Coordinated Precoding for Multicell MIMO Networks

RASMUS BRANDT

Licentiate Thesis in Electrical Engineering
Stockholm, Sweden 2014

KTH Royal Institute of Technology
School of Electrical Engineering
Department of Signal Processing
SE-100 44 Stockholm
SWEDEN

TRITA-EE 2014:023
ISSN 1653-5146
ISBN 978-91-7595-142-3

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie licentiatexamen i elektro- och systemteknik tisdagen den 3 juni 2014 klockan 10.15 i hörsal Q2, Kungliga Tekniska högskolan, Osquldas väg 10, Stockholm.

Tryck: Universitetsservice US AB

Abstract

Enabling multiple base stations to utilize the spatial dimension in a coordinated manner has been shown to be a fruitful technique for improving the spectral efficiency in wireless interference networks. This thesis considers multicell systems where base stations and mobile stations are equipped with multiple antennas. The base stations coordinate their spatial precoding, but individually serve their mobile stations with data. For such coordinated precoding systems, interference alignment (IA) is a useful theoretical tool, due to its ability to serve the maximum number of interference-free data streams. Three topics related to interference alignment and coordinated precoding are studied.

First, the feasibility of IA over a joint space-frequency signal space is studied. A necessary condition for space-frequency IA feasibility is derived, and the possible gain over space-only IA is analyzed. An upper bound on the degree of freedom gain is shown to increase in the number of subcarriers, but decrease in the number of antennas. Numerical studies, using synthetically generated channels and real-world channels obtained from indoors and outdoors channel measurements, are used for sum rate performance evaluation. The results show that although a degree of freedom gain is noticeable due to the space-frequency precoding, the sum rate of the system is mainly improved due to a power gain.

Second, distributed channel state information (CSI) acquisition techniques are proposed, which provide estimates of the information necessary to perform distributed coordinated precoding. The methods are based on pilot-assisted channel estimation in the uplink and downlink, and correspond to different tradeoffs between feedback and signaling, backhaul use, and computational complexity. Naïvely applying the existing WMMSE algorithm for distributed coordinated precoding together with the estimated CSI however results in poor performance. A robustification of the algorithm is therefore proposed, relying on the well known diagonal loading technique. An inherent property of the WMMSE solutions is derived and, when enforced onto solutions with imperfect CSI, results in diagonally loaded receive filters. Numerical simulations show the effectiveness of the proposed robustification. Further, the proposed robustified and distributed WMMSE algorithm performs well compared to existing state-of-the-art robust WMMSE algorithms. In contrast to our approach, the existing methods however rely on centralized CSI acquisition.

Third, coordinated precoding systems with hardware impairments are studied. Assuming that impairment compensation techniques have been applied, a model is used to describe the aggregate effect of the residual hardware impairments. An iterative resource allocation method accounting for the residual hardware impairments is derived, based on an existing resource allocation framework. Numerical simulations show that the proposed method outperforms all benchmarks. In particular, the gain over impairments-aware time-division multiple access is substantial.

Sammanfattning

Genom att låta flera radiobasstationer samarbeta är det möjligt att förbättra spektraleffektiviteten i trådlösa interferensnätverk. Fokus i denna licentiatavhandling ligger på multicellnätverk där både radiobasstationer och mobilenheter har flera antenner. Radiobasstationerna väljer sina spatiella förkodare gemensamt, men skickar data individuellt till sina respektive mobilenheter. För sådana system med koordinerad förkodning ('coordinated precoding') är interferensupprätning ('interference alignment') ett användbart teoretiskt verktyg, eftersom det möjliggör överföring av maximalt antal dataströmmar i nätverket. I avhandlingen studeras tre aspekter av interferensupprätning och koordinerad förkodning.

Först undersöks interferensupprätning när signalrummet består av en kombination av rymd- och frekvensdimensioner. Ett nödvändigt villkor härleds för existensen av rymd/frekvens-interferensupprätning, och prestandavinsten analyseras i jämförelse med system där enbart rymddimensionerna används för interferensupprätning. Det föreslagna systemet utvärderas med hjälp av numeriska simuleringar och uppmätta inomhus- och utomhuskanaler. Resultaten visar att rymd/frekvens-interferensupprätning ger upphov till ett ökat antal frihetsgrader, men att summadatatakten främst förbättras tack vare en upplevd effektförstärkning.

Därefter undersöks tekniker för skattning av den nödvändiga kanalkännedom som krävs för att genomföra koordinerad förkodning. Det finns flera sätt att erhålla den nödvändiga informationen, t.ex. genom olika kombinationer av kanalskattning, feedback, signalering och användning av backhaulnätverk. Speciellt söks distribuerade metoder, eftersom dessa är fördelaktiga vid praktisk implementering. Tre metoder för skattning av kanalkännedom föreslås. Dessa motsvarar olika avvägningar mellan kanalskattning och signalering, och en av metoderna är helt distribuerad. När den skattade informationen används med en existerande algoritm för koordinerad förkodning blir prestandan undermålig. Därför föreslås två förändringar av algoritmen, vilka leder till mer robusta prestanda. Förändringarna bygger på den välkända *diagonal loading*-tekniken. Utvärdering av det föreslagna systemet, som består av distribuerad erhållning av kanalkännedom samt den förbättrade algoritmen för koordinerad förkodning, genomförs med numerisk simulering. Resulterande prestanda är i nivå med ett tidigare föreslaget system, som dock kräver centraliserad tillgång till kanalskattningarna, till skillnad från vår nya lösning.

Slutligen studeras ett system med koordinerad förkodning och icke-perfekt radiohårdvara. En modell för distortionsbruset orsakad av bristerna i radiohårdvaran används, och en iterativ resurstilldelningsteknik föreslås baserad på ett existerande ramverk. Den föreslagna algoritmen kan implementeras distribuerat över mobilenheterna, men kan i allmänhet inte implementeras distribuerat över radiobasstationerna. Den föreslagna algoritmen utvärderas med numeriska simuleringar, och resultaten visar att prestanda är bättre än alla referensmetoder. Detta visar betydelsen av att hantera bristerna i radiohårdvaran i resurstilldelningen.

Sammantaget visar avhandlingen på möjligheterna att öka spektraleffektiviteten i framtida multicellnätverk med hjälp av koordinerad förkodning.

Acknowledgements

I owe my sincerest gratitude to my advisor Assoc. Prof. Mats Bengtsson for his never ceasing guidance and support in my journey as a Ph.D student. Always available for discussion, Mats has a profound way of giving direction when I think I have hit a dead end. His wide array of knowledge often amazes me, be it in the fields of mathematics, typesetting in LaTeX or debugging segfaults on the simulation computers. I would also like to thank my co-advisor Assoc. Prof. Joakim Jaldén for general advice, and for always asking the most insightful questions at the internal seminars. I would like to thank Dr. Per Zetterberg, for all the discussions about the practical issues of interference alignment.

My co-authors Henrik Asplund, Dr. Per Zetterberg, and Asst. Prof. Emil Björnson deserve a great deal of recognition for good collaboration and discussion. Hadi Ghauch, Ehsan Olfat, Hamed Farhadi, Nima Najari Moghadam, Farshad Naghibi all helped proofread the thesis, providing great feedback, for which I am very grateful. Henrik Asplund and Ericsson Research should be thanked for providing the channel measurements used in Chapter 3. Further, the European Commission FP7 FET project HIATUS is acknowledged for financial support. Tove Schwartz should be thanked for handling all the administrative issues. I would also like to thank Prof. Mikael Skoglund for acting as the quality reviewer of the thesis.

I would like to thank Dr. David Astély for taking the time to be the opponent of this thesis.

Everybody at plan 4 deserves thanks for the positive environment. In particular, I would like to thank my room mate Klas Magnusson for listening to my computer-related rants, Hadi Ghauch for the discussions on interference alignment, Farshad Naghibi for the Iranian dances, Arash Owrang for the workouts, Martin Sundin for the lunch discussions, and Dr. Satyam Dwivedi for sharing of great stories from India.

Finally, I would like to thank my brother Oskar and Johanna for all the fun, and my parents Eva and Ingvar for their unrelenting support. The obviously most important person is Melissa, thanks for sharing your life with me.

Rasmus Brandt
Stockholm, April 2014

Contents

Chapter 1

Introduction

1.1 Background

As well known to anybody owning and using a smartphone, our reliance on mobile communication as a society is rapidly increasing. The computational power of the devices in our pockets is skyrocketing, yet the experience of surfing the web on a mobile device is often constrained by the wireless connection to the base station. The number of connected devices is expected to reach 50 billion within a couple of years according to the industry [Eri11], and the amount of data transmitted over the world's wireless networks is increasing exponentially[1] [Cis13]. This results in the operators being stretched in their ability to serve the users to their demands at peak times. The capacity of wireless networks can in principle be increased by either 1) acquiring more wireless spectrum, or 2) improving the spectral efficiency of the transmissions. This thesis focuses on the latter option, in particular by applying multi-antenna transceivers such that multiple mobile devices can be served on the same time-frequency resource blocks simultaneously.

The idea of using multi-antenna techniques in wireless communication is fairly old [Win84, Fos96, RC98]. For single-user point-to-point systems, the multiple antennas can be used to increase the resilience against wireless channel variations (fading). By employing multiple, sufficiently separated, antennas at the receiver, the incident signals are independent between antennas, and a diversity gain is achieved. With multiple antennas both at the receiver and the transmitter (multiple-input multiple-output, MIMO), several *spatial data streams* can be served [Fos96, Tel99, RC98]. The added data streams lead to a multiplexing gain, improving the spectral efficiency of the system at high signal-to-noise ratios. Due to a fundamental diversity-multiplexing tradeoff [ZT03], both types of gain cannot be maximized simultaneously.

[1] In Sweden alone, the mobile data traffic grew 69% from mid-2012 to mid-2013 [Cis13].

Multicell MIMO Networks

Most interesting wireless systems have multiple users however, and are therefore not accurately described by the point-to-point model. The traditional method of serving multiple users in a system is to divide resources orthogonally between them. This can be done using e.g. time-division multiple access (TDMA), or frequency-division multiple access (FDMA). Applying these traditional methods is not spectrally efficient however. Instead, if multiple users could be served simultaneously in each time-frequency resource block, the spectral efficiency of the system would increase. This is the idea of space-division multiple access (SDMA), where the spatial separation of the receivers is used to discriminate their corresponding transmissions. There are two main incarnations of SDMA: multiuser MIMO [GKH$^+$07] and multicell MIMO [GHH$^+$10, BJ13]. In the former, one multi-antenna transmitter serves several spatially separated receivers. In the latter, several multi-antenna transmitters jointly coordinate their transmissions to their corresponding users. In this thesis, we focus on multicell MIMO networks.

Interference Alignment

In the theoretical investigations into the fundamental performance-limits of multicell MIMO networks, the discovery of *interference alignment* (IA) was a breakthrough [MAMK08, CJ08]. Interference alignment is a constructive method for serving the maximum number of spatial data streams in a multicell MIMO network, in an interference-free manner. By aligning the interference in a lower-dimensional subspace at all receivers, it can easily be removed using linear techniques. The detrimental impact of the interference is then completely removed, and the only fundamental performance-limiting factor remaining is the thermal noise. In the high-SNR regime, where interference is the main problem, applying IA can yield significantly better spectral efficiency than using orthogonalization by means of TDMA or FDMA. Indeed, there is a price to pay for the improved spectral efficiency however. IA requires channel state information (CSI) at the transmitters in order to properly align the transmissions. The CSI is estimated at the receivers, and must therefore typically be fed back to the transmitters. This can result in high overheads, which reduce the spectral efficiency gain. Furthermore, it is generally only a good idea to employ IA when the interference truly is the main performance-limiting factors. This may not always be the case in practical systems, which may have imperfect CSI, hardware distortion noises, unaligned interferers, etc.

Coordinated Precoding

In this thesis, we study the problem of how to achieve high spectral efficiencies in multicell MIMO networks from a practical standpoint. We investigate the concept of *coordinated precoding*, wherein the multiple transmitters coordinate *how* they serve their respective receivers. This can be done using e.g. IA, although other resource allocation methods might be more practical. Coordinated precoding is in

contrast to *joint transmission*, wherein the multiple transmitters jointly serve all receivers. Joint transmission has higher requirements on backhaul and synchronization compared to coordinated precoding, and is therefore less practical.

We study three main topics within the area of coordinated precoding. First, we investigate the theoretical feasibility of interference alignment over a combined space and frequency signal space. This is a practicably relevant scenario, where the precoding is performed jointly over antennas and subcarriers. We derive a necessary condition for the feasibility of IA in this scenario. Second, we study the problem of how to implement a distributed coordinated precoding system. As mentioned, the transmitters require CSI in order to design the precoders that are used for serving the receivers with data. We propose three methods for obtaining this CSI, corresponding to different tradeoffs between channel estimation, feedback, signaling and backhaul use. We also show the need to robustify an existing coordinated precoding method, since it performs poorly when naïvely coupled with the proposed CSI acquisition schemes. The findings result in a system design for a distributed joint CSI acquisition and coordinated precoding method. Thirdly, we investigate coordinated precoding with imperfect hardware. The hardware imperfections lead to distortion noises, for which compensation schemes typically are applied. The compensation is not perfect however, so some residual hardware impairments always exist. These negatively impact performance if not accounted for in the optimization. We show how a semi-distributed method for coordinated precoding can be formulated, which properly handles the residual hardware impairments.

1.2 Outline and Contributions

We now outline the thesis, and the contributions of which it consists. Many of the results have previously been published under IEEE copyright. Some sentences in this thesis may match sentences in the published works verbatim.

Chapter 2

In order to familiarize the reader with the general setting of this thesis, Chapter 2 reviews the literature and sets the stage for the forthcoming material. The foundations of wireless communication is described in general terms, and then the topic of aligning multiuser interference is discussed. The promising theoretical benefits of interference alignment are shown to be substantial, but the case is made why the weighted sum rate problem should be solved directly instead. The chapter ends with a discussion about performance-limiting transceiver impairments, some of which will be further investigated in the thesis.

Chapter 3

In Chapter 3, we investigate interference alignment for the case of a joint space-frequency signal space. Necessary conditions for the feasibility of IA for this setting

are derived, and the actual sum rate performance possible is studied using numerical simulation. First, an urban outdoors macrocell scenario is studied, where the channels were obtained from a measurement campaign. Second, a dense indoors scenario is studied, and both measured and synthetic channels are used for the performance evaluation. The sum rate results show that there is a performance improvement from precoding over a joint space-frequency signal space, rather than performing the precoding orthogonally over the different subcarriers. The performance improvement however comes as a power gain, rather than a DoF gain.

The material in this chapter has previously been published in:

[BAB12] R. Brandt, H. Asplund, and M. Bengtsson. Interference alignment in frequency – a measurement based performance analysis. In *Proc. Int. Conf. Systems, Signals and Image Process. (IWSSIP'12)*, pages 227–230, 2012. © IEEE 2012.

[BZB13] R. Brandt, P. Zetterberg, and M. Bengtsson. Interference alignment over a combination of space and frequency. In *Proc. IEEE Int. Conf. Commun. Workshop: Beyond LTE-A (ICC'13 LTE-B)*, pages 149–153, 2013. © IEEE 2013.

Chapter 4

In Chapter 3, the numerical sum rate performance results indicated that superior performance was achieved by directly trying to solve the sum rate optimization problem, rather than trying to solve the IA conditions. Therefore, in Chapter 4, we study a resource allocation method that is able to find locally optimal solutions to the weighted sum rate optimization problem. The method is known to be distributed, but requires local CSI. We show how this local CSI can be obtained in an almost fully distributed fashion, using channel estimation and uplink-downlink reciprocity. We propose three CSI acquisition methods, and analyze their feedback/signaling requirements and computational complexities. When the proposed distributed CSI acquisition is coupled with the existing resource allocation method, the resulting sum rate performance deteriorates significantly at high SNR. We therefore propose robustifying measures, resulting in a distributed and robust coordinated precoding method. The numerical sum rate performance results show that the proposed system performs excellently compared to the state-of-the-art robust coordinated precoding systems in the literature.

The material in this chapter has been submitted for possible publication in:

[BB14] R. Brandt and M. Bengtsson. Distributed CSI acquisition and coordinated precoding for TDD multicell MIMO systems. *IEEE Transactions on Signal Processing*, 2014. Submitted.

Chapter 5

The last part of Chapter 4 studied weighted sum rate optimization under imperfect CSI. This is not the only impairment that practical transceivers are affected by. In Chapter 5, we study the impairment of distortion noises from imperfect radio hardware in the transceivers. With a simple model for the distortion noises, a weighted sum rate optimization problem can be formulated. Using an existing framework, we show how an iterative algorithm for finding locally optimal solutions can be devised. Using numerical simulation, the importance of accounting for the hardware impairments in the optimization is shown. As the level of hardware impairment is increased, performance for the unaware methods deteriorate, whereas the proposed method is robust.

The material in this chapter has previously been published in:

[BBB14] R. Brandt, E. Björnson, and M. Bengtsson. Weighted sum rate optimization for multicell MIMO systems with hardware-impaired transceivers. In *IEEE Conf. Acoust., Speech, and Signal Process. (ICASSP'14)*, pages 479-483, 2014. © IEEE 2014.

Chapter 6

This chapter concludes the thesis, and an outlook on future possible research is presented.

Contributions Outside the Scope of this Thesis

The author has also produced some work which does not fall within the scope of this thesis. In [BB11], methods for approximately diagonalizing a wideband multi-antenna channel was studied. By modeling the channel as a matrix finite impulse response filter, an approximate polynomial singular value decomposition could be formed. In [BB11], the performance of applying this polynomial singular value decomposition in a wideband multi-antenna scenario is studied. Compared to the traditional approach of exactly diagonalizing the channel in a finite number of orthogonal subcarriers, the polynomial approximate decomposition was shown to have higher complexity and worse diagonalization performance.

[BB11] R. Brandt and M. Bengtsson. Wideband MIMO channel diagonalization in the time domain. In *Proc. IEEE Int. Symp. Indoor, Mobile Radio Commun. (PIMRC'11)*, pages 1958–1962, 2011. © IEEE 2011.

1.3 Notation

Bold font is used to describe matrices (e.g. $\mathbf{C}$) and vectors (e.g. $\mathbf{c}$).

$\forall$	'For all'
$\underset{\cdot}{\leftarrow}$	All uplink quantities are denoted with an arrow
$\lceil c \rceil$	Ceiling function: the smallest integer not less than c
$\|\mathbf{c}\|_2$	Euclidean norm of vector $\mathbf{c}$
$\|\mathbf{C}\|_{\mathrm{F}}^2$	Frobenius norm squared: sum of singular values of matrix $\mathbf{C}$
$[\mathbf{C}]_{:,1:m}$	Matrix formed from columns 1 through m of matrix $\mathbf{C}$
$[\mathbf{C}]_{1:n,:}$	Matrix formed from rows 1 through n of matrix $\mathbf{C}$
$\mathbf{A}_k$	Receive filter for MS k in the IC
$\mathbf{A}_{i_k}$	Receive filter for MS i_k in the IBC
α_k	Data rate weight for MS k in the IC
α_{i_k}	Data rate weight for MS i_k in the IBC
$\mathbf{B}_{i_k}$	Component precoder for MS i_k in the IBC
$\mathrm{blkdiag}\,(\cdot)$	Creates a block-diagonal matrix from the arguments
$\mathbb{C}$	Set of complex numbers
d_k	Number of data streams for MS k in the IC
d_{i_k}	Number of data streams for MS i_k in the IBC
$\mathrm{diag}\,(\cdot)$	Creates a diagonal matrix from the arguments
$\mathrm{Diag}\,(\mathbf{C})$	Diagonal matrix where the diagonal elements are taken from the diagonal of the matrix $\mathbf{C}$
$\mathbf{F}_{i_k}$	Effective downlink channel for MS i_k in the IBC
$\mathbf{\Phi}_{i_k}$	Received signal covariance matrix for MS i_k in the IBC
$\mathbf{G}_{i_k}$	Effective uplink channel for MS i_k in the IBC
$\mathbf{\Gamma}_i$	Signal plus interference covariance matrix for BS i in the IBC
M_k	Number of antennas for BS k in the IC
M_i	Number of antennas for BS i in the IBC
M_{tot}	Total dimension of precoder space
$\mathbb{N}$	Set of natural numbers

N_k	Number of antennas for MS k in the IC
N_{i_k}	Number of antennas for MS i_k in the IBC
$\mathbf{I}_n$	Identity matrix of size n
L_c	Number of subcarriers
L_e	Number of equations in a polynomial system of equations
L_f	Number of subcarriers per alignment group
L_g	Number of alignment groups
$L_{p,d}$	Number of downlink pilot symbols
$L_{p,u}$	Number of uplink pilot symbols
L_v	Number of variables in a polynomial system of equations
$\lambda_{\max}(\mathbf{C})$	The eigenvalue of matrix $\mathbf{C}$ with the largest magnitude
$\lambda_{\min}(\mathbf{C})$	The eigenvalue of matrix $\mathbf{C}$ with the smallest magnitude
$\lambda_n(\mathbf{C})$	The eigenvalue of matrix $\mathbf{C}$ with the nth largest magnitude
$\text{eigvec}_n(\mathbf{C})$	The eigenvector corresponding to the eigenvalue of matrix $\mathbf{C}$ with nth largest magnitude
p_X	Probability density function for the random variable X
$q_{i_k}(\mathbf{V})$	User utility for MS i_k in the IBC
$q_{\text{sys}}(\cdot)$	System utility
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of positive real numbers
R_k	Data rate for MS k in the IC
R_{i_k}	Data rate for MS i_k in the IBC
ρ	Robustification parameter in the RB-WMMSE algorithm
$\text{Re}(\cdot)$	Real part of the argument
$s_{\max}(\mathbf{C})$	The largest singular value of matrix $\mathbf{C}$
$s_{\min}(\mathbf{C})$	The smallest singular value of matrix $\mathbf{C}$
$s_n(\mathbf{C})$	The nth largest singular value of matrix $\mathbf{C}$
$\text{span}(\mathbf{C})$	Column span of the matrix $\mathbf{C}$
$\text{Tr}(\cdot)$	Sum of the diagonal elements of the matrix argument

$\mathbf{U}_{i_k}$	Weighted receive filter for MS i_k in the IBC
$\mathbf{V}$	Tuple of all precoders
$\mathbf{V}_k$	Precoder for MS k in the IC
$\mathbf{V}_{i_k}$	Precoder for MS i_k in the IBC
$\mathrm{vec}\,(\mathbf{C})$	Column-wise vectorized version of matrix $\mathbf{C}$
$\mathbf{W}_{i_k}$	MSE weight matrix for MS i_k in the IBC
$\mathbf{x}_k$	Transmitted signal for MS k in the IC
$\mathbf{x}_{i_k}$	Transmitted signal for MS i_k in the IBC
$\mathbf{y}_k$	Received signal for MS k in the IC
$\mathbf{y}_{i_k}$	Received signal for MS i_k in the IBC
$\mathbf{z}_k$	Noise for MS k in the IC
$\mathbf{z}_{i_k}$	Noise for MS i_k in the IBC

1.4 Acronyms

3GPP	3rd Generation Partnership Project
CDF	Cumulative Distribution Function
CSI	Channel State Information
BS	Base Station
dB	Decibel
EVM	Error Vector Magnitude
FDD	Frequency-Division Duplex
FDMA	Frequency-Division Multiple Access
HIATUS	European commission, 7th framework programme, future and emerging technologies project on enHanced Interference Alignment Techniques for Unprecedented Spectral efficiency
IA	Interference Alignment
IBC	Interfering Broadcast Channel
i.i.d.	Independent and identically distributed (random variables)
IC	Interference Channel

KKT	Karush-Kuhn-Tucker (conditions)
LTE	3GPP Long-Term Evolution (standard)
MHz	Megahertz
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MMSE	Minimum Mean Squared Error
MS	Mobile Station
MSE	Mean Squared Error
MVU	Minimum Variance Unbiased (estimator)
OFDM	Orthogonal Frequency-Division Multiplexing
RF	Radio Frequency
SISO	Single-Input Single-Output
SIMO	Single-Input Multiple-Output
SINR	Signal-to-Interference-plus-Noise Ratio
SINDR	Signal-to-Interference-plus-Noise-and-Distortions Ratio
SNR	Signal-to-Noise Ratio
TDD	Time-Division Duplex
TDMA	Time-Division Multiple Access
WMMSE	Weighted Minimum Mean Squared Error
w.r.t.	With respect to

Chapter 2

Coordinated Precoding

We begin the thesis by presenting the background knowledge that will be essential in order to appreciate the contributions that will follow in subsequent chapters. We first introduce the idea of wireless communication systems, and then move on to present some fundamental performance limits of these systems. After introducing the concept of interference alignment, we make the case for why we should attempt to solve the non-convex system-level optimization problem instead. We proceed to present some algorithms to do this, that exist in the literature. Finally, we end with a discussion about practical challenges with coordinated precoding that will be studied further along in the thesis.

2.1 Wireless Communications

Wireless communication is about transmitting a *message* from a *transmitter* to a *receiver* over the air, without connecting the nodes using fixed infrastructure, such as electrical wires or optical fibers. The general name for the transmitted message will be x in this thesis. The received signal will be denoted y. In order to mathematically analyze and design the wireless communication system, a model is needed. That is, we need some mathematical description of how y is related to x. In the spirit of Occam's razor, we would like to have models that are as simple as possible, without oversimplifying reality. A very simple model of a wireless communication system would be that the receiver receives exactly the message that was transmitted from the transmitter, that is $y = x$. Although being extremely simple, this is not an interesting model since it does not reflect reality particularly well[1]. For example, due to the temperature and electrical resistance of the wireless receiver circuitry, the constituent free electrons have some motion. This thermal motion can be measured as a voltage over the output of the circuitry, and hence

[1]There is no limit on how quickly data can be transferred error-free in this model; it has infinite information theoretical capacity. We will discuss the notion of information theoretical capacity in Section 2.2.1.

constitutes *thermal noise* [Sto06, Ch. 10]. Additionally, the medium over which the transmitted signal passes to end up at the receiver, the *wireless channel*, will affect the received signal. Taking these two effects into account, our next attempt at modeling the wireless communication link is then

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

where h is the channel and z is the thermal noise. The model is more realistic with these introduced quantities, and thus it is worth investigating further. Since the nature of the thermal noise is that it is unpredictable, z is modeled as a stochastic variable. The variable x is also modeled as a stochastic variable, since it carries information content that is a priori unknown to the receiver. The wireless channel h is given to us by nature, and in order to ease the following exposition, we assume for now that h is deterministic, fixed over the transmission period, and known at the receiver.

Our first step in analyzing (2.1) is to determine the quality of the received signal. Assume that the transmit power is $\mathbb{E}\left(|x|^2\right) = P$ [W] and that the bandwidth of the system is W [Hz]. With a noise spectral density of N_0 [W/Hz], the noise power is N_0W. We can then define a fundamental quality metric of the received signal, the *signal-to-noise ratio* (SNR):

$$\text{SNR} = \frac{\mathbb{E}\left(|hx|^2\right)}{\mathbb{E}\left(|z|^2\right)} = \frac{|h|^2P}{N_0W}. \tag{2.2}$$

It is clear that we have a good signal when either the channel h and/or the signal x are 'strong' or when the noise is 'weak'.

The SNR describes if the received signal strength is good, but as a user of a wireless link, the *data rate* is possibly a quality measure that is more directly perceived. The maximum data rate for which arbitrarily small error probabilities can be achieved is the *capacity* [TV08] of the link. For the model in (2.1), the capacity is [Sha48]

$$C = W\log_2\left(1 + \frac{|h|^2P}{N_0W}\right) \text{ [bits/s]}. \tag{2.3}$$

Notice that the SNR from (2.2) appears inside the logarithm. The capacity is the maximum achievable data rate, with arbitrarily low error probability. It relies on a set of idealistic assumptions, which will be further detailed in Section 2.2.1, and is thus a fairly optimistic performance measure.

Even though the capacity in (2.3) is an optimistic performance measure, we can use it to analyze the performance of the wireless link. We will expose the performance-limiting aspects in two extreme operating regimes. We start with the *power-limited* regime. Assume that $|h|^2P \ll N_0W$, i.e. the noise is much stronger than the desired signal. Since the natural logarithm $\log_e(1+x) \approx x$ for small x, we then have that

$$C = W\log_2\left(1 + \frac{|h|^2P}{N_0W}\right) \approx \frac{|h|^2P}{N_0}\log_2(e) \quad \text{[bits/s]}, \tag{2.4}$$

where $e = 2.718...$ is Euler's number. Thus, in this regime, the performance is improved by increasing the transmit power P. Increasing the bandwidth W will not help. On the other hand, in the *degree of freedom-limited* regime, the opposite will be true. Assume that $|h|^2 P \gg N_0 W$. Then, we have that

$$C = W \log_2 \left(1 + \frac{|h|^2 P}{N_0 W}\right) \approx W \log_2 \left(\frac{|h|^2 P}{N_0 W}\right) \approx W \log_2 \left(|h|^2 P\right) \quad \text{[bits/s]}. \quad (2.5)$$

Since $|h|^2 P$ is already large, performance will improve drastically by enlarging the bandwidth W.

With the performance characterization in (2.4) and (2.5), we have a general idea of how to design the system for good performance. If we are expecting to operate in the power-limited regime, we should increase the transmit power. Reversely, if we are in the degree of freedom-limited regime, the bandwidth should be enlarged. All practical wireless systems are however constrained in their power and bandwidth usage. For example, the regulator[2] may require that the wireless system only operates within a certain frequency band, and that the transmitted power is below some limit. Wireless operators license parts of the spectrum through spectrum auctions; increasing the bandwidth available for a wireless system may therefore be very expensive. In addition to the regulatory requirements, the radio hardware employed may only handle a certain bandwidth and power.

Multiple Antennas

Given a certain bandwidth and power budget, the ultimate performance of the system in (2.1) is determined by (2.3). If the system is operating in the degree of freedom-limited regime, and more spectrum is not available, it seems that performance cannot be increased. By exploiting the *spatial dimension*, however, the spectral efficiency can be improved. By employing multiple antennas at both the transmitter and the receiver, say N antennas at the receiver and M antennas at the transmitter, multiple spatial *data streams* can be transmitted using the same time and frequency resources. Denote the capacity of this *multiple-input multiple-output* (MIMO) system as C_{MIMO}. At high SNR, the capacity then scales linearly with the minimum number of antennas [DADSC04] such that

$$\lim_{\text{SNR}\to\infty} \frac{C_{\text{MIMO}}}{\log_2(\text{SNR})} = W \min(N, M). \quad (2.6)$$

There are several other advantages to employing multiple antennas for a wireless communication link [DADSC04]. In addition to the MIMO *multiplexing gain* described in (2.6), *diversity gains* can also be achieved using multiple antennas. In this thesis, we will focus on the types of gains described by (2.6). A similar type of gain will be shown to be important for systems where more than one user is served.

[2] In Sweden, the usage of wireless spectrum is regulated by the governmental authority Post- och telestyrelsen.

Naturally, systems where only one side of the wireless link has access to multiple antennas exist. If the transmitter has multiple antennas, this is called a multiple-input single-output (MISO) system, and the reverse case is called a single-input multiple-output (SIMO) system.

2.1.1 Multiuser Communications

In the discussion in the previous section, there was only one transmitter and one receiver; it was the *point-to-point* setting. Most interesting wireless systems—such as WiFi, cellular communication, wireless ad-hoc networks, etc.—involve multiple transmitters and multiple receivers, however. If proper *orthogonalization* [CT06, Ch. 15.3] is applied, multiple users can be served simultaneously without experiencing inter-user interference. The point-to-point model can then be used for each orthogonal resource. The orthogonalization can be performed in many domains, e.g., time, frequency or code[3]. The corresponding multiple access techniques are then called time-division multiple access (TDMA), frequency-division multiple access (FDMA), or code-division multiple access (CDMA) [TV08, Ch. 4]. In essence, the enforced separability between users enables them to be served data without interference.

In the types of wireless systems that we are concerned with in this thesis, the receivers will also be *spatially separated*. By harnessing this provided spatial diversity [DADSC04], the users can be served using space-division multiple access (SDMA). The time and frequency dimensions are naturally available to the wireless transceivers. By adding multiple antennas to the transceivers, the spatial dimension also becomes available to the transceivers. The multiple antennas can be thought of as sampling the space [TV08, Ch. 7.3].

There are two main incarnations of the described SDMA: multiuser MIMO [GKH+07] and multicell MIMO [GHH+10] (see Figure 2.1 on the facing page). In the former, one multi-antenna transmitter transmits to several receivers. In the latter, several multi-antenna transmitters transmit to several receivers. In this thesis, we are interested in the multicell MIMO approach, where the transmitters cooperate to serve the receivers in a way that is good for the system-level performance. If the transmitters jointly serve the receivers with data, the operation mode is called *joint transmission*. We are more interested in another operation mode, the *coordinated precoding*. In this mode, the transmitters each serve their receivers, while still coordinating the interference that is created towards receivers served by other transmitters. Precoding is a linear transformation technique which will be further described later in this thesis.

[3]By spreading a signal over a wide frequency band using a *code*, many users can be accommodated over the same frequency band, if the codes are orthogonal [TV08, Ch. 4].

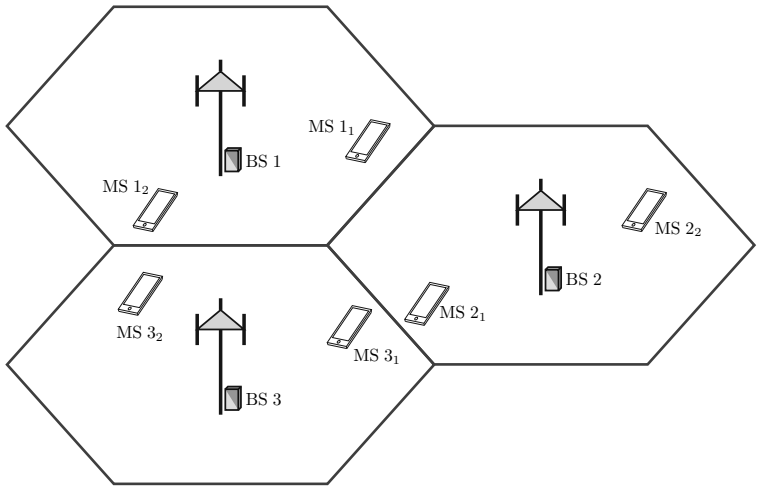


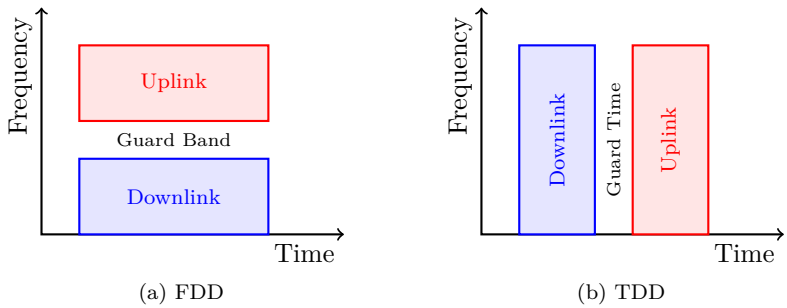
Figure 2.1. Example of a multicell system.

2.1.2 System Operation

So far, we have assumed that the data transmission only takes place in one direction: from the transmitter to the receiver. In most wireless systems, there is data to be transmitted in both directions however. In order to make the notion clear, we introduce the concept of a *base station* (BS) and the corresponding *mobile stations* (MSs). A base station is a fixed piece of hardware, which is typically connected to both the power grid and an operator backhaul network. The mobile stations on the other hand are roaming terminals, powered by battery and only connected to the network using wireless techniques. Typical MSs can for example be cell phones, tablets, portable computers, etc. At each point in time, each MS is *served* by one particular BS. The BS, together with its geographically served area, is called a *cell*. The BS and the MSs may both transmit and receive; they are *transceivers*.

The MS receives data from the BS in the *downlink*. Reversely, the MS transmits data to the BS in the *uplink*. In order for the uplink/downlink transmissions not to interfere, they must be orthogonalized. In many deployed systems, this is often done using frequency-division duplexing (FDD) [TV08, Ch. 4], where the uplink and downlink transmissions are performed on separated frequency bands. The uplink/downlink transmissions can also be orthogonalized in time using time-division duplexing (TDD) [TV08, Ch. 4]. TDD and FDD are compared in Figure 2.2 on the next page.

Although FDD traditionally has been more popular by operators, partially due to the spectrum plans set by regulators, there are some benefits of TDD over FDD. One benefit is that the ratio between the capacity of the uplink and downlink transmissions can adaptively be changed in TDD mode [HT11, Ch. 15]. Another



(a) FDD (b) TDD

Figure 2.2. Comparison of Frequency-Division Duplex (FDD) and Time-Division Duplex (TDD).

benefit of TDD is the *reciprocity* of the uplink and downlink channels. All wireless channels are reciprocal [Smi04], meaning that they are perceived the same in both the uplink and downlink. This can be exploited in channel estimation [BG06]; by pilot transmission in the uplink, the BSs can actually gain information about the channels in the downlink. This is beneficial, since downlink channel state information at the BSs is crucial for the operation of coordinated precoding. Although the wireless channel is perfectly reciprocal, the uplink/downlink RF hardware might not be. The effective channel that the coordinated precoding baseband algorithms perceive is the cascade of the transmit filter, the wireless channel, and the receive filter [GSK05]. Without proper calibration [BCK03,GSK05,RBP+13], this effective channel might not be reciprocal. In Chapter 4, we will assume a perfectly calibrated TDD system in order to achieve channel state information at the BSs, to be used in the coordinated precoding.

One drawback of using TDD is that neighbouring cells must be time-synchronized, such that the uplink transmissions in one cell are not disturbed by the unsynchronized (high power) downlink transmissions in a neighbouring cell [HT11, Ch. 15]. In terms of actually deployed cellular systems, FDD still dominates over TDD. In Sweden for example, in the 2.6 GHz band used for LTE, only 1 out of 15 frequency blocks is designated for TDD, and the rest are designated for FDD [Pos08].

Phases of System Operation

We now detail the phases of the system operation. For the most part of this thesis, we will study the downlink transmissions, and we thus describe the system operation from this perspective. The reason for mainly studying the downlink is that the traffic load experienced in the downlink is typically higher than the traffic load experienced in the uplink, due to e.g. video streaming and file downloads.

In essence, the idea of coordinated precoding is to serve multiple MSs from

multiple BSs, in a way that is beneficial for the entire system. In order to do this, the system must be aware of the current channel conditions. That is, the nodes of the network must have access to some *channel state information* (CSI). In our proposed system in Chapter 4, this information is obtained by pilot transmissions and channel estimation together with feedback and signaling. When the CSI is acquired, this information is used to select how the MSs should be served in a good way; this is the task of the *resource allocation* algorithm. Finally, data is transmitted in the way determined by the resource allocation. In summary:

CSI acquisition Channels are estimated using pilot transmissions. Other information is exchanged between the nodes through feedback and signaling.

Resource allocation Based on the obtained CSI, a resource allocation algorithm determines how the BSs should serve the MSs to maximize system performance.

Data transmission Data is transmitted in the fashion determined by the resource allocation. The estimated downlink channels are used by the MSs in their decoding of the received signals.

2.1.3 General System Models

We now introduce some general forms of mathematical system models which will be used throughout the thesis.

Point-to-Point Channel

For completeness, we first define the *point-to-point* channel. Here, one BS serves one MS in the downlink, without interference from other transmitting BSs. We assume that the MS has N receive dimensions, and that the BS has M transmit dimensions. These dimensions will often be spatial dimensions that are accessed through the use of multiple antennas, but the dimensions may also describe a combined space-frequency signal space, as elaborated on in Chapter 3. The narrowband complex valued equivalent baseband channel between the BS and the MS is then denoted $\mathbf{H} \in \mathbb{C}^{N \times M}$. The signal to be conveyed is $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{I}_d)$, and we assume that the BS uses a linear *precoder* $\bar{\mathbf{V}} \in \mathbb{C}^{M \times d}$ such that the transmitted signal is $\mathbf{s} = \bar{\mathbf{V}}\mathbf{x}$. The number of data streams that are transmitted is determined by d. By letting $d = \min(N, M)$, and using the eigenprecoding with waterfilling technique (see Section 2.3.3), the resource allocation can implicitly determine the optimal number of streams to transmit.

Under these assumptions, the received signal is modeled as

$$\mathbf{y} = \mathbf{H}\bar{\mathbf{V}}\mathbf{x} + \mathbf{z}, \tag{2.7}$$

where $\mathbf{z} \sim \mathcal{CN}\left(0, \sigma^2 \mathbf{I}_N\right)$ is some additive thermal noise. Modeling the thermal noise as a zero-mean circularly-symmetric complex white Gaussian distribution is a

very common assumption in wireless communication [CT06, Cou07, Mad08, TV08] and is basically due to the central limit theorem [Str93, Ch. 2] and the fact that there is a large number of electrons with thermal energy in the receiver circuitry.

Interference Channel

We now define the *interference channel* (IC). In this model, there are K BSs, each serving one MS in the downlink. Due to the shared medium, the intended signal for one MS will be perceived as interference at the other MSs. MS k is served by BS k for $k \in \{1, \ldots, K\}$. We assume that BS k has M_k transmit dimensions (e.g. antennas), and correspondingly that MS k has N_k receive dimensions (e.g. antennas). The narrowband complex valued equivalent baseband channel between BS l and MS k is $\mathbf{H}_{kl} \in \mathbb{C}^{N_k \times M_l}$. MS k is served d_k data streams from its corresponding BS, and its signal is $\mathbf{x}_k \sim \mathcal{CN}(0, \mathbf{I}_{d_k})$. The signals intended for different MSs are independent and identically distributed (i.i.d.). The BSs apply linear precoders $\mathbf{V}_k \in \mathbb{C}^{M_k \times d_k}$ such that the transmitted signal from BS k is $\mathbf{s}_k = \mathbf{V}_k \mathbf{x}_k$. Finally, assuming that the interference perceived over the shared medium can be described in an additive fashion, the received signal at MS k is

$$\mathbf{y}_k = \mathbf{H}_{kk}\mathbf{V}_k\mathbf{x}_k + \sum_{l \neq k} \mathbf{H}_{kl}\mathbf{V}_l\mathbf{x}_l + \mathbf{z}_k. \tag{2.8}$$

The first term is the desired signal and the second term is the sum of all the interfering signals. The last term $\mathbf{z}_k \sim \mathcal{CN}\left(0, \sigma_k^2 \mathbf{I}_{N_k}\right)$ is the additive noise, which is independent of all the transmitted signals. For this multiuser model, the covariance matrix for the received signal in (2.8) is

$$\mathbf{\Phi}_k = \mathbb{E}\left(\mathbf{y}_k\mathbf{y}_k^{\mathsf{H}}\right) = \underbrace{\mathbf{H}_{kk}\mathbf{V}_k\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{kk}^{\mathsf{H}}}_{\text{desired signal}} + \underbrace{\sum_{l \neq k} \mathbf{H}_{kl}\mathbf{V}_l\mathbf{V}_l^{\mathsf{H}}\mathbf{H}_{kl}^{\mathsf{H}}}_{\text{inter-cell interference}} + \underbrace{\sigma_k^2\mathbf{I}}_{\text{thermal noise}}. \tag{2.9}$$

The corresponding interference plus noise covariance is then

$$\mathbf{\Phi}_k^{\mathrm{i+n}} = \mathbf{\Phi}_k - \mathbf{H}_{kk}\mathbf{V}_k\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{kk}^{\mathsf{H}} = \sum_{l \neq k} \mathbf{H}_{kl}\mathbf{V}_l\mathbf{V}_l^{\mathsf{H}}\mathbf{H}_{kl}^{\mathsf{H}} + \sigma_k^2\mathbf{I}. \tag{2.10}$$

Interfering Broadcast Channel

In the interference channel of (2.8), each BS only served one MS. In order to increase the generality of the model, we now define the *interfering broadcast channel* (IBC). In this model, there are I BSs. We index the BSs as $i \in \{1, \ldots, I\}$, and now we let BS i serve K_i MSs in the downlink. In total, there are $K = \sum_{i=1}^{I} K_i$ MSs. We index the MSs served by BS i as $i_k \in \{1, \ldots, K_i\}$. MS i_k is thus the kth user in the group of users that are served by the ith BS. We call the BS and its associated MSs a 'cell'. The cells are typically geographically defined. We assume that BS i has

M_i transmit dimensions (e.g. antennas), and correspondingly that MS i_k has N_{i_k} receive dimensions (e.g. antennas). The narrowband complex valued equivalent baseband channel between BS j and MS i_k is $\mathbf{H}_{i_k j} \in \mathbb{C}^{N_{i_k} \times M_j}$. MS i_k is served d_{i_k} data streams from its corresponding BS, and its signal is $\mathbf{x}_{i_k} \sim \mathcal{CN}\left(0, \mathbf{I}_{d_{i_k}}\right)$. The signals intended for different MSs are i.i.d. The BSs apply linear precoders $\mathbf{V}_{i_k} \in \mathbb{C}^{M_i \times d_{i_k}}$ such that the transmitted signal from BS i is $\mathbf{s}_i = \sum_{k=1}^{K_i} \mathbf{V}_{i_k} \mathbf{x}_{i_k}$. Assuming that the intra-cell interference can be described similarly as the inter-cell interference, the received signal at MS i_k is modeled as

$$\mathbf{y}_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{x}_{i_k} + \sum_{(j,l) \neq (i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j_l} \mathbf{x}_{j_l} + \mathbf{z}_{i_k}. \tag{2.11}$$

The main difference between (2.11) and (2.8) is that the intra-cell interference terms are seen by the MSs as originating from the same direction as its desired signal. The noise term $\mathbf{z}_{i_k} \sim \mathcal{CN}\left(0, \sigma_{i_k}^2 \mathbf{I}_{N_{i_k}}\right)$ is complex Gaussian as before. For this multiuser model, the covariance matrix for the received signal in (2.11) is

$$\begin{aligned} \mathbf{\Phi}_{i_k} &= \mathbb{E}\left(\mathbf{y}_{i_k} \mathbf{y}_{i_k}^{\mathsf{H}}\right) \\ &= \underbrace{\mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}}}_{\text{desired signal}} + \underbrace{\sum_{(j,l) \neq (i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j_l} \mathbf{V}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}}_{\text{inter-cell and intra-cell interference}} + \underbrace{\sigma_{i_k}^2 \mathbf{I}}_{\text{thermal noise}}. \end{aligned} \tag{2.12}$$

The corresponding interference plus noise covariance is then

$$\mathbf{\Phi}_{i_k}^{\mathsf{i+n}} = \mathbf{\Phi}_{i_k} - \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} = \sum_{(j,l) \neq (i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j_l} \mathbf{V}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}^{\mathsf{H}} + \sigma_{i_k}^2 \mathbf{I}. \tag{2.13}$$

With a clear definition of the multiuser interaction models, we are prepared to analyze their performance. In that vein, we will in the next section introduce the concept of information theoretical capacity and the connection to interference alignment.

