 \\ &= \|\mathbf{K}(\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}}(\mathbf{r}-\mathbf{K}\boldsymbol{\theta})\|^2 \\ &\quad + \|(\mathbf{I}_N - \mathbf{K}(\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}})(\mathbf{r}-\mathbf{K}\boldsymbol{\theta})\|^2 \\ &= \|\mathbf{K}(\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}}(\mathbf{r}-\mathbf{K}\boldsymbol{\theta})\|^2 + \|(\mathbf{I}_N - \mathbf{K}(\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}})\mathbf{r}\|^2\end{aligned} \tag{15}$$

with respect to $\boldsymbol{\theta}$.[1] Since the second term in (15) is independent of $\boldsymbol{\theta}$ we have

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ p(\mathbf{r};\boldsymbol{\theta}) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ \|\mathbf{K}(\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}}(\mathbf{r}-\mathbf{K}\boldsymbol{\theta})\|^2 \\ &= (\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}}\mathbf{r},\end{aligned} \tag{16}$$

which is also the MVU estimator [28, Th. 7.5].

An alternative to the ML estimator, also common in practical applications, is the least squares (LS) estimator. The LS estimator also makes sense intuitively as it aims to find the parameter that minimizes the distance between the observation and what we would expect to receive in the absence of noise. Unfortunately, this easy-to-use and intuitively pleasing estimator does not assure any kind optimality. Still, it has been one of the go-to estimators, ever since it was introduced by Gauss in the late 1700s [28, Ch. 8].

[1] The superscript $^{\mathsf{H}}$ denotes Hermitian transpose (conjugate transpose).

The LS estimator for the linear model is given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\mathbf{r} - \mathbf{K}\boldsymbol{\theta}\|^2,$$

which, from (15) and (16), can be written as

$$\hat{\boldsymbol{\theta}} = (\mathbf{K}^{\mathsf{H}}\mathbf{K})^{-1}\mathbf{K}^{\mathsf{H}}\mathbf{r}.$$

For this particular model, the LS and the ML are operationally identical; however they are derived under completely different circumstances. The LS estimator makes no statistical assumptions on the data (noise), it simply aims to minimize the Euclidean distance between the observed data and the signal model. As a consequence, the LS estimator is the same regardless of the statistics of the noise vector $\mathbf{w}$, which is not the case for the ML estimator.

Bayesian Estimation

Let us now consider the Bayesian philosophy of estimation. Here, the parameter $\boldsymbol{\theta}$ is considered a random variable of which we wish to estimate a given realization. The randomness gives us the ability to incorporate any prior knowledge we have about $\boldsymbol{\theta}$ into the model, using Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{r}) = \frac{p(\mathbf{r}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{r})},$$

where the posterior is expressed in terms of the likelihood, the prior, and the evidence [35].[2] When the MVU is difficult or impossible to find, the Bayesian approach can make it easier to find a good estimator.

Again, we are faced with the decision of how to measure the performance of our estimator. One natural metric is the mean square error,

$$\mathbb{E}\left[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\right], \tag{17}$$

whose minimizer is said to be the minimum mean square error (MMSE) estimate. In classical estimation theory, the MMSE estimate is often difficult or even impossible to find [28, Ch. 2.4]. To find the MMSE estimate, we rewrite (17) as

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right] &= \mathbb{E}\left[\mathbb{E}\left[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\middle|\mathbf{r}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\theta}^{\mathsf{H}}\boldsymbol{\theta}\middle|\mathbf{r}\right] - \mathbb{E}\left[\boldsymbol{\theta}^{\mathsf{H}}\middle|\mathbf{r}\right]\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{\mathsf{H}}\mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right] + \hat{\boldsymbol{\theta}}^{\mathsf{H}}\hat{\boldsymbol{\theta}}\right],\end{aligned}$$

[2]Note that we now use the conditional PDF $p(\mathbf{r}|\boldsymbol{\theta})$, representing the prior PDF of $p(\mathbf{r})$ conditioned on a given realization of the random variable $\boldsymbol{\theta}$.

where we have used the fact that $\mathbb{E}\left[\hat{\boldsymbol{\theta}}\middle|\mathbf{r}\right] = \hat{\boldsymbol{\theta}}$. Now, with

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\theta}^{\mathsf{H}}\boldsymbol{\theta}|\mathbf{r}\right] &= \operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{H}}|\mathbf{r}\right]\right)\\ &= \operatorname{tr}\left(\mathbb{E}\left[\left(\boldsymbol{\theta}-\mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right]\right)\left(\boldsymbol{\theta}-\mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right]\right)^{\mathsf{H}}\middle|\mathbf{r}\right] + \mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right]\mathbb{E}\left[\boldsymbol{\theta}^{\mathsf{H}}|\mathbf{r}\right]\right)\\ &= \operatorname{tr}\left(\operatorname{cov}\left(\boldsymbol{\theta}|\mathbf{r}\right) + \mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right]\mathbb{E}\left[\boldsymbol{\theta}^{\mathsf{H}}\middle|\mathbf{r}\right]\right)\\ &= \operatorname{tr}\left(\operatorname{cov}\left(\boldsymbol{\theta}|\mathbf{r}\right)\right) + \mathbb{E}\left[\boldsymbol{\theta}^{\mathsf{H}}\middle|\mathbf{r}\right]\mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right]\end{aligned}$$

(17) can be written as

$$\mathbb{E}\left[\left\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right\|^2\right] = \mathbb{E}\left[\operatorname{tr}\left(\operatorname{cov}\left(\boldsymbol{\theta}|\mathbf{r}\right)\right) + \left\|\mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right]-\hat{\boldsymbol{\theta}}\right\|^2\right],$$

where $\operatorname{tr}(\cdot)$ denotes the trace and $\operatorname{cov}(\cdot)$ denotes the covariance matrix. Since the first term is independent of our choice of $\hat{\boldsymbol{\theta}}$, the MMSE estimator is given by the conditional mean of the posterior PDF,

$$\hat{\boldsymbol{\theta}} = \mathbb{E}\left[\boldsymbol{\theta}|\mathbf{r}\right].$$

In order to calculate this, we need to make assumptions on the distribution of $\boldsymbol{\theta}$. These assumptions should preferably be based on our prior knowledge of $\boldsymbol{\theta}$, perhaps originating from a physical model. Throughout, we assume $\boldsymbol{\theta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\boldsymbol{\theta}})$ as this distribution is the most relevant for this thesis.

To calculate the conditional mean, we need the posterior PDF, given by Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{r}) = \frac{p(\mathbf{r}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{r})},$$

where

$$p(\mathbf{r}|\boldsymbol{\theta}) = \frac{1}{\pi^N}\exp\left(-\left\|\mathbf{r}-\mathbf{K}\boldsymbol{\theta}\right\|^2\right),$$

$$p(\boldsymbol{\theta}) = \frac{1}{\pi^p \det\left(\mathbf{C}_{\boldsymbol{\theta}}\right)}\exp\left(-\boldsymbol{\theta}^{\mathsf{H}}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\theta}\right),$$

and

$$p(\mathbf{r}) = \frac{1}{\pi^N \det(\mathbf{K}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{K}^{\mathsf{H}}+\mathbf{I}_N)}\exp\left(-\mathbf{r}^{\mathsf{H}}\left(\mathbf{K}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{K}^{\mathsf{H}}+\mathbf{I}_N\right)^{-1}\mathbf{r}\right).$$

The resulting posterior will be Gaussian. To find its mean, we focus on the exponent and note that all terms independent of $\boldsymbol{\theta}$ can be incorporated into the normalizing constant in front of the exponent. The posterior can be written as

$$\begin{aligned}p(\boldsymbol{\theta}|\mathbf{r}) &= \text{const}\cdot\exp\left(-\left\|\mathbf{r}-\mathbf{K}\boldsymbol{\theta}\right\|^2 - \boldsymbol{\theta}^{\mathsf{H}}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\theta}\right)\\ &= \text{const}\cdot\exp\left(-\boldsymbol{\theta}^{\mathsf{H}}\left(\mathbf{K}^{\mathsf{H}}\mathbf{K}+\mathbf{C}_{\boldsymbol{\theta}}^{-1}\right)\boldsymbol{\theta} + \mathbf{r}^{\mathsf{H}}\mathbf{K}\boldsymbol{\theta} + \boldsymbol{\theta}^{\mathsf{H}}\mathbf{K}\mathbf{r}\right)\\ &= \text{const}\cdot\exp\left(-\left(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{r}}\right)^{\mathsf{H}}\left(\mathbf{K}^{\mathsf{H}}\mathbf{K}+\mathbf{C}_{\boldsymbol{\theta}}^{-1}\right)\left(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{r}}\right)\right),\end{aligned}$$

where $\boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{r}} = \left(\mathbf{K}^{\mathsf{H}}\mathbf{K} + \mathbf{C}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\mathbf{K}^{\mathsf{H}}\mathbf{r}$. The posterior is symmetric around $\boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{r}}$, hence, the posterior mean and thereby the MMSE estimate is given by

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{K}^{\mathsf{H}}\mathbf{K} + \mathbf{C}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\mathbf{K}^{\mathsf{H}}\mathbf{r}. \tag{18}$$

One attractive property of the MMSE estimator is that the estimation error, $\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, and the estimate, $\hat{\boldsymbol{\theta}}$, are uncorrelated and thereby independent for the linear model. This mutual independence can significantly simplify any further analysis.

The MMSE estimator is not the only estimator to be used in Bayesian estimation. One other popular choice is the maximum a posteriori (MAP) estimator. For the MAP estimator, the estimate is given by the mode of the posterior, as opposed to its mean in the case of MMSE. When the posterior is Gaussian, as we have here, these two estimates coincide.

2.3.2 Detection Theory

The problem of detection in digital communication boils down to deciding which message that was transmitted. Suppose the transmitter sends one of N_{M} different messages, $\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\text{M}}}$. The receiver considers N_{M} hypotheses, $\mathcal{H}_1, \ldots, \mathcal{H}_{N_{\text{M}}}$, associated with the possible messages, where

$$\mathcal{H}_m : \mathbf{x}_m \text{ was sent.}$$

The receiver must now, based on the received signal $\mathbf{r}$ decide for one of these N_{M} hypotheses. Since we have prior information about the transmitted messages that we wish to incorporate into the detection process, we focus on Bayesian detection.

The detector minimizing the probability of error is the MAP *detector* [29]. The MAP detector decides $\mathcal{H}_m$ in favor of the other hypotheses if

$$P\left(\mathcal{H}_m|\mathbf{r}\right) \geq P\left(\mathcal{H}_j|\mathbf{r}\right),$$

for all $j \neq m$. This is quite intuitive, as the detector decides for the hypothesis most probable after observing $\mathbf{r}$. From Bayes' theorem we know that the posterior density can be written in terms of the likelihood, prior, and evidence as

$$P\left(\mathcal{H}_m|\mathbf{r}\right) = \frac{p\left(\mathbf{r}|\mathcal{H}_m\right)P\left(\mathcal{H}_m\right)}{p\left(\mathbf{r}\right)}.$$

Normally, the messages are constructed in such a way that they are equally likely (equal priors) since this maximizes entropy [35, Ch. 2.4]. Moreover, the evidence (the denominator) is the same for all hypotheses. In this case, an equivalent detector is given by the ML *detector*: choosing $\mathcal{H}_m$ if

$$p\left(\mathbf{r}|\mathcal{H}_m\right) \geq p\left(\mathbf{r}\middle|\mathcal{H}_j\right),$$

for all $j \neq m$.

Under hypothesis $\mathcal{H}_m$, the received signal can be written as

$$\mathbf{r} = \mathbf{K}\mathbf{x}_m + \mathbf{w},$$

so

$$p\left(\mathbf{r}|\mathcal{H}_m\right) = \pi^{-N} \exp\left(-\|\mathbf{r} - \mathbf{K}\mathbf{x}_m\|^2\right).$$

The ML detector can then be written as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}_m}{\operatorname{argmax}} \ \exp\left(-\|\mathbf{r} - \mathbf{K}\mathbf{x}_m\|^2\right) = \underset{\mathbf{x}_m}{\operatorname{argmin}} \ \|\mathbf{r} - \mathbf{K}\mathbf{x}_m\|^2.$$

In general, the ML detector needs an exhaustive search over all N_{M} messages, which might be impractical. Suboptimal detectors are often used in practice to decrease the complexity [36].

