Multi-user MIMO wireless communications

Von der Fakultät für Elektrotechnik und Informationstechnik

der Technischen Universität Ilmenau

zur Erlangung des akademischen Grades eines

Doktoringenieurs (Dr. Ing.) genehmigte Dissertation

von

Veljko Stanković

Gutachter: Prof.Dr.-Ing. Martin Haardt

Prof.Dr.-Ing.habil. Reiner Thomä

Prof.Dr. Tadashi Matsumoto

November 2006

urn:nbn:de:gbv:ilm1-2007000068

Ovaj rad posvećujem majci Negici, ocu Ljubomiru i bratu Vladimiru.

Acknowledgments

This work evolved during my time as a research assistant at the Communications Research Laboratory at Ilmenau University of Technology, Germany. I would like to thank to Professor Martin Haardt for giving me the opportunity to work at his department and develop this thesis. He influenced my work with many helpful suggestions and fruitful discussions.

I especially thank to Professor Reiner Thomä for serving as a reviewer of my thesis and his support and help in finishing this work. The third reviewer was Professor Tadashi Matsumoto.

My thanks also go to my former colleagues at the WINNER project. The time I have spent working at WINNER was the most challenging and most exciting in my life so far. I had many interesting and helpful discussions with my colleagues from all over the world working at WINNER. I thank them for their help and pleasant discussions.

I would like to express my sincere thanks also to all the other members of the Communications Research Laboratory for the friendly, cooperative and family like atmosphere over the last years. I would like to especially thank to my colleagues Martin Fuchs and Giovanni Del Galdo for being an invaluable support and for their friendship.

Finally, I thank to my parents and my brother for being constant support in my life. They have helped me to go on whenever I had my doubts and I needed support. Thanks also to my friends Aleksandra Cvetković, Petar Spalević and Marko Milojević for words of encouragement and their help in the moments when I needed the help the most.

Veljko Stanković

Ilmenau University of Technology

November 2006

Zusammenfassung

Mehrantennensysteme sind auf Grund der erhhten Bandbreiteneffizienz und Leistung eine Schlsselkomponente von Mobilfunksystemen der Zukunft. Diese ermglichen das gleichzeitige Senden von mehreren, rumlich getrennten Datenstrmen zu verschiedenen Nutzern. Die zentrale Fragestellung in der Praxis ist, ob der ursprnglich vorausgesagte Kapazittsgewinn in realistischen Szenarios erreicht wird und welche spezifischen Gewinne durch zustzliche Antennen und das Ausnutzen von Kanalkenntnis am Sender und Empfnger erzielt werden, was andererseits einen Zuwachs an Overhead oder ntiger Rechenleistung bedeutet.

In dieser Arbeit werden neue lineare und nicht-lineare MU-MIMO Precoding-Verfahren vorgestellt. Der verfolgte Ansatz zur Bestimmung der Precoding-Matrizen ist allgemein anwendbar und die entstandenen Algorithmen knnen zur Optimierung von verschiedenen Kriterien mit beliebig vielen Antennen an der Mobilstation eingesetzt werden. Das wurde durch die Berechnung der Precoding-Matrix in zwei Schritten erreicht. Im ersten Schritt wird die berschneidung der Zeilenrume minimiert, die durch die effektiven Kanalmatrizen verschiedener Nutzer aufgespannt werden. Basierend auf mehreren parallelen Einzelnutzer-MIMO-Kanlen wird im zweiten Schritt die Systemperformanz bezglich bestimmter Kriterien optimiert.

Aus der gngigen Literatur ist bereits bekannt, dass fr Nutzer mit nur einer Antenne das MMSE Kriterium beim precoding optimal aber nicht bei Nutzern mit mehreren Antennen. Deshalb werden in dieser Arbeit zwei neue Mehrnutzer MIMO Strategien vorgestellt, die vom MSE Kriterium abgeleitet sind, nmlich sukzessives MMSE und RBD. Bei der sukzessiven Verarbeitung mit einer entsprechenden Anpassung der Sendeleistungsverteilung kann die volle Diversitt des Systems ausgeschpft werden. Die Kapazitt nhert sich dabei der maximalen Summenrate des Systems an. Bei gemeinsamer Verarbeitung der MIMO Kanle wird unabhngig vom Grad der Mehrnutzerinterferenz die maximale Diversitt erreicht.

Die genannten Techniken setzen entweder eine aktuelle oder eine ber einen lngeren Zeitraum gemittelte Kanalkenntnis voraus. Aus diesem Grund mssen die Auswirkungen von Kanal-Schtzfehlern und Einflsse des Transceiver Front-Ends auf die Verfahren nher untersucht werden.

Fr eine weitergehende Abschtzung der Mehrantennensysteme muss die Performanz des Gesamtsystems untersucht werden, da viele Einflsse auf die rumliche Signalverarbeitung bei Betrachtung eines einzelnen Links nicht erkennbar sind. Es wurde gezeigt, dass mit MIMO Precoding Strategien ein Vielfaches der Datenrate eines Systems mit nur einer Antenne erzielt werden kann, whrend der Overhead durch Pilotsymbole und Steuersignale nur geringfgig zunimmt.

Abstract

Multiple-input, multiple-output (MIMO) systems are a key component of future wireless communication systems, because of their promising improvement in terms of performance and bandwidth efficiency. An important research topic is the study of multi-user (MU) MIMO systems. Such systems have the potential to combine the high throughput achievable with MIMO processing with the benefits of space division multiple access (SDMA). The main question from a practical standpoint is whether the initially predicted capacity gains can be obtained in more realistic scenarios and what specific gains result from adding more antennas and overhead or computational power to obtain channel state information (CSI) at the transceivers.

In this thesis we introduce new linear and non-linear MU MIMO processing techniques. The approach used for the design of the precoding matrix is general and the resulting algorithms can address several optimization criteria with an arbitrary number of antennas at the user terminals (UTs). This is achieved by designing the precoding matrices in two steps. In the first step we minimize the overlap of the row spaces spanned by the effective channel matrices of different users. In the next step, we optimize the system performance with respect to the specific optimization criterion assuming a set of parallel single-user MIMO channels.

As it was previously reported in the literature, minimum mean-squared-error (MMSE) processing is optimum for single-antenna UTs. However, MMSE suffers from a performance loss when users are equipped with more than one antenna. The two MU MIMO processing techniques that result from the two different MSE criteria that are proposed in this thesis are successive MMSE and regularized block diagonalization. By iterating the closed form solution with appropriate power loading we are able to extract the full diversity in the system and empirically approach the maximum sum-rate capacity in case of

high multi-user interference. Joint processing of MIMO channels yields maximum diversity regardless of the level of multi-user interference.

As these techniques rely on the fact that there is either instantaneous or long-term CSI available at the base station to perform precoding and decoding, it was very important to investigate the influence of the transceiver front-end imperfections and channel estimation errors on their performance.

For a comprehensive assessment of multi-antenna techniques, it is mandatory to consider the performance at system level, since many effects of spatial processing are not tractable at the link level. System level investigations have shown that MU MIMO precoding techniques provide several times higher data rates than single-input single-output systems with only slightly increased pilot and control overhead.

List of Tables

List of Figures

List of Symbols and Abbreviations

$./$	-	Element wise division
$(\bullet)^+$	-	$(x)^+ := \max(0, x)$
$*$	-	Hadamard (element wise) product
$\|\boldsymbol{X}\|_{\mathcal{F}}$	-	Frobenius norm of a matrix
$\boldsymbol{X}^*$	-	Matrix conjugate
$\boldsymbol{X}^H$	-	Matrix conjugate transpose
$\boldsymbol{X}^T$	-	Matrix transpose
$\mathbf{1}$	-	Matrix with all elements equal to 1
$\boldsymbol{A}$	-	Array steering matrix
β	-	Scaling factor used to set the transmit power constraint
$\boldsymbol{B}$	-	THP feedback matrix of dimension $r \times r$
$\mathcal{CN}(a, b)$	-	Complex Gaussian random variable with mean value a and variance b
C_{BC}	-	Shannon capacity of multiple access channel
C_{MAC}	-	Shannon capacity of broadcast channel
C_{SU}	-	Single user Shannon capacity
$\boldsymbol{D}$	-	Combined decoding matrix of all users on the uplink of dimension $r \times M_T$
$\boldsymbol{D}_i$	-	The i^{th} user's decoding matrix on the uplink of dimension $r_i \times M_T$
$\boldsymbol{D}_a$	-	Combined decoding matrix for MUI suppression of all users on the uplink
$\boldsymbol{D}_b$	-	Combined decoding matrix for data decoding of all users on the uplink
$\boldsymbol{D}_{a_i}$	-	The i^{th} user's decoding matrix for MUI suppression on the uplink
$\boldsymbol{D}_{b_i}$	-	The i^{th} user's decoding matrix for data decoding on the uplink
$E\{\bullet\}$	-	Averaging function
$\boldsymbol{\Phi}$	-	Power loading matrix
$\boldsymbol{\Phi}_i$	-	The i^{th} user's power loading matrix
$\boldsymbol{\Phi}_{impD}$	-	Improved diversity power loading matrix
$\boldsymbol{\Phi}_{\text{MMSE}}$	-	MMSE power loading matrix
$\varphi_{i,i}$	-	Element on the main diagonal of the power loading matrix
f_0	-	Subcarrier spacing
$\boldsymbol{F}_a$	-	Combined precoding matrix for MUI suppression of all users on the downlink