2.2 Interference Alignment

In Section 2.1, we did a basic performance analysis of a simple point-to-point wireless link. In this section, we will provide a more thorough description of the fundamental limits of the performance of the system models in the previous section. To do so, we will introduce the information theoretical capacity and the notion of degrees of freedom of an interference network. We will also introduce interference alignment, which is a method for achieving the maximum degrees of freedom.

2.2.1 Information Theoretical Capacity

The fundamental limits of wireless communication are described using information theory [CT06]. This field was pioneered by C. E. Shannon in his formative pa-

per [Sha48][4]. In this paper, Shannon showed that a strictly positive data rate is achievable with arbitrarily low error probability, a fact that was not thought to be true before.

Information theory is thus partly concerned with finding channel capacities for different channel models. These capacities are described by *coding theorems*, which generally comprise two parts: the converse and the achievability construction. The converse describes an upper bound on performance that no channel code can surpass. The achievability construction gives a channel code that can achieve a certain performance. If the achievable performance of a particular code coincides with the upper bound, the code achieves the capacity of the system. In order to not get entangled in the details of information theory, we will in this thesis use the following operational definition of channel capacity: the channel capacity is the highest rate of information that can be transmitted over a channel with arbitrarily low error probability [CT06, p. 184]. The channel capacity is given by, in loose terms, the maximum mutual information between the transmitted signal x and the received signal y, when maximized over all possible input distributions p_X. When the noise is Gaussian, the input distribution p_X that maximizes the mutual information is the Gaussian distribution. The interested reader can find more details in some information theory textbook, e.g. [CT06].

Achievable Rate of the Point-to-Point Channel

The capacity of the (Gaussian, deterministic) point-to-point channel in (2.7) on page 17 was derived in [Tel99]. There it was shown that the optimal input distribution p_X is the multivariate Gaussian distribution, leading to the following mutual information between $\mathbf{x}$ and $\mathbf{y}$:

$$R = \log_2 \det \left(\mathbf{I}_d + \bar{\mathbf{V}}^{\mathsf{H}} \mathbf{H}^{\mathsf{H}} \mathbf{H} \bar{\mathbf{V}} \right). \tag{2.14}$$

Note that R in (2.14) can be interpreted as an *achievable rate*. The maximum $R^\star$ is thus the capacity, given by the optimal precoder $\bar{\mathbf{V}}^\star$. The precoder is found by solving a convex optimization problem; see Section 2.3.1.

Finding a code that achieves the capacity $R^\star$ hinges on a set of idealistic assumptions. First, the transmitted signal $\mathbf{x}$ must be drawn from a zero-mean circularly symmetric complex Gaussian distribution with covariance $\bar{\mathbf{V}}^\star \bar{\mathbf{V}}^{\star,\mathsf{H}}$. This maximizes the entropy of the received signal, and thus maximizes the mutual information. In practical systems, the components of $\mathbf{x}$ are often drawn from a finite constellation [Mad08, Ch. 3.3] instead. The second idealistic assumption is that the length of the codewords that are used to achieve the capacity must go to infinity for the error probability to go to zero. In practical systems, long codewords give correspondingly long decoding delays, which is not desired. Finally, the rate in (2.14) is only achievable with an optimal, and therefore high complexity, detector.

[4] As trivia, we note that this landmark paper has had 65033 citations according to Google Scholar, at the time of this writing.

Even though the formulation of (2.14) hinges on these idealistic assumptions, it is a good model for how a well-designed practical wireless system would perform given a certain precoder $\bar{\mathbf{V}}$.

Degrees of Freedom for the Interference Channel

For the point-to-point channel above, the achievable data rates were described by a single number, the capacity. For the interference channel however, all K MSs have individual data rates, and the possible data rate performance of the entire system is thus given by a K-dimensional *capacity region* [Car78]. We denote this as $\mathcal{C} \in \mathbb{R}_+^K$, and each dimension describes the achievable rate for one MS. One point on the boundary of the capacity region is the *sum capacity*, which is the point that maximizes the sum of the MS capacities.

The capacity of the point-to-point channel is easily obtained by solving a convex optimization problem (see Section 2.3.1), but for the interference channel a full characterization of the capacity region has eluded information theorists for many decades. Instead, lately a lot of focus has been on the related concept of *degrees of freedom* (DoF) of the interference channel. The corresponding DoF region is interesting since it partially characterizes the capacity region.

For the K user interference channel, let the achievable rate of MS k be R_k and define the achievable rate tuple $\mathbf{R}(\mathrm{SNR}) = (R_1(\mathrm{SNR}), \ldots, R_k(\mathrm{SNR}))$. The capacity region $\mathcal{C}(\mathrm{SNR})$ is the closure of the set of achievable rate vectors [Car78], and the DoF region is then defined as [CJ08]:

$$\mathcal{D} = \Bigg\{ (d_1, \ldots, d_K) \in \mathbb{R}_+^K : \forall\, (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}_+^K \\ \sum_{k=1}^{K} \alpha_k d_k \leq \limsup_{\mathrm{SNR}\to\infty} \left(\sup_{\mathbf{R}(\mathrm{SNR})\in\mathcal{C}(\mathrm{SNR})} \frac{1}{\log_2(\mathrm{SNR})} \left(\sum_{k=1}^{K} \alpha_k R_k(\mathrm{SNR}) \right) \right) \Bigg\}. \tag{2.15}$$

Essentially, the DoF region describes what high-SNR slope, or pre-log factor, that is possible for the sum rate. The DoF d_k can equivalently be thought of as the number of interference-free data streams, that are successfully communicated to MS k. As a system-level metric, the *sum DoF* is defined as $d_{\mathrm{sum}} = \sum_{k=1}^{K} d_k$. Consequently, the sum DoF describes the total number of interference-free data streams in the network.

In the high-SNR regime, the performance of the interference channel is limited by the DoFs. In this regime, the DoF region is therefore an interesting metric on the ultimate performance of the system.

Interference Alignment

Lately, a large body of work has been performed on finding the sum DoF for a number of different interference channels. One of the first works in this area

was [CJ08], where it was shown that the optimal sum DoF for the SISO interference channel with K MS/BS pairs and *time-varying channels* was $d_{\text{sum}} = K/2$. The achievability was shown using *linear techniques* and the concept of *interference alignment*[5] [MAMK08, CJ08, Jaf11]. Using traditional orthogonalization methods (e.g. TDMA or FDMA), the sum DoF is unity since only one interference-free data stream can be successfully transmitted. In comparison, the IA results of $K/2$ interference-free data streams was seen as a very exciting result in the wireless communication community.

The basic idea of interference alignment is to align all inter-user interference into a lower-dimensional subspace, at all receivers simultaneously. This is done by appropriately selecting the precoders. If the interference is aligned, it can easily be removed from the received signal, using e.g. linear zero-forcing. The remaining subspace will then be completely free of interference. Assuming that the desired signal is linearly independent of the interference, it can be detected by the receiver in the interference-free subspace. In most schemes, the interference is forced into a subspace of half the dimension of the full signal space, and all MSs therefore get the remaining half of the signal space without interference. Again comparing to traditional orthogonalization, this is indeed remarkable, since each MS only gets $1/K$ of the signal space interference-free using TDMA/FDMA.

Note that there still exists thermal noise in the interference-free subspace. These IA techniques are therefore only interesting in regimes where the interference is the main problem, and not the noise power.

2.2.2 Interference Alignment Conditions and Feasibility

In order to describe interference alignment mathematically, we will now introduce the idea of linear *receive filters*. These have a similar function as the linear precoder applied at the transmitter, but they are instead applied to the received signal at the receiver. For the interference channel in (2.8), each MS has a receive filter $\mathbf{A}_k \in \mathbb{C}^{N_k \times d_k}$. The received filtered signal at MS k is then

$$\hat{\mathbf{x}}_k = \mathbf{A}_k^{\mathsf{H}} \mathbf{y}_k = \underbrace{\mathbf{A}_k^{\mathsf{H}} \mathbf{H}_{kk} \mathbf{V}_k \mathbf{x}_k}_{\text{filtered desired signal}} + \underbrace{\sum_{l \neq k} \mathbf{A}_k^{\mathsf{H}} \mathbf{H}_{kl} \mathbf{V}_l \mathbf{x}_l}_{\text{filtered interference}} + \underbrace{\mathbf{A}_k^{\mathsf{H}} \mathbf{z}_k}_{\text{filtered noise}} . \tag{2.16}$$

A set of receive filters and precoders $\{\mathbf{A}_k, \mathbf{V}_k\}$ is then an *IA solution* if it satisfies

$$\mathbf{A}_k^{\mathsf{H}} \mathbf{H}_{kl} \mathbf{V}_l = \mathbf{0}, \quad \forall\, k \in \{1, \ldots, K\},\, l \in \{1, \ldots, K\},\, l \neq k \tag{2.17}$$

$$\operatorname{rank}\left(\mathbf{A}_k^{\mathsf{H}} \mathbf{H}_{kk} \mathbf{V}_k\right) = d_k, \quad \forall\, k \in \{1, \ldots, K\}. \tag{2.18}$$

The requirement of no residual interference is described by the equations in (2.17). These equations can be trivially fulfilled by letting e.g. the precoders be zero. To

[5] The work on interference alignment lead to a best paper award for [CJ08]. In [CJ09], the authors of [CJ08] muse on the impact of their work, and point out some of the subsequent work in the field.

avoid such solutions, the equations in (2.18) describe the need for the effective channel to be of sufficient rank to receive the data streams. For a set of given system parameters — such as number of users, number of antennas, number of data streams per user, etc. — the *feasibility of IA* is thus described by the solvability of the system of equations in (2.17) and (2.18). We now review some existing feasibility results from the IA literature.

The Time-Varying K User SISO Interference Channel

For the K user SISO interference channel with *varying coefficients* (either in time or frequency), the optimal sum DoF is $K/2$ [CJ08]. The method for achieving this bound is interference alignment, performed in the time domain or frequency domain by coding over L_{ext} *channel extensions*. In particular, for some $n \in \mathbb{N}$ and with $L_{\text{ext}} = (n+1)^{K^2-3K+1} + n^{K^2-3K+1}$ channel extensions, the following DoFs are achievable using IA [CJ08]:

$$d_1 = \frac{(n+1)^{K^2-3K+1}}{(n+1)^{K^2-3K+1} + n^{K^2-3K+1}}, \tag{2.19}$$

$$d_k = \frac{n^{K^2-3K+1}}{(n+1)^{K^2-3K+1} + n^{K^2-3K+1}}, \ \forall\, k \in \{2, \ldots, K\}. \tag{2.20}$$

The details of the construction of the precoders that achieve the bound can be found in [CJ08]. As $n \to \infty$, each MS achieves $1/2$ DoF, and this gives the sum DoF result. The number of channel extensions grow exponentially in the number of MS/BS pairs K, and in order to get close to the asymptotic sum DoF, a large number of channel extensions are clearly needed. Although (asymptotically) achieving the sum DoF, this is clearly not a practical precoding method. Furthermore, the constructive method requires global CSI knowledge, which also is not practical.

For a K user MIMO interference channel with time-varying channels and $M = M_k = N_k$ for all k, the above scheme can be applied as well. By treating each antenna, at each MS, as a virtual MS, it is straightforward to show that the sum DoF for this scenario is $KM/2$. Clearly, this still requires very many channel extensions. The corresponding sum DoF using TDMA is M, and thus the gain of using IA can be very large.

The sum DoF for the time-varying channel is quite remarkable, as it grows with the number of MS/BS pairs K. In order to increase the DoF region, and thus the capacity region at high SNR, more users can simply be added to the system. This result is highly idealistic however, as will be shown in the following.

The Constant $K = 3$ User Symmetric MIMO Interference Channel

We now study the $K = 3$ user MIMO interference channel whose coefficients do not vary with time. The number of antennas are $M = M_k = N_k$ for all k, and each MS is served $d = M/2$ data streams. We assume M to be even, but a similar result

holds for odd M, see [CJ08]. For this scenario, the IA conditions in (2.17) can be reformulated as [CJ08]

$$\operatorname{span}\left(\mathbf{H}_{12}\mathbf{V}_2\right) = \operatorname{span}\left(\mathbf{H}_{13}\mathbf{V}_3\right), \tag{2.21}$$

$$\mathbf{H}_{21}\mathbf{V}_1 = \mathbf{H}_{23}\mathbf{V}_3, \tag{2.22}$$

$$\mathbf{H}_{31}\mathbf{V}_1 = \mathbf{H}_{32}\mathbf{V}_2. \tag{2.23}$$

The equations (2.22) and (2.23) force the interference to arrive from the same direction at MSs 2 and 3. Equation (2.21) forces the interference at MS 1 to belong to a common subspace. For full-rank channels, these conditions are solvable almost surely, and (2.18) on page 22 holds almost surely. For even M, a solution to (2.21)–(2.23) is [CJ08]

$$\mathbf{V}_1 = \left[\mathbf{L}\right]_{:,1:M/2}, \tag{2.24}$$

$$\mathbf{V}_2 = \left(\mathbf{H}_{32}\right)^{-1}\mathbf{H}_{31}\mathbf{V}_1, \tag{2.25}$$

$$\mathbf{V}_3 = \left(\mathbf{H}_{23}\right)^{-1}\mathbf{H}_{21}\mathbf{V}_1, \tag{2.26}$$

where $\left[\mathbf{L}\right]_{:,1:M/2}$ picks out the first $M/2$ columns of $\mathbf{L}$. The columns of $\mathbf{L}$ are the eigenvectors of

$$\left(\mathbf{H}_{31}\right)^{-1}\mathbf{H}_{32}\left(\mathbf{H}_{12}\right)^{-1}\mathbf{H}_{13}\left(\mathbf{H}_{23}\right)^{-1}\mathbf{H}_{21}$$

in some arbitrary order. For this special case, IA is always feasible with $d = M/2$ for all MSs. There might be several IA solutions, and the ordering of the eigenvectors in $\mathbf{L}$ determines which solution is found. Compared to the IA construction for the K user SISO interference channel with time-varying channels, here the optimal DoF can be achieved using IA with a finite number of antennas instead. The fact that $d = M/2$, and hence $d_{\text{sum}} = KM/2$, is optimal was shown at the end of the previous section.

The Constant K User Symmetric MIMO Interference Channel

For the MIMO interference channel with a general number of MS/BS pairs K, and assuming $M = M_k$, and $N = N_k$ for all k, a necessary condition for IA feasibility is [RLL12]

$$\sum_{k=1}^{K} d_k \leq M + N - 1. \tag{2.27}$$

On less rigorous grounds, the same condition for $d = d_k = 1$ for all k, was derived in [YGJK10]. Interestingly, since $d_{\text{sum}} = \sum_{k=1}^{K} d_k$, the sum DoF is bounded as

$$d_{\text{sum}} \overset{(2.27)}{\leq} M + N - 1 \tag{2.28}$$

That is, the achievable sum DoF does not grow with K, as was the case for the time-varying channel.

For the special case of $d = d_k$ for all k, where d divides both M and N, a necessary and sufficient condition for IA feasibility is [RLL12]

$$(K+1)d \leq M + N. \tag{2.29}$$

The General MIMO-IC

For the general MIMO interference channel, there are so far no results that describe IA feasibility analytically. The only existing result is a computational framework [GBS14], which essentially shows IA feasibility by checking the rank of a matrix. This numerical test is conveniently available as a web service at [GBS].

There is also no closed-form expression for the IA solution of a general scenario. Instead, numerical iterative algorithms can be employed to seek IA solutions. One such method will be described in Section 2.3.3.

2.2.3 Fundamental Limits of Cooperation

In any large wireless system, the cooperation using coordinated precoding must be performed in *clusters* [PGH08]. Otherwise, the number of BSs that must cooperate grows quickly, as well as the number of interfering cross-channels that the MSs must estimate. As stated, the models in (2.8) and (2.11) on pages 18–19 assume that the clusters are orthogonalized, either because they are sufficiently geographically separated, or alternatively because they are using orthogonal resources for the communication.

If these assumptions do not hold, the corresponding models should incorporate a term that describes the out-of-cluster interference. By adding such a term to the models, it can be shown that the benefits of coordinated precoding is fundamentally limited [LHA13]. The results in [LHA13] essentially show that the DoF gains, as explained earlier, only apply within an SNR window. For sufficiently large transmit powers, the out-of-cluster interference becomes substantial, and the sum rate saturates [LHA13]. In this thesis, we assume that the clusters are sufficiently orthogonal, such that the models in (2.8) and (2.11) well represent the system.

2.3 Weighted Sum Rate Optimization

The DoF metric introduced in the previous section is useful in the high-SNR regime, since it then gives a partial characterization of the capacity region. For the resource allocation in our coordinated precoding system, we may thus try to achieve the optimal sum DoF directly, using interference alignment. We call this approach *pure IA*, since it is only concerned with finding some IA solution that maximizes the DoF of the system. As alluded to earlier, there may often be several IA solutions [GSB13], which may correspond to different sum rates.

In this thesis, our true objective is to maximize the MS rates. In certain scenarios, trying to solve the IA conditions in (2.17)–(2.18) on page 22 may be a fruitful

approach to finding solutions with good sum rates. However, there are several downsides with trying to solve this proxy problem.

- In case IA is not feasible, it is not meaningful to try to solve the IA conditions. IA feasibility depends on the problem dimensions, and we are interested in optimizing the MS rates for arbitrary problem dimensions.

- The sum DoF, which is provided by the achievable IA schemes, is only an interesting metric in the high-SNR regime. We are interested in optimizing the MS rates for arbitrary SNRs.

- In order to achieve the optimal sum DoF, perfect and global CSI as well as perfect transceiver hardware is typically assumed. Real-world systems have neither, and we are interested in building systems that are amenable for practical implementation.

Due to these challenges with using the pure IA precoders for coordinated precoding, in this chapter we will instead study how to directly maximize the weighted sum rate. In the high-SNR regime, when IA is feasible, the optimal solution to the weighted sum rate problem will be the IA solution with the highest weighted sum rate.

2.3.1 System Utility and Constraints

In the field of mathematical optimization [BV04, Ber06], the goal is to optimize an *objective function* subject to some *constraints* on the involved *optimization variables*. In the wireless networks that we study, each MS will have its own objective function, and the goal is to optimize all of these objective functions simultaneously. Since the MSs share the ether, they are inherently coupled however. Improving performance for one MS may therefore degrade performance for another MS. This corresponds to multi-objective optimization, and all achievable working points are described by a performance region [BJ13]. Since the performance region is not easily characterized for the multicell MIMO case [BJ13], we focus on a scalarized problem. A system-level objective is formed, as a function of all the MSs' objectives, and the system-level function is optimized. For the exposition in this section, we will use the interfering broadcast channel in (2.11) on page 19, but the results are analogous for the interference channel in (2.8) on page 18.

Optimization Variables

For each MS i_k, the serving BS i has a corresponding linear precoder $\mathbf{V}_{i_k} \in \mathbb{C}^{M_i \times d_{i_k}}$. We restrict ourselves to linear techniques due to their low complexity in terms of implementation, and due to the fact that IA was shown to be achievable using linear filters in Section 2.2. The precoders will be the optimization variables in the

forthcoming optimization problem. In order to have a compact notation, we put all the precoders $\{\mathbf{V}_{i_k}\}$ in the tuple $\mathbf{V}$ such that

$$\mathbf{V} = \left(\mathbf{V}_{1_1}, \mathbf{V}_{1_2}, \ldots, \mathbf{V}_{1_{K_1}}, \ldots, \mathbf{V}_{I_1}, \mathbf{V}_{I_2}, \ldots, \mathbf{V}_{I_{K_I}}\right) \in \mathbb{C}^{M_{\text{tot}}}. \tag{2.30}$$

In (2.30), the total dimension of the precoder space is

$$M_{\text{tot}} = \Pi_{i=1}^{I} M_i \times \left(\Pi_{k=1}^{K_i} d_{i_k}\right), \tag{2.31}$$

where the outer-most product is over the BSs, and the inner-most product is over the MSs served by each BS.

User Utilities

We now assume that each MS i_k will have some utility $q_{i_k}(\mathbf{V})$, which depends on the selected precoders $\mathbf{V}$. This utility can for example be the bit error probability, the mean squared error, or the achievable data rate for MS i_k. In principle, we could select any user utility, but in this thesis we will mainly be concerned with the achievable data rate. As mentioned in Section 2.2.1, this is the data rate that can be achieved assuming Gaussian codebooks, long codewords and optimal decoders. Although these assumptions do not hold in practical systems, the achievable data rate can be seen as an upper bound on the 'true performance'.

For mathematical tractability, and in order to decrease the complexity of the decoder, we assume that the received interference is treated as noise in the decoder. The data rate for MS i_k in the interfering broadcast channel is then

$$R_{i_k} = \log_2 \det\left(\mathbf{I}_{d_{i_k}} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\mathbf{\Phi}_{i_k}^{\text{i+n}}\right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}\right), \tag{2.32}$$

where $\mathbf{\Phi}_{i_k}^{\text{i+n}}$ is defined in (2.13) on page 19. Treating the interference as noise is in general suboptimal in capacity sense, but leads to simple and practical receiver algorithms. Note that (2.32) is non-convex in the precoders $\{\mathbf{V}_{j_l}\}$, since they show up both as an outer quadratic term, as well as inner quadratic terms inside $\mathbf{\Phi}_{i_k}^{\text{i+n}}$.

Similarly, the data rate for the MS k in the interference channel is

$$R_k = \log_2 \det\left(\mathbf{I}_{d_k} + \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{kk}^{\mathsf{H}} \left(\mathbf{\Phi}_k^{\text{i+n}}\right)^{-1} \mathbf{H}_{kk} \mathbf{V}_k\right), \tag{2.33}$$

where $\mathbf{\Phi}_k^{\text{i+n}}$ is defined in (2.9) on page 18.

System Utility

The goal is now to optimize the rates for all the MSs. That is, we want to solve the *optimization problem*

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{maximize}} \quad & \left\{R_{1_1}, R_{1_2}, \ldots, R_{I_{K_I}}\right\} \\ \text{subject to} \quad & \mathbf{V} \in \mathcal{V}. \end{aligned} \tag{2.34}$$

The precoders stacked in the tuple $\mathbf{V}$ are the optimization variables, and we wish to simultaneously optimize all data rates. The set $\mathcal{V}$ is a convex set that describes the feasible precoders. We give some examples of sets $\mathcal{V}$ that might appear in applications on the next page.

The problem (2.34) is a multi-objective optimization problem [BJ13, Ch. 1.4.2]. The *Pareto boundary* describes all operating points where the performance of some MS cannot be increased without decreasing the performance of at least some other MS [BV04, Ch. 4.7]. The Pareto boundary has been fully characterized for the MISO scenario in [JLD09], and partially characterized for the MIMO scenario [CJS13].

In order to simplify the problem, instead of searching for the Pareto boundary, we will scalarize it using a system-level objective [BV04, Ch. 4.7.4]. For some system-level objective $q_{\text{sys}}\left(R_{1_1}, R_{1_2}, \dots R_{I_{K_I}}\right)$, the scalarized optimization problem is

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{maximize}} \quad & q_{\text{sys}}\left(R_{1_1}, R_{1_2}, \dots R_{I_{K_I}}\right) \\ \text{subject to} \quad & \mathbf{V} \in \mathcal{V}. \end{aligned} \tag{2.35}$$

The system-level objective describes how the individual MS rates impact the system-level performance. The selection of the system-level objective is inherently subjective, and common choices are: a weighted sum, the harmonic mean, the minimum MS rate, etc [BJ13]. In this thesis, we choose the *weighted sum rate* as our system-level objective:

$$q_{\text{sys}}\left(R_{1_1}, R_{1_2}, \dots R_{I_{K_I}}\right) = \sum_{(i,k)} \alpha_{i_k} R_{i_k}. \tag{2.36}$$

In essence, the sum rate describes the ultimate performance of the system, when fairness is not taken into account. That is, the system will completely turn off MSs, if that benefits system performance. By suitably selecting the weights $\alpha_{i_k} \in [0, 1]$, all Pareto optimal points that coincide with the convex hull of the Pareto boundary can be found [BJ13]. The weights can also be thought to describe the relative importance of the MSs, and can e.g. be selected to achieve a proportionally fair solution [KMT98].

Finally, we state the *weighted sum rate* optimization problem as

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{maximize}} \quad & \sum_{(i,k)} \alpha_{i_k} \log_2 \det\left(\mathbf{I}_{d_{i_k}} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\mathbf{\Phi}_{i_k}^{\text{i+n}}\right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}\right) \\ \text{subject to} \quad & \mathbf{V} \in \mathcal{V}. \end{aligned} \tag{2.37}$$

This is a non-convex optimization problem, since the MS rates (2.32) are non-convex in $\{\mathbf{V}_{j_l}\}$. At least when $N_k = 1$ for all k, the problem is also NP-hard [LDL11].

In addition to the user utilities, the set of feasible precoders $\mathcal{V}$ will affect the solution to (2.37). We now detail some common feasible sets.

Total Power Constraint

If the total radiated power of the system is constrained, the feasible set would be:

$$\mathcal{V} = \left\{ \left(\mathbf{V}_{1_1}, \mathbf{V}_{1_2}, \ldots, \mathbf{V}_{I_{K_I}} \right) \in \mathbb{C}^{\prod_{i=1}^{I} M_i \times \left(\prod_{k=1}^{K_i} d_{i_k} \right)} : \sum_{(i,k)} \mathrm{Tr}\left(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \right) \leq P_{\mathrm{tot}} \right\}. \tag{2.38}$$

Per-BS Sum Power Constraint

A more common feasible set is the one corresponding to the per-BS sum power constraint. For this set, the total radiated power per BS is constrained. This constraint may be due to either hardware limitations, or regulatory requirements. The feasible set is a Cartesian product of the feasible sets corresponding to the different BSs,

$$\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_I \tag{2.39}$$

and the feasible set for each BS is described by

$$\mathcal{V}_i = \left\{ \left(\mathbf{V}_{i_1}, \ldots, \mathbf{V}_{i_{K_i}} \right) \in \mathbb{C}^{M_i \times \prod_{k=1}^{K_i} d_{i_k}} : \sum_{k=1}^{K_i} \mathrm{Tr}\left(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \right) \leq P_i \right\}. \tag{2.40}$$

Per-BS Per-Antenna Power Constraint

With the per-BS sum power constraint, in principle all power could be allocated to one antenna for a particular BS. If the transmit power per RF-chain must be constrained, this can be done with the per-BS per-antenna power constraint. Similarly to the per-BS sum power constraint, this feasible set can be written as a Cartesian product over BSs,

$$\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_I \tag{2.41}$$

and the feasible set for each BS is described by

$$\mathcal{V}_i = \left\{ \left(\mathbf{V}_{i_1}, \ldots, \mathbf{V}_{i_{K_i}} \right) \in \mathbb{C}^{M_i \times \prod_{k=1}^{K_i} d_{i_k}} : \sum_{k=1}^{K_i} \left\| [\mathbf{V}_{i_k}]_{m,:} \right\|_{\mathrm{F}}^2 \leq P_{i,m}, \quad m = 1, \ldots, M_i \right\}. \tag{2.42}$$

2.3.2 Convexity and Optimality Conditions

The weighted sum rate problem in (2.37) on the facing page is non-convex, and possibly NP-hard in general. It is therefore hard to solve, and we will only venture to find locally optimal solutions. For completeness, we will however first shortly discuss *convex optimization* problems in this section.

A function $f(x) \in \mathbb{R}$ is said to be convex if it satisfies

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}) \tag{2.43}$$

for all $\mathbf{x} \in \mathbb{R}^L$, $\mathbf{y} \in \mathbb{R}^L$ and $t \in [0,1]$. A set $\mathcal{X}$ is said to be convex if

$$t\mathbf{x} + (1-t)\mathbf{y} \in \mathcal{X} \tag{2.44}$$

for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $t \in [0,1]$. Applying these concept to mathematical optimization, we can formulate a convex optimization problem as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{2.45}$$

This problem is called convex since the objective function is convex, and the feasible set is convex. A particularly prominent feature of convex optimization problems is that any local optimum is also a global optimum [BV04]. This makes this class of problem 'easy' to solve to global optimality, a property which does not hold in general for non-convex optimization problems.

For a general non-convex optimization problem, under some regularity conditions [Ber06, Ch. 3.3], the Karush-Kuhn-Tucker (KKT) conditions [BV04, Ch. 5.5], [Ber06, Ch. 3.3], give necessary conditions for a point to be optimal. For the special case of convex optimization problems, and assuming e.g. that a strictly feasible point exists[6], the KKT conditions are both necessary and sufficient [BV04, Ch. 5.5]. Finding the optimum to a convex optimization problem therefore amounts to solving the KKT conditions. In some cases, closed-form solutions can be found, but in general, numerical methods must be used to solve them. In the last couple of decades, *interior-point methods* [BV04] have been the state-of-the-art for solving constrained convex optimization problems.

2.3.3 Algorithms

We now summarize some of the optimization algorithms that will be used to find the precoders in this thesis. The IA and coordinated precoding literature is abound with algorithms, see e.g. [SSB+13] for a comparison. We choose to highlight three algorithms in particular, which will be used further along in the thesis. We will present the WMMSE algorithm of [SRLH11], due to its rigorous construction and good performance. Further, we present the algorithms MinWLI (pure IA) and MaxSINR from [GCJ11].

For the MIMO interference channel with $K = 3$ MS/BS pairs, an achievable IA scheme was presented in closed-form in Section 2.2.2. For general configurations however, closed-form solutions do not exist, and iterative algorithms need to be applied [SSB+13]. The original iterative method for finding pure IA solutions was

[6] This is known as Slater's constraint qualification [BV04].

proposed by Gomadam et al. in [GCJ08] (journal version in [GCJ11]), where a surrogate function called the interference leakage was minimized using alternating minimization. An algorithmically identical algorithm was soon thereafter proposed in [PH09]. In [SGHP10], the leakage minimization method was combined with a step moving along the gradient of the sum rate, in order to improve sum rate of the resulting IA solution.

The interference leakage is a measure of how much interference power leaks into the subspace that should be free of interference. By instead minimizing the rank of the interference subspace, through a convex relaxation with the nuclear norm, [PD12] proposed a rank constrained rank minimization algorithm based on alternating minimization. This method was further improved by using an improved surrogate function for the rank operator, resulting in a reweighted nuclear norm minimization algorithm in [DRSP13].

Instead of minimizing the leakage, Schmidt et al. proposed an MMSE interference alignment technique in [SCB+09]. In [GCJ08], the MaxSINR method was proposed, which iteratively maximizes the SINRs of the different spatial streams of the network. The convergence of this heuristic has not been proven, but empirically it often works very well [GCJ11, BAB12]. Modified versions of the SINR maximization in [GCJ11], with proven convergence, have been proposed in [PH11] and [WV13].

Point-to-Point Channel [Tel99]

For completeness, we first show how to find the optimal precoders for the point-to-point channel in (2.7) on page 17. Under the standard assumptions of a Gaussian codebook, long codewords, and an optimal decoder, the achievable data rate can be written as

$$R = \log_2 \det \left(\mathbf{I}_d + \frac{1}{\sigma^2} \bar{\mathbf{V}}^{\mathsf{H}} \mathbf{H}^{\mathsf{H}} \mathbf{H} \bar{\mathbf{V}} \right). \tag{2.46}$$

Under a sum power constraint, the optimal precoder is found as the solution to the optimization problem

$$\begin{aligned} \underset{\bar{\mathbf{V}}}{\text{maximize}} \quad & \log_2 \det \left(\mathbf{I}_d + \frac{1}{\sigma^2} \bar{\mathbf{V}}^{\mathsf{H}} \mathbf{H}^{\mathsf{H}} \mathbf{H} \bar{\mathbf{V}} \right) \\ \text{subject to} \quad & \left\| \bar{\mathbf{V}} \right\|_{\mathrm{F}}^2 \leq P. \end{aligned} \tag{2.47}$$

This is a convex optimization problem in $\bar{\mathbf{V}}$. Telatar [Tel99] showed that the optimal precoder $\bar{\mathbf{V}}^\star$ is such that it diagonalizes the effective channel. The available transmit power should be allocated according to the water filling technique. Let $\mathbf{L}\mathbf{\Lambda}\mathbf{L}^{\mathsf{H}} =$

$\mathbf{H}^{\mathsf{H}}\mathbf{H}$ be the eigenvalue decomposition. Then it can be shown [Tel99] that

$$\bar{\mathbf{V}}^{\star} = [\mathbf{L}]_{:,1:d} \begin{pmatrix} \sqrt{p_1^{\star}} & & & \\ & \sqrt{p_2^{\star}} & & \\ & & \ddots & \\ & & & \sqrt{p_d^{\star}} \end{pmatrix}, \tag{2.48}$$

where $[\cdot]_{:,1:d}$ picks out the d first columns and p_n is the power allocation for data stream n. With this choice, note that

$$\bar{\mathbf{V}}^{\star,\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\bar{\mathbf{V}}^{\star} = \begin{pmatrix} s_1^2\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)p_1^{\star} & & & \\ & s_2^2\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)p_2^{\star} & & \\ & & \ddots & \\ & & & s_d^2\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)p_d^{\star} \end{pmatrix} \tag{2.49}$$

where $s_n^2\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)$ is the nth largest singular value of $\mathbf{H}^{\mathsf{H}}\mathbf{H}$. Consequently, the optimal rate is

$$R^{\star} = \sum_{n=1}^{d} \log_2\left(1 + \frac{s_n^2\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)p_n^{\star}}{\sigma^2}\right). \tag{2.50}$$

The optimal $p_n^{\star}$ are found via *water filling* as

$$p_n^{\star} = \max\left(0, \mu - \frac{\sigma^2}{s_n^2\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)}\right). \tag{2.51}$$

Here, μ is the water level which is selected such that $\sum_{n=1}^{d} p_n^{\star} = P$. The name water filling derives from the interpretation of (2.51), which can be seen as filling water into a container with different depths for the different subchannels. The squared singular values $s_n^2\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)$ determine the quality of the subchannels, and thus how much power should be allocated to them. It can be shown that in the low-SNR regime, only one subchannel will have non-zero power allocation. Conversely, in the high-SNR regime, the power allocation will be uniform over the subchannels.

MinWLI [GCJ11]

Next, we present the MinWLI algorithm of [GCJ11]. This algorithm tries to find solutions that satisfy (2.17) on page 22, i.e. it is a pure IA algorithm. It does so by minimizing a surrogate function, called the interference leakage. The interference leakage for MS i_k is

$$\mathrm{IL}_{i_k} = \mathrm{Tr}\left(\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{\Phi}_{i_k}^{\mathrm{int}}\mathbf{A}_{i_k}\right) = \mathrm{Tr}\left(\mathbf{A}_{i_k}^{\mathsf{H}}\left(\sum_{(j,l)\neq(i,k)}\mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}}\right)\mathbf{A}_{i_k}\right) \tag{2.52}$$

and describes the amount of interference power that leaks into the receive filtered signal space. The system-level utility function is simply the weighted sum of interference leakages. The MinWLI algorithm then tries to solve the following optimization problem, under unitary constraints for precoders are receive filters:

$$\begin{aligned} \underset{\{\mathbf{A}_{i_k}\},\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{(i,k)} \alpha_{i_k} \operatorname{Tr}\left(\mathbf{A}_{i_k}^{\mathsf{H}} \left(\sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j_l} \mathbf{V}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}^{\mathsf{H}} \right) \mathbf{A}_{i_k} \right) \\ \text{subject to} \quad & \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{A}_{i_k} = \mathbf{I}_{d_{i_k}}, \quad i = 1, \ldots, I, \ k = 1, \ldots, K_i \\ & \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{V}_{i_k} = \frac{P_i}{K_i d_{i_k}} \mathbf{I}_{d_{i_k}}, \quad i = 1, \ldots, I, \ k = 1, \ldots, K_i \end{aligned} \tag{2.53}$$

This is a non-convex optimization problem in the joint block of variables $\{\mathbf{A}_{i_k}, \mathbf{V}_{i_k}\}$. Fixing either of the blocks however, the resulting optimization problem is convex, and has a solution in closed-form. Fixing the precoders $\{\mathbf{V}_{i_k}\}$, the optimal receive filter for MS i_k is the eigenvectors corresponding to the d_{i_k} smallest eigenvalues of the interference covariance matrix

$$\mathbf{\Phi}_{i_k}^{\text{int}} = \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j_l} \mathbf{V}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}^{\mathsf{H}}. \tag{2.54}$$

By fixing $\{\mathbf{A}_{i_k}\}$, the optimal precoder for MS i_k is the eigenvectors corresponding to the d_{i_k} smallest eigenvalues of the virtual uplink interference covariance matrix

$$\mathbf{\Upsilon}_{i}^{\text{int}} = \sum_{(j,l)\neq(i,k)} \alpha_{j_l} \mathbf{H}_{j_l i}^{\mathsf{H}} \mathbf{A}_{j_l} \mathbf{A}_{j_l}^{\mathsf{H}} \mathbf{H}_{j_l i}^{\mathsf{H}}. \tag{2.55}$$

Since a unique solution is found in each alternating minimization iteration, [Ber06, Prop. 2.7.1] can be applied to show that every limit point of the iterates corresponds to a stationary point of (2.53). We summarize the MinWLI algorithm in Algorithm 2.1 on the next page.

MaxSINR [GCJ11]

We also present the MaxSINR algorithm of [GCJ11]. This is a completely ad hoc method for sum rate optimization, which iteratively maximizes the SINRs of the different streams of the network. In each step, the optimal receive filter, or precoder, for one particular stream is found. Although the precoder is optimal for the given stream, at that particular moment, it may not be optimal in the sum rate sense. No convergence proof for the original formulation in [GCJ11] has been provided, although modified versions which provenly converge have been proposed [PH11,WV13]. Empirically, for scenarios where IA is feasible, this method sometimes works extremely well however; see e.g. Section 3.4 and Section 4.4 in this thesis.

Algorithm 2.1 MinWLI [GCJ11] with Per-BS Sum Power Constraints

repeat
At MS i_k:
$\mathbf{a}_{i_k,n} = \mathrm{eigvec}_{N_{i_k}-n+1}\left(\mathbf{\Phi}_{i_k}^{\mathrm{int}}\right), \quad n = 1,\ldots,d_{i_k}$
$\mathbf{A}_{i_k} = \begin{pmatrix}\mathbf{a}_{i_k,1} & \mathbf{a}_{i_k,2} & \cdots & \mathbf{a}_{i_k,d_{i_k}}\end{pmatrix}$
At BS i:
$\mathbf{b}_{i_k,n} = \mathrm{eigvec}_{M_i-n+1}\left(\mathbf{\Upsilon}_i^{\mathrm{int}}\right), \qquad k = 1,\ldots,K_i,\ n = 1,\ldots,d_{i_k}$
$\mathbf{B}_{i_k} = \begin{pmatrix}\mathbf{b}_{i_k,1} & \mathbf{b}_{i_k,2} & \cdots & \mathbf{b}_{i_k,d_{i_k}}\end{pmatrix}, \quad k = 1,\ldots,K_i$
$\mathbf{V}_{i_k} = \sqrt{\frac{P_i}{K_i d_{i_k}}}\mathbf{B}_{i_k}, \quad k = 1,\ldots,K_i$
until convergence criterion met, or fixed number of iterations

Algorithm 2.2 MaxSINR [GCJ11] with Per-BS Sum Power Constraints

repeat
At MS i_k:
$\mathbf{a}_{i_k,n} = \frac{\mathbf{\Phi}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}}{\left\|\mathbf{\Phi}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}\right\|_2}, \quad n = 1,\ldots,d_{i_k}$
$\mathbf{A}_{i_k} = \begin{pmatrix}\mathbf{a}_{i_k,1} & \mathbf{a}_{i_k,2} & \cdots & \mathbf{a}_{i_k,d_{i_k}}\end{pmatrix}$
At BS i:
$\mathbf{b}_{i_k,n} = \frac{\sqrt{\alpha_{i_k}}\left(\mathbf{\Upsilon}_i+\varsigma_i^2\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{a}_{i_k,n}}{\left\|\sqrt{\alpha_{i_k}}\left(\mathbf{\Upsilon}_i+\varsigma_i^2\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{a}_{i_k,n}\right\|_2}$
$k = 1,\ldots,K_i,\ n = 1,\ldots,d_{i_k}$
$\mathbf{B}_{i_k} = \begin{pmatrix}\mathbf{b}_{i_k,1} & \mathbf{b}_{i_k,2} & \cdots & \mathbf{b}_{i_k,d_{i_k}}\end{pmatrix}, \quad k = 1,\ldots,K_i$
$\mathbf{V}_{i_k} = \sqrt{\frac{P_i}{K_i d_{i_k}}}\mathbf{B}_{i_k}, \quad k = 1,\ldots,K_i$
until fixed number of iterations

We do not show the details of the derivation of the original MaxSINR here, since a very similar MaxSINDR algorithm is derived in Section 5.3.2. Instead, we directly summarize the original MaxSINR in Algorithm 2.2. There,

$$\mathbf{\Upsilon}_i = \sum_{(j,l)} \alpha_{j_l}\mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{A}_{j_l}\mathbf{A}_{j_l}^{\mathsf{H}}\mathbf{H}_{j_l i}^{\mathsf{H}} \tag{2.56}$$

is the signal plus interference covariance matrix in the virtual uplink. The noise power of BS i in the virtual uplink is ς_i^2. The transmit power is uniformly allocated over streams and MSs.

WMMSE [SRLH11]

Finally, we now turn our attention to the weighted sum rate problem (2.37) on page 28. One method for finding a local optimum to this optimization problem is the WMMSE algorithm [SRLH11]. In this section, we present a trivially modified version of this algorithm. The modification lies in accepting any convex feasible

set[7] for $\mathbf{V}$. By enlarging the search space and applying a tight lower bound to the MS rates, an iterative algorithm that monotonically converges to a stationary point can be found. First, we assume that a linear receive filter is used at the MSs, similarly as in (2.16) on page 22. With the estimate of the transmitted signal $\hat{\mathbf{x}}_{i_k} = \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{y}_{i_k}$, the *mean squared error* (MSE) matrix for MS i_k is

$$\begin{aligned} \mathbf{E}_{i_k} &= \mathbb{E}\left(\left(\mathbf{x}_{i_k} - \hat{\mathbf{x}}_{i_k}\right) \left(\mathbf{x}_{i_k} - \hat{\mathbf{x}}_{i_k}\right)^{\mathsf{H}} \right) = \mathbb{E}\left(\left(\mathbf{x}_{i_k} - \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{y}_{i_k}\right) \left(\mathbf{x}_{i_k} - \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{y}_{i_k}\right)^{\mathsf{H}} \right) \\ &= \mathbf{I} - \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} - \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{A}_{i_k} + \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{\Phi}_{i_k} \mathbf{A}_{i_k}. \end{aligned} \tag{2.57}$$

The MSE is another user performance metric, and the optimal receive filter in sum MSE sense is the well known *MMSE filter*. Minimizing $\mathrm{Tr}\left(\mathbf{E}_{i_k}\right)$ w.r.t. $\mathbf{A}_{i_k}$, the optimal filter is given as

$$\mathbf{A}_{i_k}^{\mathrm{MMSE}} = \mathbf{\Phi}_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}. \tag{2.58}$$

Note that since $\mathbf{y}_{i_k}$ and $\mathbf{x}_{i_k}$ have a linear relationship, and are jointly Gaussian, the optimal receiver structure is linear [Kay93, Ch. 15]. The receiver in (2.58) is therefore *the* MSE optimal receiver, and not just the best receiver (in MSE sense) amongst all linear receivers.

The MMSE filter has an interesting connection to the data rate. By substituting (2.58) into (2.57), it can be shown that

$$\begin{aligned} \mathbf{E}_{i_k}^{\mathrm{MMSE}} &= \mathbf{E}_{i_k}\left(\mathbf{A}_{i_k}^{\mathrm{MMSE}}\right) = \mathbf{I} - \mathbf{V}_{i_k} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{\Phi}_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \\ &= \left(\mathbf{I} + \mathbf{V}_{i_k} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \right)^{-1}. \end{aligned} \tag{2.59}$$

The last equality in (2.59) is due to the matrix inversion lemma. From (2.32) on page 27, we thus note that

$$R_{i_k} = \log_2 \det\left(\left(\mathbf{E}_{i_k}^{\mathrm{MMSE}}\right)^{-1} \right) = \max_{\mathbf{A}_{i_k}} \log_2 \det\left(\left(\mathbf{E}_{i_k}\right)^{-1} \right) = -\min_{\mathbf{A}_{i_k}} \log_2 \det\left(\mathbf{E}_{i_k}\right). \tag{2.60}$$

The second equality in (2.60) can be shown by noting that the gradient [HG07] of $\mathrm{Tr}\left(\mathbf{E}_{i_k}\right)$ w.r.t. $\mathbf{A}_{i_k}$ and the gradient of $-\log_2 \det\left(\mathbf{E}_{i_k}\right)$ w.r.t. $\mathbf{A}_{i_k}$ are such that, when set to zero, they both give the same solution. Since both functions are convex w.r.t. $\mathbf{A}_{i_k}$, the corresponding unconstrained optimization problems have the same solution. This connection has long been known for the single-user MIMO scenario [PCL03], the multiuser MIMO scenario [CACC08], and it was noted for the multicell MIMO scenario in [SRLH11].