2.4 Performance Metrics

Ultimately, the goal of communicating is to transfer information from one place to another with high speed—many bits per second, and high reliability—few errors. Formally, the fundamental theoretical limits of communication is studied in the area of information theory. A wise man once told me that information theory is so complicated and requires such a meticulous attention to details that you should only talk about it when it is absolutely necessary. At any rate, I deem it necessary in the context of this thesis to at least mention the fundamental limits of communication for a few relevant channels. We will not dwell in the details here, and more rigorous explanations regarding the claims made here can be found in, e.g, [6, 7, 37–39].

Formally, *capacity* is the tight upper bound on the rate at which information can transmitted, error free, over a channel. Exactly what the capacity is depends, among other things, on the channel statistics and what the receiver/transmitter knows about the channel. The capacity can be seen a special case of $R^*(n, \epsilon)$—the maximum rate at which we can communicate for a given block length n and error probability ϵ—letting $n \to \infty$ and $\epsilon \to 0$. Similarly, the outage capacity (also called ϵ-capacity) is $R^*(n, \epsilon)$ with $n \to \infty$. In general, the expression of $R^*(n, \epsilon)$ is unknown [40, 41].

Even in the asymptotic regime where the block length tends to infinity and the error tends to zero, closed-form expressions may be difficult to obtain. As a consequence, different bounding techniques are often used, in order to bound the capacity from above or below. A lower bound on capacity is often referred to as an *achievable rate*.

2.4.1 Capacity of the AWGN Channel

Consider the received signal corrupted by real-valued additive white Gaussian noise (AWGN):

$$y = x + w,$$

where x is the transmitted symbol and $w \sim \mathcal{N}(0, \sigma^2)$ is the noise. We assume that the transmitted symbol x has zero mean, variance ρ, and is independent of the noise. Using this channel N times gives, in vector notation

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \in \mathbb{R}^N,$$

where we assume that subsequent noise samples are independent. Assume that the message $\mathbf{x}$ is drawn from a set of N_{M} equally likely messages

$$\mathbf{x} \in \mathcal{M} = \{\mathbf{x}_1, \dots, \mathbf{x}_{N_{\text{M}}}\}.$$

As the noise is Gaussian, the ML detector is

$$\hat{\mathbf{x}} = \underset{\mathbf{x}_m}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{x}_m\|^2,$$

i.e., we choose the message that is closest (in the Euclidean sense) to the received signal. If $\mathbf{x}_m$ actually was transmitted, then

$$\frac{\|\mathbf{y} - \mathbf{x}_m\|^2}{N} = \frac{\|\mathbf{w}\|^2}{N} \approx \sigma^2, \quad N \gg 1$$

by the law of large numbers. This means that when N is large, the noise starts to behave almost deterministically, it "hardens" in a sense. Let us depart slightly from the ML detector, in order to utilize the hardened noise. We choose the detector that determines that $\mathbf{x}_m$ was sent if

$$\|\mathbf{y} - \mathbf{x}_m\|^2 < N(\sigma^2 + \epsilon), \tag{19}$$

for some $\epsilon > 0$. This can be seen as having a hypersphere with radius $\sqrt{N(\sigma^2 + \epsilon)}$ surrounding the codeword $\mathbf{x}_m$. If the received signal $\mathbf{y}$ lies within this hypersphere, we decide that $\mathbf{x}_m$ was sent. If (19) is not satisfied for any message, the detection fails. Assuming that the messages in $\mathcal{M}$ are different enough that two messages cannot satisfy (19) simultaneously, erroneous detection occurs if $\mathbf{x}_m$ is sent and

$$\|\mathbf{y} - \mathbf{x}_m\|^2 > N(\sigma^2 + \epsilon),$$

which happens with probability

$$\begin{aligned}\mathbb{P}\left[\|\mathbf{y}-\mathbf{x}_m\|^2 > N(\sigma^2+\epsilon)\middle|\mathbf{x}_m \text{ sent}\right] &= \mathbb{P}\left[\|\mathbf{w}\|^2 > N(\sigma^2+\epsilon)\right] \\ &= \mathbb{P}\left[\chi^2_N > N\left(1+\frac{\epsilon}{\sigma^2}\right)\right] \\ &\approx \mathbb{P}\left[\mathcal{N}(0,1) > \frac{\sqrt{N}\epsilon}{\sqrt{2}\sigma^2}\right] \approx 0,\end{aligned} \tag{20}$$

for any $\epsilon > 0$, when N is large. In (20), χ^2_N denotes a chi-squared distributed random variable with N degrees of freedom and $\mathcal{N}(0,1)$ denotes a standard normal distributed random variable. This means that the received signal, when $\mathbf{x}_m$ was sent, lies within a hypersphere with radius $N\sigma^2$ centered around $\mathbf{x}_m$ with very high probability for large N. We call this the "message sphere"

Looking at the received signal $\mathbf{y}$ in general we note that

$$\frac{\|\mathbf{y}\|^2}{N} = \frac{\|\mathbf{x}\|^2 + \|\mathbf{w}\|^2 + 2\mathbf{x}^\mathsf{T}\mathbf{w}}{N} \approx \rho + \sigma^2,$$

again from the law of large numbers, so with a similar argument as above the received signal lies within a hypersphere with radius $N(\rho+\sigma^2)$ around the origin. We call this the "signal sphere".

In order for the transmission to be error free, the messages in $\mathcal{M}$ have to be chosen to be far enough apart so that their respective message spheres do not overlap. We are faced with a sphere-packing problem, where we want to fit as many message spheres into the signal sphere as possible; see Figure 4. Since the volume of an n-dimensional hypersphere with radius r is proportional to r^n, the total number of messages that can be sent error free over N channel uses is

$$N_{\text{M}} \approx \left(\frac{N(\rho+\sigma^2)}{N\sigma^2}\right)^{N/2},$$

giving

$$\frac{\log_2(N_{\text{M}})}{N} \approx \frac{1}{2}\log_2\left(1+\frac{\rho}{\sigma^2}\right), \tag{21}$$

bits per channel use (bpcu[3]) as the maximum rate. The right hand side of (21) is indeed the capacity of the real AWGN channel, although more rigor is needed to prove this formally.

Note that the capacity of the complex AWGN channel, is twice that of the real AWGN channel, since one complex channel use (sample) corresponds to two real ones. Moreover, it is only the signal-to-noise ratio (SNR), ρ/σ^2 that matters. Because of this, it is common to normalize the noise to have unit variance.

[3]Equivalent measures are bits per second per Hertz and bits per complex symbol.

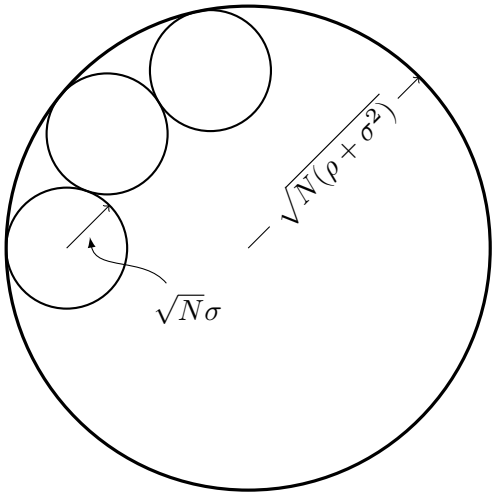


Figure 4: An illustration of the sphere packing problem. We want to fit as many non-overlapping message spheres of radius $\sqrt{N}\sigma$ into the signal sphere of radius $\sqrt{N(\rho+\sigma^2)}$ as possible.

2.4.2 The Fading Channel

In the previous section when considering the AWGN channel, the channel was constant. However, as we have seen in Section 2.2, we commonly model the channel with a random variable. Consider the following *complex* fading channel:

$$y = hx + w. \tag{22}$$

All variables in (22) are assumed to be complex, random variables with zero mean. One useful lower bound on the *ergodic capacity* that is frequently used in the massive MIMO literature is

$$\mathbb{E}\left[\log_2\left(1+\frac{\left|\mathbb{E}\left[xy^*|\Omega\right]\right|^2}{\mathbb{E}\left[|x|^2|\Omega\right]\mathbb{E}\left[|y|^2|\Omega\right]-\left|\mathbb{E}\left[xy^*|\Omega\right]\right|^2}\right)\right], \tag{23}$$

where Ω represents the receiver's knowledge about the channel. If the receiver has perfect channel knowledge, for example, $\Omega = h$. The outer expectation in (23) is taken with respect to Ω. This bound can be derived from results in [42], and can be found (in slightly different forms) in [6, Ch. 2.3.5] and [7, Cor. 1.3]. Another frequently used bound known in the literature as the "use-and-forget bound", [6,7], is given by

$$\log_2\left(1+\frac{\left|\mathbb{E}\left[xy^*\right]\right|^2}{\mathbb{E}\left[|x|^2\right]\mathbb{E}\left[|y|^2\right]-\left|\mathbb{E}\left[xy^*\right]\right|^2}\right). \tag{24}$$

The name comes from that the receiver may use the channel knowledge to preprocess the received data, but forgets it when performing the detection. We see that (24) equals (23) for the special case of a deterministic channel. Since the receiver actually disregards useful information in the use-and-forget bound, the achievable rate of (24) is lower than that of (23). The benefit of the use-and-forget bound is that it can give more tangible expressions which are easy to interpret.

2.4.3 The MIMO Fading Channel

Consider a MIMO fading channel (cf. (12)):

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{25}$$

where ρ is the transmit power, $\mathbf{H}$ is the channel, and $\mathbf{w}$ is Gaussian noise with identity covariance. The question is now how to choose the transmitted signal $\mathbf{x}$ in order to maximize the information transfer, i.e., to achieve capacity. Let

$$\mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathsf{H}}\right] \;=\; \mathbf{C}_{\mathbf{x}} \succcurlyeq\; \mathbf{0}$$

denote the covariance matrix of the transmitted symbol vector and

$$\mathrm{tr}\left(\mathbf{C}_{\mathbf{x}}\right) \leq 1,$$

in order to make the average transmit power per channel use equal to ρ. For the channel in (25), with channel state information (CSI) at the receiver and statistical CSI at the transmitter, the ergodic capacity is given by [37]

$$\max_{\mathbf{C}_{\mathbf{x}} \succcurlyeq \mathbf{0}, \mathrm{tr}(\mathbf{C}_{\mathbf{x}}) \leq 1} \mathbb{E}\left[\log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}} + \rho \mathbf{H}\mathbf{C}_{\mathbf{x}}\mathbf{H}^{\mathsf{H}}\right)\right],$$

and is achieved when $\mathbf{x}$ is a circularly-symmetric complex Gaussian random vector. If the transmitter lacks any kind of CSI, a reasonable strategy is to choose $\mathbf{C}_{\mathbf{x}} = \dfrac{1}{N_{\mathrm{T}}}\mathbf{I}_{N_{\mathrm{T}}}$, giving

$$\mathbb{E}\left[\log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}} + \frac{\rho}{N_{\mathrm{T}}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right]$$

as a lower bound. This type of bound is used in Paper C.

2.4.4 Outage Capacity

The ergodic capacity measures in Sections 2.4.2 and 2.4.3 are only valid if we consider long codewords and code over many channel realizations (coherence intervals). If this is not an option, for example if transmission only takes place during a

single coherence interval, another measure is needed. For a Rayleigh fading channel there is no rate that can be guaranteed to hold for any one channel realization; in this case, the *outage* capacity is a more meaningful measure.