$\boldsymbol{F}_b$	-	Combined precoding matrix for data precoding of all users on the downlink
$\boldsymbol{F}_{a_i}$	-	The i^{th} user's precoding matrix for MUI suppression on the downlink
$\boldsymbol{F}_{b_i}$	-	The i^{th} user's precoding matrix for data precoding on the downlink
$\boldsymbol{F}$	-	Combined precoding matrix of all users on the downlink of dimension $M_T \times r$
$\boldsymbol{F}_i$	-	The i^{th} user's precoding matrix on the downlink of dimension $M_T \times r_i$
G_{CSI}	-	Channel estimator gain
$\boldsymbol{G}$	-	Combined decoding matrix of all users on the downlink of dimension $r \times M_R$
$\boldsymbol{G}_i$	-	The i^{th} user's decoding matrix on the downlink of dimension $r_i \times M_{R_i}$
$\boldsymbol{H}$	-	Combined network MIMO channel matrix of dimension $M_R \times M_T$
$\boldsymbol{H}_{\text{est}}$	-	The estimate of the channel matrix
$\boldsymbol{H}_i$	-	The i^{th} user's MIMO channel matrix of dimension $M_{R_i} \times M_T$
$\boldsymbol{H}_i^{(k,j)}$	-	The i^{th} user's channel matrix on the j^{th} symbol of the k^{th} chunk
$\widehat{\boldsymbol{H}}_i^{(k)}$	-	The i^{th} user's equivalent channel matrix on the k^{th} chunk
$\widetilde{\boldsymbol{H}}_i$	-	Combined network MIMO channel matrix of the i^{th} user's co-channel interfering users
$\boldsymbol{H}^{(l)}$	-	Combined network MIMO channel matrix of the l^{th} channel path component
$\boldsymbol{H}_w$	-	Spatially white MIMO channel of dimension $M_R \times M_T$
$\boldsymbol{I}$	-	Identity matrix
K	-	Number of users
K_{R}	-	Ricean factor
M_R	-	Total number of antennas at the user terminals
M_{R_i}	-	Number of antennas at the i^{th} user terminal
M_T	-	Number of antennas at the base station
$\boldsymbol{n}$	-	Vector of sample of the additive Gaussian noise at the input of the receive antennas
$\bar{n}_i$	-	Filtered additive noise on the i^{th} data stream
N	-	Total number of subcarriers in OFDM system
N_{pre}	-	Length of the OFDM symbol cyclic prefix
N_c	-	Number of subcarriers used for transmission of one data block
N_{symb}	-	Number of OFDM symbols in one data block
P_T	-	Total transmit power
P_t	-	Power of one complex data symbol
$PAP(\theta)$	-	Power angle profile of the channel
$PDP(\tau)$	-	Power delay profile of the channel
$\boldsymbol{\Psi}_{T_x,UT}$	-	Complex perturbation matrix of the transmit RF chain at the user terminal
$\boldsymbol{\Psi}_{R_x,UT}$	-	Complex perturbation matrix of the receive RF chain at the user terminal
$\boldsymbol{\Psi}_{T_x,BS}$	-	Complex perturbation matrix of the transmit RF chain at the base station
$\boldsymbol{\Psi}_{R_x,BS}$	-	Complex perturbation matrix of the receive RF chain at the base station
$\boldsymbol{Q}$	-	Combined precoding matrix of all users on the uplink of dimension $M_R \times r$

$\boldsymbol{Q}_i$	-	The i^{th} user's precoding matrix on the uplink of dimension $M_{R_i} \times r_i$
$\boldsymbol{R}$	-	Correlation matrix
R_i	-	The i^{th} user's data rate
R_i^{DL}	-	The i^{th} user's data rate on the downlink
R_i^{UL}	-	The i^{th} user's data rate on the uplink
r	-	Total number of transmitted data stream in a multi-ser system
r_i	-	Number of data streams transmitted to the i^{th} user
$\boldsymbol{R}_r$	-	Receive correlation matrix
$\boldsymbol{R}_t$	-	Transmit correlation matrix
$\boldsymbol{R}_x$	-	Input data correlation matrix of dimension $r \times r$
$\boldsymbol{R}_n$	-	Noise correlation matrix
σ_e^2	-	Variance of channel estimation error
σ_i	-	The i^{th} singular value
σ_n^2	-	Variance of additive noise at the input of one antenna
σ_τ	-	RMS delay spread
σ_θ	-	RMS angle spread
$\boldsymbol{\Sigma}$	-	Diagonal matrix with singular values on main diagonal in non-increasing order
$\boldsymbol{U}$	-	Right (column) singular vectors
$\boldsymbol{V}$	-	Left (row) singular vectors
$\boldsymbol{x}$	-	Combined input data vector of all users of dimension $r \times 1$
$\boldsymbol{x}_i$	-	The i^{th} user's input data vector of dimension $r_i \times 1$
$\boldsymbol{y}$	-	Combined receive vector of all users of dimension $r \times 1$
$\boldsymbol{y}_i$	-	The i^{th} user's receive vector of dimension $r_i \times 1$
$\boldsymbol{z}_i$	-	The i^{th} user's encoded data

ADC	-	Analog digital converter
AP	-	Access point
BC	-	Broadcast channel
BER	-	Bit error rate
BD	-	Block diagonalization
BS	-	Base station
CCDF	-	Cumulative complementary distribution function
CQI	-	Channel quality indicator
CSI	-	Channel state information
DET	-	Dominant eigenmode transmission
DFE	-	Decision feedback equalizer
DFT	-	Discrete Fourier transform
DL	-	Downlink
DPC	-	Dirty-paper coding/Dirty-paper code
FDD	-	Frequency division duplex
FDMA	-	Frequency division multiple access
IQ	-	In-phase / Quadrature
IRBD	-	Iterative regularized block diagonalization
JRBD	-	Joint regularized block diagonalization
LDC	-	Linear Dispersion Code
LNA	-	Low-noise amplifier
LOS	-	Line of sight
MAC	-	Multiple access channel
MC CDMA	-	Multicarrier code division multiple access
MIMO	-	Multiple-input multiple-output
MMSE	-	Minimum mean square error
MSE	-	Mean square error
MU	-	Multi-user
MUI	-	Multi-user interference
NLOS	-	Non-line-of sight
OFDM	-	Orthogonal frequency division multiplex
OFDMA	-	Orthogonal frequency division multiple access
OSTBC	-	Orthogonal space-time block code
PA	-	Power amplifier
PAP	-	Power angle profile
PDP	-	Power delay profile
QAM	-	Quadrature amplitude modulation
QoS	-	Quality of service
RF	-	Radio frequency
RMS	-	root-mean-squared
RBD	-	Regularized block diagonalization
RSO THP	-	Regularized successive optimization Tomlinson-Harashima-precoding
QOSTBC	-	Quasi-orthogonal space time block coding
SDMA	-	Spatial division multiple access
SIC	-	Successive interference cancellation
SINR	-	Signal to interference plus noise ratio
SISO	-	Single-input single-output
SNR	-	Signal to noise ratio
SMMSE	-	Successive minimum mean square error

SMUX	-	Spatial Muliplexing
SO THP	-	Successive optimization Tomlinson-Harashima-precoding
STC	-	Space Time Coding
SU	-	Single user
SVD	-	Singular value decomposition
TDD	-	Time division duplex
TDMA	-	Time division multiple access
THP	-	Tomlinson-Harashima-precoding
UL	-	Uplink
ULA	-	Uniform Linear Array
UT	-	User Terminal
V-BLAST	-	Vertical Bell Labs Layered SpaceTime
VS	-	Very simple
WINNER	-	Wireless World Initiative New Radio
ZF	-	Zero forcing

Contents

Chapter 1

Introduction

The next generation of wireless mobile communication systems requires the reliable transmission of high-rate data under various types of channels and scenarios. Current wireless mobile, data, and fixed access communication systems are converging into a data (all IP) oriented wireless networks with high spectral efficiency. Future wireless communication systems should be flexible and adaptive to various scenarios and Quality-of-Service (QoS) requirements. The system should be robust to the influence of fading, interference, and hardware imperfections.

The very high data rates that are required for future wireless systems in reasonably large areas do not appear to be feasible with the conventional techniques and architectures. Frequency bands that are envisioned for future wireless communication systems are well above 2 GHz. The radio propagation in these bands is significantly more vulnerable to non-line-of sight (NLOS) conditions, which is typical in modern urban communications. The efficient design of wireless systems will require the use of multiple antennas, advanced adaptive modulation and coding schemes, relaying nodes, cooperative networks and users, and cross-layer design.

The goal of reaching high data rates is particularly challenging for systems that are power, bandwidth, and complexity limited. However, another domain can be exploited to significantly increase channel capacity: the use of multiple transmit and receive antennas. Pioneering work by Winters [1], Telatar [2], and Foschini [3] ignited much interest in this area by predicting remarkable spectral efficiencies for wireless systems with multiple antennas when the channel exhibits rich scattering and the channel state information (CSI) can be accurately tracked. This initial promise of exceptional spectral efficiency resulted in an explosion of research activities to characterize the theoretical and practical

issues associated with multiple-input multiple-output (MIMO) channels and to extend these concepts to multi-user systems. The main question from both a theoretical and practical standpoint is whether the enormous initially predicted capacity gains can be obtained in a more realistic operating scenarios and what specific gains result from adding more antennas and overhead or computational power to obtain CSI at the transmitter and receiver.

The large spectral efficiencies associated with MIMO channels are based on the premise that a rich scattering environment provides independent transmission paths for each transmit-receive antenna pair. Therefore, for single-user (SU) systems, a transmission and reception strategy that exploits this structure achieves capacity on approximately $\min(M_T, M_R)$ separate channels, where M_T is the number of transmit antennas and M_R is the number of receive antennas. Thus, capacity scales linearly with $\min(M_T, M_R)$ relative to a system with just one transmit and one receive antenna. The capacity increase requires a scattering environment such that the matrix of channel gains between each transmit and receive antenna pair has full rank and independent entries and that perfect estimates of these gains are available at the transmitter and receiver.

Space-time coding (STC) [4], [5], and spatial multiplexing (SMUX) [3], [6], provide full diversity and achieve high data rates over MIMO channels, respectively. Spatial multiplexing involves transmitting independent streams of data across multiple antennas to maximize throughput, whereas space-time coding maps input symbol streams across space and time for diversity and coding gain at a given data rate. Neither scheme requires CSI at the transmitter. However, to achieve the maximum information rate and/or the diversity and array gain afforded by increased computational complexity, appropriate precoding and modulation techniques are necessary.