Our goal is to find a local optimum to the weighed sum rate problem (2.37) on page 28. The first step in the WMMSE approach, as pioneered by [SRLH11], is to enlarge the search space to $\{\mathbf{A}_{i_k}, \mathbf{V}_{i_k}\}$ and use (2.60) to replace $R_{i_k} \rightarrow$

[7] One practically relevant scenario would for example be sum rate optimization for OFDM systems where the per-BS power constraints are summed over the subcarriers.

$-\log_2\det(\mathbf{E}_{i_k})$ in the objective function. Then the concave $\log_2\det(\mathbf{E}_{i_k})$ can be upper bounded by its first-order Taylor approximation, around a point $\bar{\mathbf{E}}_{i_k}$:

$$\log_2\det(\mathbf{E}_{i_k}) \leq \log_2(e)\left(-\log_e\det\left(\bar{\mathbf{E}}_{i_k}^{-1}\right) + \mathrm{Tr}\left(\bar{\mathbf{E}}_{i_k}^{-1}\mathbf{E}_{i_k}\right) - d_{i_k}\right),\ \forall\,\bar{\mathbf{E}}_{i_k} \succ 0 \tag{2.61}$$

where $e = 2.718...$ is Euler's number. Next, additional optimization variables $\{\mathbf{W}_{i_k} \succ \mathbf{0}\}$ are introduced such that $\mathbf{W}_{i_k} = \bar{\mathbf{E}}_{i_k}^{-1}$. These optimization variables thus determine the current linearization point of the MS rates. Lastly, by *minimizing* the weighted sum of the right-hand side of (2.61), we arrive at the following *weighted MMSE* problem,

$$\begin{aligned} \underset{\substack{\{\mathbf{V}_{i_k}\},\{\mathbf{A}_{i_k}\}\\ \{\mathbf{W}_{i_k}\succ\mathbf{0}\}}}{\text{minimize}} \quad & \log_2(e)\sum_{(i,k)}\alpha_{i_k}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\right) - \log_e\det\left(\mathbf{W}_{i_k}\right) - d_{i_k}\right) \\ \text{subject to} \quad & \mathbf{V}\in\mathcal{V}, \end{aligned} \tag{2.62}$$

where $\mathbf{E}_{i_k}$ is given in (2.57) on the preceding page. This optimization problem is still non-convex over the joint set $\{\mathbf{A}_{i_k}, \mathbf{W}_{i_k}, \mathbf{V}_{i_k}\}$. It is however convex in any block ($\{\mathbf{A}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$ or $\{\mathbf{V}_{i_k}\}$), when the remaining two blocks are kept fixed. Further, a stationary point can be found through alternating minimization[8] [Ber06, Ch. 2.7] over the blocks $\{\mathbf{A}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$, and $\{\mathbf{V}_{i_k}\}$ [SRLH11]. There is a one-to-one correspondence between the stationary points of (2.37) and the stationary points of (2.62) [SRLH11]. As will be shown later, in every iteration the bound in (2.61) is locally tight. Therefore, alternating minimization of (2.62) will also converge to a stationary point of (2.37) [RHL13]. Unless started from a local maximum, the WMMSE iterations will converge to a local minimum, since the objective in (2.62) is minimized in each iteration. It can also be shown that the global solutions to (2.37) and the global solutions to (2.62) coincide [SRLH11] (see similar derivation in Section 5.2.2).

The *WMMSE algorithm*, as termed by [SRLH11], follows from applying alternating minimization to (2.62) over the blocks of variables. Assuming that the nodes have perfect knowledge of their local CSI, the WMMSE algorithm is an example of a distributed resource allocation algorithm [SSB+13]. We will deliberate on this fact more in Chapter 4, where we will also discuss how to obtain the local CSI in a distributed fashion.

The first step in the WMMSE algorithm is to find the solution to (2.62) w.r.t $\{\mathbf{A}_{i_k}\}$, for fixed $\{\mathbf{W}_{i_k}, \mathbf{V}_{i_k}\}$. It is clear that it suffices to solve

$$\underset{\{\mathbf{A}_{i_k}\}}{\text{minimize}} \sum_{(i,k)}\mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\right). \tag{2.63}$$

This problem decouples naturally over the MSs, and since $\mathbf{W}_{i_k} \succ 0$, the solution to the quadratic program for MS i_k is $\mathbf{A}_{i_k}^{\star} = \mathbf{A}_{i_k}^{\mathrm{MMSE}}$. That is, the optimal receive

[8] This technique is also known as *block coordinate descent* or *block nonlinear Gauss-Seidel* in the literature.

filter is the MMSE receiver. Next, fixing $\{\mathbf{A}_{i_k}, \mathbf{V}_{i_k}\}$, the problem again decouples over the MSs. For MS i_k, we should solve

$$\underset{\mathbf{W}_{i_k}}{\text{minimize}} \ \mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\right) - \log_e \det\left(\mathbf{W}_{i_k}\right). \tag{2.64}$$

This corresponds to updating the linearization point of $\log_e \det\left(\mathbf{E}_{i_k}\right)$ and the solution for MS i_k is

$$\mathbf{W}_{i_k}^{\star} = \mathbf{E}_{i_k}^{-1} = \left(\mathbf{I} - \mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{\Phi}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}\right)^{-1} \tag{2.65}$$

where the last equality comes from plugging in $\mathbf{A}_{i_k}^{\star} = \mathbf{A}_{i_k}^{\mathrm{MMSE}}$ (cf. (2.59) on page 35).

With the new iterates for $\mathbf{A}_{i_k}$ and $\mathbf{W}_{i_k}$, it remains to solve (2.62) w.r.t $\{\mathbf{V}_{i_k}\}$ for fixed $\{\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\}$. Removing terms that are constant w.r.t. $\mathbf{V}_{i_k}$ and the scaling with $1/\log_2(e)$, this is equivalent to solving

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{(i,k)} \alpha_{i_k} \mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\right) \\ \text{subject to} \quad & \mathbf{V} \in \mathcal{V}. \end{aligned} \tag{2.66}$$

By using the property that $\mathrm{Tr}\left(\mathbf{CD}\right) = \mathrm{Tr}\left(\mathbf{DC}\right)$, and dropping constant terms, it can be shown that the following optimization problem is equivalent to (2.66):

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{(i,k)} \mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{\Gamma}_i\mathbf{V}_{i_k}\right) - 2\alpha_{i_k}\mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}\right)\right) \\ \text{subject to} \quad & \mathbf{V} \in \mathcal{V}. \end{aligned} \tag{2.67}$$

where $\mathbf{\Gamma}_i = \sum_{(j,l)} \alpha_{jl}\mathbf{H}_{jl i}^{\mathsf{H}}\mathbf{A}_{jl}\mathbf{W}_{jl}\mathbf{A}_{jl}^{\mathsf{H}}\mathbf{H}_{jl i}$ is a signal plus interference covariance matrix for BS i in the uplink. This is a convex optimization problem with a quadratic objective, which can be solved efficiently using e.g. interior-point methods [BV04, Ch. 11].

For the total power constraint in (2.38) and the per-BS sum power constraint set (2.39) on page 29, the solutions to (2.67) are particularly simple to find. For the total power constraint, the solution to (2.67) is

$$\mathbf{V}_{i_k}^{\star} = \alpha_{i_k}\left(\mathbf{\Gamma}_i + \nu^{\star}\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{A}_{i_k}\mathbf{W}_{i_k}, \quad i = 1,\ldots,I,\ k = 1,\ldots,K_i, \tag{2.68}$$

where $\nu^{\star} \geq 0$ is the optimal Lagrange multiplier for the constraint. If the constraint is satisfied for $\nu = 0$, the optimal precoders have been found. If this is not the case, ν can be found by 1D search methods such that $\sum_{(i,k)} \mathrm{Tr}\left(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^{\mathsf{H}}\right) = P_{\mathrm{tot}}$ is satisfied. Since $\sum_{(i,k)} \mathrm{Tr}\left(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^{\mathsf{H}}\right)$ can be shown to be monotonically decreasing in ν, e.g. bisection can be used.

The solution for the per-BS sum power constraint in (2.39) is very similar. The solution is

$$\mathbf{V}_{i_k}^{\star} = \alpha_{i_k}\left(\mathbf{\Gamma}_i + \nu_i^{\star}\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{A}_{i_k}\mathbf{W}_{i_k}, \quad i = 1,\ldots,I,\ k = 1,\ldots,K_i, \tag{2.69}$$

Algorithm 2.3 WMMSE Algorithm [SRLH11] with General Constraint Set

1: **repeat**
At MS i_k:
2: Find MSE weights: $\mathbf{W}_{i_k} = \mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$
3: Find MMSE receive filters: $\mathbf{A}_{i_k} = \mathbf{\Phi}_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$
At BSs:
4: Find precoders as solution to:

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{(i,k)} \alpha_{i_k} \mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{\Gamma}_i \mathbf{V}_{i_k}\right) - 2\alpha_{i_k} \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k} \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}\right)\right) \\ \text{subject to} \quad & \mathbf{V} \in \mathcal{V}. \end{aligned}$$

5: **until** convergence criterion met, or fixed number of iterations

where the $\nu_i^\star \geq 0$ are the optimal Lagrange multipliers for the I per-BS constraints. This can be found in the same manner as for the total power constraint above, i.e. if $\nu_i = 0$ satisfies the constraint for BS i, the problem is solved. Otherwise, $\nu_i > 0$ is found such that $\sum_{k=1}^{K_i} \mathrm{Tr}\left(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}}\right) = P_i$ is satisfied. This can be done using bisection, since $\sum_{k=1}^{K_i} \mathrm{Tr}\left(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}}\right)$ can be shown to be monotonically decreasing in ν_i.

When the precoders $\{\mathbf{V}_{i_k}\}$ have been found, the iterations start over by solving for $\{\mathbf{A}_{i_k}\}$ again. With each update of $\{\mathbf{A}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$ and $\{\mathbf{V}_{i_k}\}$, the objective value of (2.62) cannot increase. Since the objective value in (2.62) can be bounded, the objective value therefore converges monotonically. Unless started from a local maximum, the algorithm will find a local minimum, since it minimizes the objective function in every iteration. The convergence was shown for the per-BS sum power constraint in [SRLH11], but the convergence for the general case is a straightforward generalization that can be shown using e.g. [Ber06, Prop. 2.7.1], [RHL13] or [GS00].

We now summarize the WMMSE algorithm in Algorithm 2.3. Note that the original version in [SRLH11] only had per-BS sum power constraints, and Algorithm 2.3 is therefore slightly more general.

2.4 Practical Considerations

In the discussion up until now, the presented models expose the multiuser interaction, and the corresponding challenges of handling the negative impact of interference. Assuming perfect CSI and aligned interference, the only performance-limiting factor was the thermal noise at the receivers. In practice, several other challenges exist however. Amongst others, these can be: outdated and imperfect CSI, transmission delays, imperfect synchronization, limited and delayed backhaul, imperfect hardware, etc. We will discuss some of these issues now, as they will be further

investigated in the following chapters.

2.4.1 Distributed Techniques

Some precoding methods, such as the $K = 3$ closed-form IA solution in Section 2.2.2, require the complete network CSI to be available in one central location. For large networks, it is clearly not practical to collect all CSI in one location, due to backhaul capacity and delay constraints, as well as limited channel coherence time. Thus, distributed methods for CSI acquisition and resource allocation are needed. In the detailed resource allocation methods, both BSs and MSs need CSI. For FDD systems, the BSs can only obtain the CSI through feedback, whereas for TDD systems, the channel reciprocity [GSK05] can be used together with uplink pilot transmissions [JAWV11, SBH13]. Regarding the resource allocation, the WMMSE algorithm in Section 2.3.3 is an example of a distributed method, since the nodes can perform their part in the optimization based solely on local CSI.

Earlier work on distributed coordinated precoding methods include [SBH13], where a reciprocal channel was exploited to directly estimate the SINR-maximizing filters. Focusing on the reciprocity, and using the receive filters as transmit filters in the uplink, [GAH11] performed extensive simulations for a beam selection approach. Channel reciprocity was also used in the original paper on distributed interference alignment [GCJ11], but there the focus was on using it as an algorithmic construct. In [KTJ13], decentralized algorithms based on WMMSE ideas were proposed, achieving faster convergence than the original WMMSE in [SRLH11], in addition to CSI signaling strategies for obtaining the necessary CSI. TDD reciprocity was assumed, and the MSs used combinations of inter and intra-cell effective channel pilot transmissions. Contrary to our work in Chapter 4, perfect channel estimation was assumed, and their decentralized algorithms still require some BS backhaul.

2.4.2 Imperfect Channel State Information

In the described resource allocation methods, CSI at both BSs and MSs is essential. Although the formulations in Section 2.3.3 assume that the BSs have knowledge of CSI without errors, this will not be the case in practical systems. Since CSI often is obtained through pilot transmissions and channel estimation [BG06, JAWV11, SBH13], it is naturally imperfect. The resource allocation should take this into account, in order to be robust against the imperfections.

Earlier work on robust resource allocation was performed in [SM12, LKY13], where robustified weighted MMSE algorithms were proposed. The contribution of the channel estimation errors in the involved covariance matrices were averaged out, leading to robust but non-distributed solutions. The same approach was taken in [RBCL13], where it was mentioned that this corresponds to optimizing a lower bound on the achieved performance, and in [NGS12] where the lower bound was explicitly derived. The receive filters and precoders were in effect robustified by

diagonal loading, where the diagonal loading factors were determined by channel estimation error and transmitted power. For all the prior robust WMMSE algorithms [SM12, NGS12, LKY13, RBCL13], perfect knowledge of the receive filters, MSE weights, and precoders are assumed at all involved nodes. The methods are therefore not distributed, which is in contrast to the work presented in Chapter 4.

2.4.3 Imperfect Hardware

Another transceiver impairment is that of imperfect hardware. This may be in the form of phase noise, I/Q imbalance, power amplifier non-linearities and sampling-rate and carrier frequency offsets [Sch08]. Compensation schemes exist for all of these different hardware impairments, but since the compensation in general is not perfect, some residual hardware impairments still remain [SWB10]. These residual impairments impact the performance [SWB11, BZBO12], and should therefore be accounted for in the precoder optimization [BZB12,BJ13]. In Chapter 5, we present a simple model for the residual hardware impairments based on [BJ13, Ch. 4.3]. We also propose a WMMSE algorithm for finding locally optimal points to the corresponding weighted sum rate problem with hardware-impaired receivers. The resulting algorithm is distributed over the MSs, but the precoders must in general be found at a central BS. For a special case however, the resulting algorithm becomes fully distributed.

Chapter 3

Interference Alignment over Space and Frequency

In order to perform interference alignment, a signal space of dimension larger than one is needed. For multi-antenna systems, this requirement is naturally satisfied. For single-antenna systems, the idea of time- or frequency extensions for interference alignment was proposed in [CJ08]. By precoding over several time slots, or subcarriers, a signal space of dimension larger than one is created. Since channel state information is required at the transmitters, it seems more practical to use frequency extensions than time extensions, since the latter generally requires noncausal knowledge of the channel.

In this chapter, we study a more general model, where interference alignment is performed over a signal space composed of a *combination* of spatial and frequency dimensions. The spatial dimensions are accessed through multiple antennas at the transceivers, and the frequency dimensions are accessed through orthogonal frequency-division multiplexing[1] (OFDM). First, the system model incorporating the combined signal space is introduced. Three different types of uses of the combined signal space are presented. Then the main contribution of this chapter, the necessary condition for space-frequency IA feasibility, is presented. Finally, the performance of the system is evaluated using numerical simulation. In the simulator, both synthetically generated channels, as well as measured channels, are employed.

[1] OFDM [TV08] is a popular method for communicating over a wideband channel, reducing the complexity of the equalization at the receiver compared to single-carrier transmission. Assuming that the channel is linear and time-invariant (at least over some coherence time), it can be diagonalized using the discrete Fourier transform. The resulting parallel and narrowband *subcarriers* can then easily be equalized at the receiver. Since the discrete Fourier transform can be performed using the fast Fourier transform algorithm [OS99, Ch. 9], OFDM is amenable to practical implementation. It is being used in many wireless standards, such as WiFi (802.11g/n) and LTE.

3.1 System Model

We assume a K user MIMO-OFDM interference channel, where each MS has N antennas and each BS has M antennas. Through OFDM, the system can communicate over L_c orthogonal subcarriers. We study the downlink of the system, and each MS is served d data streams by its corresponding BS[2]. We denote the channel matrix between BS l and MS k at the nth subcarrier as $\mathbf{H}_{kl}^{(n)} \in \mathbb{C}^{N\times M}$. In the analysis, we will assume that the channel matrix entries are realizations of random variables, independent over antennas, subcarriers and transmitter-receiver links. We will mainly focus on studying the IA feasibility problem (see Sec. 2.2.2), and thus we assume that all transceivers have perfect knowledge of the necessary channel state information.

In traditional systems, only one MS is served per subcarrier. This type of orthogonalization is called OFDM multiple access (OFDMA). In the spirit of IA, we are however interested in serving multiple MSs over the same subcarrier. Since we are studying space-frequency precoding, we form a combined space-frequency space by defining the *frequency-extended* channel from BS l to MS k as

$$\mathbf{H}_{kl} = \text{blkdiag}\left(\mathbf{H}_{kl}^{(1)}, \ldots, \mathbf{H}_{kl}^{(L_c)}\right) = \begin{pmatrix} \mathbf{H}_{kl}^{(1)} & & & \\ & \mathbf{H}_{kl}^{(2)} & & \\ & & \ddots & \\ & & & \mathbf{H}_{kl}^{(L_c)} \end{pmatrix}. \tag{3.1}$$

The block-diagonality is due to the the orthogonality of the subcarriers. The frequency-extended channels can be used in different ways, and the three combined signal space methods that will be studied in this chapter are:

Space-Frequency Precoding The MSs are served jointly over the combination of spatial and frequency dimensions. Denoting the space-frequency precoder for MS l as $\mathbf{V}_l \in \mathbb{C}^{ML_c\times d}$, the received signal at MS k is

$$\mathbf{y}_k = \mathbf{H}_{kk}\mathbf{V}_k\mathbf{x}_k + \sum_{l\neq k}\mathbf{H}_{kl}\mathbf{V}_l\mathbf{x}_l + \mathbf{z}_k. \tag{3.2}$$

Space-Only Precoding The MSs are served jointly over the spatial dimension for each subcarrier, but the precoding over different subcarriers is orthogonal. Denoting the space-only precoder for MS k at the nth subcarrier as $\mathbf{V}_k^{(n)}$, this corresponds to block diagonal space-frequency precoders:

$$\mathbf{V}_k = \text{blkdiag}\left(\mathbf{V}_k^{(1)}, \ldots, \mathbf{V}_k^{(L_c)}\right). \tag{3.3}$$

[2]This is called a *symmetric* setup in the literature (e.g. [YGJK10]), since the number of antennas are the same for all transceivers, and all MSs are served the same number of data streams.

Combining (3.3) with (3.2), the received signal at MS k at the nth subcarrier is

$$\mathbf{y}_k^{(n)} = \mathbf{H}_{kk}^{(n)}\mathbf{V}_k^{(n)}\mathbf{x}_k^{(n)} + \sum_{l\neq k}\mathbf{H}_{kl}^{(n)}\mathbf{V}_l^{(n)}\mathbf{x}_l^{(n)} + \mathbf{z}_k^{(n)}. \tag{3.4}$$

Frequency-Only Precoding For the special case that $N = M = 1$, the frequency-extended channels are diagonal $\mathbf{H}_{kl} = \text{diag}\left(h_{kl}^{(1)}, \ldots, h_{kl}^{(L_c)}\right)$, and the MSs are served jointly over all subcarriers as in (3.2).

Space-only precoding is simply a method for performing independent and simultaneous precoding over the L_c subcarriers. Applying IA in this setting boils down to applying IA independently on each subcarrier. The feasibility of space-only IA is thus determined by the single-carrier IA feasibility conditions, which were elaborated on in Section 2.2.2. We remind the reader that the number of single-stream (i.e. $d = 1$) MSs that can be served interference-free over one subcarrier is [YGJK10, RLL12]

$$K = N + M - 1. \tag{3.5}$$

For frequency-only precoding, if the number of subcarriers grows exponentially fast in K, an achievable IA scheme can be set up that asymptotically achieves the optimal sum DoF of $K/2$ [CJ08]; see (2.19) and (2.20) in Section 2.2.2. In addition to requiring a large amount of subcarriers, this original scheme also relies on transmitting an exponentially large number of data streams per MS. In the computationally less demanding case[3] of single-stream transmission ($d = 1$), the number of MSs that can be served interference-free over L_c subcarriers must satisfy [SBH11]

$$K \leq 2L_c - 2. \tag{3.6}$$

Both space-only precoding and frequency-only precoding can be seen as special cases of the more general space-frequency precoding. In this chapter, we present a necessary condition for single-stream IA feasibility for the space-frequency precoding setting. We also present sum rate performance evaluations for both space-frequency precoding, as well as frequency-only precoding. For the space-frequency setting, we compare to space-only precoding using synthetically generated channels, as well as measured channels. For the frequency-only setting, we use measured channels for the performance evaluation.

3.2 Necessary Condition for Space-Frequency IA Feasibility

In order to derive the necessary condition for space-frequency IA feasibility, we first restate the IA conditions from Section 2.2.2:

$$\mathbf{A}_k^{\mathsf{H}}\mathbf{H}_{kl}\mathbf{V}_l = \mathbf{0}, \quad \forall\, k \in \{1,\ldots,K\},\, l \in \{1,\ldots,K\},\, l \neq k \tag{3.7}$$

$$\text{rank}\left(\mathbf{A}_k^{\mathsf{H}}\mathbf{H}_{kk}\mathbf{V}_k\right) = d_k, \quad \forall\, k \in \{1,\ldots,K\}. \tag{3.8}$$

[3] The complexity of the optimal detector will grow quickly in the number of data streams that are jointly decoded.

The IA conditions succinctly describe the alignment: the equations in (3.7) require all effective interfering channels to be zero, and the equations in (3.8) require that the effective desired channel have sufficient rank to decode d streams. The feasibility of IA is determined by the solvability of (3.7)–(3.8). The equations in (3.7) are a set of bilinear equations in $\{\mathbf{A}_k, \mathbf{V}_k\}$. The solvability of this type of polynomial system of equations can be analyzed using techniques from the mathematical field of algebraic geometry [CLO98].

In this chapter, we will use the *properness framework* of [YGJK10], together with *Bernstein's theorem* from algebraic geometry, to determine a necessary condition for the IA feasibility. We first define the notion of properness.

Definition 1 (Properness of symmetric systems [YGJK10])**.** *Let L_e and L_v denote the number of equations and number of complex variables, respectively, in* (3.7)*. Then the polynomial system of equations in* (3.7) *is proper iff $L_v \geq L_e$.*

In [YGJK10], the properness condition was shown to be necessary for IA feasibility for single-carrier MIMO systems with single-stream transmission. For this setting, the number of MSs K that can be served interference-free is determined by (3.5).

In our venture to find a necessary condition for IA feasibility for the space-frequency setting, we start by deriving the properness condition for the frequency-extended channels in (3.7) for arbitrary d.

Lemma 1. *The polynomial system of equations in* (3.7) *with channels from* (3.1) *is proper iff*

$$Kd\left((N+M)L_c - 2d\right) - Kd^2(K-1) - \left(L_c - \left\lceil \frac{d}{\min(N,M)} \right\rceil\right) \geq 0.$$

Proof. The number of equations in (3.7) is $L_e = Kd^2(K-1)$. In order to count the number of complex variables in (3.7), we need to use a parametrization with the fewest number of independent variables possible. In total, there are $K(N+M)L_cd$ coefficients in all filters. However, the number of independent variables in the polynomial system of equations is lower.

First, we notice that only the column spans of $\{\mathbf{A}_k\}$ and $\{\mathbf{V}_l\}$ matter in order to satisfy (3.7). There is therefore no loss in letting

$$\mathbf{A}_k = \begin{pmatrix} \mathbf{I}_d \\ \bar{\mathbf{A}}_k \end{pmatrix}, \quad \mathbf{V}_k = \begin{pmatrix} \mathbf{I}_d \\ \bar{\mathbf{V}}_k \end{pmatrix}, \quad \forall\, k \in \{1, \ldots, K\}, \tag{3.9}$$

which removes $2Kd^2$ variables. For some non-zero complex numbers $\beta^{(n)}$, $n = 1, \ldots, L_c$, let

$$\boldsymbol{\beta}_A = \text{blkdiag}\left(\beta^{(1)}\mathbf{I}_N, \ldots, \beta^{(L_c)}\mathbf{I}_N\right), \tag{3.10}$$

$$\boldsymbol{\beta}_V = \text{blkdiag}\left(\beta^{(1)}\mathbf{I}_M, \ldots, \beta^{(L_c)}\mathbf{I}_M\right). \tag{3.11}$$

Then notice that for $k \neq l$,

$$0 \overset{(a)}{=} \mathbf{A}_k^{\mathsf{H}} \mathbf{H}_{kl} \mathbf{V}_l = \mathbf{A}_k^{\mathsf{H}} \boldsymbol{\beta}_A^{-1} \boldsymbol{\beta}_A \mathbf{H}_{kl} \mathbf{V}_l \overset{(b)}{=} \mathbf{A}_k^{\mathsf{H}} \boldsymbol{\beta}_A^{-1} \mathbf{H}_{kl} \boldsymbol{\beta}_V \mathbf{V}_l = \widetilde{\mathbf{A}}_k^{\mathsf{H}} \mathbf{H}_{kl} \widetilde{\mathbf{V}}_l,$$

where (a) is due to (3.7) and (b) is due to the block diagonal structure of $\boldsymbol{\beta}_A$, $\boldsymbol{\beta}_V$ and $\mathbf{H}_{kl}$. That is, as similarly noted in [SBH11], any solution to (3.7) remains a solution after the transformation

$$\widetilde{\mathbf{A}}_k = \boldsymbol{\beta}_A^{-\mathsf{H}} \mathbf{A}_k, \ \forall k \in \{1, \ldots, K\}, \quad \widetilde{\mathbf{V}}_l = \boldsymbol{\beta}_V \mathbf{V}_l, \ \forall l \in \{1, \ldots, K\}. \tag{3.12}$$

Note that the condition in (3.8) is still satisfied after this transformation. Under this invariance, we can use the $\{\beta^{(n)}\}$ to further remove variables from the polynomial system. For instance, let

$$\widetilde{\mathbf{V}}_1 = \boldsymbol{\beta}_V \begin{pmatrix} \mathbf{I}_d \\ \bar{\mathbf{V}}_1 \end{pmatrix} = \begin{pmatrix} \beta^{(1)} \mathbf{I}_M & & \\ & \ddots & \\ & & \beta^{(L_c)} \mathbf{I}_M \end{pmatrix} \begin{pmatrix} \mathbf{I}_d \\ \bar{\mathbf{V}}_1 \end{pmatrix}. \tag{3.13}$$

Since the topmost d rows of $\widetilde{\mathbf{V}}_1$ and $\widetilde{\mathbf{A}}_1$ should not be altered, $L_c - \lceil \frac{d}{\min(N,M)} \rceil$ of the variables in $\widetilde{\mathbf{V}}_1$ can be fixed by selecting the $\{\beta^{(n)}\}$ appropriately. Subsequently, the number of variables in the polynomial system (3.7) is

$$L_v = K(N+M)L_c d - 2Kd^2 - \left(L_c - \left\lceil \frac{d}{\min(N,M)} \right\rceil \right), \tag{3.14}$$

and the lemma then follows from applying the properness condition $L_v \geq L_e$ from Definition 1. □

Although properness empirically often seems to be a sufficient indicator of IA feasibility [Gui10], the condition in Lemma 1 is not a necessary condition for IA feasibility for general d. By specializing to the $d = 1$ case however, we are able to show that the properness criterion is indeed necessary. Before stating the result, we remind the reader of the assumption that all channel coefficients are drawn from random variables, independent over antennas, subcarriers and cross-links (see Section 3.1).

Theorem 3.1. *For $d = 1$, a necessary condition for space-frequency IA feasibility is*

$$K \leq \begin{cases} (N+M) - 1, & L_c = 1 \\ (N+M)L_c - 2, & L_c \geq 2 \end{cases}. \tag{3.15}$$

Proof. For $d = 1$, after rearrangement, the properness condition in Lemma 1 is $K \leq \lfloor (N+M)L_c - 1 - \frac{L_c - 1}{K} \rfloor$, which is equivalent to (3.15). To see why this is a necessary condition for space-frequency IA feasibility when $d = 1$, we use an argument previously used in [YGJK10] and [SBH11].

Assume that (3.15) does not hold and thus $L_v < L_e$. Then by removing the last $L_e - L_v$ equations from (3.7), we end up with a L_v–by–L_v polynomial system of equations. In particular, this square system of equations is *generic* [CLO98], since all coefficients are i.i.d. random variables. Through Bernstein's theorem [CLO98, Ch. 7], it can be shown that this square generic polynomial system has a bounded number of non-zero solutions.

If there are zero solutions to the square polynomial system, there cannot be any solution to the full polynomial system in (3.7) either. Since (3.7) cannot be solved, space-frequency IA is not feasible.

If there is at least one solution to the square polynomial system, we study the remaining $L_e - L_v$ equations. These remaining equations have coefficients that are random variables independent of the coefficients of the square polynomial system. Therefore, any solution to the square polynomial system will not satisfy the eliminated $L_e - L_v$ equations with probability 1. Therefore, there are no solutions to the full polynomial system in (3.7) in this case either, and space-frequency IA is not feasible.

In conclusion, (3.15) is a necessary condition for the solvability of the full polynomial system in (3.7). Due to the construction of the receive filters and precoders in (3.9), the condition in (3.8) holds with probability 1 as well. Therefore, (3.15) is a a necessary condition for space-frequency IA feasibility. □

For the $d \geq 2$ case, the coefficients of the square polynomial system and the coefficients of the remaining $L_e - L_v$ equations might not be independent of each other. It might therefore be possible to find a solution[4] that holds for both the square polynomial system, as well as the remaining equations, even though the full system of equations is not proper.

The condition in Theorem 3.1 has the frequency-only condition in (3.6) as a special case. The condition is further compliant with the necessary part of the single-carrier condition in (3.5). From the feasibility perspective, it is clear that spatial and frequency dimensions are equivalent, save for a constant loss of one single-stream MS when using a frequency-extended channel instead of a single-carrier system.

3.2.1 Gain of Space-Frequency IA over Space-Only IA

For space-frequency precoding with $L_c \geq 2$, we know from Theorem 3.1 that the number of single-stream interference-free MSs served using IA must obey

$$K_{\text{S-F}} \leq (N + M)L_c - 2. \tag{3.16}$$

[4]This can be illustrated with the following example. The $K = 3$ user 2×2 scenario with $L_c = 1$ and $d = 1$ is IA feasible [CJ08, YGJK10, RLL12]. Thus, the same scenario with $L_c = 2$ and $d = 2$ is IA feasible through space-only precoding. The properness condition does not hold for this frequency-extended scenario, and since space-only precoding is a special case of space-frequency precoding (with block diagonal precoders), this means that the properness condition for general d is not a necessary condition.

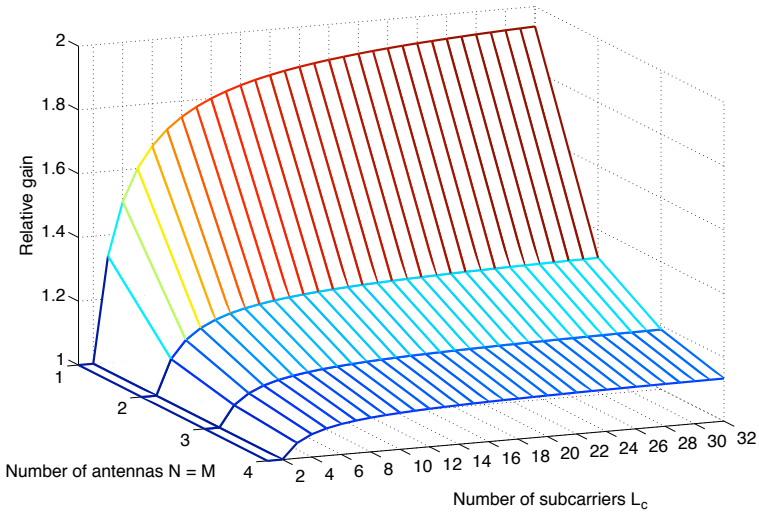


Figure 3.1. Relative gain bound of space-frequency IA over space-only IA, in terms of number of interference-free MSs served for the single-stream scenario.

With space-only precoding, we know from (3.5) that the number of single-stream interference-free MSs that can be served per subcarrier using IA is $N + M - 1$. Thus, the total number of MSs that can be served over the L_c subcarriers in a space-only configuration is

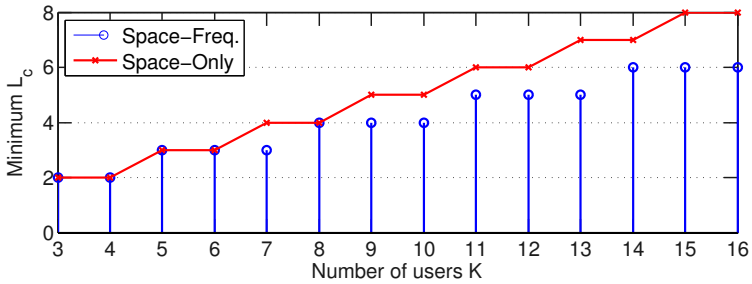
$$K_{\text{S-O}} = (N + M - 1)L_c. \tag{3.17}$$

The relative gain of space-frequency IA over space-only IA, in terms of the number of interference-free single-stream MSs served, can then be bounded as

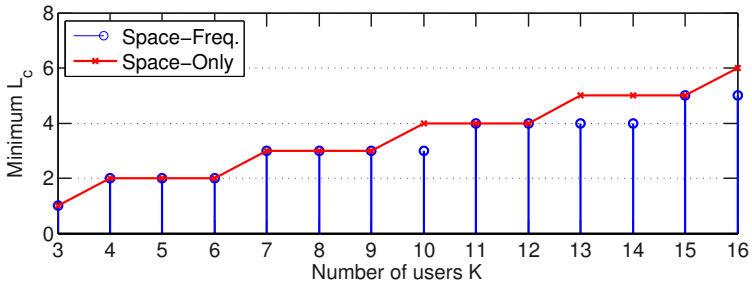
$$\frac{K_{\text{S-F}}}{K_{\text{S-O}}} \leq \frac{(N+M)L_c - 2}{(N+M-1)L_c} = 1 + \frac{1 - 2/L_c}{N+M-1} \leq 1 + \frac{1}{N+M-1}, \tag{3.18}$$

assuming $L_c \geq 2$. The relative gain improves with increasing L_c, but decreases with increasing number of antennas N, M. If $L_c = 2$, the second term in the second last expression of (3.18) is zero. Therefore, no relative gain is seen unless $L_c \geq 3$.

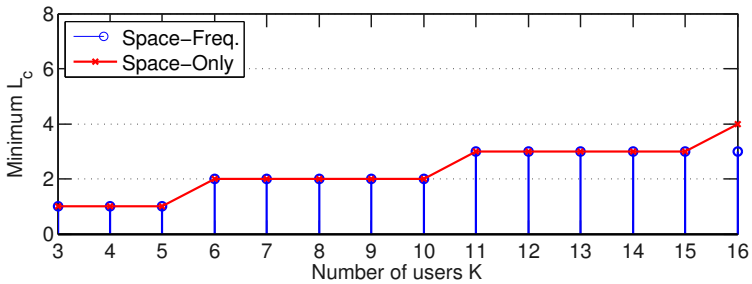
A plot of the maximum relative gain for a $N = M$ scenario can be seen in Figure 3.1. It is clear that the relative gain is the highest for the single-antenna frequency-only case, and that the relative gain drops off as the number of antennas is increased. This is because in the single-antenna scenario, only one MS can be served interference-free on each subcarrier. In the multi-antenna scenario however, using space-only IA, multiple interference-free MSs can be served on each subcarrier. This results in the decreasing relative gain of the space-frequency IA method.



(a) Number of antennas $N = 1, M = 2$



(b) Number of antennas $N = M = 2$



(c) Number of antennas $N = M = 3$

Figure 3.2. Minimum number of subcarriers needed to serve K single-stream MSs using space-frequency IA or space-only IA.

Minimum Number of Subcarriers Needed

Reversely, assuming that K single-stream MSs should be served interference-free, the feasibility conditions can be used to find the minimum number of subcarriers L_c necessary. For three different combinations of N and M, this is plotted in Figure 3.2 on the preceding page. It is clear that as the number of antennas grow, a larger number of MSs must be served in order for space-frequency to have a gain over space-only. For example, for $N = M = 2$, at least $K = 10$ MSs must be served before space-frequency IA requires fewer subcarriers than space-only IA. The reason for this effect is the same as the reason for the decreasing relative gain of space-frequency IA over space-only IA, as described in the previous section.

3.3 Aspects of Correlation and Feasibility

In order to evaluate the sum rate performance of the proposed space-frequency precoding, we first need to discuss some aspects of the combined signal space. The first issue is that in the system model in Section 3.1, the channel matrices at different subcarriers were assumed to be independent. This will in general not be the case in real-world scenarios. Therefore, we here propose a simple method for minimizing any subcarrier correlations for the subcarriers that partake in space-frequency precoding. The second issue is that of performing user selection for space-only precoding. Since each subcarrier will only be able to accommodate a certain number of MSs, these need to be selected by a user selection algorithm. Here, we will propose a simple greedy heuristic for this problem.

3.3.1 Alignment Groups

Real-world channels generally have a coherence bandwidth [TV08, Ch. 2], within which the subcarriers may be highly correlated. As required by the system model in Section 3.1, the subcarriers should be independent however. In this section we introduce the concept of *alignment groups*, to ensure that the subcarriers in one group are approximately uncorrelated. The L_c subcarriers are split into L_g alignment groups, each containing $L_f = L_c/L_g$ subcarriers. Precoding is performed orthogonally over the alignment groups, and the subcarriers in each alignment group are selected to minimize correlations. The alignment group structure that will be used here is shown in Figure 3.3 on the following page. The subcarriers belonging to one group are equidistant, and uniformly spread out over the available subcarriers. This ensures that neighbouring subcarriers belong to different alignment groups, and subcarrier correlations within one group are reduced.

3.3.2 User Selection for Space-Only Precoding

For space-frequency precoding, all $K_{\text{S-F}}$ MSs are served jointly over all subcarriers. For space-only precoding, if $K_{\text{S-O}} > N + M - 1$, not all MSs can be active on all

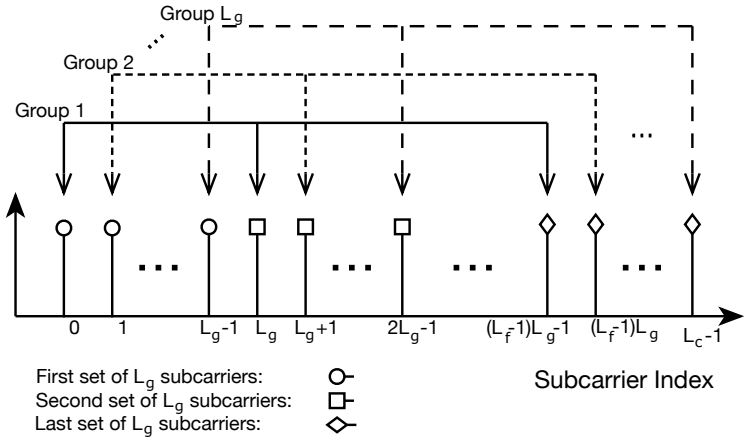


Figure 3.3. Alignment group structure. Each alignment group contains L_f subcarriers, and there are L_g alignment groups.

Algorithm 3.1 Greedy User Selection for Space-Only Precoding

Input: $\gamma_k^{(n)} = \left\| \mathbf{H}_{kk}^{(n)} \right\|_{\mathrm{F}}^2, \quad \forall\, k \in \{1, \ldots, K\},\, n \in \{1, \ldots, L_c\}$

Variables:

$\tilde{s}_k^{(n)} \in \{0, 1\}$. Equals 1 if MS k is a candidate for scheduling on subcarrier n.

$L^{(n)}$ = number of scheduled MSs on subcarrier n.

L_k = number of subcarriers which MS k is scheduled on.

Let $\tilde{s}_k^{(n)} \leftarrow 1,\ \forall\, k \in \{1, \ldots, K\},\, n \in \{1, \ldots, L_c\}$

repeat

  Find $(k^*, n^*) = \underset{\{(k,n)\,:\,\tilde{s}_k^{(n)}=1\}}{\operatorname{argmax}} \gamma_k^{(n)}$

  Schedule MS k^* on subcarrier n^* and let $\tilde{s}_{k^*}^{(n^*)} \leftarrow 0$

  // Check that this subcarrier is not overloaded w.r.t. to IA feasibility

  if $L^{(n^*)} = N + M - 1$ **then**

    Let $\tilde{s}_k^{(n^*)} \leftarrow 0, \quad \forall\, k \in \{1, \ldots, K\},$

  end if

  // Ensure the MS is not scheduled on more subcarriers than its fair share.

  if $KL_{k^*} \geq L_c(N + M - 1)$ **then**

    Let $\tilde{s}_{k^*}^{(n)} \leftarrow 0, \quad \forall\, n \in \{1, \ldots, L_c\}$

  end if

until $\gamma_k^{(n)} = 0, \quad \forall\, k \in \{1, \ldots, K\},\, n \in \{1, \ldots, L_c\}$

subcarriers. A user selection algorithm should then be employed, which picks out the MS combinations that are suitable for being served together on the different subcarriers. The selection algorithm should make sure that the IA feasibility conditions are not violated, and should at the same time try to enforce some form of fairness between MSs, in terms of the number of data streams allocated.

Since the user selection problem is combinatorial, and hence hard to solve to optimality, we propose a simple greedy approach in Algorithm 3.1 on the preceding page. In this algorithm, we let the user selection metric be the direct channel strength $\gamma_k^{(n)} = \left\| \mathbf{H}_{kk}^{(n)} \right\|_{\mathrm{F}}^2$. We let the indicator variable $\tilde{s}_k^{(n)} \in \{0, 1\}$ denote whether MS k is a candidate for being scheduled on subcarrier n, and then we sequentially schedule the available MS-subcarrier pairs based on their $\gamma_k^{(n)}$. For each scheduled MS-subcarrier pair, we let $\tilde{s}_k^{(n)} \leftarrow 0$, to ensure that this pair is not incorrectly considered for scheduling again. We verify that the IA feasibility condition for each subcarrier is not violated. We also ensure a simple form of fairness, such that each MS is not scheduled on more than its fair share of the total number of subcarriers. The method proposed in Algorithm 3.1 is very crude; a more elaborate scheduler should take into account the strength of the cross-links, which MSs are spatially compatible for being served together, etc.

3.4 Performance Evaluation

Although the derived IA feasibility conditions give insights into the performance of different types of IA systems, the optimization metric that we are mainly concerned with in this thesis is the sum rate. In this section, we investigate the sum rate performance[5] of space-frequency precoding, space-only precoding and frequency-only precoding. We use numerical simulation, and study two scenarios: one dense indoor scenario, and one urban multicell outdoor scenario. In both cases, we use measured channels[6] in the performance simulator, in order to achieve real-world correlations and path losses.

For the precoding, we use the WMMSE, MaxSINR and MinWLI algorithms as detailed in Section 2.3.3.

3.4.1 Frequency-Only IA: Outdoors Scenario

As a motivating example of the benefits of coordinated precoding, we first study a $K = 3$ user SISO interference channel with frequency-only precoding. The performance evaluation is performed using channel measurements for an outdoors setting with single-antenna transmitters. The channel measurements were taken in Kista, Northern Stockholm. A map of the measurement area is shown in Figure 3.4. The

[5] In this chapter, all MSs have unit data rate weights such that $\alpha_{i_k} = 1$ for all i_k.

[6] All measurements were performed by Ericsson, and the measured channels were provided to the author by Ericsson through the HiATUS FP7 EU project.

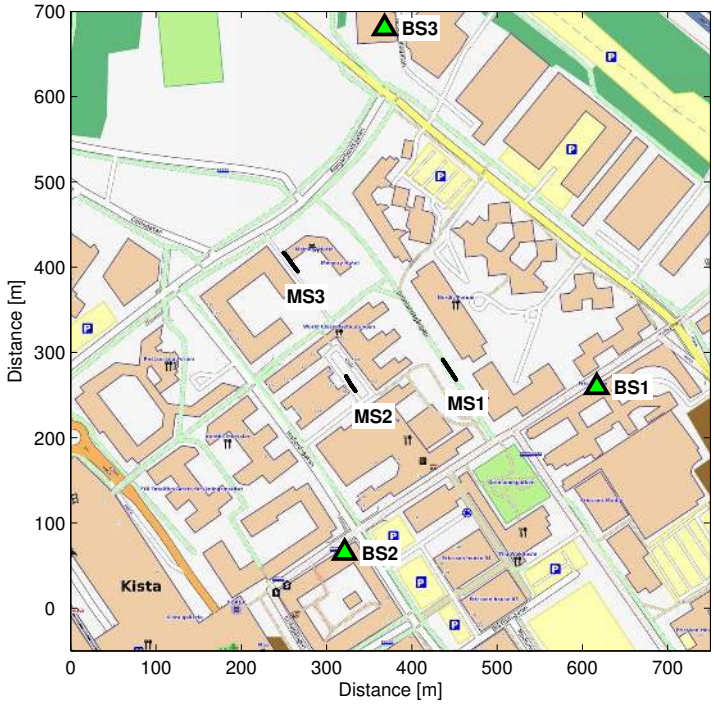


Figure 3.4. Map over Kista with BS locations marked with triangles, and possible MS locations indicated by black segments. The map is ©www.openstreetmap.org contributors, CC-BY-SA, http://creativecommons.org/licenses/by-sa/2.0

average building height was 25 m and the transmit antennas were located a couple of meter above the average rooftop level [MSKF09]. One centrally located transmit unit was used for transmission from three geographically separated sites. The inter-site distance was around 400-600 m, and the antenna sites were connected to the central transmit unit using fiber optical cables and RF-optocouplers. This enabled coherent transmission from the sites, and thus coherent channel estimation at the receiver. The receiver consisted of a measurement van that was driven along a predefined measurement trajectory. On the roof of the van, two electric dipoles and two magnetic dipoles (loops) were used as receiving antennas (9λ distance). The transmitter and receiver were part of a purpose-built channel sounder based on an LTE-like OFDM-based design [SA08]. Pilot signals were transmitted at the 2.6 GHz band, over a bandwidth of 20 MHz, and the effective 4×3 MIMO channel was estimated at the receiver. The channel was measured at 190 Hz in 432 subcarriers, which was sufficient to capture the fast fading dynamics [SA08]. The noise floor of the channel impulse response estimates was around 30 dB lower than the peak values.