Based on the channel conditions, the transmitter chooses a rate at which to transmit. If the conditions are good, a high rate with little redundancy (coding) can be used; if the conditions are poor, heavy coding, leading to a low rate, might be necessary to get the message across. If the transmitter chooses a rate that is too high, so that the channel cannot support communication at that rate, the link is said to be in outage. The outage probability of a SISO, flat fading channel, with transmit power ρ, is defined as

$$p_{\text{out}}(R) = \mathbb{P}\left[R > \log_2\left(1 + \rho|h|^2\right)\right],$$

where the effective SNR $\rho|h|^2$ will depend on the channel realization h. For any rate R, assuming Rayleigh fading, there will always be a nonzero probability that $|h|^2$ is too small to support the chosen rate. If we can tolerate an outage probability of $\epsilon > 0$, the outage capacity is defined as the largest rate R such that the outage probability is smaller than ϵ, i.e.,

$$C_\epsilon = \sup\{R : p_{\text{out}}(R) < \epsilon\}.$$

When considering short blocklengths and a fixed error probability, $R^*(n, \epsilon)$ is arguably the preferred metric. However, no closed-form expressions exist and thus we are forced to consider bounding $R^*(n, \epsilon)$ or approximating it. To this end, this thesis takes an engineering approach and considers the outage capacity, C_ϵ to be a proxy for $R^*(n, \epsilon)$, since [43]

$$R^*(n, \epsilon) = C_\epsilon + \mathcal{O}\left(\frac{\log n}{n}\right).$$

2.5 Diversity

Because of the random nature of the channel, sometimes, it might not be possible to transfer a message from the transmitter to the receiver. By using more than one channel we can improve the reliability—we increase the likelihood that at least one channel is "good enough" for us to transfer the message. This is the essence of *diversity*. Diversity can be obtained by, for example, using multiple antennas at the receiver and/or transmitter, collectively called spatial diversity, or transmitting the message in different coherence intervals, called time-frequency diversity.

The most relevant form of diversity for this thesis is spatial diversity. The spatial diversity concept is introduced in Section 2.5.1 by studying a receiver with

multiple antennas. Next, Section 2.5.2 analyzes how to achieve diversity with a single-antenna receiver, but multiple transmit antennas. Throughout we consider transmitters without CSI, since this is most relevant in the context of this thesis. To not convolute the analysis, it is assumed that the receiver has perfect CSI. A more practical setting entails estimating the channel at the receiver, as is done in Papers A and B.

2.5.1 Receive Diversity

Consider the received signal y_1 when a single-antenna transmitter transmits the signal x with unit power over the channel h_1 to a single-antenna receiver, corrupted by AWGN with unit variance:

$$y_1 = h_1 x + w_1. \tag{26}$$

Assuming that the receiver knows h_1, this is equivalent to

$$y_1' = \frac{h_1^*}{|h_1|} y_1 = |h_1| x + w_1',$$

where w_1' has the same distribution as w_1; hence, this is an AWGN channel with SNR $|h_1|^2$. The SNR will depend on the realization of the channel h_1, and if $|h_1|$ is (very) small the channel is said to be in a *deep fade*.

Now consider a receiver with $N_{\text{R}} = 2$ antennas (a SIMO channel), that collects two samples of the same signal, but with different (independent) channel and noise realizations. In addition to the received signal on antenna one, (26), the receiver obtains

$$y_2 = h_2 x + w_2.$$

Combining both of these received samples in order to maximize the SNR (maximum-ratio combining) gives

$$y = h_1^* y_1 + h_2^* y_2 = (|h_1|^2 + |h_2|^2) x + h_1^* w_1 + h_2^* w_2.$$

Since the SNR is proportional to $|h_1|^2 + |h_2|^2$, one benefit from using more antennas is clear: the effective SNR when decoding the symbol is increased. In addition, and perhaps more importantly, the SNR is now a sum of two *independent* (non-negative) random variables. As a consequence, the probability of a deep fade is significantly reduced. This increased resistance to deep fades is termed (receive) diversity.[4]

[4]Many authors define diversity (or more precisely the diversity order) more strictly, as the slope of the bit-error curve when plotted against SNR (for large SNR), in log-log scale.

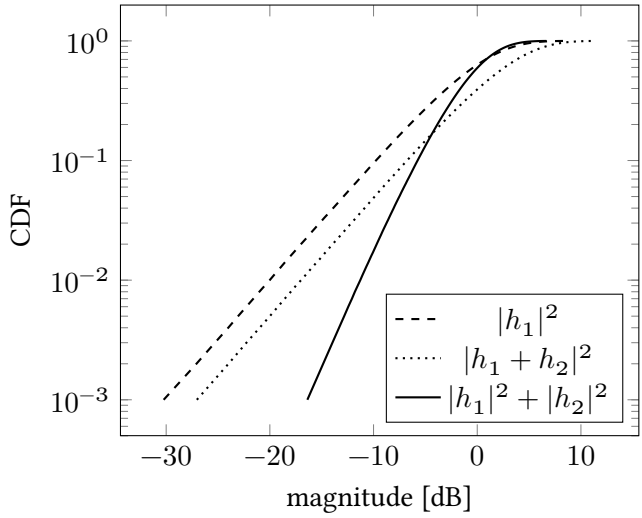


Figure 5: The effective SNR for transmission and reception with 1 or 2 antennas.

2.5.2 Transmit Diversity

When the transmitter is equipped with several antennas, $N_{\text{T}} > 1$, *transmit* diversity can be achieved. Consider a single sample at the receiver, now with $N_{\text{R}} = 1$ (a MISO channel), with the same signal transmitted from two antennas:

$$y = h_1 x + h_2 x + w = (h_1 + h_2)x + w.$$

The SNR is then proportional to

$$|h_1 + h_2|^2$$

which may be small, even if neither $|h_1|$ nor $|h_2|$ is small. We see that we cannot simply achieve transmit diversity by transmitting the same signal from the antennas.

In Figure 5 we compare the cumulative distribution function (CDF) of the SNR for the SISO, SIMO, and MISO case. We see that in the case when we have diversity, the probability of deep fades is significantly smaller than in the other two. In this example, we consider the case with $h_i \sim \mathcal{CN}(0, 1)$ for $i = 1, 2$ and h_1 and h_2 are mutually independent.

On the other hand, if we let the transmission span two channel uses, and we transmit with Antenna 1 during the first, and Antenna 2 during the second, the received signals are

$$y_1 = h_1 x + w_1$$

and

$$y_2 = h_2 x + w_2,$$

which is the same as if we would have one transmit antenna and two receiver antennas. However, in this case we spent two channel uses transmitting a single symbol, effectively cutting the rate in half compared to the example with receive diversity.

2.5.3 Space-Time Block Codes

Transmitting the same symbol from different antennas in different channel uses achieves diversity, but this technique is very inefficient—we are performing repetition coding over the antennas—leading to a low rate. The goal of a *space-time block code* (STBC) is to provide transmit diversity, without transmitter CSI, while sacrificing as little rate as possible. A STBC transmitting N_{S} symbols over N_{D} channel uses is said to have a (code) rate of $N_{\mathrm{S}}/N_{\mathrm{D}}$, where N_{D} is called the *delay* of the STBC. In the example of transmit diversity above, the rate was $1/2$ since we spent two channel uses transmitting a single symbol. The following will show that we can do better.

The Alamouti Code

Consider now a transmitter with two antennas and transmit power ρ transmitting $[x_1, x_2]^{\mathsf{T}}$ the first channel use and $[-x_2^*, x_1^*]^{\mathsf{T}}$ the second channel use. The received signals in two consecutive channel uses are then

$$y_1 = \sqrt{\rho} h_1 x_1 + \sqrt{\rho} h_2 x_2 + w_1,$$

and

$$y_2 = -\sqrt{\rho} h_1 x_2^* + \sqrt{\rho} h_2 x_1^* + w_2,$$

respectively, which in matrix notation can be written as

$$[y_1, y_2] = \sqrt{\rho}[h_1, h_2] \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix} + [w_1, w_2].$$

This can in turn be written as

$$\underbrace{\begin{bmatrix} y_1 \\ y_2^* \end{bmatrix}}_{\mathbf{y}} = \sqrt{\rho} \underbrace{\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2^* \end{bmatrix}.$$

The receiver, having full knowledge of $\mathbf{H}$, can now form

$$\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix} = \frac{1}{\|\mathbf{h}\|}\mathbf{H}^{\mathsf{H}}\mathbf{y},$$

where $\mathbf{h} = [h_1, h_2]^{\mathsf{T}}$, resulting in

$$\tilde{y}_1 = \sqrt{\rho}\|\mathbf{h}\|x_1 + \tilde{w}_1$$

and

$$\tilde{y}_2 = \sqrt{\rho}\|\mathbf{h}\|x_2 + \tilde{w}_2.$$

The two noise terms $\tilde{w}_1$ and $\tilde{w}_2$ have the same distribution as w_1 and w_2, respectively. Each of the received samples $\tilde{y}_1$ and $\tilde{y}_2$ has the same SNR, $\rho\|\mathbf{h}\|^2$, and we achieve diversity of order two since

$$\|\mathbf{h}\|^2 = |h_1|^2 + |h_2|^2.$$

What is more interesting, we transmit two symbols over two channel uses, so we achieve transmit diversity without losing any rate—a quite remarkable feat!

This transmit strategy for two transmit antennas was presented by Alamouti in the late 1990s [44]. Inspired by Alamouti's fantastic transmit strategy—providing full diversity, rate, and simple decoding—the theory of orthogonal STBCs (OSTBCs), generalizing Alamouti's code, developed with the use of orthogonal designs [45].

As the theory of OSTBCs developed, it became apparent that, in many ways the Alamouti code is not generalizable. For example, Liang showed in [46] that there are no OSTBCs with rate 1 for any number of transmit antennas larger than two. In fact, he showed that the maximum rate decreases with increasing number of transmit antennas, and that the rate approaches $1/2$ as N_{T} grows large. The delay, N_{D}, of a OSTBC is also important as this determines the number of channel uses the receiver have to wait before decoding the received message. A lower bound on the minimum delay for maximum-rate OSTBCs was derived in [47], where it was shown that the minimum delay grows, quite fast, with increasing number of antennas; see Table 2 for the maximum rate and minimum delay of a few OSTBCs.

So, sadly, there is no simple way to determine what STBC to use, without considering the specific scenario. This becomes a trade-off between diversity on the one hand, and rate and delay on the other.

Table 2: Summary of selected maximum-rate OSTBCs for up to ten antennas.

N_T	$N_\mathrm{S}/N_\mathrm{D}$	N_D
2	1	2
4	3/4	4
6	2/3	30
8	5/8	56
10	3/5	420

Chapter 3

Massive MIMO

In this chapter we introduce some fundamental concepts of massive MIMO, and cellular systems in general. In order to keep the introduction concise, we overlook some aspects that are important when considering massive MIMO as a whole; see Section 3.5 for more details. For more detailed descriptions of massive MIMO, references [6] and [7] are recommended.

3.1 Cellular Transmission

The strength of a signal decays rapidly the farther it travels, thus it is impossible for a single fixed location to provide service to all who need it. In order for the signal to be detectable at the receiver, a cellular topology was proposed [48] to make the average distance between transmitter and receiver smaller. The idea is that the area in need of service is divided into different *cells*, where each cell has a *base station* (BS), providing the terminals with service such as data for streaming music and video.

The communication in a cellular network can happen in two ways: either the BS transmits data to the terminal, called the downlink; or the terminal transmits data to the BS, called the uplink. In order for these transmissions not to interfere with each other, the uplink and downlink are placed on different time-frequency resources (they occupy different parts of the time-frequency grid). In time-division duplex (TDD), the uplink and downlink take turns transmitting on the same frequency band, while in frequency-division duplex (FDD), they transmit simultaneously but in different frequency bands. This is illustrated in Figure 6.

In TDD, the channel is *reciprocal*—the same in both directions—since the uplink and downlink use the same frequency band. This is a very useful property, as the channel only has to be estimated in one direction. If the two transceivers use dif-

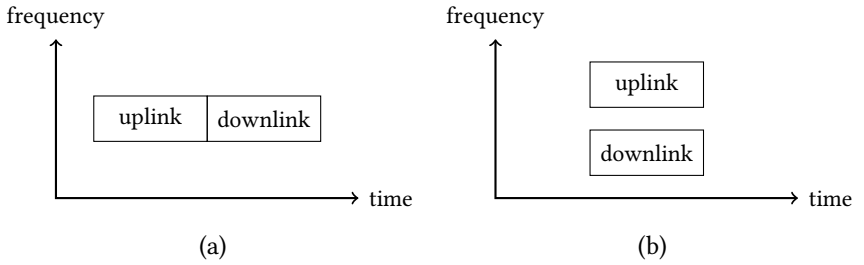


Figure 6: An illustration of the difference between TDD and FDD: (a) in TDD, the uplink and downlink share the same frequency band but take turns transmitting; (b) in FDD the uplink and downlink transmit simultaneously but occupy different frequency bands.

ferent frequency bands, like in FDD, the electromagnetic waves transmitted in the uplink and downlink will behave differently since, e.g., absorption and diffraction depend on the carrier frequency. It should be noted that even though the wireless channel is reciprocal, the hardware in the transceivers might not be. In order for the channel as a whole to be reciprocal, the hardware may need calibration [49].