Generalized designs of a jointly optimum linear precoder and decoder for a SU MIMO system, using a mean-squared error (MSE) criterion are presented in [7] and [8]. The framework presented in these papers is general and addresses several optimization criteria like minimum MSE (MMSE), minimum bit error rate (BER) and maximum information rate. It is assumed that the channel is known at the receiver as well as at the transmitter. CSI can be acquired at the transmitter either by using feedback from the receiver or by using the reciprocity principle when the transmitter and receiver operate in time division duplex (TDD) so that the time-invariant MIMO channel transfer function is the same in both ways. The optimum precoder and decoder diagonalize the MIMO channel into eigen

subchannels. The different solutions targeting different optimization criteria are obtained by using different power allocation schemes over these subchannels. For example, the optimum linear precoder and decoder that maximize the information rate, decouple the MIMO channel into eigen subchannels and allocate power to these subchannels according to the water-pouring policy [7], [9].

An important research topic is the study of multi-user (MU) MIMO systems. Such systems have the potential to combine the high capacity achievable with MIMO processing with the benefits of space division multiple access (SDMA). In the MU MIMO scenario, a base station (BS) or an access point (AP) is equipped with multiple antennas and it is simultaneously communicating with a group of users. Each of these users is also equipped with multiple antennas. Motivated by the need for cheap mobiles with low power consumption, we focus on systems where the complex signal processing is performed at the BS/AP. The BS/AP will use the CSI available at the transmitter to allow these users to share the same channel and mitigate or completely eliminate multi-user interference (MUI) in an ideal case.

In a MU scenario, capacity becomes a K-dimensional region defining the set of all rate vectors $(R_1, \ldots, R_K)$ simultaneously achievable by all K users. Two MU MIMO scenarios can be distinguished. In the first scenario, multiple non-cooperative terminals are transmitting to a single receiver. This scenario is often referred to as the MU MIMO uplink (UL) channel. In the information theory it is known as the MIMO multiple access channel (MAC). The scenario, in which a single terminal is transmitting to multiple non-cooperative receivers is referred to as MU MIMO downlink channel or broadcast channel (BC). MU MIMO downlink system is depicted in Figure 1.1 and MU MIMO uplink system is depicted in Figure 1.2.

The capacity region of a general MIMO MAC was obtained in [2], [10]. It has been shown that a linear detection with successive interference cancellation (SIC) provides the maximum sum rate capacity of a MU MAC system. However, the capacity of a MIMO BC is an open problem due to the lack of a general theory on non-degraded broadcast channels. In pioneering work by Caire and Shamai [11], a set of achievable rates for the MIMO BC was obtained by applying Costa's "dirty-paper" coding (DPC) technique at the transmitter [12]. In [12], Costa proved the surprising result that the capacity of the channel, when the non-causal additive Gaussian interference is perfectly known at the transmitter, is the same as if the interference was not present. It was also shown in [11]

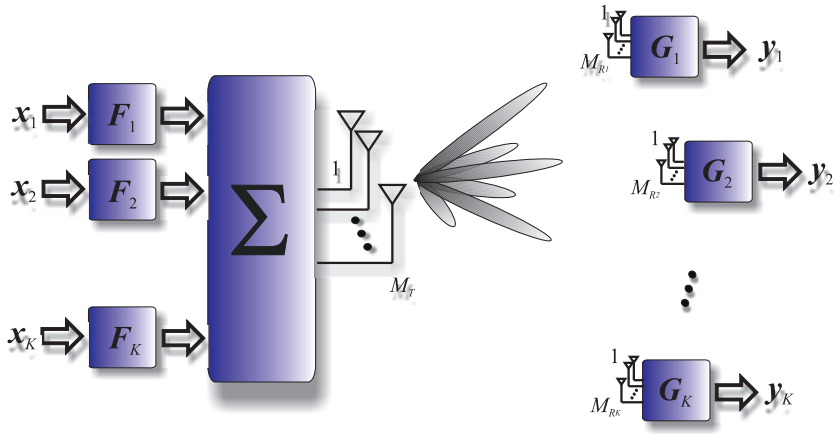

Figure 1.1: Block diagram of multi-user MIMO downlink system.

that the sum rate MIMO BC capacity equals the maximum sum rate DPC achievable region by demonstrating that the achievable rate meets the Sato upper bound [13].

DPC is a technique that allows non-causally known interference to be "pre-subtracted" at the transmitter. In [14], [15] it was shown that the achievable region of the MIMO BC obtained using DPC is equal to the capacity region of the MIMO MAC using uplink-downlink duality. This allows us to substitute the non-convex problem of finding the DPC rate region with the dual MAC problem where the rates are convex functions of the correlation matrices.

DPC can achieve the maximum sum rate of the system and provide the maximum diversity order [16], [17]. However, these techniques require the use of a complex sphere-decoder or an approximate closest-point solution, which makes them hard to implement in practice, especially when the number of users is large [17]. Tomlinson-Harashima precoding (THP) was first developed for single-input single-output (SISO) multipath channels, where it was used to overcome the error propagation problem of decision feedback filtering, [18], [19]. In [20] it is proposed for the equalization of MUI in MIMO systems. In [21] the authors propose the use of THP in combination with MMSE filtering. In [21] successive interference cancellation is performed at the transmitter, whereas the receiver still performs linear filtering. In [22], both feedforward and feedback filters are deployed at the transmitter which results in a significant reduction of the computational load at the receiver side. Although DPCs outperform THP, THP is much less computationally demanding and thus more attractive for practical implementation.

Linear MU MIMO processing techniques are less computationally demanding than DPCs, and they can use either instantaneous channel knowledge or long-term statistics of

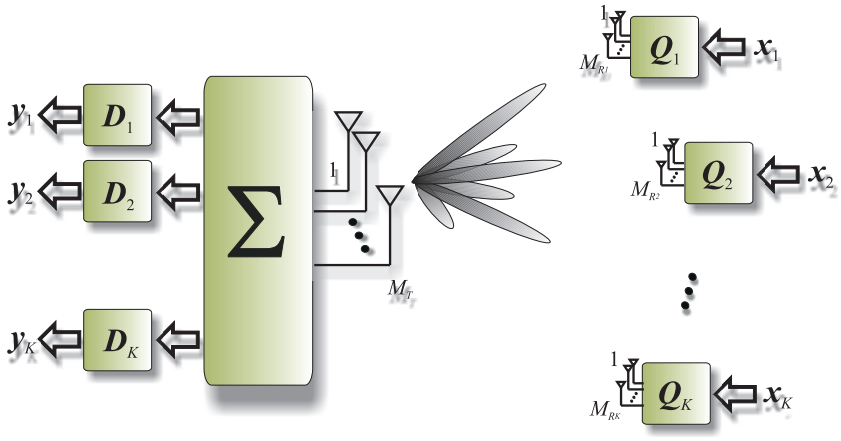

Figure 1.2: Block diagram of multi-user MIMO uplink system.

the channel to perform precoding or decoding. In general, linear MU MIMO processing techniques cannot provide the maximum sum rate capacity, but there are some cases where this is possible and where the MUI is set to a minimum by choosing semi-orthogonal users for simultaneous transmission using SDMA [23]. In this thesis it will be empirically shown that linear processing techniques reach the sum-rate capacity of the BC channel also when the total number of antennas at the user terminals is equal to or greater than the number of antennas at the base station. Non-linear MU MIMO processing techniques require the instantaneous knowledge of the channel transfer function at the BS. On the other hand, linear MU MIMO processing techniques can be used with various degrees of channel state information. Thus, linear techniques are more flexible and more favorable for practical implementation than non-linear techniques.

1.1 Scope of the thesis and contributions

In this thesis we will introduce a general framework for the design of the multi-user MIMO precoding and decoding matrices. Our goal is to define one MU MIMO algorithm that will be able to address several optimization criteria like minimum MSE (MMSE), minimum bit error rate (BER), and maximum information rate. When all users are equipped with one antenna, it has been shown in the literature that the MU MIMO precoding and decoding algorithms that are defined based on the MSE criterion have the best performance, [22], [9]. In this thesis, two new cost functions that are derived from the MSE criterion will be introduced. Based on the optimization of these two cost functions two new MU MIMO precoding techniques will be defined.

The link between the user terminals and the base station in a wireless multi-user MIMO

scenario is the wireless propagation channel. As a consequence, it is very important to accurately model the behavior of the channel in the simulations to allow realistic assessment of the performance that can be achieved. In Chapter 2 we describe the channel models that will be used in the simulations. At the end, a discrete input-output signal model will be introduced.

In Chapter 3, an overview of fundamental data rate limits of the MAC and the BC channels will be given. Numerically finding the maximum of the DPC sum-rate capacity region is not a trivial problem. Based on the iterative water-pouring algorithm for the conventional MAC problem, an iterative water-pouring algorithm was proposed in [24], to compute the correlation matrices of a dual MAC channel. These MAC correlation matrices are then transformed to the downlink correlation matrices that achieve the same rates under the same sum power constraint. This algorithm is complex and requires numerous calculations of SVDs and the water-pouring algorithm. By using the Hadamard inequality we find a novel algorithm in Chapter 3 to compute an upper bound which is not always as tight as the previous one but which requires less computational effort.

An overview of the SU MIMO processing techniques will be given in Chapter 4. First, we will review techniques that do not need any CSI at the transmitter to extract diversity gain or spatial multiplexing gain. These techniques are suboptimum because they do not require CSI at the transmitter to encode the user's data. Generalized designs of a jointly optimum linear precoder and decoder for a single user (SU) MIMO system, using a mean-squared error (MSE) criterion are given in [7] and [8]. There are numerous results in the literature which address the optimization of multi-user MIMO downlink systems using different optimization criteria. However, most of the solutions are not general like in the case of point-to-point communications.