Multiuser Emulation and Simulation Setup

The single-antenna channels were obtained from the measurement data by selecting one of the antennas on the measurement van as the active receive antenna. The $K = 3$ single-antenna MSs could then be emulated by spatially separating them along different segments of the measurement route. The possible locations of MS1, MS2 and MS3 are shown as black segments in Figure 3.4. Channel realizations were drawn with uniform probability along the defined segments. By assuming channel stationarity in time, the measured channels from the three different segments were then used to emulate the 3-user interference channel.

Although 432 measured subcarriers were available, we only used $L_c = 48$ subcarriers for the performance evaluation. The 48 subcarriers were selected uniformly over the available 432 subcarriers, such that the spacing between selected subcarriers was 0.4 MHz. In order to limit the effect of correlation between subcarriers, the subcarriers were divided into $L_g = 16$ alignment groups as described in Section 3.3.1. Each group thus consisted of $L_f = L_c/L_g = 3$ subcarriers. The precoding was performed independently over the alignment groups. For the $L_f = 3$ subcarriers in each alignment group, it is known from (2.19) and (2.20) that a feasible data stream allocation for IA is $\mathbf{d} = \begin{pmatrix} 2 & 1 & 1 \end{pmatrix}$.

The sum rate results were averaged over a number of channel realizations. The iterative algorithms (WMMSE/MaxSINR/MinWLI) were initialized with truncated DFT matrices corresponding to the data stream allocation $\mathbf{d}$. The algorithms were run for 5000 iterations. For fairness between MSs, the data stream allocation $\mathbf{d}$ was cyclically shifted for each new network realization. In addition to running the standard algorithms from Section 2.3.3, an SINR balancing method [Ben02] was also used. In each iteration of that method, the precoders were found as the solution to an SINR balancing problem, where each MS was weighted equally. This method relaxes the power constraint such that $\sum_l \mathbb{E}\left(\|\mathbf{V}_l\mathbf{x}_l\|^2\right) \leq KPL_c$.

As baselines, we use the original frequency-only IA solution from [CJ08], as well as an optimized version from [SPLL10]. For the latter, the IA solution that approximately maximized the chordal distance between the signal and interference subspace, at all MSs, was selected. We further used orthogonalization in the frequency domain, i.e. frequency-division multiple access (FDMA) with a reuse factor of three as well as uncoordinated transmission (reuse factor one). All BSs used the same transmit power P, and all MSs had the same noise power σ^2. The results were evaluated as a function of average SNR

$$\overline{\mathrm{SNR}} = \frac{P}{\sigma^2}\frac{1}{K}\sum_{l=1}^{K}\mathbb{E}\left(\left|h_{ll}^{(n)}\right|^2\right), \tag{3.19}$$

where the mean value was estimated using the sample mean over all channel realizations used.

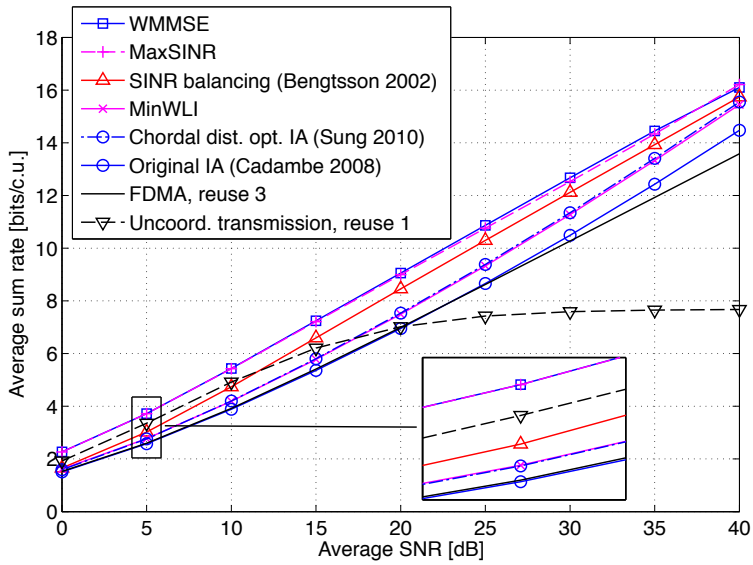


Figure 3.5. Sum rate performance for measured outdoors scenario (averaged over 500 network realizations).

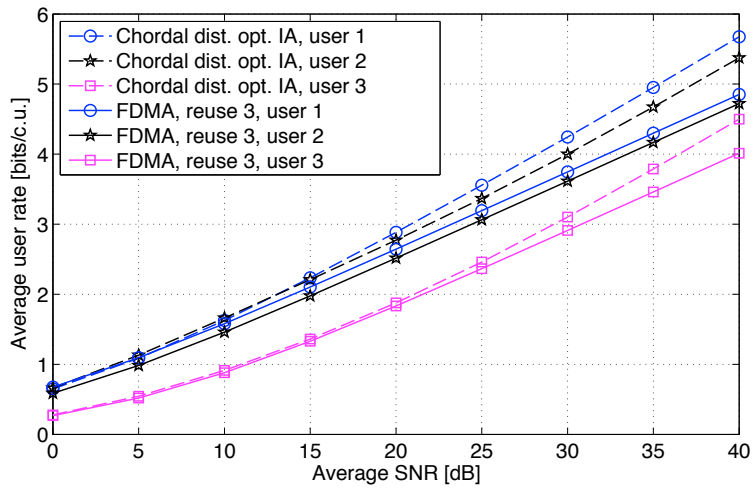


Figure 3.6. User rate performance for measured outdoors scenario (averaged over 500 network realizations).

Results

The sum rate results, as a function of $\overline{\text{SNR}}$, are shown in Figure 3.5 on the facing page. The WMMSE method is always at least as good as all other methods, with MaxSINR being a close second. The WMMSE and MaxSINR methods have particularly good performance at low SNR, but they also outperform the pure IA solution from MinWLI in the high-SNR regime. All coordinated precoding methods outperform FDMA, as well as uncoordinated transmission for sufficiently high SNR. Uncoordinated transmission performs well at low SNR, indicating that the cells are effectively decoupled then, and the thermal noise is the main performance-limiting factor.

In Figure 3.6 on the preceding page, the individual MS data rates are shown as a function of $\overline{\text{SNR}}$. The IA solution clearly achieves a better high-SNR slope than the FDMA solution, as expected from the motivation of using IA in the first place. Interestingly, the IA solution also beats FDMA in the low-SNR regime.

Estimated cumulative distribution functions (CDFs) of the MS rates are shown in Figure 3.7 on the following page. The plateauing behaviour of the curves for the coordinated precoding methods are due to the cyclic shift of the extra data stream that is always allocated to one of the MSs. The WMMSE method ensures that the two strong MSs have high data rates, at the expense of the weak MS. The SINR balancing method, on the other hand, ensures that all MSs have a minimum data rate. This can be seen from the fact that MS 1 and 2 have similar CDFs, whereas the CDF for MS 3 is shifted to the right slightly.

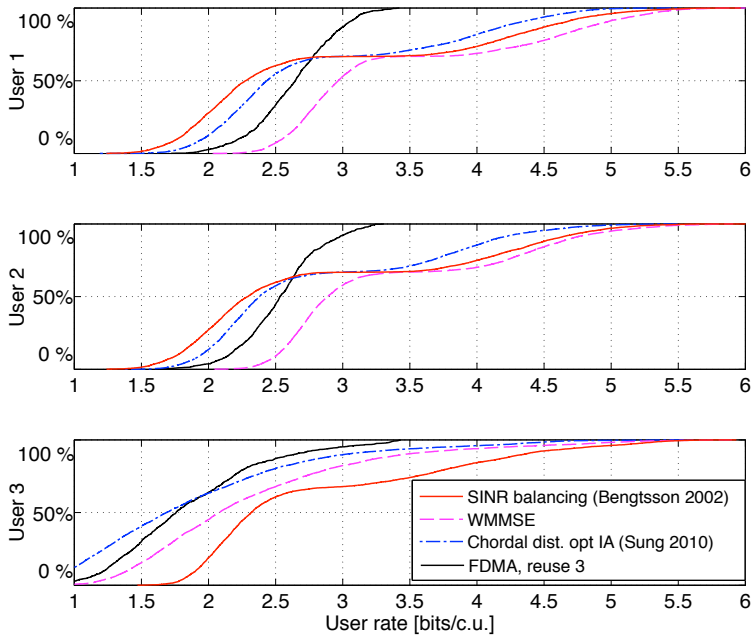


Figure 3.7. Empirical CDFs of the outdoors MS rates for $\overline{\mathrm{SNR}} = 20\,\mathrm{dB}$. The estimates were obtained from 5000 network realizations.

3.4.2 Space-Frequency IA: Indoors Scenario

We now study a full space-frequency precoding system. Again measurements are applied, but this time from an indoors scenario. As a comparison, we also perform simulations using synthetically generated channels.

The studied setting is a $K = 10$ scenario where $N = M = 2$ and $L_c = 30$. We compare space-frequency precoding with $L_g = 10$ alignment groups to space-only precoding where all 30 subcarriers are used independently, together with the user selection heuristic in Algorithm 3.1 on page 50. For space-frequency precoding, all MSs are served one data stream in all groups, i.e. the maximum sum DoF is $K/L_f = 10/3$. For space-only precoding with user selection, three single-stream MSs are accommodated per subcarrier, giving a maximum sum DoF of 3. For space-frequency precoding, the power is uniformly allocated over the alignment groups, whereas for space-only precoding, the power is uniformly allocated over the subcarriers.

Sum rate performance is evaluated for the standard methods from Chapter 2.3.3, and averaged over 100 Monte Carlo realizations. The pure IA method MinWLI is run until the relative interference leakage satisfies

$$\frac{\mathrm{Tr}\left(\sum_k \sum_{l\neq k} \mathbf{A}_k^H \mathbf{H}_{kl} \mathbf{V}_l \mathbf{V}_l^H \mathbf{H}_{kl}^H \mathbf{A}_k\right)}{\mathrm{Tr}\left(\sum_{k,l} \mathbf{H}_{kl} \mathbf{V}_l \mathbf{V}_l^H \mathbf{H}_{kl}^H\right)} \leq 10^{-9}. \tag{3.20}$$

MaxSINR is not guaranteed to converge, so we run it for 2000 iterations. WMMSE converges monotonically, and we run it until the relative change in sum rate is less than 10^{-5}. All methods are initialized with truncated DFT matrices, and the performance baselines are TDMA and uncoordinated transmission (with and without water filling).

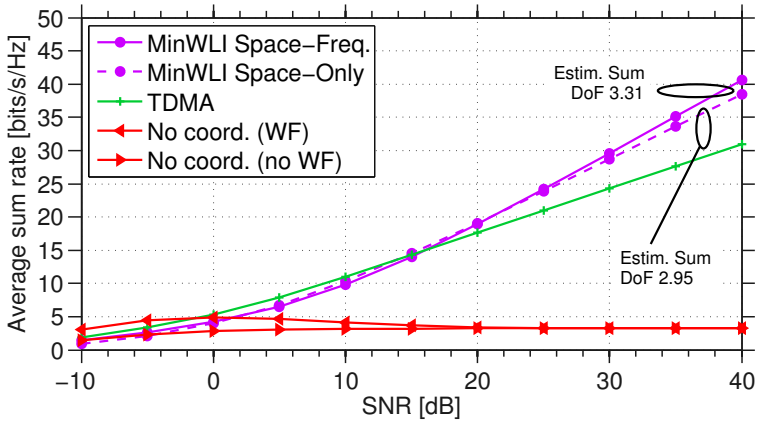
Synthetic Channels

For the synthetic channels, we used a tapped delay-line with L_t taps to generate a block fading frequency-selective channel. We assume rich scattering and an exponentially decaying power-delay profile such that the channel impulse response in one transmission period is

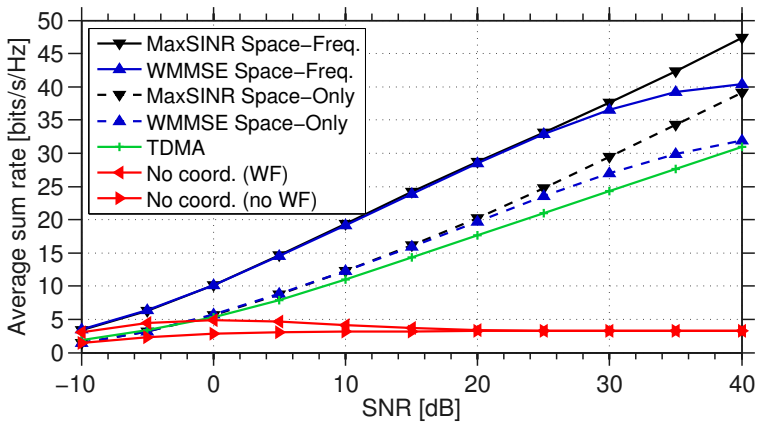
$$\mathbf{H}_{kl}[n] = c\sqrt{e^{-\frac{t}{Wt_0}}}\,\mathbf{H}_{kl}^w[n], \quad n = 0, \ldots, L_t - 1.$$

The coefficients of $\mathbf{H}_{kl}^w[n]$ are i.i.d. $\mathcal{CN}(0,1)$ and are constant within one transmission period, but vary independently between periods. The channels are normalized to $\sum_{t=0}^{L_t-1} \mathbb{E}\left(\|\mathbf{H}_{kl}[n]\|_{\mathrm{F}}^2\right) = NM$ using c, and we assume a bandwidth delay spread product $Wt_0 = 1.5$ and $L_t = 6$ taps. The transmit power was P for all transmitters, and the noise power was σ^2 for all MSs. Performance is evaluated as a function of

$$\mathrm{SNR} = \frac{P}{L_c \sigma^2}. \tag{3.21}$$



(a) Performance of approximate IA solution from MinWLI for the synthetic channels.



(b) Performance of sum rate maximization techniques for the synthetic channels.

Figure 3.8. Sum rate performance for synthetic channel.

The sum rate performance can be seen in Figure 3.8 on the facing page. In Figure 3.8a, the high-SNR slopes of space-frequency IA and space-only IA is compared, using the MinWLI algorithm. The estimated sum DoF for space-frequency IA is 3.31, close to the theoretical 10/3. The estimated sum DoF for space-only IA is 2.95, close to the theoretical 3. Comparing space-frequency IA to space-only IA, it is clear that the gain in terms of sum DoF is present, but it is obvious that the gain in terms of sum rate is very small. In Figure 3.8b, the performance of the other precoding methods is shown. WMMSE and MaxSINR performs similarly, but the high-SNR slope of WMMSE goes to zero due to its slow convergence in the high-SNR regime. It is clear that the space-frequency precoding methods in Figure 3.8b are shifted to the left with around 10 dB, compared to the space-only precoding methods. This can be interpreted as a 'power gain', or 'coding gain', of the space-frequency precoding methods. This is a consequence of there being more IA solutions to choose from in the enlarged space-frequency search space, compared to the smaller space-only search space. The conclusion is that space-frequency precoding does give an improved DoF, but that the large sum rate improvement is due to a power gain.

Measured Channels

In order to evaluate performance for real-world path losses and channel correlations, we again use channel measurements [WAF+12, WAH+13]. The measurements were taken along a 70 m long office building corridor (see map in Figure 3.9 on the next page) using an 20 MHz LTE-Advanced testbed with a carrier frequency of 2.7 GHz. The measurement antenna array consisted of four dual-polarized patch elements linearly arranged with 0.5λ spacing. Measurements were taken from three 8-antenna base stations (BSs), one located in the middle of the corridor (square, 2.3λ antenna spacing, 'indoor omni' in Figure 3.9), one located at the end of the corridor (square, 1.15λ horizontal spacing, 1.85λ vertical spacing, 'indoor panel' in Figure 3.9) and one located 65 m away on an outdoor wall facing the corridor (linear, 1.15λ antenna spacing, 'outdoor pico' in Figure 3.9). The channels were measured in 100 frequency points, out of which we select 30 equidistant points spaced 0.6 MHz apart. A typical channel response is shown in Figure 3.10 on the following page.

In order to emulate a $K = 10$ scenario, we select ten spatially separated measurement route segments, similar to the procedure in Section 3.4.1. We select one dual-polarized patch antenna from the receive array (i.e. $N = 2$), all antennas from the first BS and 6 antennas from the two other BSs. To achieve an $M = 2$ scenario, the selected BS antennas are split into 10 virtual 2-antenna transmitters. The simulated transmit powers P_k were 6 dBm for the virtual transmitters corresponding to the indoor central BS, 10 dBm for the virtual transmitters corresponding to the indoor peripheral BS and 30 dBm for the virtual transmitters corresponding to the outdoor BS. These transmit powers were selected such that the signal-to-interference ratios of the emulated cross-links were approximately 0–15 dB. Since

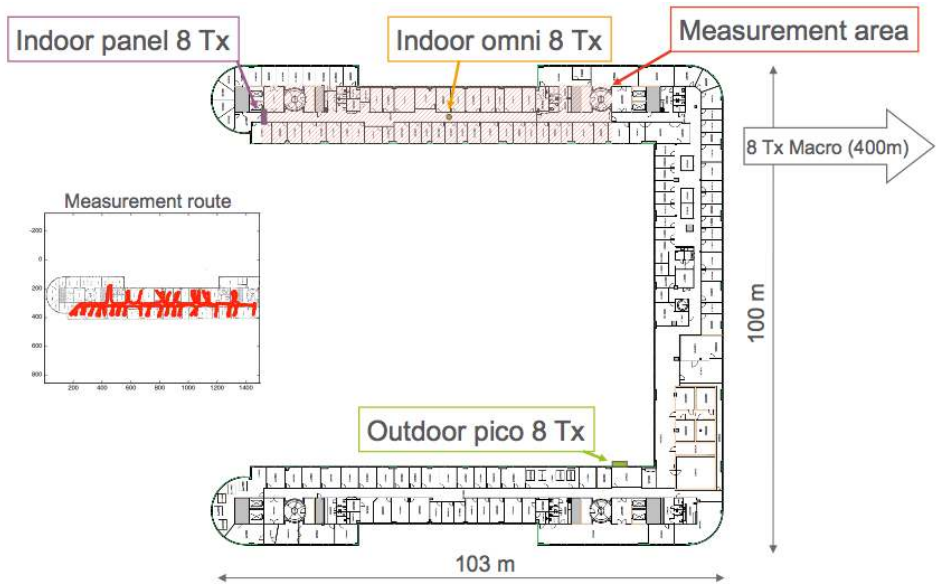


Figure 3.9. Map over indoors measurement location. Reproduced with permission from Ericsson Research.

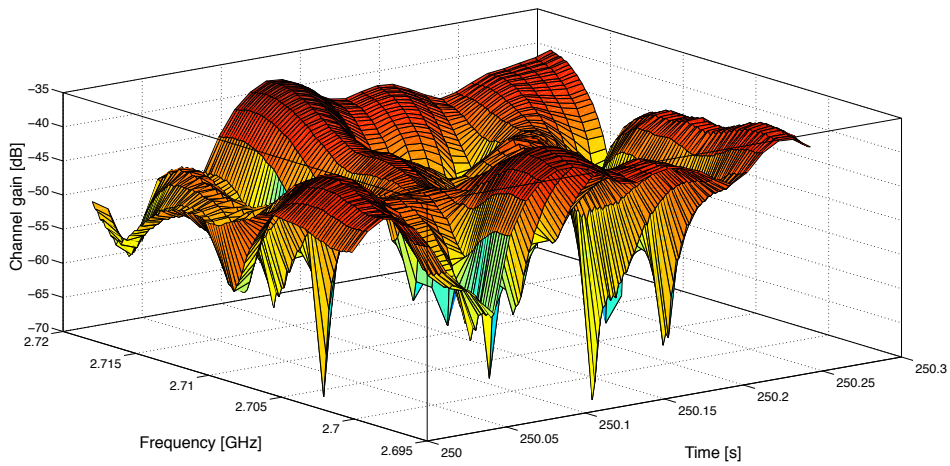


Figure 3.10. Typical channel frequency response over 300 ms from the indoors measurements.

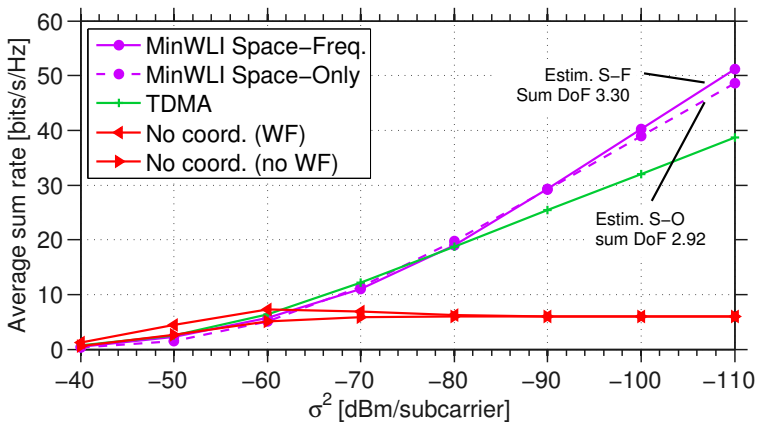
the different MSs all have different SNRs, due to their different channel conditions, we vary the received signal quality by varying the noise power σ^2. This noise power was the same for all MSs.

The sum rate results for the measured channels are shown in Figure 3.11 on the next page. The DoF gain from space-frequency IA over space-only IA in Figure 3.11a is similar to the gain for the synthetic channel. Again, the sum rate gain is very small. For the practical coordinated precoding methods in Figure 3.11b, a power gain of around 10 dB is again visible. The conclusion for the measured channels are thus the same as for the synthetic channel: although space-frequency precoding can give rise to an improved sum DoF, the large practical gain lies in the power gain.

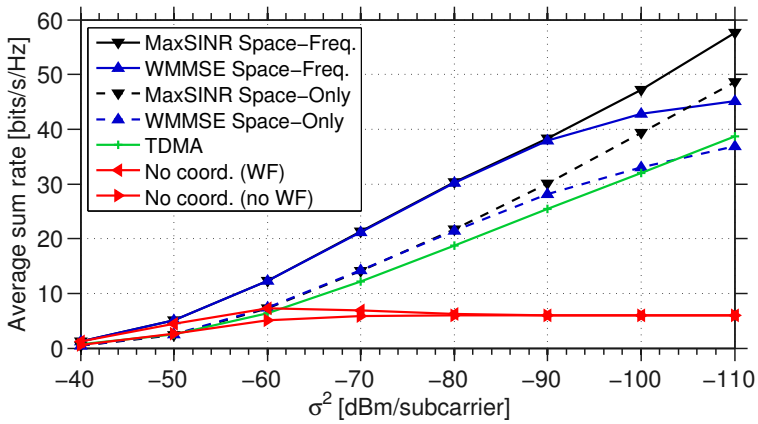
3.5 Conclusions

In this chapter, necessary conditions for space-frequency IA were derived. The theoretical gain over space-only IA was studied in terms of number of interference-free data streams served. The gain was shown to increase with the number of subcarriers L_c, but decrease with the number of antennas N and M. For the frequency-only outdoors sum rate simulations, it was clear that coordinated precoding did give a performance boost over traditional multiple access methods as such as FDMA and uncoordinated transmission. For the space-frequency indoors sum rate simulations, the conclusion was that space-frequency can indeed give an improved sum DoF performance, but the main practical benefit from applying space-frequency precoding is a large power gain. This power gain only materialized for the practical precoding methods WMMSE and MaxSINR, but not for the IA surrogate method MinWLI.

Altogether, space-frequency precoding can be an interesting approach for improving performance in interference networks, but not mainly due to its increased DoF.



(a) Performance of approximate IA solution from MinWLI for the measured channels. In the figure, 'S-F' corresponds to space-frequency IA, and 'S-O' corresponds to space-only IA.



(b) Performance of sum rate maximization techniques for the measured channels.

Figure 3.11. Sum rate performance for indoors measured channel.

Chapter 4

Distributed CSI Acquisition and Coordinated Precoding

In the previous chapter we focused on pure IA and its feasibility. The numerical results in Section 3.4 showed that directly trying to solve the weighted sum rate problem, through the WMMSE algorithm, in general yielded superior results to the pure IA methods. In this chapter, we therefore turn our attention to the weighted sum rate problem (2.37) under per-BS sum power constraints. In order to find methods amenable for practical implementation, we will investigate distributed solutions. Due to its shown excellent performance, and mathematical tractability, we base the resource allocation on the WMMSE algorithm [SRLH11]. This algorithm requires local information about effective channels and covariance matrices in each iteration. We here denote this information as *channel state information* (CSI). We take a systems perspective and propose methods for acquiring the necessary CSI at the involved nodes in a distributed fashion. Combining a proposed robustified version of the WMMSE algorithm with the proposed distributed CSI acquisition yields a robust and distributed joint coordinated precoding and CSI acquisition system.

Much of the literature on distributed coordinated precoding assumes perfect access to the CSI. In [SRLH11] for example, it is not directly evident how the CSI should be acquired for the WMMSE algorithm. In this chapter, we therefore propose distributed CSI acquisition methods to be coupled with the WMMSE algorithm. First, we succinctly describe what CSI is required by the nodes of the network to perform one iteration of the WMMSE algorithm. Many approaches can be imagined for obtaining this necessary CSI, e.g. using various combinations of channel estimation, feedback, signaling, backhaul, etc. Based on channel estimation and feedback, we propose three CSI acquisition methods. The methods correspond to different tradeoffs between signaling, feedback, and backhaul use. The key component is the effective channel estimation. Based on downlink pilot transmissions, the receivers are able to obtain local CSI for both desired and inter-

fering effective channels [SBH13]. Assuming time-division duplex (TDD) operation, and perfectly calibrated transceivers [GSK05, RBP+13], similar channel estimation can be performed in the uplink.

Second, it is shown that naïvely coupling the WMMSE algorithm with the distributed CSI acquisition methods leads to poor performance. This is since the original WMMSE algorithm was not designed with robustness against imperfect CSI in mind. We therefore propose a robustified WMMSE algorithm, which retains the distributedness of the original algorithm. We derive the local worst-case WMMSE optimization problems, and show that their solution structure is that of diagonal loading. The optimal amount of diagonal loading unfortunately depends on an unknown quantity. Therefore, we propose an implicit procedure for selecting the diagonal loading level for the precoders. At the receivers, we show an inherent property of the receive filters and MSE weights obtained from the WMMSE algorithm with perfect CSI. When this property is enforced onto the filters with imperfect CSI, this results in diagonally loaded filters, thereby robustifying them.

4.1 System Model

The system model in this chapter is the interfering broadcast channel from (2.11) on page 19. The CSI is not known a priori at the transceivers, and must therefore be estimated. In the interfering broadcast channel, the received downlink signal for the kth MS served by the ith BS is

$$\mathbf{y}_{i_k} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k}\mathbf{x}_{i_k} + \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{x}_{j_l} + \mathbf{z}_{i_k}. \tag{4.1}$$

The full details of the interfering broadcast channel models are available in Section 2.1.3. In particular, we however remind the reader that the downlink interference plus noise covariance matrix for MS i_k is

$$\mathbf{\Phi}_{i_k}^{\mathrm{i+n}} = \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}} + \sigma_{i_k}^2\mathbf{I} \tag{4.2}$$

and the corresponding received downlink signal covariance matrix is

$$\mathbf{\Phi}_{i_k} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}} + \mathbf{\Phi}_{i_k}^{\mathrm{i+n}} = \sum_{(j,l)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}} + \sigma_{i_k}^2\mathbf{I}. \tag{4.3}$$

The data rate weight $\alpha_{i_k} \in [0, 1]$ for MS i_k is known at the serving BS i.

In addition to downlink transmission, communication also takes place in the uplink. Since the downlink is described by the interfering broadcast channel, the uplink is described by the *interfering multiple access channel*. We will in the following assume that the system operates in a perfectly time-synchronized TDD mode, and that the corresponding radio hardware is perfectly calibrated

[GSK05, RBP+13]. Under this assumption, the cascade of transmit filter, wireless channel, and receiver filters are reciprocal and the uplink channel from MS j_l to BS i is[1] $\overleftarrow{\mathbf{H}}_{j_l i} = \mathbf{H}_{j_l i}^{\mathsf{T}}$. Let $\overleftarrow{\mathbf{s}}_{i_k}^* \sim \mathcal{CN}\left(\mathbf{0}, \overleftarrow{\mathbf{\Xi}}_{i_k}\right)$ be the uplink transmitted signal from MS i_k and $\overleftarrow{\mathbf{z}}_i^* \sim \mathcal{CN}\left(\mathbf{0}, \varsigma_i^2 \mathbf{I}_{M_i}\right)$ be the receiver noise at BS i. The received uplink signal at BS i is then modeled as

$$\overleftarrow{\mathbf{y}}_i^* = \sum_{k=1}^{K_i} \mathbf{H}_{i_k i}^{\mathsf{T}} \overleftarrow{\mathbf{s}}_{i_k}^* + \sum_{j \neq i} \sum_{l=1}^{K_j} \mathbf{H}_{j_l i}^{\mathsf{T}} \overleftarrow{\mathbf{s}}_{j_l}^* + \overleftarrow{\mathbf{z}}_i^* \tag{4.4}$$

For convenience, we will work with the complex conjugate version of the received signal in (4.4). The model we will use for the uplink is thus

$$\overleftarrow{\mathbf{y}}_i = (\overleftarrow{\mathbf{y}}_i^*)^* = \sum_{k=1}^{K_i} \mathbf{H}_{i_k i}^{\mathsf{H}} \overleftarrow{\mathbf{s}}_{i_k} + \sum_{j \neq i} \sum_{l=1}^{K_j} \mathbf{H}_{j_l i}^{\mathsf{H}} \overleftarrow{\mathbf{s}}_{j_l} + \overleftarrow{\mathbf{z}}_i. \tag{4.5}$$

We will assume that the MSs have individual sum power constraints such that

$$\mathbb{E}\left(\|\overleftarrow{\mathbf{s}}_{i_k}\|_{\mathrm{F}}^2\right) = \mathrm{Tr}\left(\overleftarrow{\mathbf{\Xi}}_{i_k}\right) \leq \overleftarrow{P}_{i_k}, \quad \forall\, i \in \{1, \ldots, I\},\, k \in \{1, \ldots, K_i\} \tag{4.6}$$

holds on average per transmitted symbol. For the model in (4.5), the uplink *signal plus interference covariance matrix* for BS i is

$$\mathbf{\Gamma}_i = \sum_{(j,l)} \mathbf{H}_{j_l i}^{\mathsf{H}} \overleftarrow{\mathbf{\Xi}}_{i_k} \mathbf{H}_{j_l i}^{\mathsf{H}}. \tag{4.7}$$

The MSs will estimate their local CSI using downlink pilot transmissions from the BSs. Conversely, the BSs will estimate their local CSI using uplink pilot transmissions from the MSs. Due to the assumed perfect reciprocity of the channel, this will provide the BSs with the information they need to form the precoders.

For the resource allocation, the WMMSE algorithm (see Section 2.3.3 on page 30) will be used. As will be shown below, the WMMSE algorithm is an example of a distributed resource allocation method, which only requires local CSI at the participating nodes.

4.1.1 WMMSE Algorithm with Per-BS Power Constraints

In order to lighten the forthcoming exposition, we first introduce some shorthands for the quantities involved in the WMMSE algorithm. For MS i_k, we define a *weighted receive filter* as $\mathbf{U}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and denote the *effective downlink channel* as $\mathbf{F}_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$. The *receive filter* can then be written as $\mathbf{A}_{i_k} = \mathbf{\Phi}_{i_k}^{-1} \mathbf{F}_{i_k}$. Similarly, at BS i serving MS i_k, we have the *precoder* $\mathbf{V}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{B}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and the

[1] The arrow $\overleftarrow{\cdot}$ denotes uplink quantities.

Table 4.1. Summary of CSI quantity shorthands

Downlink	$\mathbf{F}_{i_k} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k}$
	$\mathbf{\Phi}_{i_k} = \mathbf{F}_{i_k}\mathbf{F}_{i_k}^{\mathsf{H}} + \mathbf{\Phi}_{i_k}^{\mathrm{i+n}}$
	$\mathbf{\Phi}_{i_k}^{\mathrm{i+n}} = \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}} + \sigma_{i_k}^2\mathbf{I}$
Uplink	$\mathbf{G}_{i_k} = \mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{U}_{i_k}$
	$\mathbf{\Gamma}_i = \mathbf{\Gamma}_i^{\mathrm{s+i}} = \sum_{(j,l)} \mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{U}_{j_l}\mathbf{U}_{j_l}^{\mathsf{H}}\mathbf{H}_{j_l i}^{\mathsf{H}}$

Table 4.2. Quantities needed at each network node to perform one iteration of the WMMSE algorithm

	Covariance matrix	Effective channel(s)	MSE weights
MS i_k	$\mathbf{\Phi}_{i_k}$	$\mathbf{F}_{i_k}$	—
BS i	$\mathbf{\Gamma}_i$	$\{\mathbf{G}_{i_k}\}_{k=1}^{K_c}$	$\{\mathbf{W}_{i_k}^{1/2}\}_{k=1}^{K_c}$

Algorithm 4.1 WMMSE Algorithm [SRLH11] with Per-BS Sum Power Constraints (Perfect CSI)

1: **repeat**
At MS i_k:
2: $\mathbf{W}_{i_k} = \left(\mathbf{I} - \mathbf{F}_{i_k}^{\mathsf{H}}\mathbf{\Phi}_{i_k}^{-1}\mathbf{F}_{i_k}\right)^{-1}$
3: $\mathbf{A}_{i_k} = \mathbf{\Phi}_{i_k}^{-1}\mathbf{F}_{i_k}$, $\mathbf{U}_{i_k} = \sqrt{\alpha_{i_k}}\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}$
At BS i:
4: Find μ_i which satisfies $\sum_{k=1}^{K_i} \mathrm{Tr}\left(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^{\mathsf{H}}\right) \leq P_i$
5: $\mathbf{B}_{i_k} = \left(\mathbf{\Gamma}_i + \mu_i\mathbf{I}\right)^{-1}\mathbf{G}_{i_k}$, $\mathbf{V}_{i_k} = \sqrt{\alpha_{i_k}}\mathbf{B}_{i_k}\mathbf{W}_{i_k}^{1/2}, \quad k = 1, \ldots, K_i$
6: **until** convergence criterion met, or fixed number of iterations

effective uplink channel $\mathbf{G}_{i_k} = \mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{U}_{i_k}$. Finally, we have the *component precoder* $\mathbf{B}_{i_k} = \left(\mathbf{\Gamma}_i + \mu_i\mathbf{I}\right)^{-1}\mathbf{G}_{i_k}$. The shorthands are summarized in Table 4.1. With these shorthands, we restate the WMMSE algorithm with per-BS sum power constraints in Algorithm 4.1.

It is obvious from Algorithm 4.1 that the WMMSE algorithm operates in two phases: one in which the MSs form their receive filters and MSE weights, and one in which the BSs form the precoders for their served MSs. In both phases, only local CSI is required at the corresponding nodes. The MSs need estimates of $\mathbf{\Phi}_{i_k}$ and $\mathbf{F}_{i_k}$ to form their filters. Similarly, the BSs need estimates of $\mathbf{\Gamma}_i$, and estimates of $\mathbf{G}_{i_k}$ for their correspondingly served MSs. The local CSI requirements are summarized in Table 4.2. Assuming that this information is available, the filters can be formed in parallel over all MSs. The same goes for the precoders at the BSs;

no direct cooperation is needed between them, and the precoders can be formed in parallel over BSs. Given local CSI, the WMMSE algorithm is thus an example of a distributed resource allocation method. The details of how the estimation will be performed will be described in Section 4.2.

4.1.2 Weighted MaxSINR

We also take this opportunity to present a modified version of the original MaxSINR algorithm [GCJ11]. The original MaxSINR (see Algorithm 2.2 on page 34) empirically often performs well when applied to scenarios where IA is feasible (see e.g. Section 4.4). However, for cases when the number of allocated data streams exceed that for which is IA feasible, the performance of the original MaxSINR deteriorates. This can for example be seen in Figure 4.9a on page 91. In order to alleviate this problem, we propose to apply a weighting to the MaxSINR precoders. The weighting is inspired by the MSE weighting in Algorithm 4.1.

Using the shorthands defined in Table 4.1, the proposed algorithm is shown in Algorithm 4.2 on the next page. Similarly to the original MaxSINR in Algorithm 2.2, unweighted receive filters and precoders are obtained on a per-stream basis. We denote these as $\mathbf{a}'_{i_k,n}$ and $\mathbf{b}'_{i_k,n}$, respectively. The final precoders and receive filters are then obtained as normalized versions of the weighted concatenated per-stream filters. Through the multiplication with the weights $\mathbf{W}_{i_k}$, power can be spread over the data streams in the receive filters and precoders. Through this power allocation, streams can effectively be turned off as deemed necessary by the algorithm[2]. Due to the normalization of the final filters however, the same amount of total power per user is still used, and entire users can thus not be turned off. This is a major difference to the WMMSE algorithm, for which entire users can be turned off if $\mathbf{U}_{i_k} \to \mathbf{0}$.

It is easy to show that Algorithm 4.2 gives the same receive filters and precoders for MS i_k as Algorithm 2.2, if $d_{i_k} = 1$. Just as the original MaxSINR, the weighted MaxSINR qualifies as a distributed resource allocation method. The CSI requirements at the different nodes of the network are the same as for the WMMSE algorithm, as summarized in Table 4.2 on the facing page.

4.2 Distributed CSI Acquisition

According to Table 4.2, the MSs need to know their effective channel from their serving BS $\mathbf{F}_{i_k}$, as well as their signal plus interference and noise covariance matrix $\mathbf{\Phi}_{i_k}$. The BSs need to know the effective uplink channels $\mathbf{G}_{i_k}$ to the MSs they serve, the corresponding MSE weights $\mathbf{W}_{i_k}$, and the uplink signal plus interference covariance matrix $\mathbf{\Gamma}_i$. There might be several methods useful for acquiring this local

[2]Note that for nondegenerate scenarios, where the MSE matrix $\mathbf{E}_{i_k} \succ 0$, the MSE weights $\mathbf{W}_{i_k} = \mathbf{E}_{i_k}^{-1}$ are theoretically full rank. The condition number of $\mathbf{W}_{i_k}$ may however become large, and this effectively turns off data streams.

Algorithm 4.2 Weighted MaxSINR with Per-BS Sum Power Constraint (Perfect CSI)

repeat

At MS i_k:

$\mathbf{W}_{i_k} = \left(\mathbf{I} - \mathbf{F}_{i_k}^{\mathsf{H}} \boldsymbol{\Phi}_{i_k}^{-1} \mathbf{F}_{i_k}\right)^{-1}$

$\mathbf{a}'_{i_k,n} = \frac{\boldsymbol{\Phi}_{i_k}^{-1} \mathbf{f}_{i_k,n}}{\left\| \boldsymbol{\Phi}_{i_k}^{-1} \mathbf{f}_{i_k,n} \right\|_2}, \quad n = 1, \ldots, d_{i_k}$

$\mathbf{A}'_{i_k} = \begin{pmatrix} \mathbf{a}_{i_k,1} & \mathbf{a}_{i_k,2} & \cdots & \mathbf{a}_{i_k,d_{i_k}} \end{pmatrix}$

$\mathbf{A}_{i_k} = \sqrt{\overleftarrow{P}_{i_k}} \frac{\mathbf{A}'_{i_k}}{\left\| \mathbf{A}'_{i_k} \mathbf{W}_{i_k}^{1/2} \right\|_{\mathrm{F}}}, \ \mathbf{U}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}$

At BS i:

$\mathbf{b}_{i_k,n} = \frac{\sqrt{\alpha_{i_k}} \left(\boldsymbol{\Gamma}_i + \varsigma_i^2 \mathbf{I}\right)^{-1} \mathbf{g}_{i_k,n}}{\left\| \sqrt{\alpha_{i_k}} \left(\boldsymbol{\Gamma}_i + \varsigma_i^2 \mathbf{I}\right)^{-1} \mathbf{g}_{i_k,n} \right\|_2}, \qquad k = 1, \ldots, K_i, \ n = 1, \ldots, d_{i_k}$

$\mathbf{B}_{i_k} = \begin{pmatrix} \mathbf{b}_{i_k,1} & \mathbf{b}_{i_k,2} & \cdots & \mathbf{b}_{i_k,d_{i_k}} \end{pmatrix}, \quad k = 1, \ldots, K_i$

$\mathbf{B}_{i_k} = \sqrt{\frac{P_i}{K_i}} \frac{\mathbf{B}'_{i_k}}{\left\| \mathbf{B}'_{i_k} \mathbf{W}_{i_k}^{1/2} \right\|_{\mathrm{F}}}, \ \mathbf{V}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{B}_{i_k} \mathbf{W}_{i_k}^{1/2}, \quad k = 1, \ldots, K_i$

until fixed number of iterations

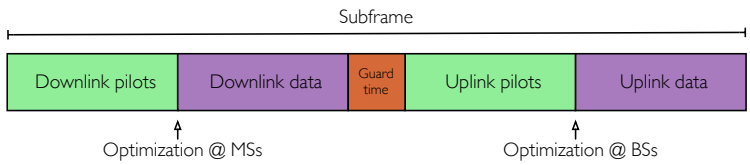


Figure 4.1. Schematic drawing of a subframe.

CSI. We will now propose an estimation framework that can be used to obtain these quantities. The estimation will be based on pilot transmissions in both the downlink and the uplink. Due to the assumed reciprocity of the network, the BSs will then be able to obtain their local CSI from the uplink pilot transmissions.

As the effective channels change between iterations in the WMMSE algorithm, a training phase always need to be performed between one WMMSE iteration and the next. A simple schematic of the TDD transmission subframes that we envision can be seen in Figure 4.1. The subframe is split between pilot transmission and data transmission, in both the uplink and downlink. Note that we assume data transmissions in all subframes. Before the iterative algorithm has converged, the data rates that are achievable in the downlink data transmission phase may be low, but not negligible, as shown by numerical results in Section 4.4. An illustration of the downlink/uplink CSI estimation process is shown in Figure 4.2 on the facing page.

We will only focus on the downlink transmission optimization, but the proposed method can in principle be used for uplink transmission optimization as well. The

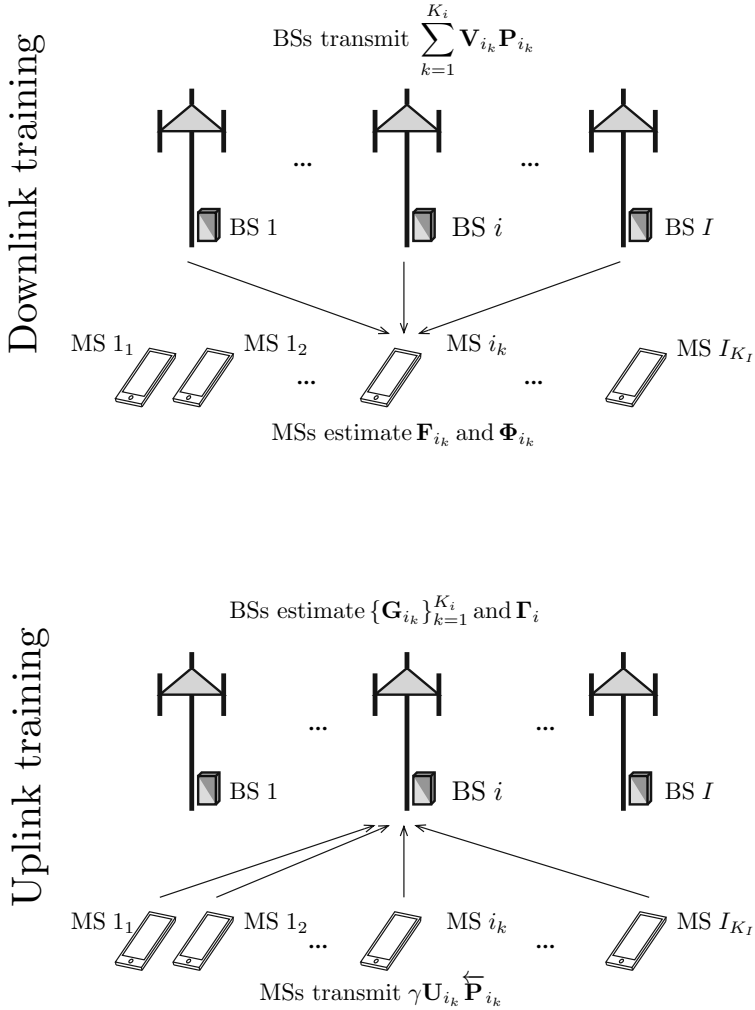


Figure 4.2. CSI estimation in one subframe (cf. Figure 4.1). In each subframe, the downlink channels are estimated using pilots from the BSs. Later, the uplink pilots are estimated using pilots from the MSs. Additionally, the MSs feed back $\mathbf{W}_{i_k}$ to their serving BS using an out-of-band feedback link.

training intervals would then need to be doubled, accommodating the corresponding uplink optimization training symbols.

4.2.1 Global Sharing of Common Scale Factor

We will present three CSI acquisition schemes of varying levels of distributedness. All schemes require some form of cooperation between the BSs, but the MSs only need to cooperate with their serving BSs. The scheme presented in this particular section is almost fully distributed over the participating nodes. The only level of cooperation that is needed is the joint selection of a common power scaling parameter for the uplink pilot transmissions.