One of the main goals of any cellular system is to increase the total throughput of the system. Historically, this has been done by allocating more bandwidth, occupying a larger part of the time-frequency grid, or through densification, placing BSs closer and closer in order to decrease the average distance to a terminal. This has worked quite well, but these methods of increasing the throughput are saturating: the frequency spectrum is a finite resource which has to be shared, and moving BSs closer together also increases interference.

Instead of increasing the total throughput by adding more BSs or using more bandwidth, massive MIMO aims to increase the *spectral efficiency*, measured in bpcu.

3.2 Why Massive MIMO?

Some of the most important benefits of massive MIMO can be seen by studying a small example. We consider a system consisting of two single-antenna terminals and one BS with M antennas. The BS is assumed to have perfect CSI. The two terminals transmit the complex symbols x_1 and x_2, respectively, with the transmit power ρ simultaneously, in the same frequency band. The described system can be modeled as a MIMO channel (12) and the received signal at the BS is

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w} = \sqrt{\rho}\mathbf{h}_1 x_1 + \sqrt{\rho}\mathbf{h}_2 x_2 + \mathbf{w}.$$

The channel vectors $\mathbf{h}_1, \mathbf{h}_2$ are independent with $\mathbf{h}_1, \mathbf{h}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ represents noise.

The BS now wishes to detect the symbol transmitted by Terminal 1, x_1. Since the BS knows $\mathbf{h}_1$, it may form

$$\frac{\mathbf{h}_1^{\mathsf{H}}\mathbf{y}}{M} = \frac{1}{M}\left(\sqrt{\rho}\mathbf{h}_1^{\mathsf{H}}\mathbf{h}_1 x_1 + \sqrt{\rho}\mathbf{h}_1^{\mathsf{H}}\mathbf{h}_2 x_2 + \mathbf{h}_1^{\mathsf{H}}\mathbf{w}\right). \tag{27}$$

When M is large, due to the law of large numbers, the scaled inner products

$$\sqrt{\rho}\frac{\mathbf{h}_1^{\mathsf{H}}\mathbf{h}_1}{M}, \tag{28}$$

$$\sqrt{\rho}\frac{\mathbf{h}_1^{\mathsf{H}}\mathbf{h}_2}{M}, \tag{29}$$

and

$$\frac{\mathbf{h}_1^{\mathsf{H}}\mathbf{w}}{M}, \tag{30}$$

are well approximated by their expected values; hence

$$\frac{\mathbf{h}_1^{\mathsf{H}}}{M}\mathbf{y} \approx \sqrt{\rho}x_1.$$

The term in front of the symbol of interest, (28), can be seen as the effective channel. For large M, this effective channel behaves almost deterministically—a phenomenon known as *channel hardening*. The term in front of the undesired symbol, (29), vanishes because the normalized inner product of the two channels approaches zero for large M—a phenomenon known as *asymptotic favorable propagation*. Lastly, we see that the term associated with the additive noise, (30), also vanishes since the effective noise is the average of many independent noise samples. All together, channel hardening, asymptotic favorable propagation, and noise averaging yields an effective channel that is almost deterministic and almost free from noise. To be more precise, an achievable rate from the use-and-forget bound (24) for Terminal 1 is given by

$$\log_2\left(1 + \frac{\rho M}{\rho + 1}\right).$$

The number of antennas shows up in the numerator; we say that we get an *array gain* of M.

For detection of the other symbol, x_2, an analogous argument can be made, giving an achievable rate of

$$\log_2\left(1 + \frac{\rho M}{\rho + 1}\right),$$

identical to the achievable rate for the first terminal. Note that these two rates can be achieved simultaneously, thus, the total spectral efficiency is

$$2\log_2\left(1+\frac{\rho M}{\rho+1}\right).$$

We call the factor in front of the logarithm the *multiplexing gain*, and it stems from that two users are served in the same time-frequency resource. Serving more users would increase the multiplexing gain but also increase the interference.

The array gain and the multiplexing gain are two major benefits of massive MIMO that make it possible to serve multiple users in the same time-frequency resource. The array gain helps the individual user by increasing the effective SNR—something particularly useful for users far from the BS—while the multiplexing gain increases the overall throughput.

Note that this is an overly simplified scenario to illustrate the main benefits of having many antennas. The two following sections will discuss massive MIMO operation in more detail.

3.3 Uplink Transmission

We consider a single cell, viewed in isolation, consisting of a BS with M antennas and K single-antenna users. We show how the base station acquires CSI and how this is used to perform efficient transmission and reception.

3.3.1 Channel Estimation

To facilitate channel estimation at the BS, the terminals transmit known reference signals, called pilot or training sequences. Each terminal in the cell is given a unique pilot sequence consisting of τ_{P} symbols. Having all K users transmit their pilots with transmit power ρ makes the received signal at the BS

$$\mathbf{Y}=\sqrt{\rho}\mathbf{H}\mathbf{X}_{\text{P}}+\mathbf{W}$$

where $\mathbf{H}\in\mathbb{C}^{M\times K}$, $\mathbf{X}_{\text{P}}\in\mathbb{C}^{K\times\tau_{\text{P}}}$, and $\mathbf{W}\in\mathbb{C}^{M\times\tau_{\text{P}}}$ are the channel from the terminals to the BS, the transmitted pilot matrix, and the additive noise, respectively. The channel matrix comprises K channel vectors:

$$\mathbf{H}=[\mathbf{h}_1,\ldots,\mathbf{h}_K],$$

where each column represents the channel between the respective user and the BS. To model the different positions of the terminals, resulting in different large-scale fading, we have $\mathbf{h}_k\sim\mathcal{CN}(\mathbf{0},\beta_k\mathbf{I}_M)$. Note that this implies that the large-scale

fading is the same, irrespective of which BS antenna we consider. It is assumed that the pilot matrix $\mathbf{X}_{\text{p}}$ is orthogonal and satisfies

$$\mathbf{X}_{\text{p}}\mathbf{X}_{\text{p}}^{\text{H}} = \tau_{\text{p}}\mathbf{I}_K,$$

implying that $\tau_{\text{p}} \geq K$. The noise has i.i.d. elements distributed as $\mathcal{CN}(0,1)$.

Since the pilot matrix is known at the BS, it can form

$$\frac{\mathbf{Y}\mathbf{X}_{\text{p}}^{\text{H}}}{\sqrt{\tau_{\text{p}}}} = \bar{\mathbf{Y}} = \sqrt{\rho\tau_{\text{p}}}\mathbf{H} + \bar{\mathbf{W}},$$

where the elements of $\bar{\mathbf{W}}$ are i.i.d. and $\mathcal{CN}(0,1)$. This operation is sometimes referred to as *de-spreading* [6, Ch. 3.1.2]. In order to estimate the channel to user k the BS considers the kth column of $\bar{\mathbf{Y}}$:

$$\bar{\mathbf{y}}_k = \sqrt{\rho\tau_{\text{p}}}\mathbf{h}_k + \bar{\mathbf{w}}_k,$$

which has the form of the linear model (13); hence, the MMSE estimate of the channel is given by (18):

$$\hat{\mathbf{h}}_k = \frac{\sqrt{\rho\tau_{\text{p}}}\beta_k}{1+\rho\tau_{\text{p}}\beta_k}\bar{\mathbf{y}}_k.$$

The variance of each element of $\hat{\mathbf{h}}_k$ is

$$\gamma_k = \frac{\mathbb{E}\left[\|\hat{\mathbf{h}}_k\|^2\right]}{M} = \frac{\rho\tau_{\text{p}}\beta_k^2}{1+\rho\tau_{\text{p}}\beta_k}.$$

3.3.2 Data Transmission

When the BS has acquired CSI through the channel estimation, the terminals can transmit data. The received signal at the BS in one channel use can be written as

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w}.$$

Let us separate the channel $\mathbf{H}$ into two different parts: the channel estimate

$$\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K]$$

and the estimation error

$$\mathbf{E} = \mathbf{H} - \hat{\mathbf{H}} = [\mathbf{e}_1, \ldots, \mathbf{e}_K]$$

The received signal can then be written as

$$\mathbf{y} = \sqrt{\rho}\hat{\mathbf{H}}\mathbf{x} + \sqrt{\rho}\mathbf{E}\mathbf{x} + \mathbf{w}.$$

To detect the symbol transmitted by user k, x_k, the BS projects the received signal $\mathbf{y}$ onto the (receive) beamforming[1] vector $\mathbf{v}_k$, resulting in

$$\mathbf{v}_k^{\mathsf{H}}\mathbf{y} = \sqrt{\rho}\mathbf{v}_k^{\mathsf{H}}\hat{\mathbf{h}}_k x_k + \sqrt{\rho}\sum_{k'\neq k}\mathbf{v}_k^{\mathsf{H}}\hat{\mathbf{h}}_{k'}x_{k'} + \sqrt{\rho}\sum_{k'=1}^{K}\mathbf{v}_k^{\mathsf{H}}\mathbf{e}_{k'}x_{k'} + \mathbf{v}_k^{\mathsf{H}}\mathbf{w}.$$

This can be seen as a symbol x_k passing through a known, fading channel $\sqrt{\rho}\mathbf{v}_k^{\mathsf{H}}\hat{\mathbf{h}}_k$ corrupted by noise. From Section 2.4.2 we know that an achievable ergodic rate is given by [7, Eq. 4.3]

$$\mathbb{E}\left[\log_2\left(1+\frac{\rho\left|\mathbf{v}_k^{\mathsf{H}}\hat{\mathbf{h}}_k\right|^2}{\rho\sum\limits_{k'\neq k}\left|\mathbf{v}_k^{\mathsf{H}}\hat{\mathbf{h}}_{k'}\right|^2+\rho\|\mathbf{v}_k\|^2\sum\limits_{k'=1}^{K}(\beta_{k'}-\gamma_{k'})+\|\mathbf{v}_k\|^2}\right)\right], \quad (31)$$

where the expectation is taken over the receiver channel knowledge, i.e. the channel estimates.

We can get a more palpable bound by setting $\mathbf{v}_k = \hat{\mathbf{h}}_k$ (maximum-ratio combining) and using the use-and-forget bound, where the decoder conveniently forgets the channel estimates [6, Eq. 3.41]:

$$\log_2\left(1+\frac{M\rho\gamma_k}{1+\rho\sum\limits_{k'=1}^{K}\beta_{k'}}\right). \quad (32)$$

By studying (32) there are a few things worth noting: the effective SNR scales with M, the number of BS antennas. Thus, in theory we could achieve any rate we desire by adding more antennas. The rate increase is rather slow though, since there is only a logarithmic increase with M. In practice, physical constraints, for example, the size of the antenna array further limits the number of antennas we can use. Also worth noting is that (32) is the achievable rate for a single terminal. The many antennas, in addition to the array gain apparent in the SNR, brings a multiplexing gain since K users can be served in the same time-frequency resource. From a system perspective, the multiplexing gain can significantly increase the spectral efficiency. Moreover, from a analytical and computational point of view, (32) is really amazing because the achievable rate no longer depends on the small-scale fading!

[1] Some authors use the term combining vector.

3.4 Downlink Transmission

In the downlink, the BS uses the channel estimates obtained from the uplink training to transmit K different symbols to the K different terminals in the same channel use. The collectively received signal at the K terminals can be written as[2]

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}^{\mathsf{T}}\mathbf{x} + \mathbf{w},$$

where $\mathbf{x} \in \mathbb{C}^M$ is the is the transmitted signal from the base station and $\mathbf{w} \in \mathbb{C}^K$ is uncorrelated noise.

In order to spatially separate the transmitted signal to user k, the BS uses the (transmit) beamforming vector[3] $\mathbf{v}_k$. The transmitted vector can be written as

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{v}_k s_k,$$

where s_k is the symbol intended for terminal k. The symbols are assumed to be uncorrelated. To ensure that the BS obeys the power constraint, we let $\mathbb{E}\left[\|\mathbf{v}_k\|^2\right] = 1$ and $\mathbb{E}\left[|s_k|^2\right] = 1/K$ for all $k = 1, \ldots, K$.