In Chapter 5 we first give a short overview of the most relevant multi-user precoding techniques that we have used as a starting point in our investigations. Each of these techniques has certain drawbacks that have significant impact on the performance and design of the multi-user MIMO systems. The minimum mean-square-error (MMSE) precoder balances the multi-user interference mitigation with noise enhancement and minimizes the total error. The drawback of this technique is that it is limited to single antenna user terminals. In a MU MIMO system employing MMSE precoding, if the user terminal is equipped with more than one antenna, the signal transmitted over each antenna needs to be precoded independently. Block diagonalization (BD) is more appropriate to be used

with user terminals with multiple number of antennas [25]. However, it has a limitation that the total number of antennas at the user terminals has to be less than or equal to the number of antennas at the base station.

BD MMSE THP is a novel technique introduced in [26] that combines MMSE THP to precode the signals to the single-antenna terminals and BD to precode the signals to the multiple-antenna terminals. The performance of this technique is better than BD or MMSE THP, however, it has the same limitation regarding the number of antennas at the user terminals and the base station as BD. Under these limitations, it has been shown by simulations that a combination of BD and THP that was proposed in [27] called successive optimization (SO) THP approaches the DPC sum-rate capacity bound.

In Section 5.5.1 we introduce a modified MSE cost function. Using this cost function we derive the successive MMSE (SMMSE) precoding filter from the linear transmit MMSE precoding optimization by neglecting the contribution between the elements of one user's channel matrix to this users' MSE, [28]. In a system using SMMSE, the number of antennas at the user terminals can be arbitrary but the total number of data streams has to be less than or equal to the rank of the combined network channel matrix of all users. The combination of SMMSE and THP was introduced in [29]. At low SNRs, SMMSE THP approaches the DPC sum-rate capacity bound in simulations. At high SNRs SMMSE THP provides higher diversity gain than SMMSE. A similar approach is used in [30] to design an uplink MU MIMO receive filter. This technique is called SMMSE successive interference cancellation (SIC) and it can provide much higher diversity than V-BLAST.

Although SMMSE processing allows a generalized design of MU MIMO precoding and decoding matrices it does not always fully extract the diversity and antenna array gains inherent in the MU MIMO system. In Section 5.6 we introduce the second cost function based on the MSE criterion. The resulting precoding technique called regularized BD (RBD) was introduced in [29] and [31]. In combination with THP or by iterating the closed form solution we reach the maximum sum rate capacity of the broadcast channel in simulations. By iterating or by joint precoding in two other dimensions, time and frequency, we are able to extract the maximum diversity in the system.

In Section 5.7 a novel approach that allows that the same MU MIMO precoding algorithms requiring instantaneous CSI at the transmitter can be used also with the long-term channel knowledge is introduced. Using this approach it is possible to combine instantaneous CSI for some users and long-term CSI for others to perform precoding. These

results were presented originally in [32]. At the end of Chapter 5 we investigate the influence of hardware impairments and channel estimation errors on the performance of the MU MIMO precoding techniques and we give a short overview of the relative complexity of these techniques.

The results of the system level investigations of MU MIMO precoding techniques are given in Chapter 6. These investigations are performed using the transmission system proposed within the WINNER (Wireless World Initiative New Radio) project. Advanced multi-antenna solutions are a very important part of the WINNER system. Since MU MIMO precoding techniques require either instantaneous or long-term channel knowledge at the transmitter, the most appropriate scenarios where these techniques can be used are indoor, hotspot or micro cellular scenario where it is possible to acquire CSI at the transmitter with low pilot and control overhead. The results show that MU MIMO systems can provide much higher data rates than single-input single-output (SISO) systems with a slightly higher pilot and control overhead.

Chapter 2

MIMO channel modeling

A profound understanding of MIMO channels is crucial in selecting proper signaling strategies in MIMO wireless systems. Simple models have been used to get the insight into the impact of propagation conditions on MIMO capacity. They assume that only spatial fading correlation is responsible for the rank structure of the MIMO channel. In practice, the realization of a high MIMO capacity in actual radio channels is sensitive not only to the fading correlation, but also to the structure of the scattering in the propagation environment. In this chapter we review the construction of wireless MIMO channels which we will use in simulations, its sampled model, as well as the input-output signal model.

2.1 The MIMO channel

Fading represents fluctuations in received signal level. Macroscopic fading is caused by shadowing effects of buildings or natural features and is determined by the local mean of the fast fading signal. Microscopic fading corresponds to rapid fluctuations of the received signal in time, frequency, and space and is caused by the signal scattering off objects between the transmitter and the receiver. The effective path loss follows an inverse n^{th} power law. In real environments the path loss exponent varies from 2.5 to 6 and is also a function of the terrain and foliage.

A MIMO channel with M_T transmit antennas and M_R receive antennas comprises of $M_T M_R$ SISO channels. The MIMO channel is given by the $M_R \times M_T$ matrix $\boldsymbol{H}(\tau;t)$ with

$$\boldsymbol{H}(\tau;t) = \begin{bmatrix} h_{1,1}(\tau;t) & \cdots & h_{1,M_T}(\tau;t) \\ \vdots & \ddots & \vdots \\ h_{M_R,1}(\tau;t) & \cdots & h_{M_R,M_T}(\tau;t) \end{bmatrix} \tag{2.1}$$

where $h_{m,n}(\tau;t)$ is the channel's impulse response between n^{th} transmit $n = 1,\ldots,M_T$ and m^{th} receive antenna $m = 1,\ldots,M_R$ as a function of delay τ and time t.

The MIMO channel is constructed under the narrowband array assumption [33]. Under the narrowband assumption, the bandwidth of the signal is assumed to be much smaller than the reciprocal of the transmit time of the signal across the antenna array. The scatterer location, antenna element patterns and geometry and the scattering model together determine the average power and the correlation between elements of $\boldsymbol{H}(\tau;t)$.

Time varying fading due to the scatterer or transmitter/receiver motion results in a Doppler spread over a finite spectral bandwidth ($f \pm f_D$), where f_D is the maximum Doppler frequency. Time selective fading is characterized by the coherence time T_C. The larger the coherence time, the slower the channel is changing.

In a multipath environment, several time-shifted and scaled versions of the transmitted signal arrive at the receiver, which cause frequency selective fading. The maximum spread of path delays is called time delay spread τ_{max}. The root-mean-squared (RMS) delay spread of the channel σ_τ, is defined as [9]

$$\sigma_\tau = \sqrt{\frac{\int_0^{\tau_{max}} (\tau - \bar{\tau})^2 \,\mathrm{PDP}(\tau)\, d\tau}{\int_0^{\tau_{max}} \mathrm{PDP}(\tau)\, d\tau}}$$

where $\mathrm{PDP}(\tau)$ is the power delay profile of the channel, i.e. the average power as a function of delay, and

$$\bar{\tau} = \frac{\int_0^{\tau_{max}} \tau \mathrm{PDP}(\tau)\, d\tau}{\int_0^{\tau_{max}} \mathrm{PDP}(\tau)\, d\tau}$$

Frequency selective fading is characterized by the coherence bandwidth B_C which is inversely proportional to the RMS delay spread and is a measure of a channel's frequency selectivity. When the coherence bandwidth is comparable or less than the signal bandwidth, the channel is said to be frequency selective.

The angle spread at the transmitter/receiver refers to the spread of angles of departure/arrival of the multipath components at the transmit/receive antenna array. The RMS angle spread, σ_θ, is defined using the angle spectrum (PAP - power angle profile), i.e., the average power as a function of angle of arrival, θ, as

$$\sigma_\theta = \sqrt{\frac{\int_{-\pi}^{\pi} \left(\theta - \bar{\theta}\right)^2 \mathrm{PAP}(\theta)\, d\theta}{\int_{-\pi}^{\pi} \mathrm{PAP}(\theta)\, d\theta}}$$

where $\bar{\theta}$ is the mean angle of arrival:

$$\bar{\theta} = \frac{\int_{-\pi}^{\pi} \theta \ \mathrm{PAP}\left(\theta\right) d\theta}{\int_{-\pi}^{\pi} \mathrm{PAP}\left(\theta\right) d\theta}$$

Angle spread causes the space selective fading which is characterized by the coherence distance, D_C. The coherence distance is the spatial separation for which the channel coefficients are significantly correlated, and it is inversely proportional to the RMS angle spread. The larger the angle spread, the shorter is the coherence distance.

The time delay spread increases with the distance between the base station and the user terminal (UT). In the rural environments σ_τ is less than 0.07 μs. In urban areas σ_τ is typically 0.8 μs, while in hilly terrains σ_τ of $2-3$ μs has been observed. In indoor scenario the average values of σ_τ are less than 200 ns, [34].

The angle spread strongly depends on the scenario and the antenna height. At the BS it varies from a fraction of a degree in a flat rural scenario to up to 20° in hilly and dense urban scenarios. The coherence distance varies from 3 to 20 λ_c, where λ_c is the wave length of the carrier. Scatterers at the UTs are distributed in all directions which yields much larger angle spreads. The coherence distance at the UT varies from $0.25\lambda_c$ to $5\lambda_c$. The azimuth angle spread in indoor scenario is in the 20° to 40° range, [9].

2.2 Sampled channel model

Let us consider a frequency flat, slowly varying MIMO channel. The matrix in equation (2.1) can be written as

$$\boldsymbol{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,M_T} \\ \vdots & \ddots & \vdots \\ h_{M_R,1} & \cdots & h_{M_R,M_T} \end{bmatrix} \tag{2.2}$$

where $h_{m,n}$ includes the effects of pulse shaping at the transmitter, matched filtering at the receiver and the physical channel.

The statistical properties of $\boldsymbol{H}$ depend on the scattering environment and array geometry at the transmitter and receiver. The classical spatially white MIMO channel $\boldsymbol{H}_w$ is characterized by

$$\begin{aligned} E\left\{h_{m,n}\right\} = 0; \quad E\left\{\left|h_{m,n}^2\right|\right\} = 1; \\ E\left\{h_{m,n} h_{i,j}^*\right\} = 0, \quad i \neq m \text{ or } j \neq n \end{aligned} \tag{2.3}$$

where the elements $h_{m,n}$ are modeled as complex Gaussian random variables. In practice, the MIMO channel can deviate significantly from the $\boldsymbol{H}_w$ behavior due to a variety of reasons which will be covered in the following sections.