When a priori statistical information about the channel to be estimated is available, the MMSE estimator [BG06] is often used. In our case, we are interested in estimating effective channels, i.e. the product of the channel and a transmit filter. It is complicated to obtain a statistical characterization of the effective channels, and therefore we choose not to assign a prior distribution to the quantities to be estimated. That is, for the estimation we regard the effective channels as deterministic but unknown. With this perspective from classical estimation theory, it is easy to find the minimum variance unbiased (MVU) estimator.

Downlink Estimation

First, we will show how to estimate the effective downlink channel $\mathbf{F}_{i_k} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k}$ using synchronous pilot transmissions. First, define orthogonal pilot sequences $\mathbf{P}_{i_k} \in \mathbb{C}^{d_{i_k} \times L_{p,d}}$, such that

$$\mathbf{P}_{i_k}\mathbf{P}_{j_l}^{\mathsf{H}} = \begin{cases} L_{p,d}\,\mathbf{I}_{d_{i_k}} & (i,k) = (j,l) \\ \mathbf{0}_{d_{i_k} \times d_{j_l}} & (i,k) \neq (j,l) \end{cases}. \tag{4.8}$$

In order to fulfill the orthogonality requirement, $L_{p,d} \geq \sum_{(j,l)} d_{j_l}$ must hold. In the downlink training phase, BS i transmits $\mathbf{s}_i = \sum_{k=1}^{K_i} \mathbf{V}_{i_k}\mathbf{P}_{i_k}$ such that the received signal at MS i_k is

$$\mathbf{Y}_{i_k} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k}\mathbf{P}_{i_k} + \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{P}_{j_l} + \mathbf{Z}_{i_k}. \tag{4.9}$$

Notice that the power allocated to the pilots in (4.9) is the same as the power allocated to the data symbols in (4.1) on page 64. This is intentional, and will enable distributed and unbiased estimation of $\mathbf{\Phi}_{i_k}$. This type of pilot transmissions are called 'UE-specific reference signals' in the LTE standard [DPSB07].

Assuming that MS i_k knows its designated pilot $\mathbf{P}_{i_k}$, the problem of estimating $\mathbf{F}_{i_k}$ is a deterministic parameter estimation problem in Gaussian noise. The MVU estimator is then [BG06]:

$$\widehat{\mathbf{F}}_{i_k} = \frac{1}{L_{p,d}}\mathbf{Y}_{i_k}\mathbf{P}_{i_k}^{\mathsf{H}} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k} + \frac{1}{L_{p,d}}\mathbf{Z}_{i_k}\mathbf{P}_{i_k}^{\mathsf{H}}. \tag{4.10}$$

The MVU estimator is an unbiased, efficient and asymptotically consistent (in $L_{p,d}$) estimator of $\mathbf{F}_{i_k}$.

In addition to knowing $\mathbf{F}_{i_k}$, MS i_k also needs knowledge of $\mathbf{\Phi}_{i_k}$. This can be obtained from the sample covariance estimator:

$$\begin{aligned}\widehat{\mathbf{\Phi}}_{i_k} &= \frac{1}{L_{p,d}}\mathbf{Y}_{i_k}\mathbf{Y}_{i_k}^{\mathsf{H}} \\ &= \sum_{(j,l)}\left(\mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}}\right) + \frac{1}{L_{p,d}}\mathbf{Z}_{i_k}\mathbf{Z}_{i_k}^{\mathsf{H}} \\ &+ \frac{1}{L_{p,d}}\sum_{(j,l)}\left(\mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{P}_{j_l}\mathbf{Z}_{i_k}^{\mathsf{H}} + \mathbf{Z}_{i_k}\mathbf{P}_{j_l}^{\mathsf{H}}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}}\right).\end{aligned} \tag{4.11}$$

Note that the pilots $\{\mathbf{P}_{j_l}\}$ are deterministic, and the channels $\{\mathbf{H}_{i_k j}\}$ are also treated as deterministic. Since the only stochastic component of $\mathbf{Y}_{i_k}$ is $\mathbf{Z}_{i_k}$, the estimator in (4.11) is unbiased.

Uplink Estimation

The uplink estimation is performed in a similar fashion as the downlink estimation. Now the MSs each transmit a signal $\overleftarrow{\mathbf{S}}_{i_k} = \gamma\mathbf{U}_{i_k}\overleftarrow{\mathbf{P}}_{i_k}$, where $\overleftarrow{\mathbf{P}}_{i_k} \in \mathbb{C}^{d_{i_k}\times L_{p,u}}$ are orthogonal pilots, such that

$$\overleftarrow{\mathbf{P}}_{i_k}\overleftarrow{\mathbf{P}}_{j_l}^{\mathsf{H}} = \begin{cases} L_{p,u}\,\mathbf{I}_{d_{i_k}} & (i,k) = (j,l) \\ \mathbf{0}_{d_{i_k}\times d_{j_l}} & (i,k) \neq (j,l) \end{cases}. \tag{4.12}$$

In order to fulfill the orthogonality requirement, $L_{p,u} \geq \sum_{(j,l)} d_{j_l}$ must hold. As will be shown in Theorem 4.1 on page 84, for MS i_k it holds that

$$\|\mathbf{U}_{i_k}\|_{\mathrm{F}}^2 = \alpha_{i_k}\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 \leq \alpha_{i_k}\frac{d_{i_k}}{\sigma_{i_k}^2}. \tag{4.13}$$

Based on this fact, the common scale factor γ can be set to make sure that the uplink transmit power constraints are satisfied for all MSs. We let

$$\gamma_{i_k} = \sqrt{\frac{\overleftarrow{P}_{i_k}\sigma_{i_k}^2}{\alpha_{i_k}d_{i_k}}}, \quad \forall\, i \in \{1,\ldots,I\},\, k \in \{1,\ldots,K_i\}, \tag{4.14}$$

and set the common scale factor as $\gamma = \min_{j_l} \gamma_{j_l}$. The sum power constraint for MS i_k is then satisfied since

$$\begin{aligned}\left\|\overleftarrow{\mathbf{S}}_{i_k}\right\|_{\mathrm{F}}^2 &= \gamma^2 \left\|\mathbf{U}_{i_k}\overleftarrow{\mathbf{P}}_{i_k}\right\|_{\mathrm{F}}^2 = L_{p,u}\gamma^2 \left\|\mathbf{U}_{i_k}\right\|_{\mathrm{F}}^2 \\ &= L_{p,u} \min_{j_l} \left(\frac{\overleftarrow{P}_{j_l}\sigma_{j_l}^2}{\alpha_{j_l} d_{j_l}}\right) \left\|\mathbf{U}_{i_k}\right\|_{\mathrm{F}}^2 \\ &\overset{(4.13)}{\leq} L_{p,u} \min_{j_l} \left(\frac{\overleftarrow{P}_{j_l}\sigma_{j_l}^2}{\alpha_{j_l} d_{j_l}}\right) \alpha_{i_k} \frac{d_{i_k}}{\sigma_{i_k}^2} \\ &\leq L_{p,u} \left(\frac{\overleftarrow{P}_{i_k}\sigma_{i_k}^2}{\alpha_{i_k} d_{i_k}}\right) \alpha_{i_k} \frac{d_{i_k}}{\sigma_{i_k}^2} = L_{p,u}\overleftarrow{P}_{i_k}.\end{aligned} \tag{4.15}$$

MS i_k will only use its full power if it has equality in (4.13) and if $\gamma_{i_k} = \gamma$. For heterogenous scenarios, potentially only one MS will use its full power, due to the $\min_{j_l}(\cdot)$ in γ. This is the price to pay for enabling distributed estimation of $\mathbf{\Gamma}_i$ in the uplink.

In order to determine γ, each MS can estimate its own γ_{i_k} and feed it back to its serving BS. The BSs can then jointly determine $\gamma = \min_{j_l} \gamma_{j_l}$. After the selected γ has been shared to the MSs from their serving BS, no more explicit cooperation is needed. For static scenarios, where the data stream allocation and transmit and noise powers do not change, γ only needs to be computed once in a transmission interval. For symmetric scenarios, where $\gamma = \gamma_{i_k}$, $\forall\, i_k$, no cooperation is necessary, and the CSI acquisition can be performed in a *fully distributed* manner.

Assuming synchronized pilot transmissions from the MSs, the received signal at BS i during the uplink training phase is

$$\overleftarrow{\mathbf{Y}}_i = \gamma \sum_{k=1}^{K_i} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{U}_{i_k} \overleftarrow{\mathbf{P}}_{i_k} + \gamma \sum_{j\neq i} \sum_{l=1}^{K_j} \mathbf{H}_{j_l i}^{\mathsf{H}} \mathbf{U}_{j_l} \overleftarrow{\mathbf{P}}_{j_l} + \overleftarrow{\mathbf{Z}}_i. \tag{4.16}$$

Similarly to the downlink, the uplink effective channel MVU estimator is

$$\widehat{\mathbf{G}}_{i_k} = \frac{1}{\gamma L_{p,u}} \overleftarrow{\mathbf{Y}}_i \overleftarrow{\mathbf{P}}_{i_k}^{\mathsf{H}} = \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{U}_{i_k} + \frac{1}{\gamma L_{p,u}} \overleftarrow{\mathbf{Z}}_i \overleftarrow{\mathbf{P}}_{i_k}^{\mathsf{H}}. \tag{4.17}$$

Furthermore, the signal plus interference *and scaled noise* covariance matrix is estimated using the sample covariance method:

$$\begin{aligned}\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}} &= \frac{1}{\gamma^2}\frac{1}{L_{p,u}} \overleftarrow{\mathbf{Y}}_i \overleftarrow{\mathbf{Y}}_i^{\mathsf{H}} \\ &= \sum_{(j,l)} \left(\mathbf{H}_{i_k j} \mathbf{U}_{j_l} \mathbf{U}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}^{\mathsf{H}}\right) + \frac{1}{\gamma^2}\frac{1}{L_{p,u}} \overleftarrow{\mathbf{Z}}_i \overleftarrow{\mathbf{Z}}_i^{\mathsf{H}} \\ &\quad + \frac{1}{\gamma}\frac{1}{L_{p,u}} \sum_{(j,l)} \left(\mathbf{H}_{i_k j}^{\mathsf{H}} \mathbf{U}_{j_l} \overleftarrow{\mathbf{P}}_{j_l} \overleftarrow{\mathbf{Z}}_i^{\mathsf{H}} + \overleftarrow{\mathbf{Z}}_i \overleftarrow{\mathbf{P}}_{j_l}^{\mathsf{H}} \mathbf{U}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}\right).\end{aligned} \tag{4.18}$$

Unless $\gamma = 1$, this is a biased estimator of the signal plus interference and noise covariance matrix in the uplink. The WMMSE algorithm needs an estimate of $\mathbf{\Gamma}_i = \mathbf{\Gamma}_i^{\mathrm{s+i}}$, without the noise covariance part of $\mathbf{\Gamma}_i^{\mathrm{s+i+n}}$. In Section 4.3.3, we resolve this issue by modifying the WMMSE algorithm.

MSE Weight Feedback

According to Algorithm 4.1 on page 66, the precoders are formed as

$$\mathbf{V}_{i_k} = \alpha_{i_k} \left(\mathbf{\Gamma}_i + \mu_i \mathbf{I} \right)^{-1} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}, \quad \forall\, i \in \{1, \ldots, I\},\, k \in \{1, \ldots, K_i\}, \quad (4.19)$$

at the BSs, assuming perfect CSI. To form the precoders as in (4.19), the product $\alpha_{i_k} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{A}_{i_k} \mathbf{W}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{G}_{i_k} \mathbf{W}_{i_k}^{1/2}$ is needed. This quantity is however not directly provided by the proposed estimation scheme. It could be independently estimated in a second uplink estimation phase, but instead we let MS i_k explicitly feed back $\mathbf{W}_{i_k}$ to its serving BS i. Together with $\widehat{\mathbf{G}}_{i_k}$ in (4.17), BS i can then form $\widehat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and use that in[3] (4.19). The reason for this procedure is to avoid *signal cancelation* [Cox73], where a small mismatch between the estimate of $\mathbf{G}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and the estimate of $\mathbf{\Gamma}_i$ can have a large detrimental impact on performance. If $\mathbf{G}_{i_k}$ and $\mathbf{\Gamma}_i$ are estimated using the same pilot transmissions, as in (4.17) and (4.18), we can decompose

$$\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}} = \widehat{\mathbf{\Gamma}}_i^{\mathrm{i+n}} + \widehat{\mathbf{G}}_{i_k} \widehat{\mathbf{G}}_{i_k}^{\mathsf{H}}. \quad (4.20)$$

Then, there is no mismatch between $\widehat{\mathbf{G}}_{i_k}$ and $\widehat{\mathbf{\Gamma}}_i$, and consequently no mismatch between $\widehat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and $\widehat{\mathbf{\Gamma}}_i$. Therefore, signal cancelation does not occur [Cox73].

Note that $R_{i_k} = \log_2 \det \left(\mathbf{W}_{i_k} \right) = \sum_n \log_2 \left(\lambda_n \left(\mathbf{W}_{i_k} \right) \right)$. Thus, feedback of the eigenvalues of $\mathbf{W}_{i_k}$ constitute a *rate request* for each stream for MS i_k, describing what rate that stream can handle under the current network conditions. This is information that is typically fed back to the serving BS in a practical system. Recall that α_{i_k} is known at BS i, so that information does not need to be fed back.

After γ has been formed at the network level, the effective channel estimation proposed in this section can be performed in a fully distributed manner. The MSs can estimate their local CSI, in the form of effective channels and covariance matrices, since the downlink pilot transmissions are precoded the same was as the data transmission. Thanks to the channel reciprocity, the BSs can estimate their local CSI in a similar way. With the feedback of $\mathbf{W}_{i_k}$ from the MSs served by BS i, the BS can then form the associated precoders. Note that with the exception of γ_{i_k}, no information needs to be shared over the BS backhaul.

Remark 4.1. *The CSI acquisition proposed in this section is fully distributed, save for the selection of γ. Due to the channel reciprocity, all nodes are able to estimate their local CSI from the pilot transmissions. MS i_k feeds back $\mathbf{W}_{i_k}$ to its serving BS, but the BSs do not need to share the MSE weights over some backhaul.*

[3] Recall that α_{i_k} is assumed a priori known at BS i.

4.2.2 Global Sharing of Individual Scale Factors

In the previous section, an almost fully distributed CSI acquisition scheme was proposed. In order to perform the estimation of $\mathbf{\Gamma}_i$ in a fully distributed manner at the BSs, all MSs had to use a common power scaling γ. For heterogenous settings, with different γ_{i_k} over the different MSs, that meant that some MSs might not satisfy their uplink pilot power constraint with equality, meaning decreased estimation performance. In this section, we relax the requirement of fully distributed estimation of $\mathbf{\Gamma}_i$ at the BSs, in order to improve the estimation SNRs. We keep the distributed downlink pilot transmission phase the same as in Section 4.2, as well as the MSE weight feedback, but modify the uplink pilot transmission phase.

Uplink Estimation

Letting $\overleftarrow{\mathbf{S}}_{i_k} = \frac{\sqrt{\overleftarrow{P}_{i_k}}}{\|\mathbf{U}_{i_k}\|_{\mathrm{F}}}\mathbf{U}_{i_k}\overleftarrow{\mathbf{P}}_{i_k}$, the uplink sum power constraint in (4.6) on page 65 is met with equality for MS i_k. The received signal at BS i is then

$$\overleftarrow{\mathbf{Y}}_i = \sum_{k=1}^{K_i} \frac{\sqrt{\overleftarrow{P}_{i_k}}}{\|\mathbf{U}_{i_k}\|_{\mathrm{F}}}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{U}_{i_k}\overleftarrow{\mathbf{P}}_{i_k} + \sum_{j\neq i}\sum_{l=1}^{K_j} \frac{\sqrt{\overleftarrow{P}_{j_l}}}{\|\mathbf{U}_{j_l}\|_{\mathrm{F}}}\mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{U}_{j_l}\overleftarrow{\mathbf{P}}_{j_l} + \overleftarrow{\mathbf{Z}}_i. \tag{4.21}$$

Assuming that the scale factors $\frac{\|\mathbf{U}_{i_k}\|_{\mathrm{F}}}{\sqrt{\overleftarrow{P}_{i_k}}}$ are fed back from the MSs to their serving BSs, and then globally shared over the BS backhaul, BS i can estimate the effective channels from MS j_l as

$$\widehat{\mathbf{G}}_{j_l i} = \frac{\|\mathbf{U}_{j_l}\|_{\mathrm{F}}}{\sqrt{\overleftarrow{P}_{j_l}}L_{p,u}}\overleftarrow{\mathbf{Y}}_i\overleftarrow{\mathbf{P}}_{j_l}^{\mathsf{H}} = \mathbf{H}_{i_k j}^{\mathsf{H}}\mathbf{U}_{j_l} + \frac{\|\mathbf{U}_{j_l}\|_{\mathrm{F}}}{\sqrt{\overleftarrow{P}_{j_l}}L_{p,u}}\overleftarrow{\mathbf{Z}}_i\overleftarrow{\mathbf{P}}_{j_l}^{\mathsf{H}}. \tag{4.22}$$

Since the scaled pilots effectively all have the same weight, the sample covariance estimator of $\mathbf{\Gamma}_i$ in (4.18) cannot be used. Instead, we rely on the biased estimator

$$\begin{aligned}\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}} &= \sum_{(j,l)}\widehat{\mathbf{G}}_{j_l i}\widehat{\mathbf{G}}_{j_l i}^{\mathsf{H}} \\ &= \sum_{(j,l)}\left(\mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{U}_{j_l}\mathbf{U}_{j_l}^{\mathsf{H}}\mathbf{H}_{j_l i}\right) + \frac{1}{L_{p,u}}\overleftarrow{\mathbf{Z}}_i\left(\frac{1}{L_{p,u}}\sum_{(j,l)}\frac{\|\mathbf{U}_{j_l}\|_{\mathrm{F}}^2}{\overleftarrow{P}_{j_l}}\overleftarrow{\mathbf{P}}_{j_l}^{\mathsf{H}}\overleftarrow{\mathbf{P}}_{j_l}\right)\overleftarrow{\mathbf{Z}}_i^{\mathsf{H}} \\ &\quad + \frac{1}{L_{p,u}}\sum_{(j,l)}\frac{\|\mathbf{U}_{j_l}\|_{\mathrm{F}}}{\sqrt{\overleftarrow{P}_{j_l}}}\left(\mathbf{H}_{i_k j}^{\mathsf{H}}\mathbf{U}_{j_l}\overleftarrow{\mathbf{P}}_{j_l}\overleftarrow{\mathbf{Z}}_i^{\mathsf{H}} + \overleftarrow{\mathbf{Z}}_i\mathbf{P}_{j_l}^{\mathsf{H}}\mathbf{U}_{j_l}\mathbf{H}_{i_k j}\right).\end{aligned} \tag{4.23}$$

The bias is determined by the factors $\frac{\|\mathbf{U}_{j_l}\|_{\mathrm{F}}^2}{\overleftarrow{P}_{j_l}}$ and the pilots length $L_{p,u}$.

This estimation scheme is similar to one proposed in [KTJ13]. There, a scaled version of $\mathbf{A}_{i_k}$ was used as the uplink precoder. Then, the MSE weights $\mathbf{W}_{i_k}$ can be calculated at the serving BSs, and do not need to be fed back. However, in order for the BSs to estimate $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}$ in that estimation scheme, they must exchange the MSE weights for their corresponding MSs over the backhaul. In essence, reduced over-the-air feedback has been traded for more backhaul use.

Remark 4.2. *The CSI acquisition proposed in this section is fully distributed over the MSs, but not over the BSs. Each BS needs knowledge of the individual scaling factors for all MSs.*

4.2.3 Global Sharing of Filters

Finally, we present a CSI acquisition scheme which relies even further on feedback and backhaul. Here, only the underlying channels are estimated exploiting the reciprocity, but the receive filters, MSE weights and precoders are fed back over an out-of-band link. With the subframe structure in Figure 4.1 on page 68, this means that consecutive training phases can be used to monotonically improve the channel estimates in one coherence block of the channel. This can be done using iterative techniques, see e.g. [Kay93, Ch. 12.6].

Downlink Estimation

Let $\mathbf{P}_j \in \mathbb{C}^{M_j \times L_{p,d}}$ be orthogonal pilots sent from BS j such that

$$\mathbf{P}_i \mathbf{P}_j^{\mathsf{H}} = \begin{cases} L_{p,d} \mathbf{I}_{M_i} & i = j \\ \mathbf{0}_{M_i \times M_j} & i \neq j \end{cases}. \tag{4.24}$$

As usual, $L_{p,d} \geq \sum_i M_i$ is needed for orthogonality reasons. For the downlink, the received training signal would be

$$\mathbf{Y}_{i_k} = \sqrt{\frac{P_i}{K_i M_i}} \mathbf{H}_{i_k i} \mathbf{P}_i + \sum_{(j,l) \neq (i,k)} \sqrt{\frac{P_j}{K_j M_j}} \mathbf{H}_{i_k j} \mathbf{P}_j + \mathbf{Z}_{i_k}. \tag{4.25}$$

This type of pilot transmissions are called 'cell-specific reference signals' in the LTE standard [DPSB07].

The receive filter is assumed to be fed back to the MSs, and the goal is therefore only to estimate the channels. We can therefore use Bayesian methods for the estimation. For example, assuming i.i.d. Rayleigh fading such that

$$\operatorname{vec}(\mathbf{H}_{i_k j}) \sim \mathcal{CN}\left(0, \mathbf{I}_{N_{i_k} \times M_i}\right), \tag{4.26}$$

the MMSE estimator of the channel from BS j to MS i_k is [BG06]

$$\widehat{\mathbf{H}}_{i_k j} = \frac{\sqrt{\frac{P_j}{K_j M_j}}}{L_{p,d} \frac{P_j}{K_j M_j} + \sigma_{i_k}^2} \mathbf{Y}_{i_k} \mathbf{P}_j^{\mathsf{H}}. \tag{4.27}$$

Then, assuming that all precoders $\{\mathbf{V}_{j_l}\}$ have been fed back to MS i_k, and that $\sigma^2_{i_k}$ is known, it can form

$$\widehat{\mathbf{F}}_{i_k} = \widehat{\mathbf{H}}_{i_k i}\mathbf{V}_{i_k}, \tag{4.28}$$

$$\widehat{\mathbf{\Phi}}_{i_k} = \sum_{(j,l)} \widehat{\mathbf{H}}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\widehat{\mathbf{H}}_{i_k j}^{\mathsf{H}} + \sigma^2_{i_k}\mathbf{I}. \tag{4.29}$$

Uplink Estimation

A similar procedure is used in the uplink. The receive filters and MSE weights for MS i_k is fed back to BS i, and the BSs then share their information over the backhaul.

Let $\mathbf{P}_{j_l} \in \mathbb{C}^{N_{i_k} \times L_{p,u}}$ be orthogonal pilots, assuming $L_{p,u} \geq \sum_{(j,l)} N_{j_l}$, such that

$$\overleftarrow{\mathbf{P}}_{i_k}\overleftarrow{\mathbf{P}}_{j_l}^{\mathsf{H}} = \begin{cases} L_{p,u}\mathbf{I}_{N_{i_k}} & (i,k) = (j,l) \\ \mathbf{0}_{N_{i_k} \times N_{j,l}} & (i,k) \neq (j,l) \end{cases}. \tag{4.30}$$

The received signal at BS i in the uplink training phase is then

$$\overleftarrow{\mathbf{Y}}_i = \sum_{k=1}^{K_i} \sqrt{\frac{\overleftarrow{P}_{i_k}}{N_{i_k}}}\mathbf{H}_{i_k i}^{\mathsf{H}}\overleftarrow{\mathbf{P}}_{i_k} + \sum_{j \neq i}\sum_{l=1}^{K_j} \sqrt{\frac{\overleftarrow{P}_{j_l}}{N_{j_l}}}\mathbf{H}_{j_l i}^{\mathsf{H}}\overleftarrow{\mathbf{P}}_{j_l} + \overleftarrow{\mathbf{Z}}_i. \tag{4.31}$$

The MMSE estimator of the uplink channel from MS j_l to BS i is

$$\widehat{\mathbf{H}}_{j_l i}^{\mathsf{H}} = \frac{\sqrt{\frac{\overleftarrow{P}_{j_l}}{N_{j_l}}}}{L_{p,u}\frac{\overleftarrow{P}_{j_l}}{N_{j_l}} + \varsigma_i^2}\overleftarrow{\mathbf{Y}}_i\overleftarrow{\mathbf{P}}_{j_l}^{\mathsf{H}}. \tag{4.32}$$

With the estimated uplink channels, together with perfect knowledge of all receive filters $\{\mathbf{U}_{j_l}\}$, BS i can form

$$\widehat{\mathbf{G}}_{i_k} = \widehat{\mathbf{H}}_{i_k i}^{\mathsf{H}}\mathbf{U}_{i_k}, \quad k = 1, \ldots, K_i \tag{4.33}$$

$$\widehat{\mathbf{\Gamma}}_i = \sum_{(j,l)} \widehat{\mathbf{H}}_{j_l i}^{\mathsf{H}}\mathbf{U}_{j_l}\mathbf{U}_{j_l}^{\mathsf{H}}\widehat{\mathbf{H}}_{j_l i}. \tag{4.34}$$

The estimation procedure presented in this section requires significant feedback and signaling of filters among BSs and MSs in every subframe in order to form the effective channels and covariance matrices. This method is still interesting however, since the state-of-the-art robust WMMSE algorithms in [NGS12, SM12, LKY13, RBCL13] require this type of channel estimation.

Remark 4.3. *The CSI acquisition proposed in this section is centralized. It requires significant feedback and sharing of filters among BSs and MSs in every subframe. In terms of estimating the underlying channels $\{\mathbf{H}_{i_k j}\}$, it is however distributed over the BSs and MSs.*

Table 4.3. Total estimation complexity, per iteration and MS

Method	Approximate number of flops
Sec. 4.2.1	$(Nd+N^2)L_{p,d}+MdL_{p,u}+M^2L_{p,u}/K_c$
Sec. 4.2.2	$(Nd+N^2)L_{p,d}+(MdL_{p,u}+M^2d+M^2)I$
Sec. 4.2.3	$NMI(L_{p,d}+L_{p,u})+(NMd+N^2d+N^2)K+(NMd+M^2d+M^2)I$

4.2.4 Feedback Requirements and Complexity

The proposed CSI acquisition schemes in Section 4.2.1 to Section 4.2.3 have vastly different feedback loads. We compare these in Table 4.4 on the following page.

The computational complexities [TI97] of forming the channel estimates are given in Table 4.3. The expressions are for the special case that all MSs have N antennas and are served d data streams. All BSs have M antennas, and each serve K_c MSs. The $NMI(L_{p,d}+L_{p,u})$ term in the Section 4.2.3 estimation method flop count dominates all other terms when the number of pilots is large. An illustration of this will be given in Section 4.4.

4.2.5 Quantized Feedback of MSE Weights

In the proposed CSI acquisition schemes in Section 4.2.1 and Section 4.2.2, feedback of the MSE weights is needed. In practical applications, quantized feedback should be employed to reduce the overhead of the feedback. Since $\mathbf{W}_{i_k}$ is Hermitian, it has an eigenvalue decomposition, which can be quantized and fed back to the serving BS. The eigenvectors can e.g. be quantized using Grassmannian subspace packing [LH05]. For the quantization of the eigenvalues, we have the following helpful lemma:

Lemma 4.1. *The eigenvalues of the MSE weight for MS i_k are bounded as* $1 \leq \lambda_n(\mathbf{W}_{i_k}) \leq 1+\frac{P_i s_{max}^2(\mathbf{H}_{i_k i})}{\sigma_{i_k}^2}$, $\forall\, n \in \{1,\ldots,d_{i_k}\}$.

Proof. For MS i_k it holds that $\mathbf{\Phi}_{i_k}^{\mathrm{i+n}} \succeq \sigma_{i_k}^2\mathbf{I}$, with equality if the MS does not experience any interference. Thus, the MSE weight for that MS satisfies

$$\mathbf{W}_{i_k} = \mathbf{I}+\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}, \tag{4.35}$$

$$\preceq \mathbf{I}+\frac{1}{\sigma_{i_k}^2}\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}. \tag{4.36}$$

Now introduce the spectral norm $|||\mathbf{D}|||_2 = \max_{\mathbf{c}}\frac{\|\mathbf{Dc}\|_2}{\|\mathbf{c}\|_2} = s_{\max}(\mathbf{D})$. Then, for all

Table 4.4. Feedback and estimation needed for the different methods

Method	Globally shared common scale parameter (Section 4.2.1)	Globally shared individual scale parameters (Section 4.2.2)	Globally shared filters (Section 4.2.3)
Estimated at MS i_k	$\mathbf{\Phi}_{i_k},\ \mathbf{F}_{i_k}$	$\mathbf{\Phi}_{i_k},\ \mathbf{F}_{i_k}$	$\{\mathbf{H}_{i_k j}\}$ [c]
BS i feedback to served MS i_k	—	—	$\{\mathbf{V}_{j_l}\}$
Estimated at BS i	$\mathbf{\Gamma}_i,\ \{\mathbf{G}_{i_k}\}$	$\{\mathbf{G}_{j_l i}\}$	$\{\mathbf{H}_{j_l i}^{\mathsf{H}}\}$ [c]
MS i_k feedback to serving BS i	$\mathbf{W}_{i_k},\ \gamma_{i_k}$	$\|\mathbf{U}_{i_k}\|_{\mathrm{F}},\ \mathbf{W}_{i_k}$	$\mathbf{U}_{i_k}, \mathbf{W}_{i_k}$
Shared over BS backhaul	$\{\gamma_{j_l}\}$ [a]	$\left\{\frac{\sqrt{\overleftarrow{P}_{j_l}}}{\|\mathbf{U}_{j_l}\|_{\mathrm{F}}}\right\}$ [b]	$\{\mathbf{U}_{j_l}\}$ [b]

[a] The quantities $\gamma_{j_l} = \sqrt{\frac{\overleftarrow{P}_{j_l}\sigma_{j_l}^2}{\alpha_{j_l} d_{j_l}}}$ only need to be shared among BSs whenever some of the involved variables change. For static conditions, this means that they only need to be shared once.

[b] These quantities must be shared over the BS backhaul in each iteration.

[c] The estimated quantities for the methods in Section 4.2.1 and Section 4.2.2 depend on the transmit and receive filters, and must therefore be re-estimated in every iteration. The estimated quantities for the method in Section 4.2.3 do not change within one coherence block, and can therefore be improved upon in every iteration.

$\mathbf{c}_{i_k}$ such that $\|\mathbf{c}_{i_k}\|_2 = 1$, we have that

$$\begin{aligned}
&\mathbf{c}_{i_k}^{\mathsf{H}} \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{c}_{i_k} \leq \|\mathbf{V}_{i_k} \mathbf{c}_{i_k}\|_2^2 \cdot \lambda_{\max}\left(\mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{H}_{i_k i}\right) \\
&= \|\mathbf{V}_{i_k} \mathbf{c}_{i_k}\|_2^2 \cdot s_{\max}^2\left(\mathbf{H}_{i_k i}\right) \leq |||\mathbf{V}_{i_k}|||_2^2 \cdot \|\mathbf{c}_{i_k}\|_2^2 \cdot s_{\max}^2\left(\mathbf{H}_{i_k i}\right) \\
&= |||\mathbf{V}_{i_k}|||_2^2 \cdot s_{\max}^2\left(\mathbf{H}_{i_k i}\right) \leq \|\mathbf{V}_{i_k}\|_{\mathrm{F}}^2 \cdot s_{\max}^2\left(\mathbf{H}_{i_k i}\right) \leq P_i s_{\max}^2\left(\mathbf{H}_{i_k i}\right).
\end{aligned} \tag{4.37}$$

Thus, $\lambda_{\max}\left(\mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}\right) \leq P_i s_{\max}^2\left(\mathbf{H}_{i_k i}\right)$, and the upper bound then directly follows from (4.36). For the lower bound, note that

$$\mathbf{I} \preceq \mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} = \mathbf{W}_{i_k}. \tag{4.38}$$

□

That is, the eigenvalues of $\mathbf{W}_{i_k}$ can be suitably quantized over

$$\left[1, 1 + \frac{P_i s_{\max}^2(\mathbf{H}_{i_k i})}{\sigma_{i_k}^2}\right]. \tag{4.39}$$

We however leave the details of how to design such quantizers for future work.

As mentioned in Section 4.2.1, $R_{i_k} = \sum_n \log_2\left(\lambda_n\left(\mathbf{W}_{i_k}\right)\right)$ can be seen as the data rate (summed over data streams) for MS i_k. Quantizing $\lambda_n\left(\mathbf{W}_{i_k}\right)$ therefore corresponds to making a set of discrete rates available to the MS, corresponding to e.g. a set of different modulation and coding schemes.

4.3 Inherent and Enforced Robustness of WMMSE Solutions

Using the methods developed in Section 4.2, we are able to perform CSI acquisition with varying levels of distributedness. We now study the robustness of the WMMSE algorithm, when applied together with these CSI acquisition schemes.

4.3.1 Naïve WMMSE Algorithm with Estimated CSI

It is straightforward to naïvely feed the WMMSE algorithm the estimated CSI from one of the presented CSI acquisition methods. An example of the resulting performance can be seen in Figure 4.3. The simulation settings are described in detail in Section 4.4. It is clear that the naïve application of the WMMSE algorithm works moderately well for the centralized CSI acquisition schemes, but performance for the fully distributed CSI acquisition scheme in Section 4.2.1 catastrophically deteriorates at high SNR. Thus, some form of robustification against CSI estimation errors is necessary.

4.3.2 Diagonal Loading as a Robustifying Structure

One approach to robustifying the WMMSE optimization problem in (2.62) on page 36 is to minimize the objective function under the worst-case estimation error

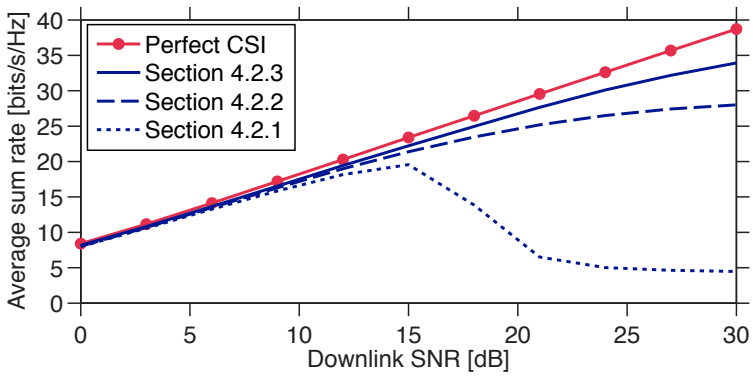


Figure 4.3. Sum rate performance when naïvely applying the WMMSE algorithm together with the CSI acquisition schemes. The scenario was a $I = 3$, $K_i = 2\ \forall i$, $M = 4$, $N = 2$ interfering broadcast channel with $d = 1$. The channels were i.i.d. Rayleigh fading, and the uplink SNR was set as $\mathrm{SNR}_{\mathrm{u}} = \overleftarrow{P}/\varsigma^2 = 10\,\mathrm{dB}$ for all links. Note that $\mathrm{SNR}_{\mathrm{d}} = P/\sigma^2$ affects both the power constraint in the WMMSE algorithm, as well as the estimation performance in the downlink estimation, since the downlink pilots are precoded with the same precoders as used in the data transmission.

conditions. This is typically formulated by letting the estimation errors belong to some known convex set, and then solving a corresponding min-max problem. The CSI acquisition methods proposed herein provide estimates of the effective channels and covariance matrices. Due to the definition of the effective channels, a worst-case optimization problem cannot be formulated for both uplink and downlink estimation errors simultaneously however[4]. Instead, each BS and MS can locally formulate a worst-case optimization problem, given the estimation errors at that particular BS or MS. In this section, we outline these local worst-case optimization problems.

Worst-Case Robust Precoders

By first studying the optimization at the BSs, we let the estimation errors for BS i be

$$\widetilde{\mathbf{\Gamma}}_i = \mathbf{\Gamma}_i - \widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i}}, \tag{4.40}$$

$$\widetilde{\mathbf{G}}_{i_k} = \mathbf{G}_{i_k} - \widehat{\mathbf{G}}_{i_k}, \quad k = 1, \ldots, K_i. \tag{4.41}$$

[4]For example, the term $\mathbf{W}_{i_k}^{1/2}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{V}_{i_k} = \mathbf{G}_{i_k}^{\mathsf{H}}\mathbf{V}_{i_k} = \mathbf{W}_{i_k}^{1/2}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{F}_{i_k}$ cannot be written in terms of $\mathbf{G}_{i_k}$ and $\mathbf{F}_{i_k}$ simultaneously.

We assume that the errors are norm bounded as

$$\left\|\widetilde{\mathbf{\Gamma}}_i\right\|_{\mathrm{F}} \leq \varepsilon_i^{(\mathrm{BS})}, \tag{4.42}$$

$$\left\|\widetilde{\mathbf{G}}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}} \leq \xi_{i_k}^{(\mathrm{BS})}, \quad k = 1, \ldots, K_i. \tag{4.43}$$

Note that $\xi_{i_k}^{(\mathrm{BS})}$ depends on $\mathbf{W}_{i_k}$. The worst-case optimization problem for BS i is then (cf. (2.67) on page 37):

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \max_{\substack{\left\|\widetilde{\mathbf{\Gamma}}_i\right\|_{\mathrm{F}} \leq \varepsilon_i^{(\mathrm{BS})} \\ \left\|\widetilde{\mathbf{G}}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}} \leq \xi_{i_k}^{(\mathrm{BS})},\, k=1,\ldots,K_i}} \sum_{k=1}^{K_c} \mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\left(\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i}} + \widetilde{\mathbf{\Gamma}}_i\right)\mathbf{V}_{i_k}\right) \\ & - 2\sqrt{\alpha_{i_k}}\mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}^{1/2}\left(\widehat{\mathbf{G}}_{i_k} + \widetilde{\mathbf{G}}_{i_k}\right)^{\mathsf{H}}\mathbf{V}_{i_k}\right)\right) \\ \text{subject to} \quad & \sum_{k=1}^{K_c} \mathrm{Tr}\left(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^{\mathsf{H}}\right) \leq P_i,\ i = 1, \ldots, I. \end{aligned} \tag{4.44}$$

The solution to the inner optimization problem of (4.44) can be found by extending the results of [SGLW03, ZSGL05] to the multiuser matrix case. By upper bounding the optimal value of the inner optimization problem using the triangle inequality[5] and the submultiplicativity of the Frobenius norm, the form of the (pessimistic) robust optimal precoder for MS i_k is

$$\mathbf{V}_{i_k}^{\mathrm{rob}} = \sqrt{\alpha_{i_k}}\left(\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i}} + \left(\varepsilon_i^{(\mathrm{BS})} + \frac{\xi_{i_k}^{(\mathrm{BS})}}{\left\|\mathbf{V}_{i_k}^{\mathrm{rob}}\right\|_{\mathrm{F}}} + \mu_i\right)\mathbf{I}\right)^{-1}\widehat{\mathbf{G}}_{i_k}\mathbf{W}_{i_k}^{1/2}. \tag{4.45}$$

As before, μ_i is the Lagrange multiplier for the sum power constraint. Note that the robust precoder in (4.45) is *diagonally loaded* by a constant factor $\varepsilon_i^{(\mathrm{BS})}$, a data dependent factor $\xi_{i_k}^{(\mathrm{BS})}/\left\|\mathbf{V}_{i_k}^{\mathrm{rob}}\right\|_{\mathrm{F}}$, and the Lagrange multiplier μ_i. Diagonal loading is well known to robustify beamformers in various settings, and a large body of literature has studied its robustifying effects; see e.g. [CZO87, Car88, WBM96, LSW03, VGL03, SGLW03, ZSGL05].

In order to construct the robust precoder in (4.45), the parameters $\varepsilon_i^{(\mathrm{BS})}$ and $\xi_{i_k}^{(\mathrm{BS})}$ must be known. For the fully distributed CSI estimation in Section 4.2.1, the effective channel error $\widetilde{\mathbf{G}}_{i_k}$ follows a zero-mean Gaussian distribution with known covariance, and $\xi_{i_k}^{(\mathrm{BS})}$ can thus be selected such that $\left\|\widetilde{\mathbf{G}}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}} \leq \xi_{i_k}^{(\mathrm{BS})}$ holds with some probability. The statistics of the covariance error $\widetilde{\mathbf{\Gamma}}_i$ however depend on the filters $\{\mathbf{U}_{i_k}\}$, which are unknown at BS i. Since the optimal amount of diagonal loading is unknown, we therefore propose to disregard $\varepsilon_i^{(\mathrm{BS})}$ and $\xi_{i_k}^{(\mathrm{BS})}$, and let the

[5] This effectively relaxes the problem such that the $\widetilde{\mathbf{\Gamma}}_i$ is the worst for each user simultaneously.

factor μ_i handle all the diagonal loading. To compensate for the missing $\varepsilon_i^{(\mathrm{BS})}$ and $\xi_{i_k}^{(\mathrm{BS})}$, we will implicitly amplify μ_i using a scaling procedure, to be described in Section 4.3.3.

Worst-Case Robust Receive Filters

The development of the worst-case robust receive filters follow similarly as for the worst-case robust precoders. For completeness, we however briefly discuss this case as well.

The estimation errors for MS i_k are

$$\widetilde{\mathbf{\Phi}}_{i_k} = \mathbf{\Phi}_{i_k} - \widehat{\mathbf{\Phi}}_{i_k}, \tag{4.46}$$

$$\widetilde{\mathbf{F}}_{i_k} = \mathbf{F}_{i_k} - \widehat{\mathbf{F}}_{i_k}. \tag{4.47}$$

We assume that the errors are norm bounded as

$$\left\|\widetilde{\mathbf{\Phi}}_{i_k}\right\|_{\mathrm{F}} \leq \varepsilon_i^{(\mathrm{MS})}, \tag{4.48}$$

$$\left\|\widetilde{\mathbf{F}}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}} \leq \xi_{i_k}^{(\mathrm{MS})}. \tag{4.49}$$

The worst-case optimization problem for the receive filter of MS i_k is then:

$$\begin{aligned}\underset{\{\mathbf{A}_{i_k}\}}{\text{minimize}} \quad & \max_{\substack{\left\|\widetilde{\mathbf{\Phi}}_{i_k}\right\|_{\mathrm{F}} \leq \varepsilon_i^{(\mathrm{UE})} \\ \left\|\widetilde{\mathbf{F}}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}} \leq \xi_{i_k}^{(\mathrm{UE})}}} \quad \mathrm{Tr}\left(\mathbf{W}_{i_k}\left(\mathbf{I} + \mathbf{A}_{i_k}^{\mathsf{H}}\left(\widehat{\mathbf{\Phi}}_{i_k} + \widetilde{\mathbf{\Phi}}_{i_k}\right)\mathbf{A}_{i_k}\right)\right) \\ & \qquad -2\mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}\left(\widehat{\mathbf{F}}_{i_k} + \widetilde{\mathbf{F}}_{i_k}\right)^{\mathsf{H}}\mathbf{A}_{i_k}\right)\right),\end{aligned} \tag{4.50}$$

Using similar techniques as in the previous section, it can be shown that the form of the (pessimistic) worst-case robust receive filter for MS i_k is

$$\mathbf{A}_{i_k}^{\mathrm{rob}} = \left(\widehat{\mathbf{\Phi}}_{i_k} + \left(\varepsilon_i^{(\mathrm{MS})} + \frac{\xi_{i_k}^{(\mathrm{MS})}}{\left\|\mathbf{A}_{i_k}^{\mathrm{rob}}\right\|_{\mathrm{F}}}\right)\mathbf{I}\right)^{-1}\widehat{\mathbf{F}}_{i_k}. \tag{4.51}$$

Again, the robust solution is diagonally loaded by a quantity that is unknown, in this case $\varepsilon_i^{(\mathrm{MS})}$. Our proposed solution to this conundrum is to enforce a particular property onto the receive filters. This will be detailed in Section 4.3.4.

4.3.3 Precoder Robustness

From Section 4.3.2 we know that the structure of the optimal worst-case robust precoder is that of diagonal loading. Unfortunately, the optimal level of diagonal loading is unknown however. We now detail a heuristic for selecting the diagonal loading level.

Given estimates $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}$, $\widehat{\mathbf{G}}_{i_k}$ and fed back $\mathbf{W}_{i_k}$, the naïve precoders from the original WMMSE algorithm are formed like (cf. (4.1) on page 66)

$$\mathbf{V}_{i_k} = \sqrt{\alpha_{i_k}} \left(\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}} + \mu_i \mathbf{I} \right)^{-1} \widehat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}. \tag{4.52}$$

Note that the form of (4.52) and (4.45) on page 81 are similar, and that μ_i alone acts as the diagonal loading for the naïve WMMSE precoder. The level of diagonal loading is determined by $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}$, $\widehat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and P_i. As seen in Figure 4.3, this level of diagonal loading may be too low when it is solely determined by the sum power constraint. We therefore propose to artificially inflate the level, using a scaling technique.

Common Downlink Sum Power Constraint Scale Factor

We let $0 \leq \rho \leq 1$ be a scaling factor, and modify the WMMSE algorithm with the following steps:

1. In the precoder optimization at BS i, let the sum power constraint be ρP_i. The resulting precoders from (4.52) are denoted $\{\mathbf{V}_{i_k}^{(\rho)}\}$, and will have equal or higher diagonal loading level than the original precoder in (4.52), since μ_i is nonincreasing in the sum power constraint value.

2. Form scaled precoders $\mathbf{V}_{i_k} = \frac{1}{\sqrt{\rho}} \mathbf{V}_{i_k}^{(\rho)}$, and use these for downlink pilot and data transmission. This scaling ensures that all allowable transmit power can be used.

3. At the MSs, perform the estimation given the precoders $\{\mathbf{V}_{i_k}\}$, giving $\{\widehat{\mathbf{F}}_{i_k}^{(\rho)}\}$ and $\{\widehat{\mathbf{\Phi}}_{i_k}^{(\rho)}\}$.