The received symbol at user k can be written as

$$y_k = \sqrt{\rho}\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k s_k + \sqrt{\rho}\sum_{k' \neq k} \mathbf{h}_k^{\mathsf{T}}\mathbf{v}_{k'} s_{k'} + w_k, \tag{33}$$

which can be seen as the desired signal s_k passing through the fading channel $\sqrt{\rho}\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k$, corrupted by noise. Now, unlike the uplink where the receiver (the BS), is assumed to have estimated the channel, no such assumption is made for the downlink.[4] Instead, the user approximates the instantaneous channel $\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k$ with its mean value. Rewriting (33), we see that

$$y_k = \sqrt{\rho}\mathbb{E}\left[\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k\right] s_k + \sqrt{\rho}\sum_{k' \neq k} \mathbf{h}_k^{\mathsf{T}}\mathbf{v}_{k'} s_{k'} + \sqrt{\rho}\left(\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k - \mathbb{E}\left[\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k\right]\right) s_k + w_k,$$

which can be interpreted as the desired signal s_k passing through a known, deterministic channel $\sqrt{\rho}\mathbb{E}\left[\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k\right]$, corrupted by noise. A lower bound on capacity,

[2] With an abuse of notation, we will denote the received signal, the transmit power, the transmitted signal, the noise, and the beamforming vector with the same variables as in the previous section.

[3] Some authors call this the precoding vector.

[4] There is actually not much to gain from using downlink pilots in this scenario, if the BS has enough antennas [50].

obtained from using the use-and-forget bound, is given by [7, Eq. 4.6]

$$\log_2\left(1+\frac{\frac{\rho}{K}\left|\mathbb{E}\left[\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k\right]\right|^2}{\frac{\rho}{K}\sum\limits_{k'=1}^{K}\mathbb{E}\left[\left|\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_{k'}\right|^2\right]-\frac{\rho}{K}\left|\mathbb{E}\left[\mathbf{h}_k^{\mathsf{T}}\mathbf{v}_k\right]\right|^2+1}\right). \tag{34}$$

Evaluating the bound for maximum-ratio transmission,

$$\mathbf{v}_k=\frac{\hat{\mathbf{h}}_k^*}{\sqrt{\mathbb{E}\left[\|\hat{\mathbf{h}}_k\|^2\right]}}$$

gives [6, Eq. 3.63]

$$\log_2\left(1+\frac{M}{K}\frac{\rho\gamma_k}{\rho\beta_k+1}\right). \tag{35}$$

The bounds for uplink and downlink, (35) and (32) are strikingly similar, but two differences are worth noting. First, in order to serve the K users, the BS has to split the total transmit power among them, this brings the array gain from M in the uplink to M/K in the downlink, that is, if the BS uses the same amount of power as each terminal. Second, all interference experienced by the user in the downlink passes through the same channel, thus only the large-scale fading coefficient β_k, associated with the user's own channel affects the effective SNR.

3.5 Caveats

As mentioned in the beginning of this chapter, a comprehensive introduction to massive MIMO is out of scope for this thesis, so we instead focused on a small example to convey the basics of massive MIMO. In doing so, we have overlooked some important details, which we will comment on here.

Everything described in Sections 3.3 and 3.4 takes place within a single coherence interval. That is, channel estimation (pilots), uplink data transmission, and downlink data transmission need to share the τ_c channel uses available in the coherence interval. As a consequence, the more channel uses spent on pilots, the fewer can be spent on transmitting data.[5] This is illustrated in Figure 7. To take the number of pilots into account, the spectral efficiency expression in (31), (32),

[5]There are strategies that transmit pilots and data in the same channel use, known as superimposed pilots [51, 52]. In this thesis we focus on the classical approach where the data and pilot transmissions are separated.

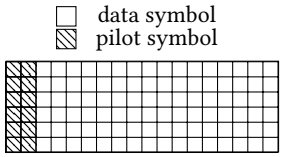


Figure 7: To enable coherent communication (beamforming), part of the coherence interval is used for pilots. If τ_{p} channel uses are used for pilots in a coherence interval of τ_{c} channel uses, only $\tau_{\mathrm{c}} - \tau_{\mathrm{p}}$ can be used for data.

(34), and (35) should be multiplied with the fraction of the coherence interval that is used for data, for example

$$\left(1 - \frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{c}}}\right) \log_2 \left(1 + \frac{M}{K} \frac{\rho \gamma_k}{\rho \beta_k + 1}\right) \tag{36}$$

in lieu of (35). Moreover, (36) implicitly assumes that all data symbols are allocated to the downlink. If uplink and downlink share the data symbols, an additional factor—representing the fraction of data symbols allocated to downlink transmission—is needed in (36) to get the net spectral efficiency [6, Ch. 3.6].

So far we studied a single, isolated cell, while in reality a cell may be surrounded by other cells. Let us call the cell under consideration the "home cell". If the surrounding cells use the same time-frequency resource as the home cell, the transmitted signals in the surrounding cells will cause interference, both during uplink and downlink transmission.

We mentioned earlier in Section 3.3.1 that in order to have orthogonal pilots within a cell with K users, we need to spend $\tau_{\mathrm{p}} \geq K$ channel uses on pilots. Preferably, to avoid interference during the channel estimation, we would in a multi-cell scenario like all cells to have mutually orthogonal pilots. If the number of cells is N_{c} and each cell serves K users, the total number of channel uses used for training would then have to be at least KN_{c}. Recalling that the coherence interval consists of τ_{c} channel uses, having orthogonal pilots across the cells is infeasible if $KN_{\mathrm{c}} > \tau_{\mathrm{c}}$. When having cells with non-orthogonal pilots, the de-spreading of the pilots does not fully separate all users' channels. As a consequence, channel estimates in the home cell will be contaminated with user channels from other cells. This phenomenon is called *pilot contamination* and can severely impact the system performance [3, 53].

To decrease the impact of pilot contamination, *pilot reuse* can be used. This means that all cells are divided into disjoint groups, where cells in one group share the same pilots, but the pilots across different groups are mutually orthogonal. This comes at a cost of spectral efficiency, as the pilot overhead is increased. Other,

more sophisticated strategies for how to avoid pilot contamination, known as *pilot decontamination*, have also been investigated [54–57]. Recently, it has been argued that pilot contamination is not a fundamental issue, but rather a consequence of the assumptions commonly made to increase analytical tractability [58].

We only treat suboptimal, linear processing. Nonlinear processing, such as successive interference cancellation could be used, but when $M \gg K$, the benefit of any nonlinear processing vanishes, as linear processing is almost optimal [6, Ch. 3.5], [7, Ch. 1.3]. Even though the bounds (31) and (34) are general in the sense that they work for any linear processing, we only explicitly calculated the bound for the simplest case, when maximum-ratio beamforming is used. There are, however, many more sophisticated beamforming alternatives, such as zero-forcing or MMSE, discussed in detail in [7].

To derive the bound in (35) it was assumed that the BS allocates the same amount of power to all users, which is not optimal in general. Power control—or how to distribute the power among the users—can greatly impact the system performance both in the uplink and in the downlink [59–61]. One popular method of power control in massive MIMO is the so-called max-min fairness power control, in which the power is allocated in such a way to make the effective SNR the same for all users [6, 7, 62]. This ensures that all users get the same performance—no one gets left behind.

An implicit assumption in the models used in this thesis is that the hardware is perfectly linear. However, massive MIMO requires many antennas and transceiver chains, so to be commercially viable, cheap, low-end hardware will need to be used [8]. Non-ideal components such as low-end power amplifiers and low-bit analogue-to-digital converters cause distortions. Some types of distortion combine non-coherently and average out [18, 63] while other types combine coherently and prevail as the number of antennas grows large [64].

3.6 Cell-Free Massive MIMO

So far, we have focused on "conventional" massive MIMO, in which the service area is divided into cells, and the users in each cell are served by a single BS with many (collocated) antennas. In contrast, cell-free massive MIMO aims to serve all users in an area with many smaller, simple BSs with only one or a few antennas, hereafter called access points (APs). The APs are distributed geographically over the area, but are connected to a central processing unit (CPU) which in turn is connected to the core network; see Figure 8. As the name suggests, the service area is not dividend into cells, instead all APs within the area serve the users in a coordinated fashion. Paper B considers APs distributed in this way.

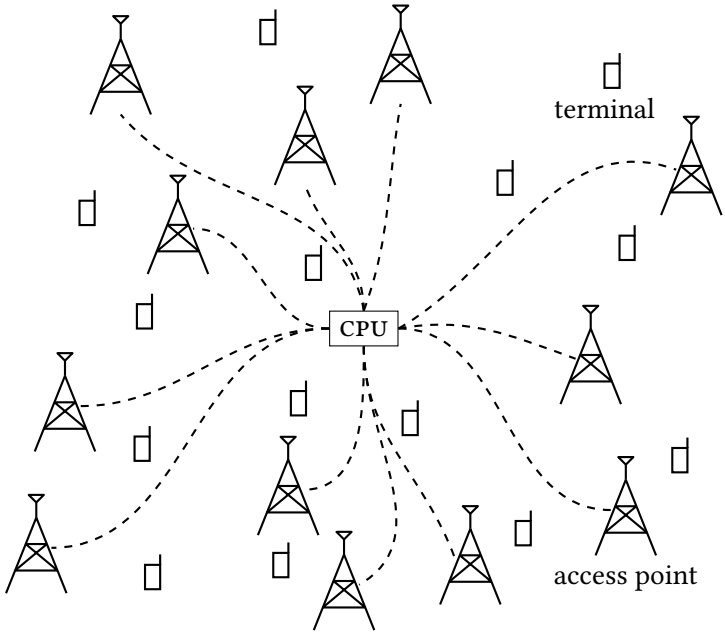


Figure 8: A cell-free massive MIMO system where all the access points are connected to a central processing unit (CPU). All access points serve all terminals, in the same time-frequency resource.

Although the cell-free concept is rather new [65–69], the idea of distributing APs and having them serve the users jointly to enhance the spectral efficiency is not [70, 71]. What is new in cell-free massive MIMO is the signal processing: the principles of massive MIMO carry over and make distributed channel estimation and data transmission possible. The massive MIMO framework enables the APs to, for example, beamform without sharing the CSI among the APs, which is an attractive feature since sharing all information in real time may not be practically feasible [72].

Distributing the antennas has the advantage of increasing the macro diversity, in the sense that the probability of having an AP close to a terminal increases. This in turn increases the average large-scale fading and SNR experienced at a terminal. A side effect of distributing the antennas geographically is that the amount of channel hardening decreases as the macro diversity increases [73]. As a result, the expected value of the downlink channel used by the terminal might be a poor estimate of the instantaneous channel; hence, downlink pilots might be more useful in cell-free massive MIMO than in conventional massive MIMO [74].

3.7 System Information

Up until this point, when discussing uplink and downlink transmission we have focused on so-called closed-loop operation, where the BS performs beamforming based on CSI acquired from uplink pilots. However, some downlink signals need to be transmitted without CSI at the BS, before a communication link has been initiated. This information includes, in particular, public broadcast channels that contain *system information* such as cell identifiers and instructions on how to perform random access. In LTE[6], the most important parts of the system information are located in the so-called master information block (MIB) [22]. Every user needs to decode this system information in order to connect to and function within the network. The challenge with the transmission of this information is that without CSI at the BS, coherent beamforming is impossible and no array gain can be harnessed. The beamforming particularly improves cell-edge coverage since the effective signal-to-interference-and-noise ratio (SINR) scales with the number of BS antennas, disregarding coherent interference (which can be kept small in practice through appropriate pilot reuse) [6]. Figure 9 illustrates the coverage with CSI (and closed-loop beamforming), and the coverage without CSI.

Broadcasting of public channels that contain system information is critical in any wireless system. Transmission of this information with good coverage is a challenge because no CSI is available at the BS so no link adaptation can be performed. In massive MIMO, the difficulties are amplified by the facts that i) closed-loop beamforming is impossible due to the unavailability of CSI at the BS; ii) space-time coding must be used to achieve sufficiently "omnidirectional" transmission; and iii) decoding of these codes requires CSI at the users which in turn necessitates the transmission of downlink pilots. In this section, methods for limiting the downlink pilot overhead, ensuring omnidirectional transmission, and exploiting spatial diversity are presented. Moreover, new possible solutions for system information broadcast, made possible by the abundance of antennas, are also discussed. The majority of the discussion is based on broadcasting in LTE.