2.2.1 Spatial fading correlation

Spatial fading correlation can be modeled by

$$\mathrm{vec}\left(\boldsymbol{H}\right) = \boldsymbol{R}^{1/2}\mathrm{vec}\left(\boldsymbol{H}_w\right) \tag{2.4}$$

where $\boldsymbol{H}_w$ is the spatially white MIMO channel and $\boldsymbol{R}$ is the spatial correlation matrix defined as

$$\boldsymbol{R} = E\left\{\mathrm{vec}\left(\boldsymbol{H}\right)\mathrm{vec}\left(\boldsymbol{H}\right)^H\right\} \tag{2.5}$$

The operation $\mathrm{vec}\left(\boldsymbol{H}\right)$ stacks all elements of the matrix $\boldsymbol{H}$ column by column in a column vector. If the SVD of matrix $\boldsymbol{R}$ is defined as $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H$, then $\boldsymbol{R}^{1/2}$ is defined as $\boldsymbol{R}^{1/2} = \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^H$.

In many applications, a simpler and less general model, known as Kronecker model, is more adequate and is given by

$$\boldsymbol{H} = \boldsymbol{R}_r^{1/2}\boldsymbol{H}_w\boldsymbol{R}_t^{1/2} \tag{2.6}$$

where $\boldsymbol{R}_t \in \mathbb{C}^{M_T \times M_T}$ is the transmit correlation matrix and $\boldsymbol{R}_r \in \mathbb{C}^{M_R \times M_R}$ is the receive correlation matrix. Both $\boldsymbol{R}_t$ and $\boldsymbol{R}_r$ are positive semi-definite matrices. The model in equation (2.6) has less degrees of freedom than (2.4), which is capable of capturing any correlation effects between the elements of $\boldsymbol{H}$.

Modeling of frequency selective MIMO channel

In case of frequency selective channels, antenna correlation is modeled in the delay domain using the Kronecker model. The l^{th} channel path component is modeled as

$$\boldsymbol{H}^{(l)} = {\boldsymbol{R}_r^{(l)}}^{1/2}\boldsymbol{H}_w^{(l)}{\boldsymbol{R}_t^{(l)}}^{1/2} \tag{2.7}$$

where $\boldsymbol{H}_w^{(l)}$ is a spatially white unit variance flat fading MIMO channel of dimension $M_R \times M_T$, whereas $\boldsymbol{R}_r^{(l)} = \mathrm{E}\left\{\boldsymbol{H}^{(l)}\boldsymbol{H}^{(l)\ H}\right\}/M_T$ and $\boldsymbol{R}_t^{(l)} = \mathrm{E}\left\{\boldsymbol{H}^{(l)\ H}\boldsymbol{H}^{(l)}\right\}/M_R$ are the receive and the transmit correlation matrices with $\mathrm{tr}\left(\boldsymbol{R}_r^{(l)}\right) = M_R$ and $\mathrm{tr}\left(\boldsymbol{R}_t^{(l)}\right) = M_T$.

Let us consider a scenario where the receiver is surrounded by a rich scattering envi-

ronment and the transmitter antennas are separated by less than the coherence distance. These propagation conditions correspond to a cellular communication systems typically characterized by a low angular spread at the transmitter. On the other hand, the angular spread at the mobile is often very large and thus low spatial correlation can be achieved with relatively small antenna separation. Hence, we can write

$$\boldsymbol{R}_r^{(l)} = \boldsymbol{I}_{M_R}, \quad \boldsymbol{R}_t^{(l)} = \frac{M_T}{\mathrm{tr}\left(\boldsymbol{A}^{(l)^*}\boldsymbol{A}^{(l)^T}\right)}\boldsymbol{A}^{(l)^*}\boldsymbol{A}^{(l)^T} \tag{2.8}$$

and the l^{th} channel path component is modeled as

$$\boldsymbol{H}^{(l)} = \sqrt{\frac{M_T}{\mathrm{tr}\left(\boldsymbol{A}^{(l)^*}\boldsymbol{A}^{(l)^T}\right)}}\boldsymbol{H}_w^{(l)}\boldsymbol{A}^{(l)^T} \tag{2.9}$$

where $\boldsymbol{A}^{(l)} \in \mathbb{C}^{M_T \times N}$ is an array steering matrix containing N array response vectors of the transmitting antenna array corresponding to N directions of departure [35], and $\boldsymbol{H}_{w_i}^{(l)} \in \mathbb{C}^{M_R \times N}$ is a spatially white unit variance flat fading MIMO channel.

2.2.2 Line-of-sight component

In the presence of a line-of-sight (LOS) component between the transmitter and the receiver, the MIMO channel may be modeled approximately as the sum of a fixed component and a scattered component as follows [9]

$$\boldsymbol{H} = \sqrt{\frac{K_R}{K_R+1}}\boldsymbol{H}_{LOS} + \sqrt{\frac{1}{K_R+1}}\boldsymbol{H}_w \tag{2.10}$$

where $E\{\boldsymbol{H}\} = \sqrt{\frac{K_R}{K_R+1}}\boldsymbol{H}_{LOS}$ is the LOS component of the channel and $\sqrt{\frac{1}{K_R+1}}\boldsymbol{H}_w$ is the fading component assuming uncorrelated fading. The elements of $\boldsymbol{H}_{LOS}$ are assumed to have unit power. K_R is the Ricean factor, which is defined as a ratio of the power of the LOS component and the power of the scattered component.

2.2.3 Cross-polarized antennas

So far, we have assumed that the antennas at the transmitter and the receiver have identical polarization. The use of antennas with different polarizations at the transmitter and the receiver leads to a gain and correlation imbalance between the elements of the channel $\boldsymbol{H}$. As a consequence the elements of $\boldsymbol{H}$ show more complex behavior.

Assuming correlated Rayleigh fading, the channel is modeled approximately as

$$\boldsymbol{H} = \boldsymbol{X} * \left(\boldsymbol{R}_r^{1/2} \boldsymbol{H}_w \boldsymbol{R}_t^{1/2} \right) \tag{2.11}$$

where

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{1}_{M_R/2 \times M_T/2} & \sqrt{\alpha}\mathbf{1}_{M_R/2 \times M_T/2} \\ \sqrt{\alpha}\mathbf{1}_{M_R/2 \times M_T/2} & \mathbf{1}_{M_R/2 \times M_T/2} \end{bmatrix}. \tag{2.12}$$

Moreover, $*$ is the Hadamard product and $\mathbf{1}_{M_R/2 \times M_T/2}$ is a matrix with all elements equal to one. The parameter $0 \leq \alpha \leq 1$ is related to the separation of orthogonal polarizations. The values of α, $\boldsymbol{R}_r$ and $\boldsymbol{R}_t$ depend on variety of factors including cross-polarization discrimination, cross-polarization coupling and antenna spacing. If antennas are capable of perfectly separating polarizations then $\alpha = 0$, otherwise if antennas use the same polarization or the rich scattering environment changes the polarization of the signal, α is close to or equal to 1.

2.3 Input-output signal model

In the following we will assume that the transmission of data is carried out using OFDM modulation, that the cyclic prefix is longer than the channel time delay spread and that the maximum Doppler frequency is much smaller than the subcarrier spacing. In this case we can neglect the intercarrier interference caused by the Doppler spread and assume that the channel is constant during one OFDM symbol. The discrete input-output signal relation for the MIMO system on the k^{th} subcarrier can be written as

$$\boldsymbol{y}(k) = \boldsymbol{G}(k) \left(\boldsymbol{H}(k) \boldsymbol{F}(k) \boldsymbol{x}(k) + \boldsymbol{n}(k) \right) \tag{2.13}$$

where $\boldsymbol{x}(k)$ is the data vector to be transmitted, $\boldsymbol{y}(k)$ is the data vector at the output of the channel and $\boldsymbol{n}(k)$ is the vector containing samples of complex, zero mean additive white Gaussian noise at the input of the receive antenna array. The matrices $\boldsymbol{F}(k)$ and $\boldsymbol{G}(k)$ are precoding and decoding matrices, respectively. In the following we will drop the index of the subcarrier k for simplicity reasons, except where necessary to denote the joint processing of a group of subcarriers.

In a MU MIMO scenario M_T antennas are located at the base station and M_{R_i} antennas are located at the i^{th} user terminal, $i = 1, 2, \ldots, K$. There are K users (or UTs) in the

system. The total number of antennas at the user terminals is equal to

$$M_R = \sum_{i=1}^{K} M_{R_i}.$$

We will use the notation $\{M_{R_1}, \ldots, M_{R_K}\} \times M_T$ to describe the antenna configuration of the system. The MIMO channel to user i is denoted as $\boldsymbol{H}_i \in \mathbb{C}^{M_{R_i} \times M_T}$.

We will use this notation for the users' channel matrices, the number of antennas at the base station and the number of antennas at the user terminals for both the uplink and the downlink.