4. Scale the estimates as $\widehat{\mathbf{\Phi}}_{i_k} = \rho \widehat{\mathbf{\Phi}}_{i_k}^{(\rho)}$, $\widehat{\mathbf{F}}_{i_k} = \sqrt{\rho} \widehat{\mathbf{F}}_{i_k}^{(\rho)}$, and use $\{\widehat{\mathbf{\Phi}}_{i_k}\}$ and $\{\widehat{\mathbf{F}}_{i_k}\}$ to form receive filters and MSE weights. This scaling is necessary in order for the WMMSE algorithm at the MSs to be aware of what the original precoders $\mathbf{V}_{i_k}^{(\rho)}$ were.

The same scaling ρ is used at all BSs, and therefore the signal-to-interference ratios of the cross-links are not affected. In Figure 4.4 on the following page, we plot the impact of selecting different ρ. A simple selection that appears to work well is $\rho = \min(\overleftarrow{P}/P, 1)$.

Removing the Noise Component of $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}$

Comparing (4.52) on the current page with the version with perfect CSI in (2.69) on page 37, it can be noted that the covariance matrix should be $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i}}$, and not $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}$. The noise portion of $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}$ will on average be $\frac{\sigma_t^2}{\gamma^2}$, but simply subtracting

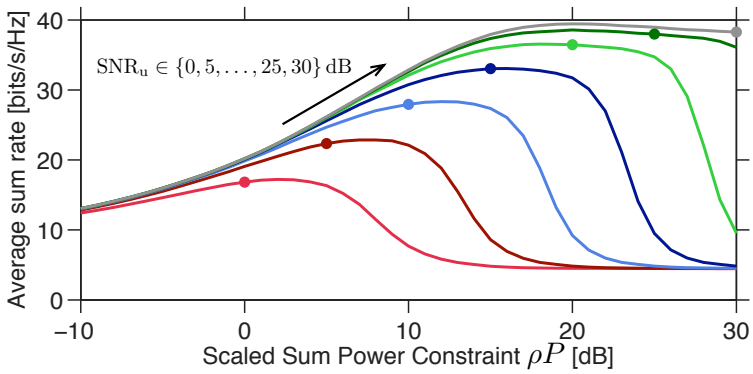


Figure 4.4. Sum rate performance when varying ρ, for $P = 1000$ and $\sigma^2 = 1$ and varying SNR_u. The solid markers represent the performance for $\rho = \min(\overleftarrow{P}/P, 1)$. The scenario is the same as in Figure 4.3.

that might make the resulting matrix indefinite. Instead, we modify μ_i to allow for *negative* values; this is the same as seeing μ_i as the difference of a non-negative Lagrange multiplier with an estimate of the noise power. Specifically, we let $\mu_i \geq -\min\left(\frac{\sigma_t^2}{\gamma^2}, \lambda_{\min}\left(\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}\right) - \zeta\right)$ where ζ is some value determining how close to singular $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}} + \mu_i \mathbf{I}$ can be.

4.3.4 Receive Filter and MSE Weight Robustness

The precoders determined at the BS side are naturally robustified due to the diagonal loading from the downlink sum power constraint. By artificially inflating the loading level, the robustness level can be increased. On the MS side however, where $\mathbf{A}_{i_k}$ and $\mathbf{W}_{i_k}$ are found, so such natural robustification takes place. Again, the optimal worst-case robustness structure is that of diagonal loading however. Since the optimal level of diagonal loading is unknown, in this section we propose a method for selecting this level. This procedure is based on the following observation:

Theorem 4.1. *The receive filter $\mathbf{A}_{i_k}$ and MSE weight $\mathbf{W}_{i_k}$ obtained for MS i_k in the WMMSE algorithm with perfect CSI (Algorithm 4.1 on page 66) satisfies*

$$\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_F^2 = \mathrm{Tr}\left(\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1} - \mathbf{\Phi}_{i_k}^{-1}\right) \leq \frac{d_{i_k}}{\sigma_{i_k}^2}.$$

If the effective channel $\mathbf{F}_{i_k}$ is fully contained in an interference-free subspace of dimension $\widetilde{d}_{i_k} \leq d_{i_k}$, then asymptotically $\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_F^2 \to \widetilde{d}_{i_k}/\sigma_{i_k}^2$ as the SNR in the interference-free subspace grows large.

Proof. The proof is given in Appendix 4.A. □

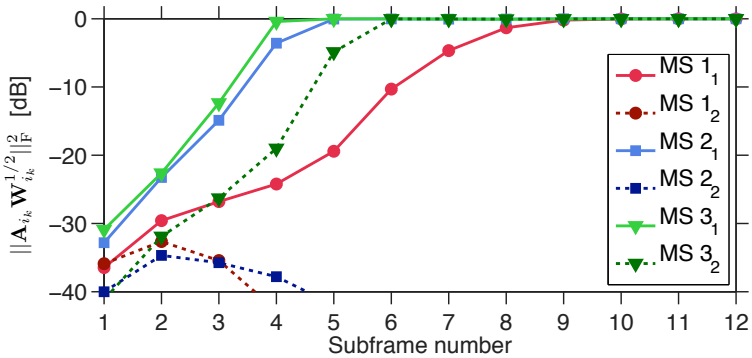


Figure 4.5. Example of convergence of $\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2$ for $d = d_{i_k} = 1$, $\forall i_k$, $\sigma^2 = \sigma_{i_k}^2 = 1$, $\forall i_k$. Note how 4 MSs get close to achieving the bound, and how the remaining 2 MSs are effectively turned off.

Remark 4.4. *The first part of this theorem has an important connection to the uplink training phase in Section 4.2.1. In the uplink training, the transmitted signal from MS i_k is $\overleftarrow{\mathbf{S}}_{i_k} = \gamma \mathbf{U}_{i_k} \overleftarrow{\mathbf{P}}_{i_k} = \gamma\sqrt{\alpha_{i_k}}\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\overleftarrow{\mathbf{P}}_{i_k}$. Since γ, α_{i_k} and $\overleftarrow{\mathbf{P}}_{i_k}$ are fixed, then $\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}$ directly determines the effective MS transmit power, and hence the uplink estimation SNR. The second part of the theorem shows that $\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_F^2$ also indicates whether 'perfect IA' is achieved for MS i_k.*

An example of how $\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2$ converges, as a function of subframe number for a $I = 3, K = 2$ network at $P/\sigma^2 = 30\,\mathrm{dB}$, can be seen in Figure 4.5.

Enforcing Theorem 4.1 onto WMMSE Solutions with Imperfect CSI

Theorem 4.1 relates to perfect CSI, but the inequality may not hold for the naïve solutions in (2.58) and (2.65) on pages 35–37 with imperfect CSI. In order to robustify the algorithm, we therefore explicitly impose the constraint on the MS side optimization problem with imperfect CSI. Fixing $\{\mathbf{V}_{i_k}\}$ in (2.62) on page 36, the resulting optimization problem decouples over users. Dropping constant terms, the optimization problem that MS i_k should solve is then

$$\begin{aligned} \underset{\mathbf{A}_{i_k},\mathbf{W}_{i_k}\succ\mathbf{0}}{\text{minimize}} \quad & \mathrm{Tr}\left(\mathbf{W}_{i_k}\widehat{\mathbf{E}}_{i_k}\right) - \log_e\det\left(\mathbf{W}_{i_k}\right) \\ \text{subject to} \quad & \left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 \leq \frac{d_{i_k}}{\sigma_{i_k}^2} \end{aligned} \tag{4.53}$$

where $\widehat{\mathbf{E}}_{i_k} = \mathbf{I} - \mathbf{A}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k} - \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{A}_{i_k} + \mathbf{A}_{i_k}^{\mathsf{H}}\widehat{\mathbf{\Phi}}_{i_k}\mathbf{A}_{i_k}$.

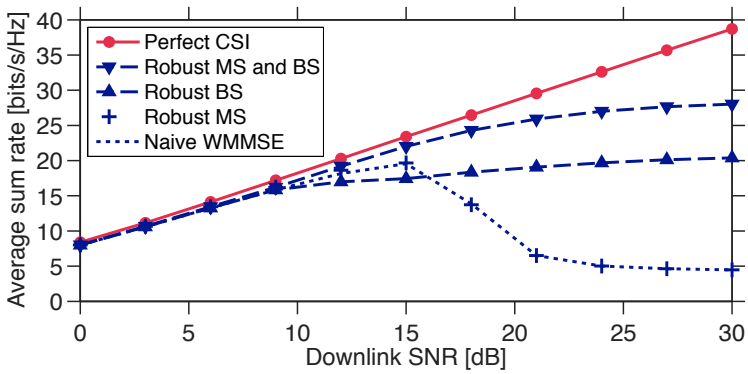


Figure 4.6. Sum rate performance when selectively applying the robustifying measures, together with the fully distributed CSI acquisition in Section 4.2.1. For comparison purposes, the scenario is the same as in Figure 4.3 on page 80.

Theorem 4.2. *The solution to* (4.53) *is*

$$\begin{aligned}\mathbf{A}_{i_k}^{\star} &= \left(\widehat{\boldsymbol{\Phi}}_{i_k} + \nu_{i_k}^{\star}\mathbf{I}\right)^{-1}\widehat{\mathbf{F}}_{i_k} \\ \mathbf{W}_{i_k}^{\star} &= \left(\mathbf{I} - \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\left(\widehat{\boldsymbol{\Phi}}_{i_k} + \nu_{i_k}^{\star}\mathbf{I}\right)^{-1}\widehat{\mathbf{F}}_{i_k}\right)^{-1} \\ &= \mathbf{I} + \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\left(\widehat{\boldsymbol{\Phi}}_{i_k}^{\mathsf{i+n}} + \nu_{i_k}^{\star}\mathbf{I}\right)^{-1}\widehat{\mathbf{F}}_{i_k}.\end{aligned}$$

If $\left\|\mathbf{A}_{i_k}^{\star}\left(\mathbf{W}_{i_k}^{\star}\right)^{1/2}\right\|_F^2 \leq d_{i_k}/\sigma_{i_k}^2$ *holds for* $\nu_{i_k}^{\star} = 0$, *the constraint is not active and the solution has the same form as the original solution in* (2.58) *and* (2.65). *Otherwise,* $\nu_{i_k}^{\star}$ *can be found by bisection over* $\left(0, \sigma_{i_k}^2\right]$ *such that* $\left\|\mathbf{A}_{i_k}^{\star}\left(\mathbf{W}_{i_k}^{\star}\right)^{1/2}\right\|_F^2 = d_{i_k}/\sigma_{i_k}^2$.

Proof. The proof is given in Appendix 4.B. □

Interestingly, explicitly imposing the results of Theorem 4.1 as a constraint in (4.53) corresponds to diagonal loading of the receive filter, thereby robustifying it in a similar fashion as the robustification of the precoder in Section 4.3.3. By increasing ν_{i_k}, the requested rate $\log_2\det\left(\mathbf{W}_{i_k}\right)$ is decreased. A large ν_{i_k}, would occur when there are obvious discrepancies in the estimated CSI, such that $\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 \leq d_{i_k}/\sigma_{i_k}^2$ is far from being fulfilled without the diagonal loading. Imposing the constraint in the optimization can be seen as a 'sanity check' on the solution.

We visualize the robustifying effects in Figure 4.6, using the same simulation settings as in Figure 4.3 on page 80. The robustifying measures are effective, and

Algorithm 4.3 RB-WMMSE Algorithm (Estimated CSI)

Input: BS robustification parameter $\rho \in [0,1]$.
repeat
At MSs:
  Pilot transmission from BSs: estimate $\widehat{\mathbf{\Phi}}_{i_k}^{(\rho)}$ and $\widehat{\mathbf{F}}_{i_k}^{(\rho)}$
  Rescale $\widehat{\mathbf{\Phi}}_{i_k} = \rho\widehat{\mathbf{\Phi}}_{i_k}^{(\rho)}$, $\widehat{\mathbf{F}}_{i_k} = \sqrt{\rho}\widehat{\mathbf{F}}_{i_k}^{(\rho)}$
  Find ν_{i_k} to satisfy $\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 \leq d_{i_k}/\sigma_{i_k}^2$
  $\mathbf{W}_{i_k} = \left(\mathbf{I} - \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\left(\widehat{\mathbf{\Phi}}_{i_k} + \nu_{i_k}\mathbf{I}\right)^{-1}\widehat{\mathbf{F}}_{i_k}\right)^{-1}$
  $\mathbf{A}_{i_k} = \left(\widehat{\mathbf{\Phi}}_{i_k} + \nu_{i_k}\mathbf{I}\right)^{-1}\widehat{\mathbf{F}}_{i_k}$, $\mathbf{U}_{i_k} = \sqrt{\alpha_{i_k}}\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}$
At BSs:
  Pilot transmission from MSs: estimate $\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}$ and $\widehat{\mathbf{G}}_{i_k}$.
  Obtain $\mathbf{W}_{i_k}^{1/2}$ through feedback.
  Find $\mu_i \geq -\min\left(\frac{\varsigma_i^2}{\gamma^2}, \lambda_{\min}\left(\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}}\right) - \epsilon\right)$
  to satisfy $\sum_{k=1}^{K_i}\mathrm{Tr}\left(\mathbf{V}_{i_k}^{(\rho)}\mathbf{V}_{i_k}^{(\rho),\mathsf{H}}\right) \leq \rho P_i$
  $\mathbf{B}_{i_k}^{(\rho)} = \left(\widehat{\mathbf{\Gamma}}_i^{\mathrm{s+i+n}} + \mu_i\mathbf{I}\right)^{-1}\widehat{\mathbf{G}}_{i_k}$, $\mathbf{V}_{i_k}^{(\rho)} = \sqrt{\alpha_{i_k}}\mathbf{B}_{i_k}^{(\rho)}\mathbf{W}_{i_k}^{1/2}$
  Scale $\mathbf{V}_{i_k} = \frac{1}{\sqrt{\rho}}\mathbf{V}_{i_k}^{(\rho)}$
until fixed number of iterations

when combined results in a factor 5 sum rate gain over the naïve WMMSE algorithm at high SNR.

4.3.5 Robustified WMMSE Algorithm

We now combine the diagonal loading robustifications in Section 4.3.3 and Section 4.3.4 to form a *RoBustified WMMSE algorithm* (RB-WMMSE); see Algorithm 4.3 on this page. This algorithm can be combined with any of the CSI acquisition schemes outlined in Section 4.2, and the joint system is fully distributed if the CSI acquisition is distributed.

In the performance evaluation to come, we will compare to the existing robust WMMSE algorithms in [NGS12, SM12, LKY13, RBCL13]. These also gain their robustness from diagonal loading, but their diagonal loading comes from optimizing a lower bound on performance. These methods require the CSI acquisition method in Section 4.2.3, and the joint coordinated precoding system can hence not be implemented in a fully distributed fashion.

4.4 Performance Evaluation

We evaluate the performance of the proposed RB-WMMSE algorithm, together with the distributed CSI acquisition systems, using numerical simulations. Synthetic channels were used, and the channel model was i.i.d. Rayleigh fading on all channels such that $[\mathbf{H}_{i_k j}]_{mn} \sim \mathcal{CN}(0,1)$. The results were averaged over 1000 Monte Carlo independent realizations. We study scenarios where I BSs each serve K_c MSs (i.e. $K_c = K_i,\ \forall\, i$), for a total of $K = IK_c$ users. Further, all BSs have the same number of antennas M and all MSs have the same number of antennas N. The number of served data streams are $d = d_{i_k}$ for all MSs. All MSs have the same noise power $\sigma^2 = \sigma^2_{i_k},\ \forall\, i_k$, and all BSs have the same noise power $\varsigma^2 = \varsigma^2_i,\ \forall\, i$. All BSs have the same maximum downlink transmit power P, and all MSs have the same maximum uplink transmit power $\overleftarrow{P}$. We let all users have the same data rate weights, i.e. $\alpha_{i_k} = 1,\ \forall\, i_k$. Unless otherwise stated, the BS power scaling is set as $\rho = \min(\overleftarrow{P}/P, 1)$ based on the results in Figure 4.4. Due to the symmetry of the setup, $\gamma = \gamma_{i_k}$ for the CSI acquisition in Section 4.2.1, and γ can thus be assumed to be known a priori at the BSs. In this setup, the CSI acquisition in Section 4.2.1 is therefore fully distributed. For all estimation schemes, we let $L_{p,d} = IM$ and $L_{p,u} = KN$. For the uplink and downlink pilots, truncated DFT matrices of appropriate dimensions were used. We always initialize the algorithms with precoders corresponding to truncated DFT matrices.

As a baseline performance measure, we use single-user eigenprecoding and waterfilling. For the single-user processing, we show the performance under TDMA, as well as under uncoordinated concurrent transmissions (both with water filling).

4.4.1 Convergence

First, we investigate the average convergence behaviour of the RB-WMMSE algorithm for an interfering broadcast channel with $I = 3$ BSs. Each BS serves $K_c = 2$ users with $d = 1$ data stream each. The number of antennas is $M = 4$ and $N = 2$. The downlink and uplink SNRs are $\mathrm{SNR_d} = P/\sigma^2 = 20\,\mathrm{dB}$ and $\mathrm{SNR_u} = \overleftarrow{P}/\varsigma^2 = 10\,\mathrm{dB}$, respectively. The performance of RB-WMMSE is compared for the CSI acquisition schemes in Section 4.2.1 and Section 4.2.2. The results are shown in Figure 4.7 on the next page, where it can be seen that the robustifying measures in RB-WMMSE are needed for good performance. For the estimation with global sharing of individual scale factors (Section 4.2.2), convergence is slightly faster than for the fully distributed estimation (Section 4.2.1). The WMMSE algorithms need on the order of 1000 iterations to converge, which is consistent with the findings of [SSB+13]. We do however note that a significant fraction of the final performance is achieved after just around 10 to 20 iterations.

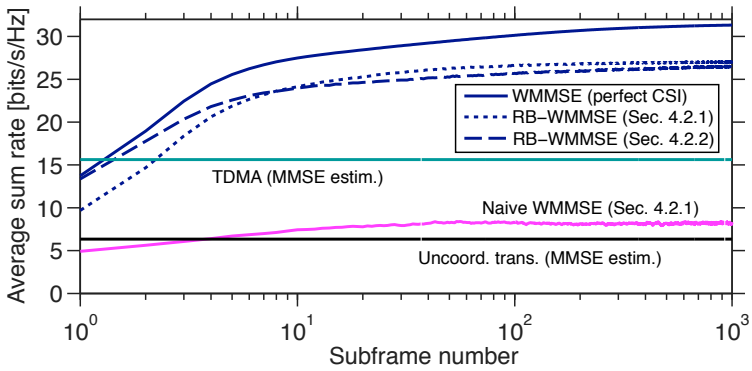


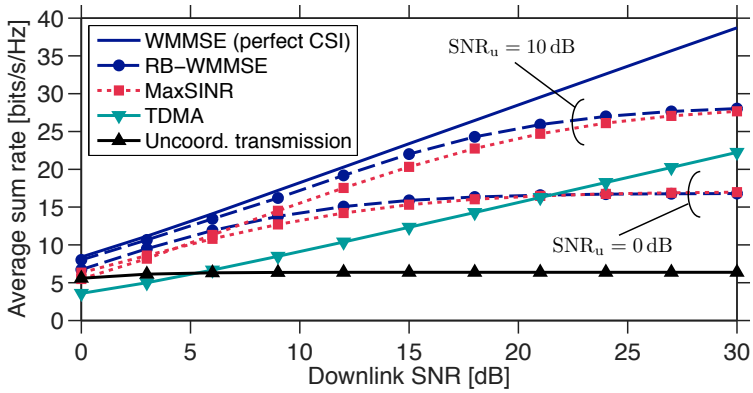
Figure 4.7. Convergence comparison of the different methods for $I = 3, K_c = 2, M = 4, N = 2, d = 1$, $\mathrm{SNR_d} = 20\,\mathrm{dB}$ and $\mathrm{SNR_u} = 10\,\mathrm{dB}$.

4.4.2 Fixed SIR, Varying SNR

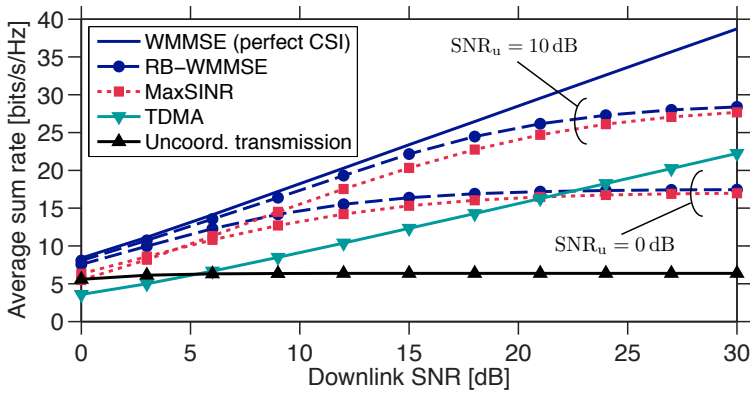
Next, we study the sum rate when varying downlink and uplink SNRs. Recall that the downlink SNR affects both the downlink power constraint in the resource allocation, as well as the downlink estimation performance (see Section 4.2.1). The uplink SNR only affects the uplink estimation performance.

Interfering Broadcast Channel

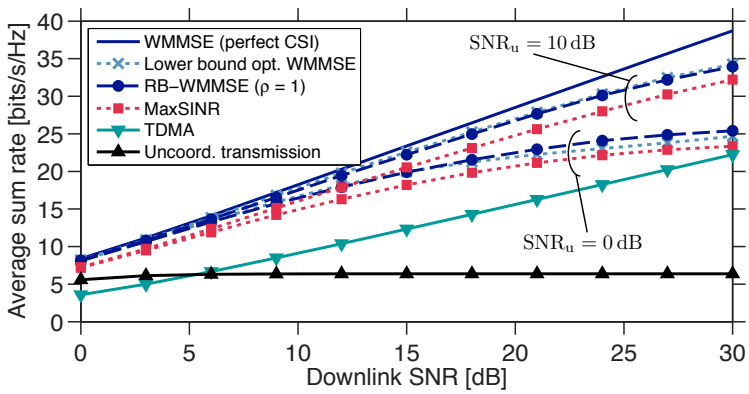
We first study an interference broadcast channel with the same system parameters as in the convergence study above. Based on the behaviour in Figure 4.7, we let the algorithms iterate for 20 iterations. The averaged achieved sum rate is plotted in Figure 4.8 on the following page. We compare to MaxSINR, for which we actively turn off two users in order not to overload the algorithm. Note that MaxSINR and Weighted MaxSINR are equivalent here, since $d = 1$. The results for the CSI acquisition with globally shared common scaling parameter (Section 4.2.1) are shown in Figure 4.8a. RB-WMMSE consistently performs slightly better than MaxSINR, and better than TDMA for sufficiently high uplink SNR. The results for the CSI acquisition with globally shared individual scaling parameter (Section 4.2.2) are shown in Figure 4.8b. The results are almost identical to those of Figure 4.8a. The results for the CSI acquisition with global sharing of filters (Section 4.2.3) are shown in Figure 4.8c. We now let $\rho = 1$, since the heuristic for selecting the diagonal loading level seems to be too conservative in this case. We compare to the (centralized) optimized lower bound method of [NGS12, SM12, LKY13, RBCL13], and it can be seen that RB-WMMSE exhibits similar performance. Furthermore, comparing to Figure 4.8a and Figure 4.8b, the corresponding sum rates are significantly higher in Figure 4.8c. This is because of the assumed perfect feedback of filters, and that the estimates of the channels can be improved in every iteration, as described in



(a) Estimation with globally shared common scaling (Section 4.2.1)

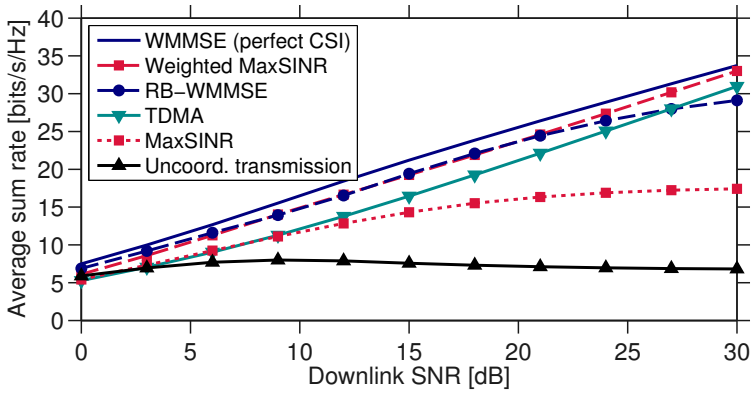


(b) Estimation with globally shared individual scaling (Section 4.2.2)

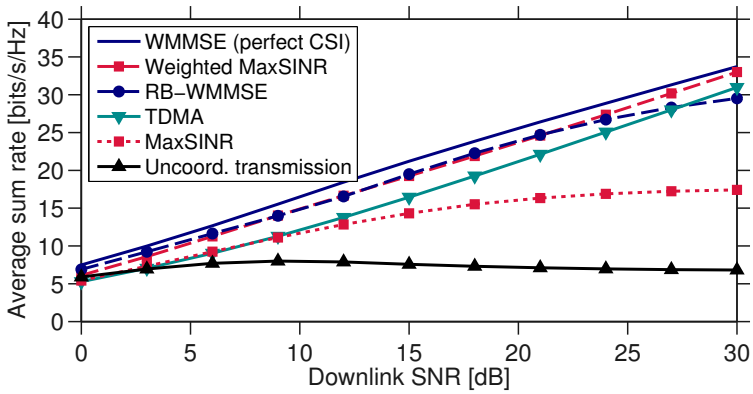


(c) Estimation with globally shared filters (Section 4.2.3)

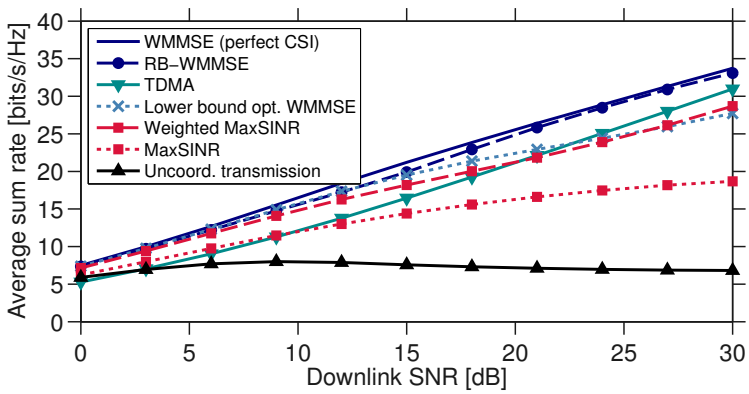
Figure 4.8. Sum rate after the 20th iteration for $I = 3, K_c = 2, M = 4, N = 2$ interfering broadcast channel with $d = 1$.



(a) Estimation with globally shared common scaling (Section 4.2.1)



(b) Estimation with globally shared individual scaling (Section 4.2.2)



(c) Estimation with globally shared filters (Section 4.2.3)

Figure 4.9. Sum rate after the 20th iteration for $I = 3, K_c = 1, M = 3, N = 3$ interference channel with $d = 2$ and $\mathrm{SNR_u} = 10\,\mathrm{dB}$.

Section 4.2.3.

Interference Channel

For the interfering broadcast scenario above, the weighted MaxSINR in Algorithm 4.2 was identical to the original MaxSINR of [GCJ11] since $d = 1$. In order to evaluate performance for the weighted MaxSINR version, we now let $d = 2$. We also change the scenario into an interference channel, where the $I = 3$ BSs each serve $K_c = 1$ user, in order to be able to check the IA feasibility using the test in [GBS14]. With $M = 3$ and $N = 3$, the test in [GBS14] gives that IA is feasible if two MSs are served with one stream, and one MS is served with two streams. We initialize the algorithms with $d = 2$ however, and let them turn off the streams as necessary. We let $\mathrm{SNR_u} = 10\,\mathrm{dB}$ and vary $\mathrm{SNR_d}$. The sum rate results are shown in Figure 4.9 on the previous page. With the CSI acquisition from Section 4.2.1, it can be seen that for low to intermediate SNR, RB-WMMSE and weighted MaxSINR perform similarly. At high SNR however, the performance of RB-WMMSE flattens out. This is due to its slow convergence behaviour at high SNR. The unweighted MaxSINR performs poorly due to being overloaded with respect to an IA feasible data stream allocation. With the CSI acquisition from Section 4.2.3, the results are almost identical. When the centralized CSI acquisition from Section 4.2.3 is applied however, RB-WMMSE performs significantly better than the weighted MaxSINR. This is possibly due to the ad hoc nature of the weighted MaxSINR.

4.4.3 Fixed SNR, Varying SIR

Next, we investigate sum rate performance for varying signal-to-interference ratio (SIR). We study the interfering broadcast channel scenario described earlier, with $\mathrm{SNR_d} = 30\,\mathrm{dB}, \mathrm{SNR_u} = 10\,\mathrm{dB}$. Now, we do not turn off any users for MaxSINR, since that hinders performance at high SIR. The results are shown in Figure 4.10 on the facing page. RB-WMMSE performs better than MaxSINR, and gets close to the perfect CSI case when the cells are effectively uncoupled. MaxSINR is limited from its overloading at high SIR. Inter- and intra-cell TDMA (as used before) do not gain from the cell separation. If TDMA is only applied for the intra-cell interference however, performance improves at high SIR, but deteriorates at low SIR. The uncoordinated transmission approach is still limited by intra-cell interference at high SIR.

4.4.4 Sum Rate and Complexity vs. Flop Count

We also study the performance and complexity of the system, as a function of the number of pilots used. For the interfering broadcast channel scenario, with $\mathrm{SNR_d} = 20\,\mathrm{dB}$, and $\mathrm{SNR_u} = 10\,\mathrm{dB}$, we vary $L_{p,d} = L_{p,u}$ and show the results in Figure 4.11 on page 94. The CSI acquisition from Section 4.2.3 performs slightly better in the sum rate sense than the CSI acquisition from Section 4.2.1, but the difference in

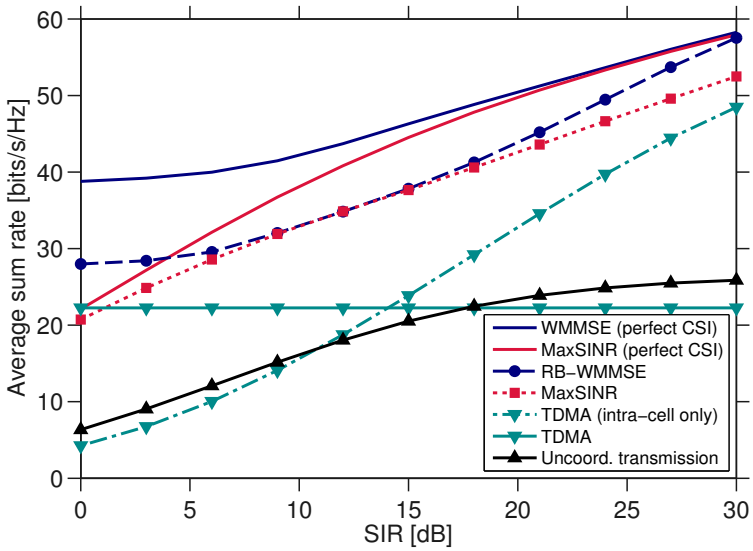


Figure 4.10. Sum rate after the 20th iteration for $I = 3, K_c = 2, M = 4, N = 2$ interfering broadcast channel with $d = 1$ and $\mathrm{SNR_d} = 30\,\mathrm{dB}, \mathrm{SNR_u} = 10\,\mathrm{dB}$.

computational complexity is significant. The reason for the the centralized CSI acquisition from Section 4.2.3 requiring particularly many flops is that it needs to estimate all interfering channels, at both MSs and BSs. The computational complexity of the RB-WMMSE algorithm is independent of the number of pilots used, and is displayed as the black line in Figure 4.11.

4.4.5 Quantized MSE Weight Feedback

Finally, we evaluate performance with quantized MSE weights. For the interfering broadcast channel scenario with fixed uplink $\mathrm{SNR_u} = 40\,\mathrm{dB}$, we vary the $\mathrm{SNR_d}$ and the number of quantization bits. Each MS had an individual codebook with MSE weights uniformly quantized on

$$\left[0, 10\log_{10}\left(1 + \frac{P_i s^2_{\max}(\mathbf{H}_{i_k i})}{\sigma^2_{i_k}}\right)\right]\ \mathrm{dB}. \tag{4.54}$$

Using the channel estimation from Section 4.2.1, the performance is shown in Figure 4.12 on the following page. For higher downlink SNR, more bits are needed for good performance. For high resolution quantization, the performance is equal to that of perfect feedback.

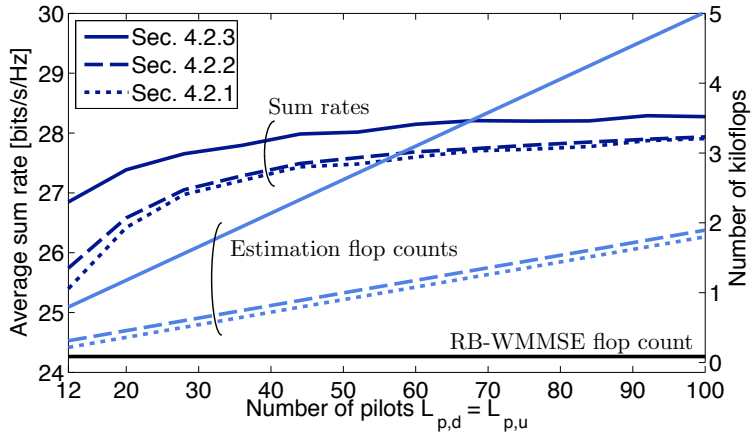


Figure 4.11. Comparison of complexity and sum rate performance after the 20th iteration for $I = 3, K_c = 2, M = 4, N = 2$ interfering broadcast channel with $d = 1$, $\mathrm{SNR_d} = 20\,\mathrm{dB}$, and $\mathrm{SNR_u} = 10\,\mathrm{dB}$.

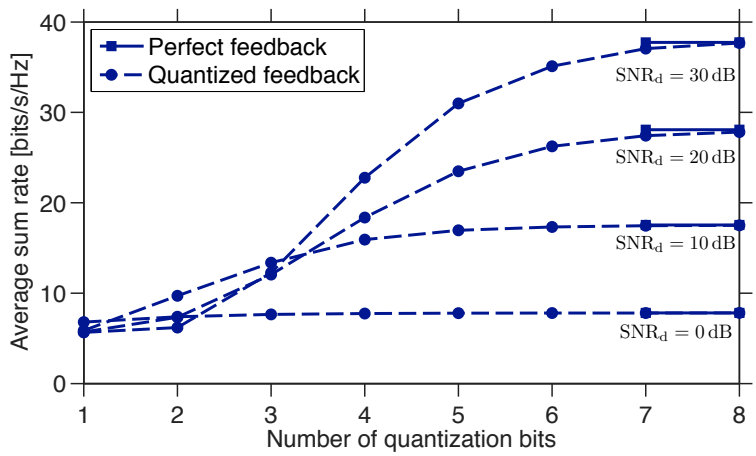


Figure 4.12. Sum rate as a function of quantization accuracy for $I = 3, K_c = 2, M = 4, N = 2$ interfering broadcast channel with $d = 1$ and $\mathrm{SNR_u} = 10\,\mathrm{dB}$. The channel estimation from Section 4.2.1 was used.

4.5 Conclusions

In this chapter, we focused directly on solving the weighted sum rate problem. In order to find a practical solution, we sought a distributed and robust coordinated precoding method. In that venture, first three CSI acquisition schemes of varying level of distributedness were proposed. The most distributed method only required joint selection of one real-valued parameter. When this parameter was fixed, all further steps could be performed in a fully distributed manner. Directly using the estimates into the WMMSE algorithm developed for perfect CSI, was shown to yield inferior performance. Instead, robustifying measures at both the BS as well as the MS sides were proposed. At the MS side, enforcing some properties of the solutions with perfect CSI, to the solutions with imperfect CSI, resulted in diagonally loaded receive filters. At the BS side, some diagonal loading was provided by the sum power constraint. This effect was amplified using a common power scaling parameter.

4.A Proof of Theorem 4.1

Proof. Decompose $\mathbf{\Phi}_{i_k} = \mathbf{F}_{i_k}\mathbf{F}_{i_k}^{\mathsf{H}} + \mathbf{\Phi}_{i_k}^{\mathrm{i+n}}$ and let $\mathbf{C}_{i_k} = \mathbf{F}_{i_k}^{\mathsf{H}} \left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1} \mathbf{F}_{i_k}$ and $\mathbf{D}_{i_k} = \mathbf{F}_{i_k}^{\mathsf{H}} \left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-2} \mathbf{F}_{i_k}$. We have that $\mathbf{A}_{i_k} = \mathbf{\Phi}_{i_k}^{-1}\mathbf{F}_{i_k}$ and $\mathbf{W}_{i_k} = \mathbf{I} + \mathbf{C}_{i_k}$. Plugging in,

$$\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(\mathbf{A}_{i_k}\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\right) = \mathrm{Tr}\left(\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{A}_{i_k}\mathbf{W}_{i_k}\right) = \mathrm{Tr}\left(\mathbf{F}_{i_k}^{\mathsf{H}}\mathbf{\Phi}_{i_k}^{-2}\mathbf{F}_{i_k}\left(\mathbf{I}+\mathbf{C}_{i_k}\right)\right).$$

Applying the matrix inversion lemma to $\mathbf{\Phi}_{i_k}^{-1}$, it can be shown that

$$\mathbf{F}_{i_k}^{\mathsf{H}}\mathbf{\Phi}_{i_k}^{-2}\mathbf{F}_{i_k} = \left(\mathbf{I}+\mathbf{C}_{i_k}\right)^{-1}\mathbf{D}_{i_k}\left(\mathbf{I}+\mathbf{C}_{i_k}\right)^{-1}$$

after simplifications. Thus,

$$\begin{aligned}\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 &= \mathrm{Tr}\left(\left(\mathbf{I}+\mathbf{C}_{i_k}\right)^{-1}\mathbf{D}_{i_k}\right)\\ &= \mathrm{Tr}\left(\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\mathbf{F}_{i_k}\left(\mathbf{I}+\mathbf{F}_{i_k}^{\mathsf{H}}\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\mathbf{F}_{i_k}\right)^{-1}\mathbf{F}_{i_k}^{\mathsf{H}}\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\right)\\ &= \mathrm{Tr}\left(\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1} - \mathbf{\Phi}_{i_k}^{-1}\right),\end{aligned}$$

where the last equality comes from applying the matrix inversion lemma backwards. Further,

$$\begin{aligned}\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 &= \mathrm{Tr}\left(\left(\mathbf{I}+\mathbf{C}_{i_k}\right)^{-1}\mathbf{D}_{i_k}\right) \overset{\text{(a)}}{\leq} \mathrm{Tr}\left(\mathbf{C}_{i_k}^{-1}\mathbf{D}_{i_k}\right)\\ &= \mathrm{Tr}\left(\left(\mathbf{F}_{i_k}^{\mathsf{H}}\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\mathbf{F}_{i_k}\right)^{-1}\mathbf{F}_{i_k}^{\mathsf{H}}\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-2}\mathbf{F}_{i_k}\right)\\ &= \left[\text{Let } \widetilde{\mathbf{F}}_{i_k} = \left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1/2}\mathbf{F}_{i_k}\right]\\ &= \mathrm{Tr}\left(\left(\widetilde{\mathbf{F}}_{i_k}^{\mathsf{H}}\widetilde{\mathbf{F}}_{i_k}\right)^{-1}\widetilde{\mathbf{F}}_{i_k}^{\mathsf{H}}\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\widetilde{\mathbf{F}}_{i_k}\right)\\ &= \mathrm{Tr}\left(\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\widetilde{\mathbf{F}}_{i_k}\left(\widetilde{\mathbf{F}}_{i_k}^{\mathsf{H}}\widetilde{\mathbf{F}}_{i_k}\right)^{-1}\widetilde{\mathbf{F}}_{i_k}^{\mathsf{H}}\right)\\ &\overset{\text{(b)}}{\leq} \max_{\mathrm{rank}\left(\mathbf{\Pi}_{i_k}\right)=d_{i_k}} \mathrm{Tr}\left(\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\mathbf{\Pi}_{i_k}\right)\\ &= \sum_{n=1}^{d_{i_k}}\lambda_n\left(\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1}\right) = \sum_{n=1}^{d_{i_k}}\frac{1}{\lambda_{N_{i_k}-n+1}\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)} \overset{\text{(c)}}{\leq} \frac{d_{i_k}}{\sigma_{i_k}^2}\end{aligned}$$

where $\mathbf{\Pi}_{i_k}$ is a rank-d_{i_k} projection matrix. The inequality (a) is due to the trace being an increasing function on the cone of positive definite matrices and the fact that $\mathbf{D}_{i_k}^{1/2}\left(\mathbf{I}+\mathbf{C}_{i_k}\right)^{-1}\mathbf{D}_{i_k}^{1/2} \preceq \mathbf{D}_{i_k}^{1/2}\mathbf{C}_{i_k}^{-1}\mathbf{D}_{i_k}^{1/2}$. The inequality (b) holds since

$\widetilde{\mathbf{F}}_{i_k}\left(\widetilde{\mathbf{F}}_{i_k}^{\mathsf{H}}\widetilde{\mathbf{F}}_{i_k}\right)^{-1}\widetilde{\mathbf{F}}_{i_k}^{\mathsf{H}}$ is a rank-d_{i_k} projection matrix. The inequality (c) is due to the fact that $\mathbf{\Phi}_{i_k}^{\mathrm{i+n}} \succeq \sigma_{i_k}^2\mathbf{I}$.

Now assume that there are $\widetilde{d}_{i_k} \leq d_{i_k}$ interference-free dimensions, and that the effective channel is fully contained in those. Let the eigenvalues of $\mathbf{\Phi}_{i_k}$ be $\{\tau_{i_k,n}^{\mathrm{s+i+n}}\}$ and the eigenvalues of $\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}$ be $\{\tau_{i_k,n}^{\mathrm{i+n}}\}$. Let the eigenvalues be ordered such that

$$\tau_{i_k,n}^{\mathrm{s+i+n}} = \begin{cases} \tau_{i_k,n}^{\mathrm{i+n}}, & n \in \{1,\ldots,N_{i_k}-\widetilde{d}_{i_k}\} \\ \tau_{i_k,n}^{\mathrm{s}} + \sigma_{i_k}^2, & n \in \{N_{i_k}-\widetilde{d}_{i_k}+1,\ldots,N_{i_k}\} \end{cases}, \tag{4.55}$$

where the first $N_{i_k}-\widetilde{d}_{i_k}$ eigenvalues are for the interference subspace, and the remaining eigenvalues are for the interference-free subspace. The $\{\tau_{i_k,n}^{\mathrm{s}}\}$ are the signal powers of the effective channel in the interference-free subspace. Further, note that $\tau_{i_k,n}^{\mathrm{i+n}} = \sigma_{i_k}^2$ for all $n \in \{N_{i_k}-\widetilde{d}_{i_k}+1,\ldots,N_{i_k}\}$. Then,

$$\begin{aligned} \mathrm{Tr}\left(\left(\mathbf{\Phi}_{i_k}^{\mathrm{i+n}}\right)^{-1} - \mathbf{\Phi}_{i_k}^{-1}\right) &= \sum_{n=1}^{N_{i_k}}\left(\frac{1}{\tau_{i_k,n}^{\mathrm{i+n}}} - \frac{1}{\tau_{i_k,n}^{\mathrm{s+i+n}}}\right) \\ &= \underbrace{\sum_{n=1}^{N_{i_k}-\widetilde{d}_{i_k}}\left(\frac{1}{\tau_{i_k,n}^{\mathrm{i+n}}} - \frac{1}{\tau_{i_k,n}^{\mathrm{i+n}}}\right)}_{\text{Interference dimensions}} + \underbrace{\sum_{n=N_{i_k}-\widetilde{d}_{i_k}+1}^{N_{i_k}}\left(\frac{1}{\sigma_{i_k}^2} - \frac{1}{\tau_{i_k,n}^{\mathrm{s}}+\sigma_{i_k}^2}\right)}_{\text{Interference-free dimensions}} \\ &= \sum_{n=N_{i_k}-\widetilde{d}_{i_k}+1}^{N_{i_k}} \frac{\tau_{i_k,n}^{\mathrm{s}}}{\sigma_{i_k}^2\left(\tau_{i_k,n}^{\mathrm{s}}+\sigma_{i_k}^2\right)} \to \sum_{n=N_{i_k}-\widetilde{d}_{i_k}+1}^{N_{i_k}} \frac{1}{\sigma_{i_k}^2} = \frac{\widetilde{d}_{i_k}}{\sigma_{i_k}^2} \end{aligned}$$

as the $\{\tau_{i_k,n}^{\mathrm{s}}\}$ grow large w.r.t. $\sigma_{i_k}^2$. □

4.B Proof of Theorem 4.2

Proof. We first denote the objective function as $g_0\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right)$ and the constraint function as $g_1\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right)$ such that

$$g_0\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) = \mathrm{Tr}\left(\left(\mathbf{I} - \mathbf{A}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k} - \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{A}_{i_k} + \mathbf{A}_{i_k}^{\mathsf{H}}\widehat{\mathbf{\Phi}}_{i_k}\mathbf{A}_{i_k}\right)\mathbf{W}_{i_k}\right) - \log_e\det\left(\mathbf{W}_{i_k}\right), \tag{4.56}$$

$$g_1\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) = \left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 - \frac{d_{i_k}}{\sigma_{i_k}^2} = \mathrm{Tr}\left(\mathbf{A}_{i_k}\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\right) - \frac{d_{i_k}}{\sigma_{i_k}^2}. \tag{4.57}$$

For future convenience, we also note that the complex partial gradients [HG07] of the functions are

$$\nabla_{\mathbf{A}_{i_k}^*} g_0\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) = \widehat{\mathbf{\Phi}}_{i_k}\mathbf{A}_{i_k}\mathbf{W}_{i_k} - \widehat{\mathbf{F}}_{i_k}\mathbf{W}_{i_k}, \tag{4.58}$$

$$\nabla_{\mathbf{W}_{i_k}^*} g_0\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) = \mathbf{I} - \mathbf{A}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k} - \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{A}_{i_k} + \mathbf{A}_{i_k}^{\mathsf{H}}\widehat{\mathbf{\Phi}}_{i_k}\mathbf{A}_{i_k} - \mathbf{W}_{i_k}^{-1}, \tag{4.59}$$

$$\nabla_{\mathbf{A}_{i_k}^*} g_1\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) = \mathbf{A}_{i_k}\mathbf{W}_{i_k}, \tag{4.60}$$

$$\nabla_{\mathbf{W}_{i_k}^*} g_1\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) = \mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{A}_{i_k}. \tag{4.61}$$

The full gradients $\nabla\left(\cdot\right) = \nabla_{\mathbf{A}_{i_k}^*, \mathbf{W}_{i_k}^*}\left(\cdot\right)$ are the partial gradients stacked row-wise.