3.7.1 Initial Access

The first thing a device does when it is powered on is to perform a cell search to find and synchronize to a cell within the network. The device needs to synchronize both in frequency and on a symbol level to get the correct frame structure of the network. This is necessary as there will always be static information on fixed time-frequency resources conveying the most basic settings of the network: the

[6]Long-Term Evolution, a common standard for mobile communication in contemporary networks.

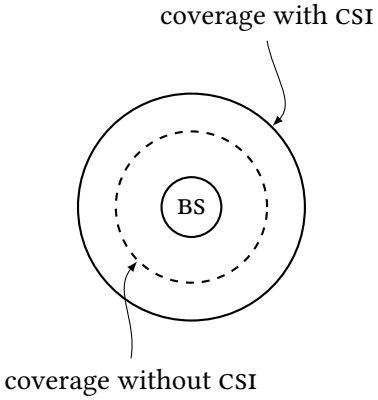


Figure 9: The area of coverage, where the effective SINR is large enough to facilitate communication, is limited by the non-coherent transmission.

system information.

After the initial cell search and successfully decoding the system information the terminal enters an idle state. In the idle state, where the terminal spends most of its time, it periodically scans for incoming transmissions according to a predetermined pattern; this enables the terminal to sleep for extended periods to conserve battery and save power. When the terminal wishes to transmit data, or if the network has data for the terminal, radio resources are allocated to the terminal and it enters the connected state. Collectively, we call a terminal in either idle or connected state *active*, meaning it has successfully decoded the system information; see Figure 10.[7] It should be noted that an active terminal also performs cell search to facilitate cell re-selection and handover [22, Ch. 14].

3.7.2 Distribution of System Information

The system information must be available to any user wanting to connect to the network. To this end, the BS needs to broadcast the system information periodically on predefined time-frequency resources, without any CSI to the users. This makes parts of the system information inevitably static, in the sense that it is the same no matter if any user is listening or not. To limit this static allocation of resources, some of the system information in NR is therefore envisioned to be transmitted on dynamically allocated time-frequency resources. Specifically, since the coverage without CSI is less than the coverage with CSI, the system information

[7] The names idle and connected are taken from LTE and corresponds to the states "RRC_IDLE" and "RRC_CONNECTED", respectively.

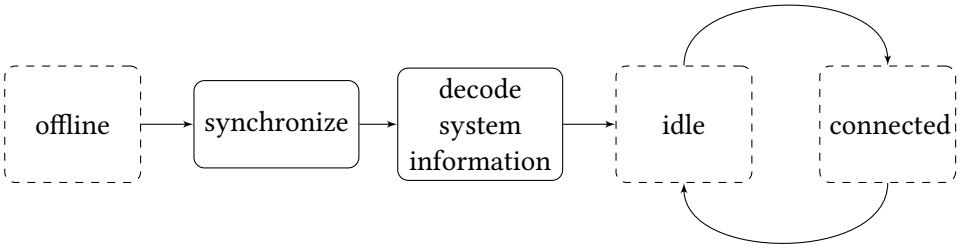


Figure 10: An illustration of initial access—how a terminal goes from being inactive (offline) to being active (idle/connected). The bridge from inactive to active is successful decoding of the system information.

may be restructured such that only a bare minimum of it is transmitted without CSI, and the rest is sent on demand through closed-loop beamforming. The main goal of this reconstruction for NR is to have as few bits as possible broadcast on statically allocated resources.

Exactly how to segment the system information into a static part and an on-demand part, and how to transmit these segments in NR is not obvious. One approach, detailed in [75], divides the system information into two parts. The first part is transmitted with maximum coverage and the subsequent part is transmitted with lower coverage, to save resources. The users that are not able to decode the second part can then request a retransmission by sending a reference signal (which was enabled through decoding of the first part). Multicast techniques that send the same information to several users in the same time-frequency resource [76] may then be used for the part that is transmitted on demand. These techniques exploit the pilot contamination effect that arises if more than one user uses the same pilot. The resulting channel estimate formed at the BS is then a linear combination of the channels to these users and a multicast beamforming effect automatically appears, such that the array gain is divided among the users.

3.7.3 Omnidirectional Transmission

When the BS has no information of the users' positions or channels, the system information should be transmitted omnidirectionally into the cell. For example, in an unsectorized system, the BS should transmit in all directions with equal strength, such that the received signal power only depends on the channel's magnitude; see Figure 11. Put differently, the user's position within the cell is not important—only the large-scale fading of the channel between the BS and the user matters.

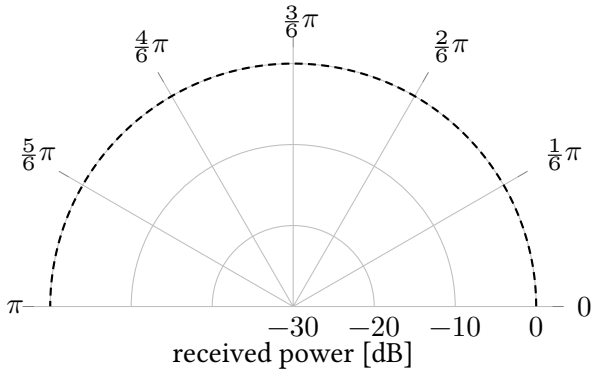


Figure 11: The radiation pattern for (instantaneous) omnidirectional transmission—all angles have the same radiated power. This is the preferred way to transmit system information when the user position is unknown.

Omnidirectional transmission would be automatically achieved if the BS transmitted through a single antenna that had an isotropic gain pattern. The BS could for instance have a dedicated system information antenna, possibly operating in a different frequency band. A special case would be to use a legacy system, such as LTE, to send the system information. This is a short-sighted solution, since we might want to turn off the legacy system eventually. A closely related solution would be to turn off all but one antenna in the array. This would, however, require a more expensive RF chain, as the power of the transmitted signal would be much larger during the system information broadcasting than in the data transmission. Consequently, it would be beneficial to use the entire array and not rely on special equipment for the transmission of system information [77].

When transmitting from an array, the transmitted signal in one channel use will unavoidably be beamformed: some angular directions will receive more power than others. Random beamforming, where the BS randomly selects the beamforming vector in each channel use, will on average result in omnidirectional transmission but suffers from large power variations over the antennas and between channel uses. However, several more sophisticated methods for omnidirectional transmission using a large array exist:

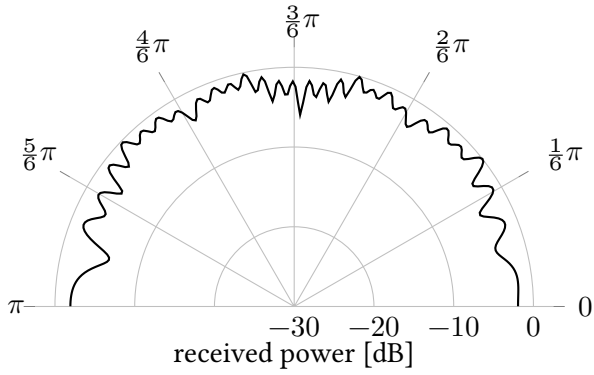
1. In [23], the authors formulate an approximation of omnidirectional transmission that works well for uniform linear arrays (ULAs) when the number of antennas, M, is large. First, the omnidirectional transmission requirement is relaxed by considering the received signal power in M predetermined discrete angular directions. Second, the requirement of *instantaneous* omnidirectional transmission is relaxed by considering the *average* of the

received signal powers. The (average) radiation pattern is illustrated in Figure 12a. The authors show that in addition to an equal average received signal power in the M discrete angular directions, equal average output power on all antennas is also achieved.

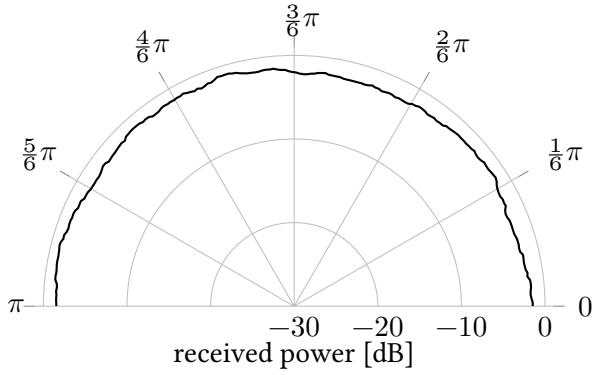
2. In [78], the authors relax the requirement of instantaneous omnidirectional transmission by considering the average of the received signal powers. The authors further allow for the signal power to vary slightly between different angles. This is illustrated in Figure 12b. Both ULAs and uniform rectangular arrays (URAs) are considered, and the solution is obtained by solving an optimization problem. For the URA, both azimuth and elevation angles are taken into account. In addition, the transmitted signal is optimized to minimize its dynamic range over the antennas. Because the dynamic range over the antennas is minimized, the power amplifiers in the RF chains are used more efficiently. This is illustrated by the, on average, higher gain in Figure 12b compared to Figure 12a.

3. Reference [79] relaxes the requirement of instantaneous omnidirectional transmission only slightly and shows that omnidirectional transmission on the average is possible with a ULA, when considering a few consecutive channel uses. This is illustrated in Figure 12c. In particular, the authors show that the received signal power over just two channel uses can be made constant for any angle. Moreover, the transmitted signal power is constant over the antennas (0 dB peak-to-average-ratio) in discrete symbol-sampled time, which makes this solution utilize the power amplifiers more efficiently (allows for reduced back-off) than the solutions in [23] and [78].

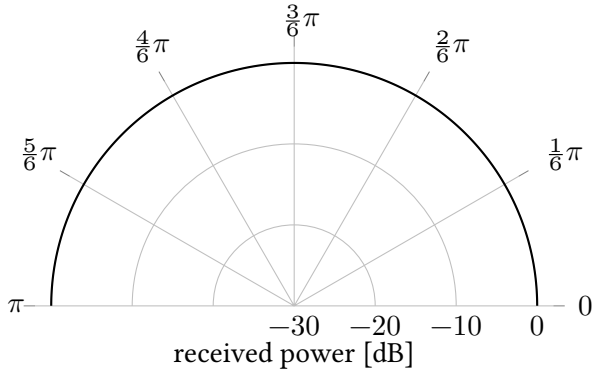
4. Another way of effectively achieving omnidirectional transmission is to use beam sweeping, where the BS beamforms in different directions in each channel use, hoping to eventually "hit" the user with a beam. Since the same information is transmitted several times, but in different beams, beam sweeping can be seen as spatial repetition coding. When using beam sweeping, there is a trade-off between latency and utilized array gain. The BS could form M orthogonal beams, each one having the array gain of M. Doing this would require M channel uses to search the entire cell, and there is no guarantee that any of the beams hits perfectly. The BS could also form more than M non-orthogonal beams to increase the chance that one beam hits the terminal, at the cost of increased time overhead (reduced net spectral efficiency). If the BS is able to use prior knowledge to limit the search space, beam sweeping might be useful, as noted in [80]. However, if no information is available at the BS, beam sweeping is best suited for applications with



(a)



(b)



(c)

Figure 12: Conceptual comparison of the different radiation patterns. As opposed to Figure 11, which shows instantaneous radiation, these plots show *average* radiation: a) approximate omnidirectional transmission on average; b) omnidirectional transmission on average, with small fluctuations; c) omnidirectional transmission on average.

lax latency constraints.

As a complement to the omnidirectional techniques listed above, it is also possible to transmit the system information and the payload on the same time-frequency resources, and separate them spatially [14]. Here, the system information is transmitted in a subspace orthogonal to that spanned by the payload-user channels—some beams are used for payload, and the unused beams are used for system information. This way the system information can benefit from maximum frequency diversity at the cost of increased interference, since it is no longer confined to separate time-frequency resources.

3.7.4 Transmit Diversity and Pilot Overhead

Increasing coverage is not only about making the signals reach farther, but also improving reliability (decreasing the probability of outage). In order to improve reliability, wireless systems can make use of diversity in time, frequency, or space. Depending on the situation, one or more diversity methods might be unavailable since time diversity requires laxer latency constraints and channel time variations; frequency diversity requires a wideband, frequency-selective channel; and spatial diversity requires multiple antennas at the receiver and/or the transmitter.