The combined MU MIMO channel matrix is given by

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_1^T & \boldsymbol{H}_2^T & \cdots & \boldsymbol{H}_K^T \end{bmatrix}^T \in \mathbb{C}^{M_R \times M_T}. \tag{2.14}$$

A block diagram of a MU MIMO downlink system is depicted in Fig. 1.1. The downlink input-output signal model can be written as

$$\boldsymbol{y} = \boldsymbol{G}\left(\boldsymbol{H}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{n}\right) \tag{2.15}$$

where $\boldsymbol{y}$, $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{n}$ are given by

$$\begin{aligned}
\boldsymbol{x} &= \begin{bmatrix} \boldsymbol{x}_1^T & \cdots & \boldsymbol{x}_K^T \end{bmatrix}^T \in \mathbb{C}^{r \times 1} \\
\boldsymbol{y} &= \begin{bmatrix} \boldsymbol{y}_1^T & \cdots & \boldsymbol{y}_K^T \end{bmatrix}^T \in \mathbb{C}^{r \times 1} \\
\boldsymbol{n} &= \begin{bmatrix} \boldsymbol{n}_1^T & \cdots & \boldsymbol{n}_K^T \end{bmatrix}^T \in \mathbb{C}^{M_R \times 1} \\
\boldsymbol{G} &= \begin{bmatrix} \boldsymbol{G}_1 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{G}_K \end{bmatrix} \in \mathbb{C}^{r \times M_R} \\
\boldsymbol{F} &= \begin{bmatrix} \boldsymbol{F}_1 & \cdots & \boldsymbol{F}_K \end{bmatrix} \in \mathbb{C}^{M_T \times r}
\end{aligned} \tag{2.16}$$

The vectors $\boldsymbol{x}_i \in \mathbb{C}^{r_i \times 1}$ and $\boldsymbol{y}_i \in \mathbb{C}^{r_i \times 1}$ are the transmitted and the received data vectors of the i^{th} user, respectively. $r_i \leq \text{rank}(\boldsymbol{H}_i) \leq \min(M_{R_i}, M_T)$ is the number of spatially multiplexed information data streams transmitted to the i^{th} user. The total number of data streams transmitted is $r = \sum_{i=1}^{K} r_i$. Samples of the white additive Gaussian noise at the input of the i^{th} user's antenna array are given in $\boldsymbol{n}_i \in \mathbb{C}^{M_{R_i} \times 1}$. The matrices $\boldsymbol{F}_i \in \mathbb{C}^{M_T \times r_i}$ and $\boldsymbol{G}_i \in \mathbb{C}^{r_i \times M_{R_i}}$ are the i^{th} user's precoding and decoding matrices.

The discrete input-output data model on the uplink, depicted in Fig. 1.2, is given by:

$$\boldsymbol{y} = \boldsymbol{D}\left(\boldsymbol{H}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{n}\right) \tag{2.17}$$

Similar to the downlink, the vectors $\boldsymbol{x} \in \mathbb{C}^{r\times 1}$ and $\boldsymbol{y} \in \mathbb{C}^{r\times 1}$ are the transmit and receive data vectors, and $\boldsymbol{n} \in \mathbb{C}^{M_T\times 1}$ is the vector of sampled additive white Gaussian noise at the input of the BS antenna array. The matrices $\boldsymbol{D}$ and $\boldsymbol{Q}$ are given by

$$\begin{aligned} \boldsymbol{D} &= \begin{bmatrix} \boldsymbol{D}_1{}^T & \cdots & \boldsymbol{D}_K{}^T \end{bmatrix}^T \in \mathbb{C}^{r\times M_T} \\ \boldsymbol{Q} &= \begin{bmatrix} \boldsymbol{Q}_1 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{Q}_K \end{bmatrix} \in \mathbb{C}^{M_R\times r} \end{aligned} \tag{2.18}$$

where $\boldsymbol{Q}_i \in \mathbb{C}^{M_{R_i}\times r_i}$ and $\boldsymbol{D}_i \in \mathbb{C}^{r_i\times M_T}$ are the i^{th} user's precoding and decoding matrices on the uplink.

Chapter 3

Capacity region of multi-user MIMO channels

In this chapter we focus on single-user and multi-user MIMO channel capacities in the Shannon theoretic sense. The Shannon capacity of a time-invariant channel is defined as the maximum mutual information between the channel input and output. This is the maximum data rate that can be transmitted over the channel with arbitrarily small error probability. When the CSI is perfectly known at both the transmitter and the receiver, the transmitter can adapt its transmission strategy relative to the instantaneous channel state. If the channel is time variant, the ergodic capacity is the maximum mutual information averaged over all channel states. The ergodic capacity is typically achieved using an adaptive transmission policy where the power and data rate vary relative to the channel state variations.

In a multiple user scenario, MU MIMO allows the reuse of time and frequency resources. Due to the scattering in different scenarios, the users' wavefronts may have large angle spreads and random signatures. Therefore, even users that are well separated in angle may have potentially overlapping subspaces spanned by left singular vectors of their channel matrices. Separability of their subspaces is much more difficult to achieve.

In a single-user MIMO system the link is point-to-point with a defined capacity. In a multi-user MIMO system, the link is a multiple access channel on the uplink and broadcast channel on the downlink. The achievable rates are characterized in this case in terms of a sum rate region. SU MIMO suffers only a small penalty in information rate without CSI at the transmitter. MU MIMO has a much larger penalty on the downlink. In a SU MIMO system, precoding at the transmitter and decoding at the receiver can be done with

full cooperation between the collocated antennas. In a MU MIMO system, the antennas can cooperate at the base station for precoding on the downlink and for decoding on the uplink. However, the users cannot cooperate in decoding on the downlink or during the precoding on the uplink. In a MU MIMO system, cooperation between the users may be possible in terms of power rates assigned to the users. In a SU MIMO system, the information rate is identical on the uplink and downlink for the same transmit power if the channel is known at the transmitter and the receiver.

3.1 Single-user MIMO capacity

When the channel is constant and known perfectly at the transmitter and the receiver, the capacity of the system defined by (2.13) is

$$C_{SU} = \max_{\boldsymbol{F}:\mathrm{Tr}(\boldsymbol{F}\boldsymbol{R}_x\boldsymbol{F}^H)\leq P_T} \log \frac{\det\left(\boldsymbol{R}_n + \boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_x\boldsymbol{F}^H\boldsymbol{H}^H\right)}{\det\left(\boldsymbol{R}_n\right)} \tag{3.1}$$

where $\boldsymbol{R}_x$ and $\boldsymbol{R}_n$ are the input data correlation matrix and the noise correlation matrix. P_T is the maximum transmit power.

It was shown in [2], that the optimum strategy to achieve maximum information rate is to convert the MIMO channel to parallel, non-interfering SISO channels through a singular value decomposition (SVD) of the channel matrix. The SVD yields $\min\left(M_R, M_T\right)$ parallel channels with gains corresponding to the singular values of $\boldsymbol{H}$.

Let us assume from now on that the elements of $\boldsymbol{x}$ and $\boldsymbol{n}$ in (2.13) are independent, identically distributed (i.i.d.) random variables and

$$E\left\{x_i\right\} = 0,\ \frac{1}{2}E\left\{\left|x_i\right|^2\right\} = 1 \tag{3.2}$$

$$E\left\{n_i\right\} = 0,\ \frac{1}{2}E\left\{\left|n_i\right|^2\right\} = \sigma_n^2 \tag{3.3}$$

If the SVD of the channel matrix is $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, the channel is decomposed into a set of parallel subchannels by choosing $\boldsymbol{F}$ as

$$\boldsymbol{F} = \boldsymbol{M}\boldsymbol{\Phi} \tag{3.4}$$

where $\boldsymbol{M} \in \mathbb{C}^{M_T \times M_T}$ is a unitary matrix and $\boldsymbol{M} = \boldsymbol{V}$, [2], [9]. The matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M_T \times r}$ is a non-negative power loading matrix. The optimal solution for the power loading matrix

$\boldsymbol{\Phi}$ is found via the water-pouring algorithm

$$\varphi_{i,i}^2 = \left(\mu - \frac{1}{\sigma_i^2}\right)_+, \; i = 1, \ldots, r \tag{3.5}$$

such that

$$\sum_{i=1}^{r} \varphi_{i,i}^2 = P_T$$

where μ is a constant, $r \leq \min(M_R, M_T)$ is the rank of the channel matrix and σ_i is the i^{th} singular value of $\boldsymbol{H}$. The values of $\varphi_{i,i}$ are calculated using an iterative algorithm [9]. The important point to note is that the modal decomposition achieves the maximum information rate when CSI is available at the transmitter and the receiver.

If $\boldsymbol{H}$ is random, the channel capacity is a random variable too, and can vary from zero to infinity. The statistics of the channel capacity are captured by its cumulative distribution function (CDF). The $X\%$ outage capacity is the rate that the channel can support with $(100 - X)\,\%$ probability. If we use very large block (packet) size, and capacity achieving codes, the block error probability (BLER) will be always binary. The block is always decoded successfully if the rate is at or below actual instantaneous capacity, and is always in error if the rate exceeds the instantaneous capacity. Therefore, if the transmitter does not know the CSI, the BLER will equal the outage probability for that signaling rate, i.e. outage capacity.

If the exact CSI is not known at the transmitter, the information rate maximization can now be performed only in terms of the outage or ergodic capacity. The ergodic capacity of a MIMO channel (2.13) is given by

$$C_{SU} = \max_{\boldsymbol{F}:\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^H) \leq P_T} E\left\{\log\det\left(\boldsymbol{I}_{M_R} + \sigma_n^{-2}\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{H}^H\right)\right\} \tag{3.6}$$

It has been shown that the maximum information rate can be achieved by Gaussian signals aligned along the eigenvectors of the correlation matrix $\boldsymbol{R}_t = E\left\{\boldsymbol{H}^H\boldsymbol{H}\right\}$, i.e., $\boldsymbol{M} = \boldsymbol{V}$ where $\boldsymbol{R}_t = \boldsymbol{V}^H\boldsymbol{\Lambda}\boldsymbol{V}$. The capacity achieving power allocation is harder to compute. It was shown that depending on the transmit correlation matrix $\boldsymbol{R}_t$, there is a range of signal-to-noise ratios (SNRs) for which the optimal strategy is to direct all the power in only the dominant eigenmode of $\boldsymbol{R}_t$, [9].

3.2 Capacity region of MAC channel

The union of achievable rates under all transmission strategies is called the capacity region af the multi-user system. It defines the limit of error-free communications given certain channel characteristics and it is used as the ultimate measure of channel capacity.