Any regular stationary point $\left(\mathbf{A}_{i_k}^\star, \mathbf{W}_{i_k}^\star\right)$ of the optimization problem in (4.53) must satisfy the KKT conditions [Ber06, Ch. 3.3.1],

$$\nabla g_0\left(\mathbf{A}_{i_k}^\star, \mathbf{W}_{i_k}^\star\right) + \nu_{i_k}^\star \nabla g_1\left(\mathbf{A}_{i_k}^\star, \mathbf{W}_{i_k}^\star\right) = 0, \tag{4.62}$$

$$\mathbf{W}_{i_k} \succ \mathbf{0}, \tag{4.63}$$

$$g_1\left(\mathbf{A}_{i_k}^\star, \mathbf{W}_{i_k}^\star\right) \leq 0, \tag{4.64}$$

$$\nu_{i_k}^\star \geq 0, \tag{4.65}$$

$$\nu_{i_k}^\star g_1\left(\mathbf{A}_{i_k}^\star, \mathbf{W}_{i_k}^\star\right) = 0. \tag{4.66}$$

For the single-constraint problem at hand, a feasible point $\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right)$ is said to be *regular* if [Ber06, Ch. 3.3.1],

$$g_1\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) < 0, \quad \text{or} \quad \begin{cases} g_1\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) & = 0 \\ \nabla g_1\left(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\right) & \neq \mathbf{0} \end{cases}. \tag{4.67}$$

To find a potentially irregular point $\left(\mathbf{A}_{i_k}^0, \mathbf{W}_{i_k}^0\right)$, we note that since

$$\nabla g_1\left(\mathbf{A}_{i_k}^0, \mathbf{W}_{i_k}^0\right) = \begin{pmatrix} \mathbf{A}_{i_k}^0\mathbf{W}_{i_k}^0 \\ \mathbf{A}_{i_k}^{0,\mathsf{H}}\mathbf{A}_{i_k}^0 \end{pmatrix}, \tag{4.68}$$

and $\mathbf{W}_{i_k}^0 \succ \mathbf{0}$, we must have $\mathbf{A}_{i_k}^0\mathbf{W}_{i_k}^0 = \mathbf{0}$ in order to have $\nabla g_1\left(\mathbf{A}_{i_k}^0, \mathbf{W}_{i_k}^0\right) = \mathbf{0}$. From the rank-nullity theorem [HJ85], we know that

$$\mathrm{nullity}\left(\mathbf{A}_{i_k}^0\right) = d_{i_k} - \mathrm{rank}\left(\mathbf{A}_{i_k}^0\right), \tag{4.69}$$

and since $\mathbf{A}_{i_k}^0 \mathbf{W}_{i_k}^0 = \mathbf{0}$ requires nullity $\left(\mathbf{A}_{i_k}^0\right) \geq d_{i_k}$, we then must have rank $\left(\mathbf{A}_{i_k}^0\right) = 0$ for our potentially irregular point. The only point satisfying this is $\mathbf{A}_{i_k}^0 = \mathbf{0}$, and since $g_1\left(\mathbf{0}, \mathbf{W}_{i_k}\right) = -d_{i_k}/\sigma_{i_k}^2 < 0$, this is a regular point according to (4.67). Concluding, all feasible points are regular for this problem, and therefore the global minimizer is among the points described by the KKT conditions in (4.62)–(4.66).

We now venture to solve the KKT conditions. From (4.62), together with the fact that $\mathbf{W}_{i_k} \succ \mathbf{0}$, we get the expressions

$$\mathbf{A}_{i_k}^\star = \left(\widehat{\mathbf{\Phi}}_{i_k} + \nu_{i_k}^\star \mathbf{I}\right)^{-1} \widehat{\mathbf{F}}_{i_k} \tag{4.70}$$

$$\mathbf{W}_{i_k}^\star = \left(\mathbf{I} - \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}} \left(\widehat{\mathbf{\Phi}}_{i_k} + \nu_{i_k}^\star \mathbf{I}\right)^{-1} \widehat{\mathbf{F}}_{i_k}\right)^{-1} \tag{4.71}$$

$$= \mathbf{I} + \widehat{\mathbf{F}}_{i_k}^{\mathsf{H}} \left(\widehat{\mathbf{\Phi}}_{i_k}^{\mathrm{i+n}} + \nu_{i_k}^\star \mathbf{I}\right)^{-1} \widehat{\mathbf{F}}_{i_k}, \tag{4.72}$$

where the last equality is due to the matrix inversion lemma [HJ85]. It now remains to find the optimal $\nu_{i_k}^\star \geq 0$. If the constraint is satisfied for $\nu_{i_k}^\star = 0$, the problem is solved and the form is identical to the solutions in (2.58) and (2.65).

Otherwise, let $\widehat{\mathbf{\Phi}}_{i_k} = \mathbf{L}_{i_k} \mathbf{\Lambda}_{i_k} \mathbf{L}_{i_k}^{\mathsf{H}}$ and $\widehat{\mathbf{\Phi}}_{i_k}^{\mathrm{i+n}} = \mathbf{L}_{i_k}^{\mathrm{i+n}} \mathbf{\Lambda}_{i_k}^{\mathrm{i+n}} \mathbf{L}_{i_k}^{\mathrm{i+n,H}}$ be eigenvalue decompositions. Then, as can be seen in (4.73), $\left\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2$ is decreasing in ν_{i_k}, and the $\nu_{i_k}^\star$ which satisfies the inequality constraint with equality can be found by bisection. A natural starting point for the lower value in the bisection is $\nu_{i_k}^{\mathrm{lower}} = 0$. Using the same argument as in the proof for Theorem 4.1, we have that

$$\left\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 \leq \frac{d_{i_k}}{\lambda_{\min}\left(\widehat{\mathbf{\Phi}}_{i_k}^{\mathrm{i+n}} + \nu_{i_k} \mathbf{I}\right)}$$

and we can thus enforce

$$\left\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 \Bigg|_{\nu_{i_k} = \nu_{i_k}^{\mathrm{upper}}} \leq \frac{d_{i_k}}{\sigma_{i_k}^2}$$

with $\nu_{i_k}^{\mathrm{upper}} = \sigma_{i_k}^2$. The optimal $\nu_{i_k}^\star$ can now be found using bisection on (4.73), given the bounds. With $\nu_{i_k}^\star$ found, the minimizer can be identified as unique. The fact that $\left(\mathbf{A}_{i_k}^\star, \mathbf{W}_{i_k}^\star\right)$ is indeed a minimizer is clear, since each variable minimizes the objective function when the other variable is kept fixed. □

$$
\begin{aligned}
&\left\|\mathbf{A}_{i_k}\mathbf{W}_{i_k}^{1/2}\right\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(\mathbf{A}_{i_k}\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\right)\\
&= \mathrm{Tr}\left(\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\left(\widehat{\mathbf{\Phi}}_{i_k}+\nu_{i_k}\mathbf{I}\right)^{-2}\widehat{\mathbf{F}}_{i_k}\left(\mathbf{I}+\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\left(\widehat{\mathbf{\Phi}}_{i_k}^{\mathrm{i+n}}+\nu_{i_k}\mathbf{I}\right)^{-1}\widehat{\mathbf{F}}_{i_k}\right)\right) =\\
&\mathrm{Tr}\Bigg(\mathbf{L}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k}\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{L}_{i_k}\left(\mathbf{\Lambda}_{i_k}+\nu_{i_k}\mathbf{I}\right)^{-2}\\
&+\left(\mathbf{\Lambda}_{i_k}^{\mathrm{i+n}}+\nu_{i_k}\mathbf{I}\right)^{-1/2}\mathbf{L}_{i_k}^{\mathrm{i+n,H}}\widehat{\mathbf{F}}_{i_k}\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{L}_{i_k}\left(\mathbf{\Lambda}_{i_k}+\nu_{i_k}\mathbf{I}\right)^{-2}\mathbf{L}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k}\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{L}_{i_k}^{\mathrm{i+n}}\left(\mathbf{\Lambda}_{i_k}^{\mathrm{i+n}}+\nu_{i_k}\mathbf{I}\right)^{-1/2}\Bigg)\\
&=\sum_{n=1}^{N_{i_k}}\frac{\left[\mathbf{L}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k}\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{L}_{i_k}\right]_{nn}}{\left(\left[\mathbf{\Lambda}_{i_k}\right]_{nn}+\nu_{i_k}\right)^2}+\left\|\left(\mathbf{\Lambda}_{i_k}+\nu_{i_k}\mathbf{I}\right)^{-1}\mathbf{L}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k}\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{L}_{i_k}^{\mathrm{i+n}}\left(\mathbf{\Lambda}_{i_k}^{\mathrm{i+n}}+\nu_{i_k}\mathbf{I}\right)^{-1/2}\right\|_{\mathrm{F}}^2\\
&=\sum_{n=1}^{N_{i_k}}\frac{\left[\mathbf{L}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k}\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{L}_{i_k}\right]_{nn}}{\left(\left[\mathbf{\Lambda}_{i_k}\right]_{nn}+\nu_{i_k}\right)^2}\\
&+\sum_{n=1}^{N_{i_k}}\sum_{p=1}^{N_{i_k}}\frac{1}{\left(\left[\mathbf{\Lambda}_{i_k}\right]_{nn}+\nu_{i_k}\right)^2}\frac{1}{\left(\left[\mathbf{\Lambda}_{i_k}^{\mathrm{i+n}}\right]_{pp}+\nu_{i_k}\right)}\left|\left[\mathbf{L}_{i_k}^{\mathsf{H}}\widehat{\mathbf{F}}_{i_k}\widehat{\mathbf{F}}_{i_k}^{\mathsf{H}}\mathbf{L}_{i_k}^{\mathrm{i+n}}\right]_{np}\right|^2
\end{aligned}
\tag{4.73}
$$

Chapter 5

Coordinated Precoding with Hardware-Impaired Transceivers

At this point in the thesis, we have both obtained results for the IA feasibility of certain scenarios, as well as designed practical methods for CSI acquisition and coordinated precoding. In Chapter 3, where space-frequency IA was studied, the only receiver impairment was the thermal noise. The system model was then further extended in Chapter 4, to allow for imperfect CSI as well. In both of these chapters however, a common assumption was that the radio hardware was ideal. In this chapter, we relax this assumption, and investigate coordinated precoding under imperfect hardware. In order to fully focus on this impairment, we instead assume perfect knowledge of CSI at all the participating nodes.

All physical radio transceivers suffer from various impairments, such as phase noise, I/Q imbalance, power amplifier non-linearities, etc [Sch08]. A large amount of previous work has focused on these individual impairment, and proposing corresponding compensation schemes; see e.g. [Sch08] and references therein. Here we aim to find a simple model for the hardware impairments, to be used in precoder optimization. A step in the direction of finding a simple model was [SWB10], where the *aggregate* effect of the residual hardware impairments, after compensation, was studied. For their hardware setup, it was shown that the distortion noises from the residual hardware impairments were Gaussian [SWB10]. This type of residual hardware impairments was shown to fundamentally limit performance in the high-SNR regime for a MIMO point-to-point system [BZBO12]. A pure simulation study was performed in [GGLF08], for a more complicated system with path loss. With a generalized model for the residual impairments compared to the one proposed in [SWB10], the optimal beamforming problem for a MISO multicell network was solved in [BZBO12] and [BJ13, Ch. 4.3].

5.1 System Model

We study a wideband system, where OFDM is used to transform the wideband channel into a set of orthogonal narrowband channels (subcarriers). The subcarriers are orthogonal, and we will thus study them separately[1]. The multiuser interaction in the downlink will be described by the interfering broadcast channel in (2.11) on page 19, but we will augment the model with some transceiver distortion noises coming from the hardware impairments. At a given subcarrier, the received signal at MS i_k will be

$$\mathbf{y}_{i_k} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k}\mathbf{x}_{i_k} + \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{x}_{j_l} + \sum_{j=1}^{I} \mathbf{H}_{i_k j}\mathbf{z}_j^{(\mathrm{BS})} + \mathbf{z}_{i_k}^{(\mathrm{MS})}. \tag{5.1}$$

As before, the received signal contains terms for the desired signal, as well as the received interference. The two last terms in (5.1) are however new compared to the model in (2.11). They represent the additive transmitter distortion noises, $\mathbf{z}_j^{(\mathrm{BS})}$, as well as the additive receiver distortion noise $\mathbf{z}_{i_k}^{(\mathrm{MS})}$. We assume that the system uses compensation techniques [Sch08] for the hardware impairments, and thus $\mathbf{z}_j^{(\mathrm{BS})}$ and $\mathbf{z}_{i_k}^{(\mathrm{MS})}$ are the distortion noises coming from the *residual* hardware impairments. As discussed in Section 2.4.3, the compensation techniques applied are necessarily imperfect, and the hardware impairment can therefore not be completely eliminated.

In the model with ideal hardware in (2.11), the *desired* transmitted signal of BS i is $\mathbf{s}_i = \sum_{k=1}^{K_i} \mathbf{V}_{i_k}\mathbf{x}_{i_k}$. Under the residual hardware impairments however, the *actual* transmitted signal of BS i is $\widetilde{\mathbf{s}}_i = \sum_{k=1}^{K_i} \mathbf{V}_{i_k}\mathbf{x}_{i_k} + \mathbf{z}_i^{(\mathrm{BS})}$.

Not all hardware impairments can be described using the model in (5.1). For example, the common phase error due to phase noise can be seen as a rotation of the perceived channel [Sch08], and should thus appear as a multiplicative error. We however leave the extension of the current work to such models for future research.

5.1.1 Hardware Impairments

With the extended system model in (5.1), the goal is now to solve the corresponding weighted sum rate problem. In order to formulate that problem, we introduce models for the transceiver distortion noises. We will use a model proposed in [BJ13, Ch. 4.3], where the distortion noises are modeled as zero-mean circularly symmetric complex Gaussian random variables. The rationale for this modeling assumption is the fact that there are generally many residual hardware impairments. Their sum, after compensation, will then behave as Gaussian [Sch08,SWB10]. This fact applies to both transmitters and receivers, and has been verified using measurements on a wireless testbed [SWB10].

[1] For notational simplicity, we will not include the subcarrier index in the expressions.

Transmitter Distortions

With the Gaussian assumption of the transmitter distortions, it now remains to find the mean and covariance of the signal. The mean is assumed to be zero, otherwise further compensation could be applied that reduces the power of the distortions.

Modeling the covariance, the distortion noises are assumed to be uncorrelated over the antennas. This is a reasonable assumption for systems where the antennas are served with individual RF chains, if the transmitted signal is independent over the antennas. The point of the precoding is however to introduce correlations between the transmitted signals at different antennas. The effect on the cross-correlations of the corresponding transmitter distortion noises was studied in [MZHH12], for a 3rd order non-linear memoryless system. The analysis showed that the cross-correlation coefficient between the distortion noises scaled as the cross-correlation coefficient between desired transmitted signals, to the 3rd power. The correlation of the distortion noises is thus small, and in the forthcoming modeling we will approximate it as zero, for tractability. Concluding, the model for the transmitter distortion noise at BS i is

$$\mathbf{z}_i^{(\mathrm{BS})} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_i^{(\mathrm{BS})}\right), \quad \mathbf{C}_i^{(\mathrm{BS})} = \mathrm{diag}\left(c_{i,1}^{(\mathrm{BS}),2}, \ldots, c_{i,M_i}^{(\mathrm{BS}),2}\right). \tag{5.2}$$

The transmitter distortion noise $\mathbf{z}_i^{(\mathrm{BS})}$ is further assumed to be independent of the desired transmitted signal $\mathbf{s}_i$. If the compensation schemes are reasonably effective, any distortion noises from the residual hardware impairments should be independent of the desired transmitted signal, and this motivates our independence assumption. The *power* of the $\mathbf{z}_i^{(\mathrm{BS})}$ will however be a function of the *power* of $\mathbf{s}_i$. To allow for a large class of relations between these powers, we model the power of the transmitter distortion noise at the mth antenna branch of BS i as

$$c_{i,m}^{(\mathrm{BS}),2} = \eta_i^2\left(\sqrt{\sum_{k=1}^{K_i}\left\|[\mathbf{V}_{i_k}]_{m,:}\right\|_{\mathrm{F}}^2}\right), \tag{5.3}$$

where $\sum_{k=1}^{K_i}\left\|[\mathbf{V}_{i_k}]_{m,:}\right\|_{\mathrm{F}}^2$ is the power of the desired signal allocated to antenna m. In this model, the transmitter *impairment functions* $\eta_i(\cdot)$ are convex, nonnegative, and nondecreasing functions describing how the magnitude of the desired signal maps to the magnitude of the distortions. These assumptions cover a large class of functions, and crucially, will enable the optimization in Section 5.2.

In the literature, the level of distortion noises in a radio transmitter is typically measured using the *error vector magnitude* (EVM) metric. In essence, this metric describes the relation of the distortion noise power, to the desired signal power. For

the proposed model, the EVM at transmitter antenna m at BS i is

$$\mathrm{EVM}_{i,m}^{(\mathrm{BS})} \triangleq \sqrt{\frac{c_{i,m}^{(\mathrm{BS}),2}}{\sum_{k=1}^{K_i}\left\|[\mathbf{V}_{i_k}]_{m,:}\right\|_{\mathrm{F}}^2}} = \frac{\eta_i\left(\sqrt{\sum_{k=1}^{K_i}\left\|[\mathbf{V}_{i_k}]_{m,:}\right\|_{\mathrm{F}}^2}\right)}{\sqrt{\sum_{k=1}^{K_i}\left\|[\mathbf{V}_{i_k}]_{m,:}\right\|_{\mathrm{F}}^2}}. \tag{5.4}$$

The first equality defines the transmitter EVM as the square root of the distortion noise power relative to the desired signal power. The second equality shows how $\eta_i(\cdot)$ affects the EVM.

Depending on the required spectral efficiency, a typical maximum transmit-EVM range in the 3GPP LTE standard [HT11] is $[0.08, 0.175]$.

Receiver Distortions

The modeling of the receiver distortion noises is performed in a similar manner. Again, we base our model on the one proposed in [BJ13, Ch. 4.3]. That model however only considered single-antenna receivers. Here we extend the model in a simple way to allow for multi-antenna receivers. The receiver distortion noises are assumed to be uncorrelated over the antennas, for the same reason as given for the transmitter distortion noises. The receiver distortion noise at MS i_k can then be modeled as

$$\mathbf{z}_{i_k}^{(\mathrm{MS})} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{i_k}^{(\mathrm{MS})}\right), \quad \mathbf{C}_{i_k}^{(\mathrm{MS})} = \mathrm{diag}\left(c_{i_k,1}^{(\mathrm{MS}),2}, \ldots, c_{i_k,N_{i_k}}^{(\mathrm{MS}),2}\right). \tag{5.5}$$

Assuming reasonably effective compensation schemes for the receiver hardware impairments, the distortion noise $\mathbf{z}_{i_k}^{(\mathrm{MS})}$ is assumed to be independent of the received signal $\sum_{(j,l)}\mathbf{H}_{i_kj}\mathbf{V}_{jl}\mathbf{x}_{jl}$. Again, the *power* of the receiver distortion noise will however depend on the power of the received signal. We model that as

$$c_{i_k,n}^{(\mathrm{MS}),2} = \sigma_{i_k}^2 + \zeta_{i_k}^2\left(\sqrt{\sum_{(j,l)}\left\|[\mathbf{H}_{i_kj}\mathbf{V}_{jl}]_{n,:}\right\|_{\mathrm{F}}^2}\right). \tag{5.6}$$

where $\sum_{(j,l)}\left\|[\mathbf{H}_{i_kj}\mathbf{V}_{jl}]_{n,:}\right\|_{\mathrm{F}}^2$ is the power of the received signal[2] at antenna n. The receiver impairment functions $\zeta_{i_k}(\cdot)$ are convex, nonnegative, and nondecreasing functions describing how the magnitude of the received signal maps to the magnitude of the receiver distortions. The $\sigma_{i_k}^2$ represents the power of the thermal noise, as a part of the total receiver distortion noise.

[2]Note that we neglect the received transmitter distortion noises here. In any well-designed system, the part of the receiver distortion noise power which is directly dependent on the received transmitter distortion noise power should be small. Because of this model simplification, some expressions will simplify in the impending exposition.

As in the transmitter case, we define a receiver EVM. We choose to define it in terms of the receiver distortion noise, excluding the thermal noise, relative to the desired received signal. At receive antenna m at MS i_k, it is

$$\mathrm{EVM}_{i_k,n}^{(\mathrm{MS})} \triangleq \frac{\zeta_{i_k}\left(\sqrt{\sum_{(j,l)}\left\|[\mathbf{H}_{i_kj}\mathbf{V}_{j_l}]_{m,:}\right\|_{\mathrm{F}}^2}\right)}{\sqrt{\sum_{(j,l)}\left\|[\mathbf{H}_{i_kj}\mathbf{V}_{j_l}]_{m,:}\right\|_{\mathrm{F}}^2}} \tag{5.7}$$

Signal Covariances

With the proposed model for the distortion noises, the covariance matrix of the received signal in (5.1) can be formed. In order to distinguish it from the covariance matrix for the interfering broadcast channel (without hardware-impairments) in (2.12), we denote it $\widetilde{\mathbf{\Phi}}_{i_k}$ for MS i_k. It is defined as

$$\begin{aligned}\widetilde{\mathbf{\Phi}}_{i_k} = \mathbb{E}\left(\mathbf{y}_{i_k}\mathbf{y}_{i_k}^{\mathsf{H}}\right) = \underbrace{\mathbf{H}_{i_ki}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_ki}^{\mathsf{H}}}_{\text{desired signal}} + \underbrace{\sum_{(j,l)\neq(i,k)}\mathbf{H}_{i_kj}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_kj}^{\mathsf{H}}}_{\text{inter-cell and intra-cell interference}} \\ + \underbrace{\sum_{j=1}^{I}\mathbf{H}_{i_kj}\mathbf{C}_j^{(\mathrm{BS})}\mathbf{H}_{i_kj}^{\mathsf{H}}}_{\text{impact of transmitter distortions}} + \underbrace{\mathbf{C}_{i_k}^{(\mathrm{MS})}}_{\substack{\text{receiver thermal noise}\\ \text{and distortions}}}.\end{aligned} \tag{5.8}$$

The corresponding interference plus distortions covariance matrix is then

$$\begin{aligned}\widetilde{\mathbf{\Phi}}_{i_k}^{\mathrm{int+dist}} &= \widetilde{\mathbf{\Phi}}_{i_k} - \mathbf{H}_{i_ki}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_ki}^{\mathsf{H}} \\ &= \sum_{(j,l)\neq(i,k)}\mathbf{H}_{i_kj}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_kj}^{\mathsf{H}} + \sum_{j=1}^{I}\mathbf{H}_{i_kj}\mathbf{C}_j^{(\mathrm{BS})}\mathbf{H}_{i_kj}^{\mathsf{H}} + \mathbf{C}_{i_k}^{(\mathrm{MS})}.\end{aligned} \tag{5.9}$$

5.2 Weighted Sum Rate Optimization

Given the model of the transceiver distortion noises, the goal is now to maximize the weighted sum rate. First, we must modify our definition of the user rates to take into account the distortion noises. Assuming that the receivers treat the received distortion noises as additive Gaussian noise in the decoder, the achievable data rate for MS i_k is

$$\widetilde{R}_{i_k} = \log_2\det\left(\mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_ki}^{\mathsf{H}}\left(\widetilde{\mathbf{\Phi}}_{i_k}^{\mathrm{int+dist}}\right)^{-1}\mathbf{H}_{i_ki}\mathbf{V}_{i_k}\right). \tag{5.10}$$

We now want to solve the weighted sum rate problem with hardware-impaired transceivers

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{maximize}} \quad & \sum_{(i,k)} \alpha_{i_k} \log_2 \det \left(\mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}} \right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \right) \\ \text{subject to} \quad & \mathbf{V} \in \widetilde{\mathcal{V}}, \end{aligned} \tag{5.11}$$

where $\widetilde{\mathcal{V}}$ is some convex set describing the constraints on the precoders. This may be different from the sets described in Section 2.3.1, due to the hardware impairments. For example, assuming that the transmitted distortion noises count towards the total power budget of a BS, the per-BS sum power constraint set for BS i would be

$$\widetilde{\mathcal{V}}_i = \left\{ \left(\mathbf{V}_{i_1}, \ldots, \mathbf{V}_{i_{K_i}} \right) \in \mathbb{C}^{M_i \times \Pi_{k=1}^{K_i} d_{i_k}} : \operatorname{Tr}\left(\mathbf{C}_i^{(\text{BS})} \right) + \sum_{k=1}^{K_i} \operatorname{Tr}\left(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \right) \leq P_i \right\}. \tag{5.12}$$

The full constraint set for (5.11) is then described by the Cartesian product of the per-BS constraint sets:

$$\widetilde{\mathcal{V}} = \widetilde{\mathcal{V}}_1 \times \widetilde{\mathcal{V}}_2 \times \cdots \times \widetilde{\mathcal{V}}_I. \tag{5.13}$$

5.2.1 Weighted MMSE Minimization

The weighted sum rate problem with hardware-impaired transceivers in (5.11) is non-convex since (5.10) is non-convex in $\{\mathbf{V}_{i_k}\}$. We will therefore apply the WMMSE approach [SRLH11] as described in Section 2.3.3. Assuming linear receive filters $\{\mathbf{A}_{i_k}\}$ at all MSs, the MSE matrix for MS i_k is

$$\begin{aligned} \widetilde{\mathbf{E}}_{i_k} &= \mathbb{E}\left(\left(\mathbf{x}_{i_k} - \hat{\mathbf{x}}_{i_k} \right) \left(\mathbf{x}_{i_k} - \hat{\mathbf{x}}_{i_k} \right)^{\mathsf{H}} \right) = \mathbb{E}\left(\left(\mathbf{x}_{i_k} - \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{y}_{i_k} \right) \left(\mathbf{x}_{i_k} - \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{y}_{i_k} \right)^{\mathsf{H}} \right) \\ &= \mathbf{I} - \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} - \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{A}_{i_k} + \mathbf{A}_{i_k}^{\mathsf{H}} \widetilde{\mathbf{\Phi}}_{i_k} \mathbf{A}_{i_k}. \end{aligned} \tag{5.14}$$

Notice the similarity to the MSE matrix in (2.57) on page 35 for the system without hardware impairments. The impact of the distortion noises appear inside $\widetilde{\mathbf{\Phi}}_{i_k}$, and therefore the WMMSE approach can be applied directly. Next, we seek the MMSE receiver

$$\mathbf{A}_{i_k}^{\text{MMSE}} = \underset{\mathbf{A}_{i_k}}{\arg\min} \operatorname{Tr}\left(\widetilde{\mathbf{E}}_{i_k} \right) = \widetilde{\mathbf{\Phi}}_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \tag{5.15}$$

and note that

$$\begin{aligned} \widetilde{\mathbf{E}}_{i_k}^{\text{MMSE}} &= \widetilde{\mathbf{E}}_{i_k}\left(\mathbf{A}_{i_k}^{\text{MMSE}} \right) = \mathbf{I} - \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \widetilde{\mathbf{\Phi}}_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \\ &= \left(\mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}} \right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \right)^{-1}. \end{aligned} \tag{5.16}$$

The last equality in (5.16) is due to the matrix inversion lemma. By applying the WMMSE approach, we arrive at the weighted MMSE problem for hardware-impaired transceivers,

$$\begin{aligned} \underset{\substack{\{\mathbf{V}_{i_k}\},\{\mathbf{A}_{i_k}\} \\ \{\mathbf{W}_{i_k} \succ \mathbf{0}\}}}{\text{minimize}} \quad & \log_2(e) \sum_{(i,k)} \alpha_{i_k} \left(\operatorname{Tr}\left(\mathbf{W}_{i_k} \widetilde{\mathbf{E}}_{i_k} \right) - \log_e \det\left(\mathbf{W}_{i_k} \right) - d_{i_k} \right) \\ \text{subject to} \quad & \mathbf{V} \in \widetilde{\mathcal{V}}. \end{aligned} \tag{5.17}$$

Comparing to the WMMSE problem for unimpaired transceivers in (2.62) on page 36, it can be seen that the two problems have the same structure, but different MSE matrices. In a similar vein as the WMMSE algorithm without hardware impairments, we can thus try to find a local optimum to (5.17) using alternating minimization.

5.2.2 Optimality Conditions

Before applying alternating minimization to (5.17), we derive the first-order necessary optimality conditions [Ber06] for two of the blocks of variables. These are found by setting the partial complex gradients [HG07] of the objective function to zero.

Since $\mathbf{W}_{i_k} \succ \mathbf{0}$, the first-order optimality conditions for the receive filters are

$$\widetilde{\mathbf{\Phi}}_{i_k} \mathbf{A}_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k}, \quad \forall\, i \in \{1, \ldots, I\},\, k \in \{1, \ldots, K_i\}. \tag{5.18}$$

The first-order optimality conditions for the MSE weights are

$$\left(\mathbf{W}_{i_k} \right)^{-1} = \widetilde{\mathbf{E}}_{i_k}, \quad \forall\, i \in \{1, \ldots, I\},\, k \in \{1, \ldots, K_i\}. \tag{5.19}$$

We will now show that (5.17) has the same global solutions as (5.11). This follows analogously to the problem considered in [SRLH11], but we show the steps here for completeness.

Substituting the necessary conditions in (5.19) into (5.17) and changing the base of the logarithm, the resulting equivalent optimization problem is

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\},\{\mathbf{A}_{i_k}\}}{\text{minimize}} \quad & \sum_{(i,k)} \alpha_{i_k} \log_2 \det\left(\widetilde{\mathbf{E}}_{i_k}^{-1} \right) \\ \text{subject to} \quad & \mathbf{V} \in \widetilde{\mathcal{V}}. \end{aligned} \tag{5.20}$$

Then further substituting (5.18) into (5.20), using (5.16), the resulting equivalent optimization problem is

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{maximize}} \quad & \sum_{(i,k)} \alpha_{i_k} \log_2 \det\left(\mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}} \right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \right) \\ \text{subject to} \quad & \mathbf{V} \in \widetilde{\mathcal{V}}. \end{aligned} \tag{5.21}$$

Clearly, (5.21) is identical to (5.11), and thus (5.17) and (5.11) have the same globally optimal solutions.

Furthermore, it can be shown using [SRLH11, Thm. 3] that (5.17) and (5.11) have the same locally optimal solutions as well. This is done by noting that the stationarity condition of (5.17) w.r.t. $\{\mathbf{V}_{i_k}\}$, for optimal $\{\mathbf{A}_{i_k}^\star, \mathbf{W}_{i_k}^\star\}$, is the same as the stationarity condition of (5.11). This fact is key in the convergence of the alternating minimization in the WMMSE algorithm.

5.2.3 Alternating Minimization

Although we have shown that (5.17) has the same global optimal solutions as (5.11), we will only be able to constructively find local optima. In order to do that, we now apply alternating minimization over the blocks of variables $\{\mathbf{A}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$ and $\{\mathbf{V}_{i_k}\}$ to (5.17).

By fixing $\{\mathbf{W}_{i_k}, \mathbf{V}_{i_k}\}$, the remaining unconstrained optimization problem is convex in $\{\mathbf{A}_{i_k}\}$. The first-order necessary condition in (5.18) is thus both necessary and sufficient, and the solution for MS i_k is

$$\mathbf{A}_{i_k}^\star = \widetilde{\mathbf{\Phi}}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{V}_{i_k} = \mathbf{A}_{i_k}^{\mathrm{MMSE}}, \tag{5.22}$$

where the last equality is identified from (5.15). Similarly, fixing $\{\mathbf{A}_{i_k}, \mathbf{V}_{i_k}\}$, the remaining optimization problem is convex in $\{\mathbf{W}_{i_k}\}$. The necessary and sufficient first-order condition in (5.19) then gives the solution for MS i_k as

$$\mathbf{W}_{i_k}^\star = \left(\widetilde{\mathbf{E}}_{i_k}\right)^{-1} = \mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\left(\widetilde{\mathbf{\Phi}}_{i_k}^{\mathrm{int+dist}}\right)^{-1}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}, \tag{5.23}$$

where the last equality is from substituting (5.22) and (5.16). Since $\mathbf{W}_{i_k}^\star \succ \mathbf{0}$, this is indeed the solution.

Finally, we fix $\{\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\}$ and optimize over $\{\mathbf{V}_{i_k}\}$. By dropping constant terms, and rearranging the remaining terms using properties of the trace, the following problem should be solved:

$$\begin{aligned}\underset{\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{i=1}^{I}\Bigg[\mathrm{Tr}\left(\mathbf{\Gamma}_i\mathbf{C}_i^{(\mathrm{BS})}\right) + \sum_{k=1}^{K_i}\Big[\mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{\Gamma}_i\mathbf{V}_{i_k}\right) \\ & - 2\alpha_{i_k}\mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}\right)\right) + \alpha_{i_k}\mathrm{Tr}\left(\mathbf{A}_{i_k}\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{C}_{i_k}^{(\mathrm{MS})}\right)\Big]\Bigg] \\ \text{subject to} \quad & \mathbf{V} \in \widetilde{\mathcal{V}}.\end{aligned} \tag{5.24}$$

Recall that $\mathbf{\Gamma}_i = \sum_{j,l}\alpha_{j_l}\mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{A}_{j_l}\mathbf{W}_{j_l}\mathbf{A}_{j_l}^{\mathsf{H}}\mathbf{H}_{j_l i}$ is the signal plus interference covariance matrix for a virtual uplink[3]. Compared to (2.67) on page 37, there are two

[3]In this chapter, we do not assume reciprocal channels, and $\mathbf{\Gamma}_i$ may therefore be a quantity which is not related to the true uplink.

Algorithm 5.1 WMMSE Algorithm for Hardware-Impaired Transceivers

1: **repeat**

At MS i_k:

2: Find MSE weights: $\mathbf{W}_{i_k} = \mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}} \right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$

3: Find MMSE receive filters: $\mathbf{A}_{i_k} = \widetilde{\mathbf{\Phi}}_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$

At Central BS Unit:

4: Find precoders as solution to:

$$\begin{aligned} \underset{\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{i=1}^{I} \Bigg[\operatorname{Tr}\left(\mathbf{\Gamma}_i \mathbf{C}_i^{(\text{BS})} \right) + \sum_{k=1}^{K_i} \Big[\operatorname{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{\Gamma}_i \mathbf{V}_{i_k} \right) \\ & - 2\alpha_{i_k} \operatorname{Re}\left(\operatorname{Tr}\left(\mathbf{W}_{i_k} \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \right) \right) + \alpha_{i_k} \operatorname{Tr}\left(\mathbf{A}_{i_k} \mathbf{W}_{i_k} \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{C}_{i_k}^{(\text{MS})} \right) \Big] \Bigg] \\ \text{subject to} \quad & \mathbf{V} \in \widetilde{\mathcal{V}} \end{aligned}$$

5: **until** convergence criterion met, or fixed number of iterations

additional terms. The term $\operatorname{Tr}\left(\mathbf{\Gamma}_i \mathbf{C}_i^{(\text{BS})} \right)$ describes the impact of the transmitter distortion noises generated by BS i on the total performance. Similarly, the term $\alpha_{i_k} \operatorname{Tr}\left(\mathbf{A}_{i_k} \mathbf{W}_{i_k} \mathbf{A}_{i_k}^{\mathsf{H}} \mathbf{C}_{i_k}^{(\text{MS})} \right)$ describes the impact of the receiver distortion noises generated by MS i_k on the total performance.

Note that $\mathbf{C}_i^{(\text{BS})}$ and $\mathbf{C}_{i_k}^{(\text{MS})}$ are functions of $\{\mathbf{V}_{i_k}\}$, and since $\eta_i^2(\cdot)$ and $\zeta_{i_k}^2(\cdot)$ are convex, the problem in (5.24) is convex in $\{\mathbf{V}_{i_k}\}$. For the general case, it can then be solved efficiently using general interior-point methods [BV04, Ch. 11].

The alternating minimization procedure now consists of iteratively applying (5.22), (5.23), and solving (5.24). The resulting algorithm is presented in Algorithm 5.1. The iterations continue until convergence, or until a fixed number of steps is reached.

Theorem 5.1. *The alternating minimization of* (5.17) *monotonically converges. Every limit point of the alternating minimization iterates is a stationary point of* (5.11).

Proof. In each substep, solving for either $\{\mathbf{A}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$, or $\{\mathbf{V}_{i_k}\}$, the objective value of (5.17) can never increase. Since the objective function of (5.17) can be lower-bounded, the objective value monotonically converges.

It now remains to show that the alternating minimization iterates reach a stationary point of the problem. If any of the $c_{i,m}^{(\text{BS}),2}$ or $c_{i_k,n}^{(\text{MS}),2}$ are non-differentiable w.r.t. $\{\mathbf{V}_{i_k}\}$, introduce auxiliary optimization variables $d_{i,m}^{(\text{BS})}$ and $d_{i_k,n}^{(\text{MS})}$ to (5.17).

Replace $c_{i,m}^{(\mathrm{BS}),2} \to d_{i,m}^{(\mathrm{BS})}$, $c_{i_k,n}^{(\mathrm{MS}),2} \to d_{i_k,n}^{(\mathrm{MS})}$ and introduce inequality constraints

$$c_{i,m}^{(\mathrm{BS}),2} = \qquad \eta_i^2 \left(\sqrt{\sum_k \left\| [\mathbf{V}_{i_k}]_{m,:} \right\|_{\mathrm{F}}^2} \right) \leq d_{i,m}^{(\mathrm{BS})}, \quad \forall\, i, m$$

$$c_{i_k,n}^{(\mathrm{MS}),2} = \sigma_{i_k}^2 + \zeta_{i_k}^2 \left(\sqrt{\sum_{(j,l)} \left\| [\mathbf{H}_{i_k j} \mathbf{V}_{j_l}]_{n,:} \right\|_{\mathrm{F}}^2} \right) \leq d_{i_k,n}^{(\mathrm{MS})}, \ \forall\, i_k, n$$

to (5.17), in order to get the squared impairment functions on epigraph form. For this equivalent problem, the objective function is continuously differentiable and the extended feasible set is convex. Then, since the subproblem for $\{\mathbf{A}_{i_k}\}$ is strictly convex, [GS00, Prop. 5] gives that every limit point of the alternating minimization iterates is a stationary point of (5.17). That this is also a stationary point of (5.11) follows directly from the proof of Theorem 3 in [SRLH11]. □

5.3 Constant-EVM Transceivers

One interesting special case is that of *constant-EVM* transceivers. For these, the EVMs are

$$\mathrm{EVM}_{i,m}^{(\mathrm{BS})} = \kappa_i^{(\mathrm{BS})}, \quad \forall\, i, m \tag{5.25}$$

$$\mathrm{EVM}_{i_k,n}^{(\mathrm{MS})} = \kappa_{i_k}^{(\mathrm{MS})}, \quad \forall\, i_k, n. \tag{5.26}$$

which for our model with impairment functions corresponds to

$$\eta_i(x) = \kappa_i^{(\mathrm{BS})} x, \quad \forall\, i, \tag{5.27}$$

$$\zeta_{i_k}(x) = \kappa_{i_k}^{(\mathrm{MS})} x, \quad \forall\, i_k. \tag{5.28}$$

With these impairment functions, the distortion noise covariance matrices are

$$\mathbf{C}_i^{(\mathrm{BS})} = \left(\kappa_i^{(\mathrm{BS})} \right)^2 \sum_{k=1}^{K_i} \mathrm{Diag}\left(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \right), \tag{5.29}$$

$$\mathbf{C}_{i_k}^{(\mathrm{MS})} = \sigma_{i_k}^2 \mathbf{I} + \left(\kappa_{i_k}^{(\mathrm{MS})} \right)^2 \sum_{(j,l)} \mathrm{Diag}\left(\mathbf{H}_{i_k j} \mathbf{V}_{j_l} \mathbf{V}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}^{\mathsf{H}} \right). \tag{5.30}$$

Note that $\mathrm{Diag}\,(\cdot)$ is the operator which retains the diagonal elements, and lets the non-diagonal elements be zero.

Two features of this special case is that the proposed WMMSE algorithm will become distributed over the BSs, and that we can propose a MaxSINDR method for resource allocation.

5.3.1 Distributed WMMSE Algorithm

Algorithm 5.1 on page 109 is naturally distributed over the MSs, but the BS side optimization problem (5.24) must be solved centralized in general. Only if the term

$$\sum_{(i,k)} \alpha_{i_k} \mathrm{Tr}\left(\mathbf{A}_{i_k}\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{C}_{i_k}^{(\mathrm{MS})}\right) \tag{5.31}$$

decomposes, the resulting problem can be solved in parallel over the BSs. This is indeed the case for the constant-EVM transceivers.

It can easily be shown that $\mathrm{Tr}\left(\mathbf{F}\,\mathrm{Diag}\left(\mathbf{G}\right)\right) = \mathrm{Tr}\left(\mathrm{Diag}\left(\mathbf{F}\right)\,\mathbf{G}\right)$, for arbitrary square matrices $\mathbf{F}$ and $\mathbf{G}$. Using this fact, together with the covariance matrix in (5.30), we can rewrite (5.31) as

$$\sum_{(i,k)} \alpha_{i_k} \mathrm{Tr}\left(\mathbf{A}_{i_k}\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{C}_{i_k}^{(\mathrm{MS})}\right) = \sum_{(i,k)} \left(\mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\bar{\mathbf{\Gamma}}_i\mathbf{V}_{i_k}\right) + \alpha_{i_k}\sigma_{i_k}^2\,\mathrm{Tr}\left(\mathbf{A}_{i_k}\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\right)\right), \tag{5.32}$$

where

$$\bar{\mathbf{\Gamma}}_i = \sum_{(j,l)} \alpha_{j_l}\left(\kappa_{j_l}^{(\mathrm{MS})}\right)^2 \mathbf{H}_{j_l i}^{\mathsf{H}}\mathrm{Diag}\left(\mathbf{A}_{j_l}\mathbf{W}_{j_l}\mathbf{A}_{j_l}^{\mathsf{H}}\right)\mathbf{H}_{j_l i}. \tag{5.33}$$

Using the same trick, we can show that

$$\mathrm{Tr}\left(\mathbf{\Gamma}_i\mathbf{C}_i^{(\mathrm{BS})}\right) = \left(\kappa_i^{(\mathrm{BS})}\right)^2 \sum_{k=1}^{K_i} \mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\mathrm{Diag}\left(\mathbf{\Gamma}_i\right)\mathbf{V}_{i_k}\right). \tag{5.34}$$

Substituting (5.32) and (5.34) into (5.24), an equivalent problem is

$$\begin{aligned}\underset{\{\mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{(i,k)} \Bigg[\mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\left(\mathbf{\Gamma}_i + \bar{\mathbf{\Gamma}}_i + \left(\kappa_i^{(\mathrm{BS})}\right)^2 \mathrm{Diag}\left(\mathbf{\Gamma}_i\right)\right)\mathbf{V}_{i_k}\right) \\ & - 2\alpha_{i_k}\mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{A}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}\right)\right)\Bigg] \\ \text{subject to} \quad & \mathbf{V} \in \widetilde{\mathcal{V}}.\end{aligned} \tag{5.35}$$

Under the per-BS sum power constraint defined in (5.12) on page 106, the problem in (5.35) decomposes over BSs, and the optimal precoder for MS i_k is

$$\mathbf{V}_{i_k}^{\star} = \alpha_{i_k}\left(\mathbf{\Gamma}_i + \bar{\mathbf{\Gamma}}_i + \left(\kappa_i^{(\mathrm{BS})}\right)^2 \mathrm{Diag}\left(\mathbf{\Gamma}_i\right) + \widetilde{\mu}_i^{\star}\mathbf{I}\right)^{-1} \mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{A}_{i_k}\mathbf{W}_{i_k}. \tag{5.36}$$

If the optimal Lagrange multiplier for BS i is $\widetilde{\mu}_i^{\star} = 0$, the solution is given. Otherwise, $\widetilde{\mu}_i^{\star} > 0$ can be found by bisection such that

$$\left(1 + \left(\kappa_i^{(\mathrm{BS})}\right)^2\right) \sum_{k=1}^{K_i} \left\|\mathbf{V}_{i_k}^{\star}\right\|_{\mathrm{F}}^2 = P_i. \tag{5.37}$$

Algorithm 5.2 WMMSE Algorithm for Constant-EVM Transceivers and Per-BS Sum Power Constraint

1: **repeat**

At MS i_k:

2: Find MSE weights: $\mathbf{W}_{i_k} = \mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \left(\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}} \right)^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$

3: Find MMSE receive filters: $\mathbf{A}_{i_k} = \widetilde{\mathbf{\Phi}}_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$

At BS i:

4: Find $\widetilde{\mu}_i$ which satisfies $\left(1 + \left(\kappa_i^{(\text{BS})} \right)^2 \right) \sum_{k=1}^{K_i} \text{Tr} \left(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \right) \leq P_i$

5: Find precoders:

$$\mathbf{V}_{i_k} = \alpha_{i_k} \left(\mathbf{\Gamma}_i + \bar{\mathbf{\Gamma}}_i + \left(\kappa_i^{(\text{BS})} \right)^2 \text{Diag} \left(\mathbf{\Gamma}_i \right) + \widetilde{\mu}_i \mathbf{I} \right)^{-1} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}, \quad k = 1, \ldots, K_i \tag{5.38}$$

6: **until** convergence criterion met, or fixed number of iterations

The algorithm for the constant-EVM transceivers and per-BS sum power constraint is described in Algorithm 5.2.