In the following, we distinguish between *antenna ports*, which are logical entities, and *antenna elements*, which are physical entities; see Figure 13. In the downlink, the number of antenna ports is defined by the number of reference signals (pilots) transmitted from the antenna array. Signals that are transmitted through one antenna port will experience a scalar channel to the receiver. The number of antenna ports and the mapping from the antenna ports to the antenna elements can be changed quickly in a fully digital system, while the number of antenna elements is fixed. Massive MIMO with M antenna elements is equipped with M RF chains, and thus supports up to M antenna ports. In contrast, a BS in LTE typically has an order-of-magnitude more antenna elements than RF chains and it is the latter that determines the maximum number of antenna ports. The many antenna elements, which are individually phase shifted, allows the LTE BS to beamform even with relatively few RF chains.

Spatial diversity will be important in NR, for applications that operate over channels with small bandwidths, most notably machine-to-machine communications and Internet of Things. In addition, the packets containing system information might also shrink because of the possible segmentation discussed in Section 3.7.2. The short packets and increased reliability constraints call for increased diversity [81], and with massive MIMO, ample spatial diversity is available.

To achieve spatial diversity without CSI at the BS and with a single-antenna user, a STBC can be used, as mentioned in Section 2.5.2. Specifically, OSTBCs

Figure 13: The number of antenna ports can be thought of as the number of streams transmitted from the array: (a) if the same signal is transmitted from all antenna elements we have one antenna port; (b) if two signals are mapped to the antenna elements and transmitted, we have two antenna ports.

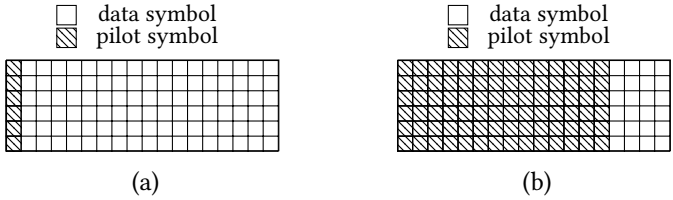


Figure 14: An illustration of finite coherence interval and the pilot overhead problem: (a) a handful of pilots are used, leaving adequate room left for data; (b) many antenna ports require many pilots, leaving little room for data.

achieve full diversity, meaning a diversity order equal to the number of antenna ports. OSTBCs also ensure omnidirectional transmission of the symbols, on the average, over any channel: the received signal power at the user depends only on the magnitude of the channel, not its direction. To see this, let $\mathbf{X} \in \mathbb{C}^{\tau \times m}$ be an OSTBC spanning m antenna ports and τ channel uses. Since $\mathbf{X}$ is an OSTBC, it satisfies $\mathbf{X}^{\mathsf{H}}\mathbf{X} = \alpha \mathbf{I}_m$, where α is a positive constant. With $\mathbf{h}$ denoting the channel from the BS antenna ports to the user, the received power at the user can be written as

$$\|\mathbf{X}\mathbf{h}\|^2 = \mathbf{h}^{\mathsf{H}}\mathbf{X}^{\mathsf{H}}\mathbf{X}\mathbf{h} = \alpha \mathbf{h}^{\mathsf{H}}\mathbf{h} = \alpha \|\mathbf{h}\|^2,$$

which does not depend on the channel direction.

In order for the user to decode the OSTBC, downlink pilots must be used. This is very different from closed-loop massive MIMO, which only requires uplink pilots. The BS has M antenna elements, and thus a potential diversity order of M. Since the minimum required overhead is proportional to the number of antenna ports and the channel is only static within a coherence interval, a large number of antenna ports implies significant pilot overhead, as depicted in Figure 14. There is a fundamental trade-off between diversity and pilot overhead [23, 82].

One way of reducing the pilot overhead, without turning off antennas, is to use fewer than M antenna ports, say m, for broadcasting in massive MIMO. In

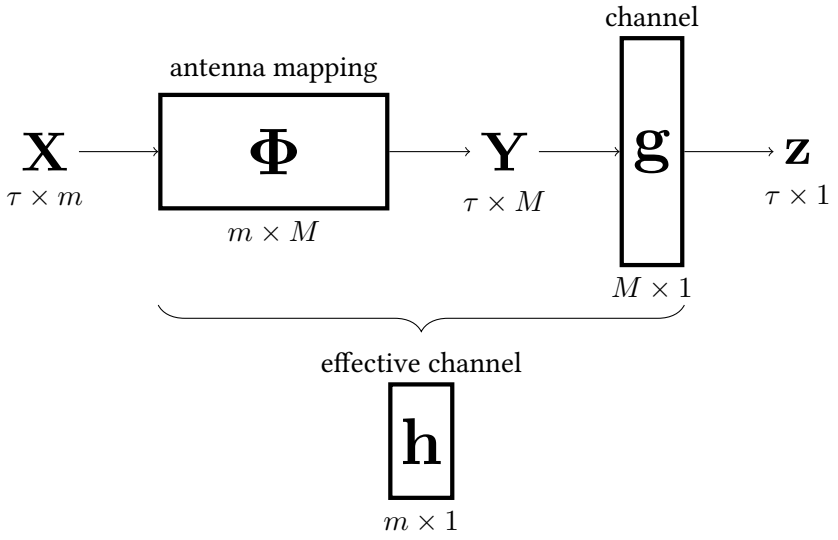


Figure 15: An illustration of how the antenna mapping can shrink the effective channel from the BS to the user. The dimension of the effective channel $\mathbf{h} = \mathbf{\Phi g}$ depends only on the number of antenna ports, m, not on the number of antenna elements M. This mapping allows for less pilot overhead, at a cost of lower potential spatial diversity. In this figure, $\mathbf{X}$ is the output of the antenna ports (for example an STBC), $\mathbf{Y}$ is the output of the antenna elements, $\mathbf{z}$ is the output of the (MISO) channel, and τ is the decoding delay—the length of the STBC.

other words, we can reduce the dimension of the *effective* channel between the BS and the user [23, 79], as shown in Figure 15. This can be done by choosing the mapping from the antenna ports to the antenna elements appropriately, and keep it fixed within each coherence interval. The combined effect of this mapping and the physical channel to the user is contained in the effective channel. Since the number of antenna ports can be changed easily in a fully digital system, so can the mapping, and the effective channel can be made smaller and thus easier for the user to estimate. This is discussed in more detail in Paper A.

The cost of decreasing the dimension of the effective channel is that parts of the (potential) spatial diversity is lost: the maximum possible diversity order is equal to the number of antenna ports, not the number of antenna elements. However, increasing the diversity beyond a certain point suffers from diminishing returns; hence, the decrease in pilot overhead will be more beneficial than the extra diversity branches when the BS has many antennas.

The trade-off between diversity and channel estimation overhead in the downlink is conceptually illustrated in Figure 16 for a coherence interval of 200 samples.

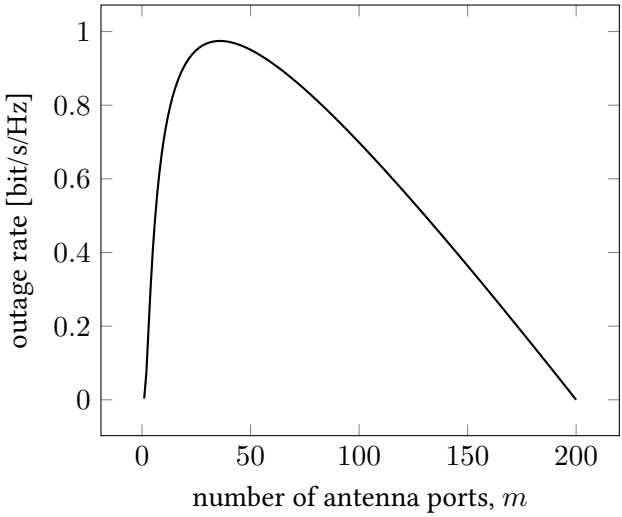


Figure 16: A conceptual illustration of the outage rate for a coherence interval of 200 channel uses for different number of antenna ports. For few antenna ports (effective channels of small dimension), the extra diversity is helpful. When many antenna ports (effective channels of large dimension) are used, the extra diversity is not enough to compensate for the additional pilot overhead.

At first, increasing the number of antenna ports (diversity order) pays off, but at a certain point, the cost of extra pilots outweighs the benefit of the extra diversity. Using all available spatial diversity may severely limit the number of channel uses that can be used to transmit actual data. In particular, if 200 or more antenna ports are used, there is no room left for data at all.

Diversity can be very valuable, but has to be balanced with the increase in pilot overhead. A massive MIMO BS can in principle, on a per coherence interval basis, choose what diversity order to use and choose a STBC accordingly, in order to transmit with the optimal diversity order. This is studied in detail in Paper A.

3.8 Physical Layer Security

Wireless communication is inherently vulnerable to attacks because the medium (air) is open to everyone. Anyone with the proper equipment can transmit signals to interfere with another transmission, or measure signals to try to decode and access information. *Physical layer security* is a collection techniques used to increase security in wireless systems as a complement to cryptographic techniques [83]. It leverages the unpredictability of the wireless channels, such as fading and noise,

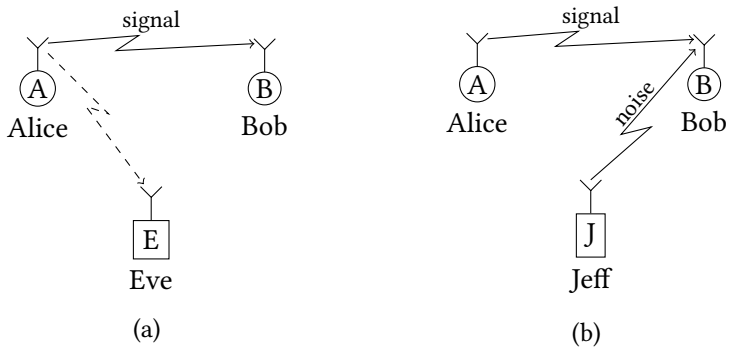


Figure 17: An illustration of the two main issues in physical layer security: (a) part of the signal intended for Bob is simultaneously received at Eve; (b) the reception at Bob is disrupted by noise transmitted by Jeff.

to ensure secure and reliable communication [24].

In physical layer security, one commonly considers three parties: a transmitter-receiver pair—called Alice and Bob—and an adversary. The communication link between Alice and Bob is termed the *legitimate* link. The adversary may behave differently depending on its agenda and we classify it as an *eavesdropper* (and call it Eve) if the goal is to extract information sent over the legitimate link, and as a *jammer* (and call it Jeff) if it aims to disrupt the legitimate communication between Alice and Bob.

Eavesdroppers do not aim to destroy or disturb the legitimate communication between Alice and Bob, but want to know what information is transmitted. There are two main categories of eavesdroppers: passive and active. A passive eavesdropper only listens, and does not transmit anything; it relies on simply being close enough to the transmitter to pick up parts of the legitimate signal, illustrated in Figure 17a. By being silent, the passive eavesdropper does not arouse any unnecessary suspicion. An active eavesdropper is allowed to transmit signals to aid its mission and is often more potent [84], but easier to detect, than its passive counterpart.

A jammer does not care about the information content of the transmitted signal, but is dedicated to make sure that information can not be conveyed between the transmitter and the receiver of the legitimate link. In contrast to eavesdroppers, there are no passive jammers (a jammer that does not transmit anything is useless). Jeff, a malicious transmitter, deliberately tries to deny the receiver from decoding the message. Jeff transmits noise in order to confuse the receiver (Bob) and make the noise level so high that the legitimate signal cannot be decoded; see Figure 17b.

If the legitimate channel between Alice and Bob is known, and orthogonal to the channel between Alice and Eve, secure transmission is easy: if Alice transmits using maximum ratio beamforming towards Bob, Eve is unable to detect anything. If the channel to Eve is known, Alice could transmit in the nullspace of the channel (zero forcing) and completely null the signal towards Eve or even transmit artificial noise in her direction [24, 83, 85]. However, the channel to the eavesdropper, especially a passive one, is difficult to obtain in practice. Also, what is to say that the two channels would be orthogonal? The law of large numbers, that's what! If Alice has a large number of antennas, the law of large number tells us that the two channel vectors (assuming that asymptotic favorable propagation holds), actually are (almost) orthogonal. As a consequence, massive MIMO is almost completely immune to passive eavesdroppers [25, 86]. A similar argument also shows that massive MIMO is very resilient to uplink jamming in the data phase.