Let us denote the rate that can be reliably, i.e., error free, transmitted for the i^{th} user by R_i in bits per second per Hertz (bps/Hz) and assume Gaussian signaling for each user. We consider joint decoding of the users' signals. Joint decoding means that decoding of all signals is performed simultaneously. The MU MAC capacity region with joint decoding and with individual power constraints $P_1, \ldots, P_K$ on each user has been shown to satisfy [36], [37]

$$\begin{aligned} \sum_{i=1}^{K} R_i \leq & \max_{\mathrm{tr}(\boldsymbol{Q}_i \boldsymbol{Q}_i^H) \leq P_i} \log\det\left(\boldsymbol{I}_{M_T} + \sigma_n^{-2} \boldsymbol{H}^T \boldsymbol{Q} \boldsymbol{Q}^H \boldsymbol{H}^*\right) \\ = & \max_{\mathrm{tr}(\boldsymbol{Q}_i \boldsymbol{Q}_i^H) \leq P_i} \log\det\left(\boldsymbol{I}_{M_T} + \sigma_n^{-2} \sum_{i=1}^{K} \boldsymbol{H}_i^T \boldsymbol{Q}_i \boldsymbol{Q}_i^H \boldsymbol{H}_i^*\right) \end{aligned} \tag{3.7}$$

While maximum likelihood (ML) decoding is optimal, the MU MAC sum-rate capacity can also be achieved via an MMSE receiver with successive interference cancellation (SIC). This can be seen if we rewrite the equation (3.7) as

$$\begin{aligned} & \log\det\left(\boldsymbol{I}_{M_T} + \sigma_n^{-2} \boldsymbol{H}_1^T \boldsymbol{Q}_1 \boldsymbol{Q}_1^H \boldsymbol{H}_1^* + \sigma_n^{-2} \sum_{i=2}^{K} \boldsymbol{H}_i^T \boldsymbol{Q}_i \boldsymbol{Q}_i^H \boldsymbol{H}_i^*\right) \\ = & \log\det\left(\boldsymbol{I}_{M_T} + \sigma_n^{-2} \sum_{i=2}^{K} \boldsymbol{H}_i^T \boldsymbol{Q}_i \boldsymbol{Q}_i^H \boldsymbol{H}_i^*\right) + \\ & \log\det\left(\boldsymbol{I}_{M_T} + \left(\sigma_n^2 \boldsymbol{I}_{M_T} + \sum_{i=2}^{K} \boldsymbol{H}_i^T \boldsymbol{Q}_i \boldsymbol{Q}_i^H \boldsymbol{H}_i^*\right)^{-1} \boldsymbol{H}_1^T \boldsymbol{Q}_1 \boldsymbol{Q}_1^H \boldsymbol{H}_1^*\right) \end{aligned} \tag{3.8}$$

The objective function in (3.7) is a convex function of the uplink precoding matrices $\boldsymbol{Q}_i$ and the constraints are separable because there is an individual trace constraint on each correlation matrix $\boldsymbol{Q}_i \boldsymbol{Q}_i^H$. In such situations, it is generally sufficient to optimize with respect to the first variable while holding all other variables constant, then optimize with respect to the second variable, etc., in order to reach a globally optimum point. This is referred to as the block-coordinate ascent algorithm and convergence can be shown under relatively general conditions [24].

Successive interference cancellation means that users are decoded sequentially, and that the user to be decoded treats all the other users to be decoded as interference, and subtracts out the symbols transmitted by the users already decoded from the received codeword. Since each user is transmitting at an arbitrarily small bit error rate, the users

already decoded can be subtracted without introducing additional errors.

The optimal iterative water pouring algorithm was first proposed in [10]. At each step of the algorithm, one user optimizes his precoding matrix while treating the signals from all other users as interference. In the next step, the next user in numerical order optimizes his precoding matrix while treating all other users, including the updated precoding matrix of the previous user, as interference. The optimal ordering of users is independent of the channel state [38].

3.3 Capacity region of BC channel

The MU MIMO downlink channel in general belongs to the class of non-degraded Gaussian channels. The sum-rate capacity of a Gaussian broadcast channel, for multiple-users each having multiple antennas, has been shown to satisfy [13]

$$C_{BC} = \min_{\boldsymbol{R}_n>0,[\boldsymbol{R}_n]_{k,k}=\sigma_n^2} \left(\max_{\mathrm{tr}(\boldsymbol{F}\boldsymbol{F}^H)\leq P_T} \log \frac{\det\left(\boldsymbol{R}_n + \boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{H}^H\right)}{\det\left(\boldsymbol{R}_n\right)} \right) \tag{3.9}$$

This is Sato's upper bound on the capacity region of general broadcast channels, which is the capacity of a system where the users in the downlink can cooperate. The Sato bound is not tight in general, but by introducing noise correlation at the different receivers, we can get a much stronger bound [39], [40].

The downlink problem at the BS is to broadcast the user signals with appropriate processing and spatial weighting, such that each user receives a maximum or desired signal-to-interference and noise ratio (SINR), information rate or BER. Antennas at the base station can cooperate during the encoding phase. Cooperation between the users might either entail cooperative management of the rates or SINR at each user.

The capacity region of the general non-degraded broadcast channels is unknown. However, in [11] it was shown that Costa's "dirty-paper" coding is optimal in achieving the sum-rate capacity, by demonstrating that the achievable rate meets the Sato upper bound. The basic premise of DPCs is that if the transmitter has perfect, non-causal knowledge of additive Gaussian interference in the channel, then the capacity of the channel is the same as if there was no additive interference. DPC allow non-causally interference to be "pre-subtracted" at the transmitter, but in such a way that the transmit power is not increased.

Let $\pi(\cdot)$ denote a permutation of the user indices and $\left(\boldsymbol{F}_k\boldsymbol{F}_k^H\right), k = 1, \ldots, K$, is a set

of positive semi-definite correlation matrices with $\text{tr}\left(\sum_{k=1}^{K} \boldsymbol{F}_k \boldsymbol{F}_k^H\right) \leq P_T$, where P_T is maximum total transmit power. Under DPC, if the $\pi(1)^{th}$ user signal is encoded first, followed by the $\pi(2)^{th}$ user, etc., then the following rate is achievable:

$$R_{\pi(i)} = \log \frac{\det\left(\boldsymbol{I}_{M_R} + \sigma_n^{-2} \boldsymbol{H}_{\pi(i)} \left(\sum_{k \geq i} \boldsymbol{F}_{\pi(k)} \boldsymbol{F}_{\pi(k)}^H\right) \boldsymbol{H}_{\pi(i)}^H\right)}{\det\left(\boldsymbol{I}_{M_R} + \sigma_n^{-2} \boldsymbol{H}_{\pi(i)} \left(\sum_{k > i} \boldsymbol{F}_{\pi(k)} \boldsymbol{F}_{\pi(k)}^H\right) \boldsymbol{H}_{\pi(i)}^H\right)}, \; i = 1, \ldots, K \tag{3.10}$$

The capacity region is the convex hull of the union of all such rates over all permutations and all positive semi-definite correlation matrices satisfying the sum power constraint:

$$C_{DPC}\left(P_T, \boldsymbol{H}\right) = \max_{\sum_{k=1}^{K} \text{tr}\left(\boldsymbol{F}_{\pi(k)} \boldsymbol{F}_{\pi(k)}^H\right) \leq P_T} \sum_{i=1}^{K} R_{\pi(i)} \tag{3.11}$$

where $R_{\pi(i)}$ is given in the previous equation. The DPC implies that the user signals are uncorrelated.

3.3.1 Achievable sum rate BC capacity and UL/DL duality

It is easily seen that the objective function for the DPC sum-rate capacity region is not a convex function of the correlation matrices. Thus, numerically finding the maximum is not a trivial problem and requires a brute force search over the entire space of correlation matrices that meet the power constraint. However, by establishing the duality between the uplink and the downlink, it was shown that it is possible to obtain the maximum achievable sum-rate capacity of the broadcast channel from the dual uplink channel [37].

The channel capacity is different for the uplink and the downlink due to the fundamental differences between these channels. However, the fact that the downlink and the uplink channels look like mirror images of each other implies that there is a duality between these channels that allows the capacity region of either channel to be obtained from the capacity region of the other.

The equivalence between the performance of receive and transmit strategies when the roles of transmitters and receivers are reversed for vector Gaussian channels has been observed in many different situations. In a point-to-point communication, the capacity is unchanged when the role of transmitters and receivers is interchanged. In case of downlink linear processing followed by SU receivers at the UTs, the choice of transmit and receive matrices is closely related to a virtual uplink problem. Finally, the capacity region of degraded Gaussian channels is the same as the capacity region of the corresponding MAC

with the transmit power constraint of the BC translated to the sum of powers in the MAC [41], [15].

The difference between the uplink and the downlink channel is that on the downlink there is an additive noise term associated with each user terminal, while on the uplink there is only one. Another important difference is that on the downlink there is a single power constraint associated with the transmitter, whereas on the uplink there is a different power constraint associated with each user. Finally, on the downlink both the signal and interference associated with each user travel through the same channel, whereas on the uplink these signals travel through the different channels.