5.3.2 Distributed MaxSINDR Algorithm

For the case of constant-EVM transceivers, with the covariances in (5.29) and (5.30), a MaxSINDR algorithm can be devised. This is done by modifying the original MaxSINR [GCJ11] described in Section 2.3.3, by taking into account the constant-EVM distortions.

First, we define the virtual uplink signal for the MaxSINDR algorithm as

$$\mathbf{\Upsilon}_i = \sum_{(j,l)} \alpha_{j_l} \mathbf{H}_{i_k j}^{\mathsf{H}} \mathbf{A}_{j_l} \mathbf{A}_{j_l}^{\mathsf{H}} \mathbf{H}_{i_k j}, \tag{5.39}$$

and a corresponding 'diagonalized' signal plus interference covariance matrix

$$\bar{\mathbf{\Upsilon}}_i = \sum_{(j,l)} \alpha_{j_l} \left(\kappa_{j_l}^{(\text{MS})} \right)^2 \mathbf{H}_{j_l i}^{\mathsf{H}} \text{Diag} \left(\mathbf{A}_{j_l} \mathbf{A}_{j_l}^{\mathsf{H}} \right) \mathbf{H}_{j_l i}. \tag{5.40}$$

Receive Filter Optimization

At the MSs, the receive filters $\{\mathbf{A}_{i_k}\}$ are found on a per-stream basis, such that the instantaneous per-stream signal-to-interference-distortions-and-noise ratio (SINDR)

is maximized. The columns of the receive filters are then given by

$$\begin{aligned}\mathbf{a}_{i_k,n}^{\star} &= \underset{\mathbf{a}_{i_k,n}^{\mathsf{H}}\mathbf{a}_{i_k,n}=1}{\arg\max} \frac{\mathbf{a}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}\mathbf{v}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{a}_{i_k,n}}{\mathbf{a}_{i_k,n}^{\mathsf{H}}\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}}\mathbf{a}_{i_k,n}} \\ &= \frac{\left(\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}}\right)^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}}{\left\|\left(\widetilde{\mathbf{\Phi}}_{i_k}^{\text{int+dist}}\right)^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}\right\|_2} \\ &= \frac{\widetilde{\mathbf{\Phi}}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}}{\left\|\widetilde{\mathbf{\Phi}}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}\right\|_2}, \quad \forall\, i_k,\, n\end{aligned} \tag{5.41}$$

where the last equality is due to the matrix inversion lemma. The receive filters can be found in parallel over the MSs, and even in parallel over the streams for each MS. Note that the solution in (5.41) has the same form as $\mathbf{a}_{i_k,n}^{\star}$ for the original MaxSINR [GCJ11] as described in Section 2.3.3.

Precoder Optimization

At the BSs, the precoders $\{\mathbf{V}_{i_k}\}$ are also found on a per-stream basis. These are however selected to maximize a metric which is related to the ratio between the received desired signal power due to the per-stream beamformer, and the received interference and distortion signal power due to the per-stream precoder. We call this metric a 'quasi-SINDR'. The intuition is that each beamformer $\mathbf{v}_{i_k,n}$ creates a desired signal at MS i_k, but it will also create interference at all other MSs, as well as distortions at all MSs. The per-stream beamformers should thus be selected to balance the positive and detrimental effects. In order to find the per-stream beamformer for stream n to MS i_k, we first note that the weighted received desired signal power can be written as

$$\alpha_{i_k}\mathbf{a}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}\cdot\mathbf{v}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{a}_{i_k,n} = \alpha_{i_k}\mathbf{v}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{a}_{i_k,n}\cdot\mathbf{a}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}. \tag{5.42}$$

This quantity will end up in the numerator of the quasi-SINDR.

We now study the quantities that will end up in the denominator of the quasi-SINDR. The total weighted interference power due to $\mathbf{v}_{i_k,n}$ is

$$\begin{aligned}&\sum_{(j,l)}\alpha_{j_l}\operatorname{Tr}\left(\mathbf{A}_{j_l}^{\mathsf{H}}\mathbf{H}_{j_l i}\mathbf{v}_{i_k,n}\mathbf{v}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{A}_{j_l}\right) - \alpha_{i_k}\left|\mathbf{a}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}\right|^2 \\ &\quad = \mathbf{v}_{i_k,n}^{\mathsf{H}}\left(\mathbf{\Upsilon}_i - \alpha_{i_k}\mathbf{H}_{i_k i}\mathbf{a}_{i_k,n}\mathbf{a}_{i_k,n}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\right)\mathbf{v}_{i_k,n}.\end{aligned} \tag{5.43}$$

For the constant-EVM transmitters, the transmitter distortion power is linear in the power of the per-stream beamformers (cf. (5.29)). Therefore, the total weighted (received) transmitter distortion power due to $\mathbf{v}_{i_k,n}$ is

$$\left(\kappa_i^{(\text{BS})}\right)^2\operatorname{Tr}\left(\mathbf{\Upsilon}_i\operatorname{Diag}\left(\mathbf{v}_{i_k,n}\mathbf{v}_{i_k,n}^{\mathsf{H}}\right)\right) = \left(\kappa_i^{(\text{BS})}\right)^2\mathbf{v}_{i_k,n}^{\mathsf{H}}\operatorname{Diag}\left(\mathbf{\Upsilon}_i\right)\mathbf{v}_{i_k,n}. \tag{5.44}$$

Finally, using a similar relation as in (5.32) on page 111, the total weighted receiver distortion power due to $\mathbf{v}_{i_k,n}$ is

$$\mathbf{v}_{i_k,n}^{\mathsf{H}} \bar{\boldsymbol{\Upsilon}}_i \mathbf{v}_{i_k,n}. \tag{5.45}$$

Putting together the desired signal power with the interference and distortion terms in (5.43), (5.44), and (5.45), the quasi-SINDR to optimize is

$$\frac{\alpha_{i_k} \mathbf{v}_{i_k,n}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{a}_{i_k,n} \mathbf{a}_{i_k,n}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{v}_{i_k,n}}{\mathbf{v}_{i_k,n}^{\mathsf{H}} \left(\boldsymbol{\Upsilon}_i - \alpha_{i_k} \mathbf{H}_{i_k i} \mathbf{a}_{i_k,n} \mathbf{a}_{i_k,n}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} + \bar{\boldsymbol{\Upsilon}}_i + \left(\kappa_i^{(\mathrm{BS})} \right)^2 \mathrm{Diag}\left(\boldsymbol{\Upsilon}_i \right) + \frac{\sigma_{i_k}^2 K_i}{P_i} \mathbf{I} \right) \mathbf{v}_{i_k,n}}. \tag{5.46}$$

Notice that a white term relating to the SNR of MS i_k was also added to the denominator. This term is needed at low SNR, since the interference is negligible then.

Concluding, the per-stream beamformers are now given by

$$\begin{aligned}
\mathbf{v}_{i_k,n}^{\star} &= \underset{\mathbf{v}_{i_k,n}^{\mathsf{H}} \mathbf{v}_{i_k,n} = \frac{P_i}{K_i d_{i_k}}}{\arg\max} \ \text{quasi-SINDR}_{i_k,n} \quad \text{in (5.46)} \\
&= \sqrt{\frac{P_i}{K_i d_{i_k}}} \frac{\sqrt{\alpha_{i_k}} \left(\boldsymbol{\Upsilon}_i - \alpha_{i_k} \mathbf{H}_{i_k i} \mathbf{a}_{i_k,n} \mathbf{a}_{i_k,n}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} + \bar{\boldsymbol{\Upsilon}}_i + \left(\kappa_i^{(\mathrm{BS})} \right)^2 \mathrm{Diag}(\boldsymbol{\Upsilon}_i) + \frac{\sigma_{i_k}^2 K_i}{P_i} \mathbf{I} \right)^{-1} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{a}_{i_k,n}}{\left\| \sqrt{\alpha_{i_k}} \left(\boldsymbol{\Upsilon}_i - \alpha_{i_k} \mathbf{H}_{i_k i} \mathbf{a}_{i_k,n} \mathbf{a}_{i_k,n}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} + \bar{\boldsymbol{\Upsilon}}_i + \left(\kappa_i^{(\mathrm{BS})} \right)^2 \mathrm{Diag}(\boldsymbol{\Upsilon}_i) + \frac{\sigma_{i_k}^2 K_i}{P_i} \mathbf{I} \right)^{-1} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{a}_{i_k,n} \right\|_2} \\
&= \sqrt{\frac{P_i}{K_i d_{i_k}}} \frac{\sqrt{\alpha_{i_k}} \left(\boldsymbol{\Upsilon}_i + \bar{\boldsymbol{\Upsilon}}_i + \left(\kappa_i^{(\mathrm{BS})} \right)^2 \mathrm{Diag}(\boldsymbol{\Upsilon}_i) + \frac{\sigma_{i_k}^2 K_i}{P_i} \mathbf{I} \right)^{-1} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{a}_{i_k,n}}{\left\| \sqrt{\alpha_{i_k}} \left(\boldsymbol{\Upsilon}_i + \bar{\boldsymbol{\Upsilon}}_i + \left(\kappa_i^{(\mathrm{BS})} \right)^2 \mathrm{Diag}(\boldsymbol{\Upsilon}_i) + \frac{\sigma_{i_k}^2 K_i}{P_i} \mathbf{I} \right)^{-1} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{a}_{i_k,n} \right\|_2}, \ \forall\, i_k,\, n
\end{aligned} \tag{5.47}$$

where the last equality is due to the matrix inversion lemma. The full algorithm is summarized in Algorithm 5.3 on the next page. Since the algorithm does not optimize a single global objective, it is unclear whether it is guaranteed to converge or not. A pragmatic approach is therefore to perform a predetermined fixed number of iterations before quitting. In the numerical performance evaluation in Section 5.4, it will be shown that the algorithm seems to converge on average.

5.4 Performance Evaluation

We study the performance of the proposed method using numerical simulation. In the simulation study, we let the impairment functions be

$$\eta_i(x) = \kappa_{\mathrm{t}}\, x \left(1 + \left(\frac{x}{\kappa^{(\mathrm{NL})}} \right)^2 \right), \quad \forall\, i, \tag{5.48}$$

$$\zeta_{i_k}(x) = \kappa_{\mathrm{r}}\, x, \qquad \forall\, i_k. \tag{5.49}$$

Algorithm 5.3 MaxSINDR with Per-BS Sum Power Constraints

1: **repeat**

At MS i_k:

2: $\mathbf{a}_{i_k,n} = \dfrac{\widetilde{\mathbf{\Phi}}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}}{\left\|\widetilde{\mathbf{\Phi}}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{v}_{i_k,n}\right\|_2}, \quad n = 1,\ldots,d_{i_k}$

3: $\mathbf{A}_{i_k} = \begin{pmatrix}\mathbf{a}_{i_k,1} & \mathbf{a}_{i_k,2} & \cdots & \mathbf{a}_{i_k,d_{i_k}}\end{pmatrix}$

At BS i:

4: $$\mathbf{b}_{i_k,n} = \frac{\sqrt{\alpha_{i_k}}\left(\mathbf{\Upsilon}_i + \bar{\mathbf{\Upsilon}}_i + \left(\kappa_i^{(\mathrm{BS})}\right)^2 \mathrm{Diag}(\mathbf{\Upsilon}_i) + \frac{\sigma_{i_k}^2 K_i}{P_i}\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{a}_{i_k,n}}{\left\|\sqrt{\alpha_{i_k}}\left(\mathbf{\Upsilon}_i + \bar{\mathbf{\Upsilon}}_i + \left(\kappa_i^{(\mathrm{BS})}\right)^2 \mathrm{Diag}(\mathbf{\Upsilon}_i) + \frac{\sigma_{i_k}^2 K_i}{P_i}\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{a}_{i_k,n}\right\|_2}$$
$$k = 1,\ldots,K_i,\ n = 1,\ldots,d_{i_k}$$

5: $\mathbf{B}_{i_k} = \begin{pmatrix}\mathbf{b}_{i_k,1} & \mathbf{b}_{i_k,2} & \cdots & \mathbf{b}_{i_k,d_{i_k}}\end{pmatrix}, \quad k = 1,\ldots,K_i$

6: $\mathbf{V}_{i_k} = \sqrt{\frac{P_i}{K_i d_{i_k}}}\mathbf{B}_{i_k}, \quad k = 1,\ldots,K_i$

7: **until** fixed number of iterations

The receivers are thus constant-EVM receivers (with constant EVM of κ_{r}), and the transmitters have a 3rd order non-linearity. For low transmit powers, the EVM of the transmitters is κ_{t}. At a transmit power of $\kappa^{(\mathrm{NL})}$, the EVM has doubled. With this choice of impairment functions, the covariance matrices $\mathbf{C}_{i_k}^{(\mathrm{MS})}$ and $\mathbf{C}_i^{(\mathrm{BS})}$ are differentiable w.r.t. $\{\mathbf{V}_{i_k}\}$. The BSs have individual sum power constraints, including the created distortions, as described by (5.12) on page 106. We use the modeling language YALMIP [LÖ4] together with the Gurobi solver [Gur14] to solve (5.24).

We study a scenario with $I = 3$ BSs, each serving $K_c = K_i = 2$ MSs. The BSs have $M = M_i = 4$ antennas each. The MSs have $N = N_{i_k} = 2$ antennas each, and the user priorities are $\alpha_{i_k} = 1$, $\forall\, i_k$. The BSs are located at the corner of an equilateral triangle with an inter-site distance of 500 m, and their antenna boresights aimed towards the center of the triangle. For each Monte Carlo realization of the network, the MSs were dropped with uniform probability in the cells belonging to their serving BS. They were however never closer than 35 m to the BS. The other simulation parameters are described in Table 5.1.

We compare the performance of the proposed Algorithm 5.1 to the case of impairments-ignoring BSs and MSs using Algorithm 2.3 on page 38 with per-BS sum power constraints. We also compare it to the case of having impairments-aware MSs which use the impairments-aware MMSE receiver in (5.15), but impairments-ignoring BSs which use Algorithm 2.3. The case of having aware MSs but ignorant BSs could occur if the MSs estimate their covariances $\widetilde{\mathbf{\Phi}}_{i_k}$ over the air (using e.g. the techniques in Section 4.2), without having a specific model for the impairments. The impact of the distortions is then picked up by the MSs, and that knowledge is implicitly distributed to the BSs in the WMMSE iterations. Effectively, the ignorant BSs let $\mathbf{C}_i^{(\mathrm{BS})} = \mathbf{0}$ and $\mathbf{C}_{i_k}^{(\mathrm{MS})} = \mathbf{0}$ in their optimization of (5.24).

Table 5.1. Simulation parameters

Path loss	$\mathrm{PL}_{\mathrm{dB}} = 15.3 + 37.6 \log_{10}(\text{distance [m]})$
Penetration loss	20 dB
BS antenna gain	$12 \left(\frac{\theta}{35^\circ}\right)^2$ dB
MS antenna gain	0 dB
Small scale fading	i.i.d. $\mathcal{CN}(0, 1)$
Bandwidth	15 kHz
Transmit power	$P = P_i, \ \forall\, i$
Noise power	$\sigma^2 = \sigma_{i_k}^2 = -127\,\mathrm{dBm}, \ \forall\, i_k$

As baselines, we use the proposed MaxSINDR algorithm in Algorithm 5.3, as well as TDMA. MaxSINDR is only aware of the linear impairments, and is thus unaware of the non-linearity in $\eta_i(x)$. For TDMA, we use Algorithm 5.1 to find the impairments-aware precoders. For TDMA with impairments-ignoring BSs and MSs, we use eigenprecoding with water filling as described in Section 2.3.3.

5.4.1 Convergence

First we investigate the convergence behaviour of the proposed algorithms. We generate one user drop, and show sum rate performance as a function of iteration number. The specific user geography for this user drop is shown in Figure 5.1.

We let $d = 2$ for Algorithm 5.1 and $d = 1$ for Algorithm 5.3[4]. The power constraint per BS is $P = 18.2$ dBm and the impairments parameters are $\kappa_{\mathrm{t}} = \kappa_{\mathrm{r}} = \frac{10}{100}$ and $20 \log_{10}(\kappa^{(\mathrm{NL})}) = 15.2$ dBm. The sum rate evolution for these parameters is shown in Figure 5.2. The proposed impairments-aware WMMSE converges within a couple of tens of iterations. Interestingly, the WMMSE algorithm with impairments-aware MSs but impairments-ignoring BSs also seems to converge, but to a lower sum rate performance. The performance of the impairments-ignoring WMMSE algorithm actually slowly deteriorates as the number of iterations grow large. This shows that it is clearly important to take the hardware impairments into account when performing the resource allocation. Finally, the MaxSINDR also seems to converge, but to a sum rate around 25 % lower than the proposed WMMSE algorithm.

5.4.2 Varying Impairment Levels

Next we study sum rate performance when varying the levels of impairments. We fix the transmit power at $P = 18.2\,\mathrm{dBm}$. We generated 100 user drops, and 10 small scale fading realizations per user drop. The iterative methods were run with

[4] Using $d = 2$ for Algorithm 5.3 decreased performance.

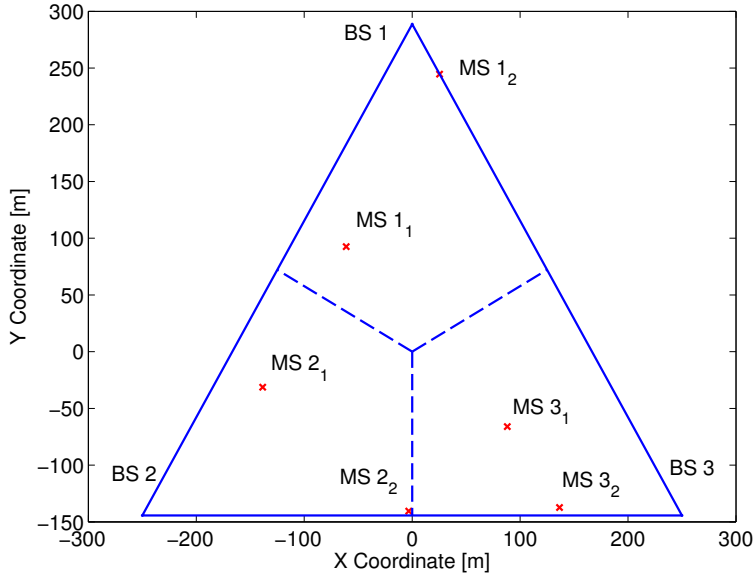


Figure 5.1. User geography for the convergence simulation in Figure 5.2.

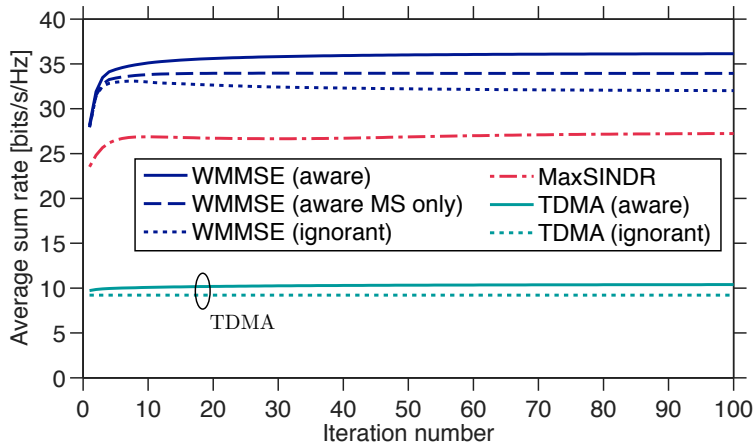


Figure 5.2. Sum rate evolution (one realization) for $P = 18.2$ dBm, $\kappa_{\mathrm{t}} = \kappa_{\mathrm{r}} = \frac{10}{100}$ and $20\log_{10}(\kappa^{(\mathrm{NL})}) = 15.2$ dBm.

a stopping criterion of 10^{-3} relative difference in increased sum rate. The sum rate results, averaged over Monte Carlo realizations, are presented in Figure 5.3. Clearly, the fully impairments-aware Algorithm 5.1 performs the best, but performance drops as the severity of the impairments increases. The same trend holds for the WMMSE algorithm with and without impairments-aware MSs. MaxSINDR performs significantly worse than all other methods.

5.4.3 Varying Transmit Powers

Lastly, we study performance as a function of available transmit power. In order to have a reasonable simulation scenario, we specialize to $\eta(x) = \kappa_{\mathrm{t}} x$ and vary $\kappa_{\mathrm{t}} = \kappa_{\mathrm{r}}$ and the transmit power P. The sum rate results, averaged over Monte Carlo realizations, can be seen in Figure 5.4. The sum rates saturate at high transmit powers, due to the residual hardware impairments. Clearly, the high-SNR scaling of the curves are zero, but the gain for coordinated precoding over TDMA is significant, as predicted by [BZBO12]. Interestingly, in this case the WMMSE algorithm with impairments-ignoring BSs performs almost equally well as Algorithm 5.1. For TDMA, there is barely any difference in taking the hardware impairments into account or not.

5.5 Conclusions

The studies in earlier chapters had assumed ideal hardware, but in this chapter we studied the weighted sum rate optimization problem with hardware-impaired transceivers. Applying the WMMSE approach to the weighted sum rate problem, an alternating minimization technique was proposed. Convergence of the algorithm was shown, and sum rate performance was evaluated using numerical methods. These showed that the high-SNR scaling was zero due to the hardware impairments, which essentially indicates that pure IA is not an interesting concept for wireless networks with hardware-impaired transceivers. However, the relative gain in performance for coordinated precoding over TDMA was still large, so it is highly interesting to apply coordinated precoding techniques based on weighted sum rate optimization, albeit not pure IA techniques.

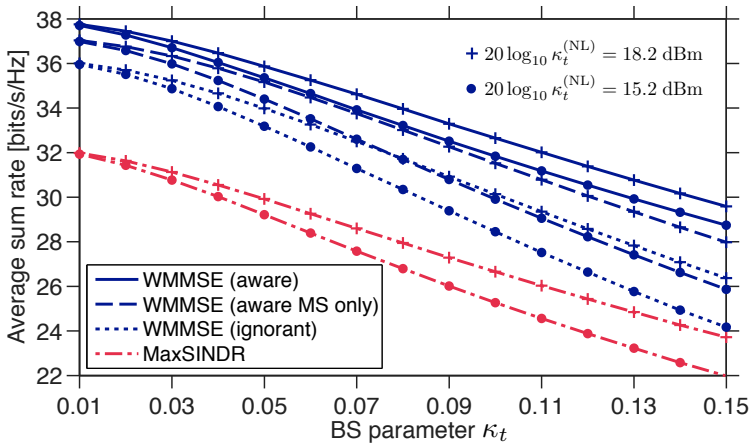


Figure 5.3. Sum rate for $P = 18.2$ dBm when varying impairment parameters. Note the scale of the vertical axis.

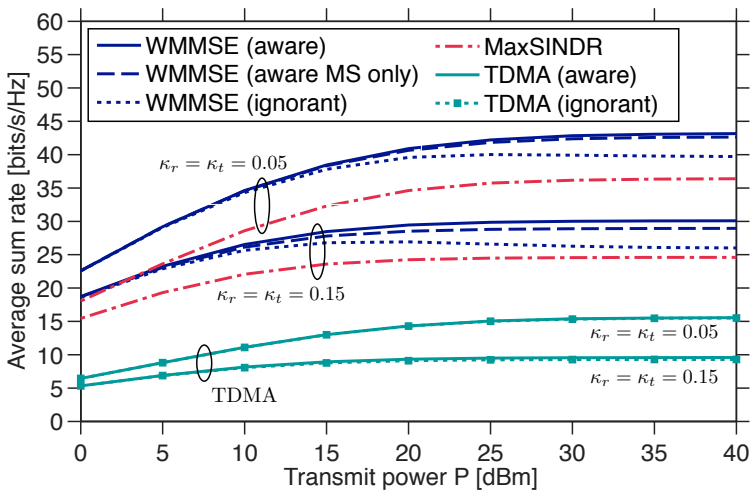


Figure 5.4. Sum rate for $\eta(x) = \kappa_t x$ when varying transmit power and impairment parameters.

Chapter 6

Conclusions and Future Research

6.1 Conclusions

This thesis studied some aspects of coordinated precoding for multicell MIMO networks: IA feasibility, distributed CSI acquisition and robust precoding, and imperfect hardware. It was argued that IA is a useful tool for finding the achievable DoF limits of the network, and in certain cases a useful tool for finding good precoders (in the weighted sum rate sense). In general, an attempt to solve the weighted sum rate optimization problem should be done however. Contrarily to IA, the weighted sum rate problem formulation is relevant for all SNRs and system scenarios. In the high-SNR regime, solving the weighted sum rate problem corresponds to finding the IA solution with the best weighted sum rate.

The first contribution of the thesis was the derivation of a necessary condition for the feasibility of space-frequency IA. Using a bound on the gains over space-only IA, it was shown that the gain increased in the number of subcarriers, and decreased in the number of antennas. Numerical sum rate studies show the existence of a DoF gain, but the performance improvement was largely due to a power gain. The conclusion was the same both for synthetically generated channels, as well as measured channels. Furthermore, the numerical results also showed that coordinated precoding significantly outperformed traditional orthogonalization.

Leaving the theoretical realm, the next focus of the thesis was on practical CSI acquisition. The need for coupling distributed coordinated precoding with distributed CSI acquisition, in order to achieve a distributed joint system, was argued. The CSI requirements of the WMMSE algorithm were studied, and three CSI acquisition methods were proposed. The methods correspond to different tradeoffs between channel estimation, feedback and signaling, backhaul use, and computational complexity. Naïvely coupling the WMMSE algorithm with the proposed CSI acquisition methods is straightforward, but was shown to yield poor performance. Therefore, robustifying measures were proposed, which resulted in a robust and distributed WMMSE algorithm. Numerical studies showed that the proposed sys-

tem had similar performance to the (centralized) state-of-the-art methods, with the added benefit that our system can be implemented in a fully distributed fashion.

The access to, and quality of, CSI is not the only practical aspect that should be taken into account. Any practical RF transceiver will have some hardware impairments, which give rise to distortion noises. The final contribution of the thesis was to study the weighted sum rate problem with these hardware-impaired transceivers. Using an extended WMMSE approach, a semi-distributed resource allocation was proposed for this scenario. The numerical results clearly showed the need for modeling and optimizing over the hardware impairments, and the proposed algorithm was shown to be robust against increasing impairment levels.

Concluding, this thesis has shown the benefits of coordinated precoding for improving the spectral efficiency of future multicell MIMO networks. Various practical aspects were studied, and enabling and robustifying measures were proposed.

6.2 Future Research

We now highlight some items that might be worthy of further study:

- The result in Chapter 3 would be strengthened if it could be expanded to include sufficiency as well. This would possibly require a deeper application of algebraic geometry. It would also be good with a condition for the interfering broadcast channel. In that case, several interfering streams arrive over the same channel matrix, and the statistical dependence between the interfering links must be handled.

- In Chapter 4, a perfectly reciprocal channel was assumed. Practical calibration schemes will not be perfect however, leading to reciprocity errors. By modeling these errors, better estimators could possibly be found. Such reciprocity errors should also be taken into account in the resource allocation.

- In Chapter 5, the resource allocation was performed on a per-subcarrier basis, even though a multicarrier system was studied. With phase noise for instance, the distortion noise at one subcarrier will be affected by the signal power of neighbouring subcarriers. By modeling these types of effects, a joint optimization over all subcarriers could be performed. Further, models involving more than additive distortion noises should be studied. For example, phase noise gives rise to a complex rotation of all subcarriers, an effect that was not studied in Chapter 5.

- In some cases at high SNR, the numerical results in this thesis have shown that the WMMSE algorithms converge slowly. In cases where IA is feasible, the objective function could possibly be regularized in a way to drive the solution towards an IA solution. This would possibly improve convergence speed.

- A basic assumption in this thesis was that the user association is fixed. By introducing the user association as a variable in the optimization problem, the WMMSE approach might be able to give a constructive algorithm for joint precoder and user selection. This would possibly involve relaxing the binary decision variables that would model the user associations.
- Another basic assumption is that the continuous data rate function is a good model for a practical system. In practical systems however, rates must be selected from a finite set determined by the modulation and coding schemes available. A tractable problem formulation can be obtained by formulating a weighted *discrete* sum rate problem, linearizing the continuous rate constraints, and rewriting the objective function using an indicator function for the selection of the discrete rates. By relaxing the indicator function, a constructive method for finding good precoders is obtained.

Bibliography

[BAB12] R. Brandt, H. Asplund, and M. Bengtsson. Interference alignment in frequency – a measurement based performance analysis. In *Proc. Int. Conf. Systems, Signals and Image Process. (IWSSIP'12)*, pages 227–230, 2012.

[BB11] R. Brandt and M. Bengtsson. Wideband MIMO channel diagonalization in the time domain. In *Proc. IEEE Int. Symp. Indoor, Mobile Radio Commun. (PIMRC'11)*, pages 1958–1962, 2011.

[BB14] R. Brandt and M. Bengtsson. Distributed CSI acquisition and coordinated precoding for TDD multicell MIMO systems. *IEEE Transactions on Signal Processing*, 2014. Submitted.

[BBB14] R. Brandt, E. Björnson, and M. Bengtsson. Weighted sum rate optimization for multicell MIMO systems with hardware-impaired transceivers. In *Proc. IEEE Conf. Acoust., Speech, and Signal Process. (ICASSP'14)*, pages 479–483, 2014.

[BCK03] A. Bourdoux, B. Come, and N. Khaled. Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation. In *Proc. Radio Wireless Conf. (RAWCON'03)*, pages 183–186, 2003.

[Ben02] M. Bengtsson. A pragmatic approach to multi-user spatial multiplexing. In *Proc. IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM'02)*, pages 130–134, 2002.

[Ber06] D. Bertsekas. *Nonlinear programming.* Athena Scientific, 2006.

[BG06] M. Biguesh and A. B. Gershman. Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals. *IEEE Transactions on Signal Processing*, 54(3), 2006.

[BJ13] E. Björnson and E. Jorswieck. Optimal resource allocation in coordinated multi-cell systems. *Foundations and Trends in Communications and Information Theory*, 9(2-3):113–381, 2013.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[BZB12] E. Björnson, P. Zetterberg, and M. Bengtsson. Optimal coordinated beamforming in the multicell downlink with transceiver impairments. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'12)*, pages 4775–4780, 2012.

[BZB13] R. Brandt, P. Zetterberg, and M. Bengtsson. Interference alignment over a combination of space and frequency. In *Proc. IEEE Int. Conf. Commun. Workshop: Beyond LTE-A (ICC'13 LTE-B)*, pages 149–153, 2013.

[BZBO12] E. Björnson, P. Zetterberg, M. Bengtsson, and B. Ottersten. Capacity limits and multiplexing gains of MIMO channels with transceiver impairments. *IEEE Communications Letters*, 17(1):91–94, 2012.

[CACC08] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. *IEEE Transactions on Wireless Communications*, 7(12):4792–4799, December 2008.

[Car78] A. Carleial. Interference channels. *IEEE Transactions on Information Theory*, 24(1):60–70, 1978.

[Car88] B. D. Carlson. Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Transactions on Aerospace and Electronic Systems*, 24(4):397–401, July 1988.

[Cis13] Cisco. Cisco visual networking index: Global mobile data traffice forecast update, 2013–2018. White paper, Cisco, 2013.

[CJ08] V. R. Cadambe and S. A. Jafar. Interference alignment and degrees of freedom of the K-user interference channel. *IEEE Transactions on Information Theory*, 54(8):3425–3441, 2008.

[CJ09] V. R. Cadambe and S. A. Jafar. Reflections on interference alignment and the degrees of freedom of the k user interference channel. *IEEE Inform. Theory Soc. Newslett.*, 59:1–32, 2009.

[CJS13] P. Cao, E. A. Jorswieck, and S. Shi. Pareto boundary of the rate region for single-stream MIMO interference channels: Linear transceiver design. *IEEE Transactions on Signal Processing*, 61(20):4907–4922, 2013.

[CLO98] D. Cox, J. Little, and D. O'Shea. *Using algebraic geometry*. Springer, 1998.

[Cou07] L. W. Couch. *Digital and Analog Communication Systems*. Pearson Prentice Hall, 7th edition, 2007.

[Cox73] H. Cox. Resolving power and sensitivity to mismatch of optimum array processors. *J. Acoustical Soc. America*, 54(3):771–785, 1973.

[CT06] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006.

[CZO87] H. Cox, R. M. Zeskind, and M. M. Owen. Robust adaptive beamforming. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(10):1365–1376, 1987.

[DADSC04] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank. Great expectations: The value of spatial diversity in wireless networks. *Proceedings of the IEEE*, 92(2):219–270, 2004.

[DPSB07] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming. *3G Evolution: HSPA and LTE for Mobile Broadband*, chapter 16. Academic Press, 2007.

[DRSP13] H. Du, T. Ratnarajah, M. Sellathurai, and C. B. Papadias. Reweighted nuclear norm approach for interference alignment. *IEEE Transactions on Communications*, 61(9):3754–3765, 2013.

[Eri11] Ericsson. More than 50 billion connected devices. White paper 284 23-3149 Uen, 2011.

[Fos96] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Syst. Tech. J*, 1996.

[GAH11] B. Ghimire, G. Auer, and H. Haas. Busy burst enabled coordinated multipoint network with decentralized control. *IEEE Transactions on Wireless Communications*, 10(10):3310–3320, 2011.

[GBS] Ó. González, C. Beltrán, and I. Santamaría. Feasibility test for linear interference alignment. Accessible at: `http://gtas.unican.es/IAtest`.

[GBS14] Ó. González, C. Beltrán, and I. Santamaría. A feasibility test for linear interference alignment in MIMO channels with constant coefficients. *IEEE Transactions on Information Theory*, 60(3):1840–1856, 2014.

[GCJ08] K. Gomadam, V. R. Cadambe, and S. A. Jafar. Approaching the capacity of wireless networks through distributed interference alignment. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'08)*, pages 1–6, 2008.

[GCJ11] K. Gomadam, V. R. Cadambe, and S. Jafar. A distributed numerical approach to interference alignment and applications to wireless intererence networks. *IEEE Transactions on Information Theory*, 57(6):3309–3322, 2011.

[GGLF08] B. Göransson, S. Grant, E. Larsson, and Z. Feng. Effect of transmitter and receiver impairments on the performance of MIMO in HSDPA. In *Proc. IEEE Workshop Signal Process. Advances Wireless Commun. (SPAWC'08)*, pages 496–500, 2008.

[GHH+10] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu. Multi-cell MIMO cooperative networks: A new look at interference. *IEEE Journal on Selected Areas in Communications*, 28(9):1380–1408, 2010.

[GKH+07] D. Gesbert, M. Kountouris, R. Heath, C.-b. Chae, and T. Salzer. Shifting the MIMO paradigm. *IEEE Signal Processing Magazine*, 24(5):36–46, 2007.

[GS00] L. Grippo and M. Sciandrone. On the convergence of the block nonlinear Gauss-Seidel method under convex constraints. *Operations Research Letters*, 26(3):1–10, 2000.

[GSB13] O. Gonzalez, I. Santamaria, and C. Beltran. Finding the number of feasible solutions for linear interference alignment problems. In *Proc. IEEE Int. Symp. Inform. Theory (ISIT'13)*, pages 384–388, 2013.

[GSK05] M. Guillaud, D. T. M. Slock, and R. Knopp. A practical method for wireless channel reciprocity exploitation through relative calibration. In *Proc. 8th Int. Symp. Signal Process. Applicat. (ISSPA'05)*, pages 403–406, 2005.

[Gui10] M. Guillaud. Receive diversity and ergodic performance of interference alignment on the MIMO gaussian interference channel. In *Ann. Allerton Conf. Commun., Control, Computing (Allerton'10)*, pages 208–214, 2010.

[Gur14] Gurobi. Gurobi optimizer reference manual. 2014.

[HG07] A. Hjørungnes and D. Gesbert. Complex-valued matrix differentiation: Techniques and key results. *IEEE Transactions on Signal Processing*, 55(6):2740–2746, 2007.

[HJ85] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[HT11] H. Holma and A. Toskala. *LTE for UMTS: Evolution to LTE-Advanced*. Wiley, 2nd edition edition, 2011.

[Jaf11] S. A. Jafar. *Interference alignment: a new look at signal dimensions in a communication network*, volume 7 of *Foundations and Trends in Communications and Information Theory*. Now Publishers, 2011.

[JAWV11] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath. Channel estimation and linear precoding in multiuser multiple-antenna TDD systems. *IEEE Transactions on Vehicular Technology*, 60(5):2102–2116, 2011.

[JLD09] E. A. Jorswieck, E. G. Larsson, and D. Danev. Complete characterization of the Pareto boundary for the MISO interference channel. *IEEE Transactions on Signal Processing*, 56(10):5292–5296, 2009.

[Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[KMT98] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *J. Operational Research Soc.*, 49(3):237–252, 1998.

[KTJ13] P. Komulainen, A. Tölli, and M. Juntti. Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems. *IEEE Transactions on Signal Processing*, 61(9):2204–2218, May 2013.

[LÖ4] J. Löfberg. Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proc. CACSD Conf.*, Taipei, Taiwan, 2004.

[LDL11] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo. Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms. *IEEE Transactions on Signal Processing*, 59(3):1142–1157, March 2011.

[LH05] D. Love and R. Heath. Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Transactions on Information Theory*, 51(8):2967–2976, 2005.

[LHA13] A. Lozano, R. Heath, and J. Andrews. Fundamental limits of cooperation. *IEEE Transactions on Information Theory*, 59(9):5213–5226, 2013.

[LKY13] H.-H. Lee, Y.-C. Ko, and H.-C. Yang. On robust weighted sum rate maximization for MIMO interfering broadcast channels with imperfect channel knowledge. *IEEE Communications Letters*, 17(6):1156–1159, 2013.

[LSW03] J. Li, P. Stoica, and Z. Wang. On robust Capon beamforming and diagonal loading. *IEEE Transactions on Signal Processing*, 51(7):1702–1715, 2003.

[Mad08] U. Madhow. *Fundamentals of Digital Communication*. Cambridge University Press, 2008.

[MAMK08] M. Maddah-Ali, A. Motahari, and A. Khandani. Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis. *IEEE Transactions on Information Theory*, 54(8):3457–3470, 2008.

[MSKF09] J. Medbo, I. Siomina, A. Kangas, and J. Furuskog. Propagation channel impact on LTE positioning accuracy: A study based on real measurements of observed time difference of arrival. In *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC'09)*, pages 2213–2217, 2009.

[MZHH12] N. N. Moghadam, P. Zetterberg, P. Händel, and H. Hjalmarsson. Correlation of distortion noise between the branches of MIMO transmit antennas. In *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC'12)*, pages 2079–2084, 2012.

[NGS12] F. Negro, I. Ghauri, and D. T. M. Slock. Sum rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE. In *Proc. Int. Symp. Wireless Commun. Systems (ISWCS'12)*, pages 576–580, 2012.

[OS99] A. V. Oppenheim and R. W. Schafer. *Discrete-time signal processing*. Prentice Hall, 2nd edition, 1999.

[PCL03] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas. Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization. *IEEE Transactions on Signal Processing*, 51(9):2381–2401, 2003.

[PD12] D. S. Papailiopoulos and A. G. Dimakis. Interference alignment as a rank constrained rank minimization. *IEEE Transactions on Signal Processing*, 60(8):4278–4288, 2012.

[PGH08] A. Papadogiannis, D. Gesbert, and E. Hardouin. A dynamic clustering approach in wireless networks with multi-cell cooperative processing. In *Proc. IEEE Int. Conf. Commun. (ICC'08)*, pages 4033–4037, 2008.

[PH09] S. Peters and R. Heath. Interference alignment via alternating minimization. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP'09)*, pages 2445–2448, 2009.

[PH11] S. W. Peters and R. W. Heath. Cooperative algorithms for MIMO interference channels. *IEEE Transactions on Vehicular Technology*, 60(1):206–218, 2011.

[Pos08] Post- och telestyrelsen. Tillstånd att använda radiosändare i frekvensbandet 2500-2690 MHz enligt lagen (2003:389) om elektronisk kommunikation. Beslut, Post- och telestyrelsen, May 2008.

[RBCL13] M. Razaviyayn, M. Baligh, A. Callard, and Z.-Q. Luo. Robust transceiver design. *Patent WO 2013/044824 A1*, April 2013.

[RBP+13] R. Rogalin, O. Bursalioglu, H. Papadopoulos, G. Caire, A. Molisch, A. Michaloliakos, V. Balan, and K. Psounis. Scalable synchronization and reciprocity calibration for distributed multiuser MIMO. *arXiv:1310.7001v2 [cs.NI]*, October 2013.

[RC98] G. G. Raleigh and J. M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Transactions on Communications*, 46(3):357–366, March 1998.

[RHL13] M. Razaviyayn, M. Hong, and Z. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM J. Optimization*, 23(2):1126–1153, 2013.

[RLL12] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo. On the degrees of freedom achievable through interference alignment in a MIMO interference channel. *IEEE Transactions on Signal Processing*, 60(2):812–821, 2012.

[SA08] Y. Selen and H. Asplund. 3G LTE simulations using measured MIMO channels. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'08)*, pages 1–5, 2008.

[SBH11] C. Shi, R. Berry, and M. Honig. Interference alignment in multi-carrier interference networks. In *Inform. Theory (ISIT'11), Proc. IEEE Int. Symp.*, pages 26–30, 2011.

[SBH13] C. Shi, R. Berry, and M. Honig. Bi-directional training for adaptive beamforming and power control in interference networks. *IEEE Transactions on Signal Processing*, 2013. Accepted.

[SCB+09] D. A. Schmidt, S. Changxin, R. A. Berry, M. L. Honig, and W. Utschick. Minimum mean squared error interference alignment. In *Conf. Rec. 43rd Asilomar Conf. Signals, Systems and Computers*, pages 1106–1110, 2009.

[Sch08] T. Schenk. *RF Imperfections in High-rate Wireless Systems: Impact and Digital Compensation*. Springer, 2008.

[SGHP10] I. Santamaria, O. Gonzalez, R. W. Heath, and S. W. Peters. Maximum sum-rate interference alignment algorithms for MIMO channels. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'10)*, pages 1–6, 2010.

[SGLW03] S. Shahbazpanahi, A. Gershman, Z.-Q. Luo, and K. M. Wong. Robust adaptive beamforming for general-rank signal models. *IEEE Transactions on Signal Processing*, 51(9):2257–2269, Sept 2003.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[SM12] J. Shin and J. Moon. Weighted-sum-rate-maximizing linear transceiver filters for the K-user MIMO interference channel. *IEEE Transactions on Communications*, 60(10):2776–2783, October 2012.

[Smi04] G. S. Smith. A direct derivation of a single-antenna reciprocity relation for the time domain. *IEEE Transactions on Antennas and Propagation*, 52(6):1568–1577, 2004.

[SPLL10] H. Sung, S. Park, K. Lee, and I. Lee. Linear precoder designs for K-user interference channels. *IEEE Transactions on Wireless Communications*, 9(1):291–301, 2010.

[SRLH11] Q. Shi, M. Razavivayn, Z. Luo, and C. He. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Transactions on Signal Processing*, 59(9):4331–4340, 2011.

[SSB+13] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick. Comparison of distributed beamforming algorithms for MIMO interference networks. *IEEE Transactions on Signal Processing*, 61(13):3476–3489, 2013.

[Sto06] N. Storey. *Electronics: A Systems Approach*. Pearson Prentice Hall, 3rd edition, 2006.

[Str93] D. W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, 1993.

[SWB10] C. Studer, M. Wenk, and A. Burg. MIMO transmission with residual transmit-RF impairments. In *Int. ITG Workshop Smart Antennas (WSA'10)*, pages 189–196, 2010.

[SWB11] C. Studer, M. Wenk, and A. Burg. System-level implications of residual transmit-RF impairments in MIMO systems. In *Proc. European Conf. Antennas Propagation (EUCAP'11)*, pages 2686–2689, 2011.

[Tel99] E. Telatar. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, 1999.

[TI97] L. N. Trefethen and D. B. III. *Numerical Linear Algebra*. SIAM, 1997.

[TV08] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2008.

[VGL03] S. Vorobyov, A. Gershman, and Z.-Q. Luo. Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem. *IEEE Transactions on Signal Processing*, 51(2):313–324, Feb 2003.

[WAF+12] K. Werner, H. Asplund, D. V. P. Figueiredo, N. Jalden, and B. Halvarsson. LTE-Advanced 8×8 MIMO measurements in an indoor scenario. In *Antennas and Propagation. (ISAP'12), Proc. Int. Symp.*, pages 750–753, 2012.

[WAH+13] K. Werner, H. Asplund, B. Halvarsson, N. Jalden, D. V. P. Figueiredo, and A. K. Kathrein. LTE-A field measurements: 8×8 MIMO and carrier aggregation. In *Veh. Tech. Conf. (VTC'13 Spring), Proc. IEEE*, pages 750–753, 2013.

[WBM96] R. Wu, Z. Bao, and Y. Ma. Control of peak sidelobe level in adaptive arrays. *IEEE Transactions on Antennas and Propagation*, 44(10):1341–1347, 1996.

[Win84] J. H. Winters. Optimum Combining in Digital Mobile Radio with Cochannel Interference. *IEEE Journal on Selected Areas in Communications*, 2:528–539, July 1984.

[WV13] C. Wilson and V. Veeravalli. A convergent version of the MaxSINR algorithm for the MIMO interference channel. *IEEE Transactions on Wireless Communications*, 12(6):2952–2961, June 2013.

[YGJK10] C. Yetis, T. Gou, S. A. Jafar, and A. Kayran. On feasibility of interference alignment in MIMO interference networks. *IEEE Transactions on Signal Processing*, 58(9):4771–4782, 2010.

[ZSGL05] K. Zarifi, S. Shahbazpanahi, A. Gershman, and Z.-Q. Luo. Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver. *IEEE Transactions on Signal Processing*, 53(1):295–305, Jan 2005.

[ZT03] L. Zheng and D. Tse. Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, May 2003.