The benefit of the beamforming with many antennas can also work against the legitimate link, if Eve is "smart". If, for example, Eve has access to the pilot sequence and frame structure of the legitimate link, she can transmit the same pilot sequence concurrently as Bob. In the channel estimation phase, when Alice aims to estimate Bob's channel, she instead estimates a linear combination of the channels to Bob and Eve. We have seen this phenomenon before—pilot contamination (c.f. Section 3.5). This can cause a severe degradation of the performance, as Alice is, unknowingly, beamforming towards Eve [84, 87, 88]. Because massive MIMO is very susceptible towards this kind of pilot-contamination attack, efforts have been made to find efficient ways of detecting if an active eavesdropper/jammer is present and how to mitigate the attack [25, 86, 89, 90].

In Paper C the jamming problem is discussed in more detail and it is shown that Jeff does not need much a priori information about the legitimate link to be able to disrupt Alice and Bob. In particular, there is no assumption on knowledge of any of the signals sent between Alice and Bob, or any of the channels. This is in stark contrast to many other "smart" jammers, which either need to leverage the knowledge of the legitimate signals to estimate the channel to the target, or is given partial (or full) knowledge of the channel a priori. In Paper C, the channel is estimated[8] blindly from Jeff's received signal and its covariance matrix is analyzed to find when Alice and Bob transmits, respectively.

[8]To be precise, the channel is estimated up to a complex multiplicative factor ambiguity, similar to what was done in [91]. This ambiguity is unimportant to Jeff since he does not need to decode the received symbols.

Chapter 4

Future Work

At this point the fundamentals of massive MIMO are well understood, preliminary sites have deployed and tested, and the phase of convincing skeptics that it actually does work in practice is over. Now it is time to develop the technology even further. Just as previous standards, such as LTE, are continuously evolving, so will massive MIMO. Lately, the focus of research is switching to more practical challenges: how do deal with non-ideal hardware and how to implement efficient algorithms that work in real time.

Cell-free massive MIMO, as mentioned in the text, is not as well developed as conventional massive MIMO, and still needs time to grow. Even though many of the principles are the same as in conventional massive MIMO, it will be a great technical challenge to enable coordination and synchronization between the access points in real time. Skeptics argue that a distributed, cooperating system such as cell-free massive MIMO is not practically viable. This is not entirely unlike the discussion regarding the feasibility of massive MIMO a handful of years ago, and now it is one of the corner stones in next generations cellular networks. It is probably wise to not make any definitive statements about topics not yet fully understood. What may seem infeasible today might not be a few years down the road.

For system information, it is clear that some kind of diversity is needed to improve reliability and that omnidirectional transmission is preferable if the user location is unknown to the base station; however, exactly how to best achieve diversity and omnidirectional transmission is not. One interesting, more practical aspect, would be to consider transmission with non-ideal hardware. For example, to investigate the effect of non-isotropic antennas on the omnidirectional transmission or the effect of coarse quantization on the outage rates. It would be interesting to find and compare alternatives to the space-time block codes.

Bibliography

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

[2] C. X. Wang, F. Haider, X. Gao, X. H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

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

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] A. Nordrum, "5G Researchers Set New World Record For Spectrum Efficiency," http://spectrum.ieee.org/tech-talk/telecom/wireless/5g-researchers-achieve-new-spectrum-efficiency-record, Dec. 2016.

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

[7] 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.

[8] C. Mollén, "High-End Performance with Low-End Hardware : Analysis of Massive MIMO Base Station Transceivers," *DIVA*, 2017.

[9] T. L. Marzetta, "Massive MIMO: An Introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

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

[11] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multiuser MIMO for energy efficiency: When is massive MIMO the answer?" in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2014, pp. 242–247.

[12] S. K. Mohammed and E. G. Larsson, "Per-Antenna Constant Envelope Precoding for Large Multi-User MIMO Systems," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

[13] J. Hoydis, K. Hosseini, S. T. Brink, and M. Debbah, "Making smart use of excess antennas: Massive MIMO, small cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 5–21, Sep. 2013.

[14] E. G. Larsson and H. V. Poor, "Joint beamforming and broadcasting in massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 3058–3070, Apr. 2016.

[15] E. Björnson, E. de Carvalho, E. G. Larsson, and P. Popovski, "Random access protocol for massive MIMO: Strongest-user collision resolution (SUCR)," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[16] C. Mollen, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink Performance of Wideband Massive MIMO with One-Bit ADCs," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2016.

[17] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[18] 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, 2014.

[19] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 53–64.

[20] 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 *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 287–293.

[21] 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 *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–2.

[22] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*, 1st ed. Academic Press, 2011.

[23] X. Meng, X. Gao, and X. G. Xia, "Omnidirectional precoding based transmission in massive MIMO systems," *IEEE Transactions on Communications*, vol. 64, no. 1, pp. 174–186, Jan. 2016.

[24] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 20–27, Apr. 2015.

[25] Y. O. Basciftci, C. E. Koksal, and A. Ashikhmin, "Securing massive MIMO at the physical layer," in *2015 IEEE Conference on Communications and Network Security (CNS)*, Sep. 2015, pp. 272–280.

[26] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press, 2005.

[27] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, 1st ed. Englewood Cliffs, N.J: Prentice Hall, Apr. 1993.

[29] S. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, 1st ed. Englewood Cliffs, N.J: Prentice Hall, Feb. 1998.

[30] H. L. Van Trees and K. L. Bell, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. New Jersey: John Wiley & Sons, Inc., 2013.

[31] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[32] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[33] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[34] D. Slepian, "On bandwidth," *Proceedings of the IEEE*, vol. 64, no. 3, pp. 292–300, Mar. 1976.

[35] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. New York, NY, USA: Cambridge University Press, 2002.

[36] E. G. Larsson, "MIMO Detection Methods: How They Work," *IEEE signal processing magazine (Print)*, vol. 26, no. 3, pp. 91–95, 2009.

[37] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[38] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[39] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[40] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[41] W. Yang, "Fading Channels: Capacity and Channel Coding Rate in the Finite-Blocklength Regime," Ph.D. dissertation, Chalmers University of Technology, 2015.

[42] M. Medard, "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.

[43] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

[44] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[45] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[46] X.-B. Liang, "Orthogonal designs with maximal rates," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2468–2503, Oct. 2003.

[47] S. S. Adams, N. Karst, and J. Pollack, "The minimum decoding delay of maximum rate complex orthogonal space time block codes," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2677–2684, Aug. 2007.

[48] V. H. M. Donald, "Advanced mobile phone service: The cellular concept," *The Bell System Technical Journal*, vol. 58, no. 1, pp. 15–41, Jan. 1979.

[49] 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.

[50] H. Q. Ngo and E. G. Larsson, "No Downlink Pilots are Needed in TDD Massive MIMO," *arXiv:1606.02348 [cs, math]*, Jun. 2016.

[51] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed Pilots Are Superior for Mitigating Pilot Contamination in Massive MIMO," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2917–2932, Jun. 2017.

[52] D. Verenzuela, "Analysis of Alternative Massive MIMO Designs : Superimposed Pilots and Mixed-ADCs," *DIVA*, 2018.

[53] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A Comprehensive Survey of Pilot Contamination in Massive MIMO #x2014;5G System," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 905–923, Secondquarter 2016.

[54] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[55] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[56] 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.

[57] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot Reuse for Massive MIMO Transmission over Spatially Correlated Rayleigh Fading Channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[58] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Has Unlimited Capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[59] Q. Zhang, S. Jin, M. McKay, D. Morales-Jimenez, and H. Zhu, "Power Allocation Schemes for Multicell Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 5941–5955, Nov. 2015.

[60] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-Efficient Power Control: A Look at 5G Wireless Technologies," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1668–1683, Apr. 2016.

[61] H. V. Cheng, "Optimizing Massive MIMO : Precoder Design and Power Allocation," *DIVA*, 2018.

[62] H. Yang and T. L. Marzetta, "Massive MIMO With Max-Min Power Control in Line-of-Sight Propagation Environment," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4685–4693, Nov. 2017.

[63] A. Pitarokoilis, "Phase Noise and Wideband Transmission in Massive MIMO," *DIVA*, 2016.

[64] C. Mollén, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Impact of Spatial Filtering on Distortion from Low-Noise Amplifiers in Massive MIMO Base Stations," *IEEE Transactions on Communications*, pp. 1–1, 2018.

[65] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-free massive MIMO systems," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, Nov. 2015, pp. 695–699.

[66] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and Power Optimization in Cell-Free Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[67] 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.

[68] H. Q. Ngo, L. N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the Total Energy Efficiency of Cell-Free Massive MIMO," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[69] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous Cell-Free Massive MIMO Communications," *arXiv:1804.03421 [cs, math]*, Apr. 2018.

[70] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *IEEE VTS 53rd Vehicular Technology Conference*, Rhodes, 2001, pp. 1745–1749.

[71] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: A new architecture for future public wireless access," *IEEE Communications Magazine*, vol. 41, no. 3, pp. 108–113, Mar. 2003.

[72] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at Interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[73] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Transactions on Communications*, pp. 1–1, 2018.

[74] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, "How Much Do Downlink Pilots Improve Cell-Free Massive MIMO?" in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2016, pp. 1–7.

[75] S. Sörman, "System information distribution in massive MIMO systems," Master Thesis, 2016, available on diva-portal.org.

[76] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[77] Ericsson, "On forming wide beams," Ericsson, Spokane, WA, USA, Tech. Rep. R1-1700772, Jan. 2017.

[78] D. Qiao, H. Qian, and G. Y. Li, "Broadbeam for massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2365–2374, May 2016.

[79] X. G. Xia and X. Gao, "A space-time code design for omnidirectional transmission in massive MIMO systems," *IEEE Wireless Communications Letters*, vol. PP, no. 99, pp. 1–1, 2016.

[80] C. Shepard, A. Javed, and L. Zhong, "Control channel design for many-antenna MU-MIMO," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: ACM, 2015, pp. 578–591.

[81] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[82] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," in *2012 IEEE Information Theory Workshop*, Sep. 2012, pp. 572–576.

[83] X. Chen, D. W. K. Ng, W. H. Gerstacker, and H. H. Chen, "A Survey on Multiple-Antenna Techniques for Physical Layer Security," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 1027–1053, Secondquarter 2017.

[84] X. Zhou, B. Maham, and A. Hjorungnes, "Pilot Contamination for Active Eavesdropping," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 903–907, Mar. 2012.

[85] J. Zhu, R. Schober, and V. K. Bhargava, "Linear Precoding of Data and Artificial Noise in Secure Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2245–2261, Mar. 2016.

[86] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 21–27, Jun. 2015.

[87] S. Tomasin, I. Land, and F. Gabry, "Pilot Contamination Attack Detection by Key-Confirmation in Secure MIMO Systems," *arXiv:1609.02197 [cs, math]*, Sep. 2016.

[88] H. Pirzadeh, S. M. Razavizadeh, and E. Björnson, "Subverting massive MIMO by smart jamming," *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 20–23, Feb. 2016.

[89] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure Massive MIMO Transmission With an Active Eavesdropper," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 3880–3900, Jul. 2016.

[90] T. T. Do, H. Q. Ngo, T. Q. Duong, T. J. Oechtering, and M. Skoglund, "Massive MIMO Pilot Retransmission Strategies for Robustification Against Jamming," *IEEE Wireless Communications Letters*, vol. 6, no. 1, pp. 58–61, Feb. 2017.

[91] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 3249–3252.

Included 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-149898

Other Recently Published Theses From

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

Daniel Verenzuela, *Analysis of Alternative Massive MIMO Designs: Superimpsed Pilots and Mixed-ADCs*, Linköping Studies in Science and Technology. Licentiate Thesis, No. 1803, 2018.

Victor Hei Cheng, *Optimizing Massive MIMO: Precoder Design and Power Allocation*, Linköping Studies in Science and Technology. Licentiate Thesis, No. 1929, 2018.

Christopher Mollén, *High-End Performance with Low-End Hardware: Analysis of Massive MIMO Base Station Transceivers*, Linköping Studies in Science and Technology. Dissertations, No. 1896, 2017.

Trịnh Văn Chiến, *Resource Allocation for Max-Min Fairness in Multi-Cell Massive MIMO*, Linköping Studies in Science and Technology. Licentiate Thesis, No. 1797, 2018.

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

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.