We say that the downlink and uplink channels are duals of each other if the channel impulse responses for each user are the same in the downlink and the uplink, each receiver in the downlink has the same noise statistics and these statistics are the same as those of the receiver noise in the uplink, and the power constraint P_T on the downlink equals the sum of individual power constraints $P_{T_k}, k = 1, \ldots, K$ on the uplink. The set of BC correlation matrices $\boldsymbol{F}_k \boldsymbol{F}_k^H,\ k = 1, \ldots, K$, is found using the duality principle from the dual MAC channel that use the same sum power constraint [41]. We assume that in the uplink the first user is decoded first, then the second, etc. In the downlink we assume that the users are precoded in the reverse order, i.e, the K^{th} user is precoded first, then the $(K-1)^{th}$, etc. Then the rate achieved by the k^{th} user in the uplink is given by

$$R_k^{\mathrm{UL}} = \log\det\left(\boldsymbol{I}_{M_T} + \left(\sigma_n^2 \boldsymbol{I}_{M_T} + \sum_{i>k} \boldsymbol{H}_i^T \boldsymbol{Q}_i \boldsymbol{Q}_i^H \boldsymbol{H}_i^*\right)^{-1} \boldsymbol{H}_k^T \boldsymbol{Q}_k \boldsymbol{Q}_k^H \boldsymbol{H}_k^*\right) \tag{3.12}$$

and in the downlink by

$$R_k^{\mathrm{DL}} = \log\det\left(\boldsymbol{I}_{M_{R_k}} + \left(\sigma_n^2 \boldsymbol{I}_{M_{R_k}} + \boldsymbol{H}_k \left(\sum_{i<k} \boldsymbol{F}_i \boldsymbol{F}_i^H\right) \boldsymbol{H}_k^H\right)^{-1} \boldsymbol{H}_k \boldsymbol{F}_k \boldsymbol{F}_k^H \boldsymbol{H}_k^H\right) \tag{3.13}$$

Let us introduce the following auxiliary matrices [41]

$$\boldsymbol{A}_k = \left(\sigma_n^2 \boldsymbol{I}_{M_{R_k}} + \boldsymbol{H}_k \left(\textstyle\sum_{i<k} \boldsymbol{F}_i \boldsymbol{F}_i^H\right) \boldsymbol{H}_i^H\right)$$
$$\boldsymbol{B}_k = \left(\sigma_n^2 \boldsymbol{I}_{M_T} + \textstyle\sum_{i>k} \boldsymbol{H}_i^H \boldsymbol{Q}_i^* \boldsymbol{Q}_i^T \boldsymbol{H}_i\right)$$

The equation (3.13) can be now rewritten as

$$\begin{aligned} R_k^{\mathrm{DL}} &= \log\det\left(\boldsymbol{I}_{M_{R_k}} + \boldsymbol{A}_k^{-1}\boldsymbol{H}_k\boldsymbol{F}_k\boldsymbol{F}_k^H\boldsymbol{H}_k^H\right) \\ &= \log\det\left(\boldsymbol{I}_{M_{R_k}} + \boldsymbol{A}_k^{-1/2}\boldsymbol{H}_k\boldsymbol{F}_k\boldsymbol{F}_k^H\boldsymbol{H}_k^H\boldsymbol{A}_k^{-1/2}\right) \\ &= \log\det\left(\boldsymbol{I}_{M_{R_k}} + \boldsymbol{A}_k^{-1/2}\boldsymbol{H}_k\boldsymbol{B}_k^{-1/2}\boldsymbol{B}_k^{1/2}\boldsymbol{F}_k\boldsymbol{F}_k^H\boldsymbol{B}_k^{1/2}\boldsymbol{B}_k^{-1/2}\boldsymbol{H}_k^H\boldsymbol{A}_k^{-1/2}\right) \end{aligned} \tag{3.14}$$

and equation (3.12) using the property $\det(\boldsymbol{X}) = \det\left(\boldsymbol{X}^T\right)$ as

$$\begin{aligned} R_k^{\mathrm{UL}} &= \log\det\left(\boldsymbol{I}_{M_T} + \boldsymbol{B}_k^{-1}\boldsymbol{H}_k^H\boldsymbol{Q}_k^*\boldsymbol{Q}_k^T\boldsymbol{H}_k\right) \\ &= \log\det\left(\boldsymbol{I}_{M_T} + \boldsymbol{B}_k^{-1/2}\boldsymbol{H}_k^H\boldsymbol{Q}_k^*\boldsymbol{Q}_k^T\boldsymbol{H}_k\boldsymbol{B}_k^{-1/2}\right) \\ &= \log\det\left(\boldsymbol{I}_{M_T} + \boldsymbol{B}_k^{-1/2}\boldsymbol{H}_k^H\boldsymbol{A}_k^{-1/2}\boldsymbol{A}_k^{1/2}\boldsymbol{Q}_k^*\boldsymbol{Q}_k^T\boldsymbol{A}_k^{1/2}\boldsymbol{A}_k^{-1/2}\boldsymbol{H}_k\boldsymbol{B}_k^{-1/2}\right) \end{aligned} \tag{3.15}$$

The definition of the matrix square root is the same as in Section 2.2.1.

Treating $\boldsymbol{A}_k^{-1/2}\boldsymbol{H}_k\boldsymbol{B}_k^{-1/2}$ as the effective channel of the system, we note that when we take the Hermitian of this channel we have the effective channel of the uplink channel $\boldsymbol{B}_k^{-1/2}\boldsymbol{H}_k^H\boldsymbol{A}_k^{-1/2}$. This suggests that in this case we can use the same logic as in case of point-to-point system where the capacity on the uplink and the downlink channel are the same under the previously given conditions, i.e., we can write that $R_k^{\mathrm{DL}} = R_k^{\mathrm{UL}}$. Therefore, we can now use the same transformation of correlation matrices as for the point-to-point system in order to transform the MAC channel correlation matrices into BC correlation matrices. Let us define the SVD of the effective channel as $\boldsymbol{A}_k^{-1/2}\boldsymbol{H}_k\boldsymbol{B}_k^{-1/2} = \boldsymbol{U}_{e_k}\boldsymbol{\Sigma}_{e_k}\boldsymbol{V}_{e_k}^H$, then it follows [41]

$$\boldsymbol{F}_k\boldsymbol{F}_k^H = \boldsymbol{B}_k^{-1/2}\boldsymbol{V}_{e_k}\boldsymbol{U}_{e_k}^H\boldsymbol{A}_k^{1/2}\boldsymbol{Q}_k^*\boldsymbol{Q}_k^T\boldsymbol{A}_k^{1/2}\boldsymbol{U}_{e_k}\boldsymbol{V}_{e_k}^H\boldsymbol{B}_k^{-1/2} \tag{3.16}$$

Thus, the achievable sum-rate capacity of BC is equal to the sum-rate capacity of the dual MAC, i.e.

$$C_{\mathrm{MAC}} = \max_{\sum_i \mathrm{tr}\left(\boldsymbol{Q}_i\boldsymbol{Q}_i^H\right)\leq P_T} \log\det\left(\boldsymbol{I}_{M_T} + \sigma_n^{-2}\sum_{i=1}^{K}\boldsymbol{H}_i^T\boldsymbol{Q}_i\boldsymbol{Q}_i^H\boldsymbol{H}_i^*\right) \tag{3.17}$$

where the optimization is performed over uplink correlation matrices $\boldsymbol{Q}_k\boldsymbol{Q}_k^H, k = 1,\ldots,K$ subject to the same sum power constraint P_T. This allows us to substitute non-convex rate functions of the user correlation resulting from the BC region with the dual MAC where the rates are convex functions of the covariance matrices. Using the transformation given in (3.16) we map the transmit uplink correlation matrices $\boldsymbol{Q}_k\boldsymbol{Q}_k^H, k = 1,\ldots,K$ to

the downlink transmit correlation matrices $\boldsymbol{F}_k\boldsymbol{F}_k^H, k = 1, \ldots, K$ that achieve the same rates under the same sum power constraint.

An iterative water-pouring algorithm was proposed in [24] to compute the correlation matrices of a dual MAC channel. This algorithm is based on the iterative water-pouring algorithm for the conventional MAC problem, which finds the sum-rate capacity of MAC with individual power constraints on each user [42]. The difference from (3.17) is only in the structure of the power constraint. In such situations, the optimization is performed using a block-coordinate ascent algorithm, i.e., by optimizing with respect to the first variable while holding all other variables constant, then optimize with respect to the second variable, etc., in order to reach a globally optimum point. In other words, at each step of the algorithm one user optimizes his correlation matrix while treating the signals from all other users as noise including the users with previously updated correlation matrices. In case of the dual MAC there is a sum power constraint, i.e., the water level of all users must be equal. Unlike in the conventional MAC, with sum power constraint we must update all correlation matrices simultaneously to maintain a constant water-level.

This algorithm is complex and requires numerous calculations of SVDs and the water-pouring algorithm. By using the Hadamard inequality we find an upper bound which is not always as tight as the previous one but which requires less computational effort.

The equation (3.17) can be rewritten as

$$C_{\text{MAC}} = \max_{\text{tr}\left(\boldsymbol{Q}_i\boldsymbol{Q}_i^H\right)\geq 0, \sum_i \text{tr}\left(\boldsymbol{Q}_i\boldsymbol{Q}_i^H\right)\leq P_T} \log\det\left(\boldsymbol{I}_r + \sigma_n^{-2}\boldsymbol{Q}^H\boldsymbol{H}^*\boldsymbol{H}^T\boldsymbol{Q}\right) \tag{3.18}$$

The expression in equation (3.18) can be written in a block matrix form as

$$\boldsymbol{Q}^H\boldsymbol{H}^*\boldsymbol{H}^T\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_1^H\boldsymbol{H}_1^*\boldsymbol{H}_1^T\boldsymbol{Q}_1 & \cdots & \boldsymbol{Q}_1^H\boldsymbol{H}_1^*\boldsymbol{H}_K^T\boldsymbol{Q}_K \\ \vdots & \ddots & \vdots \\ \boldsymbol{Q}_K^H\boldsymbol{H}_K^*\boldsymbol{H}_1^T\boldsymbol{Q}_1 & \cdots & \boldsymbol{Q}_K^H\boldsymbol{H}_K^*\boldsymbol{H}_K^T\boldsymbol{Q}_K \end{bmatrix} \tag{3.19}$$

Using the Hadamard inequality $\det(\boldsymbol{A}) \leq \prod_i a_{i,i}$, where $a_{i,i}$ are the diagonal elements of

$\boldsymbol{A}$, we can write

$$\begin{aligned} &\log\det\left(\boldsymbol{I}_r + \sigma_n^{-2}\boldsymbol{Q}^H\boldsymbol{H}^*\boldsymbol{H}^T\boldsymbol{Q}\right) \\ &\leq \log\prod_{k=1}^{K}\det\left(\boldsymbol{I}_{r_k} + \sigma_n^{-2}\boldsymbol{Q}_k^H\boldsymbol{H}_k^*\boldsymbol{H}_k^T\boldsymbol{Q}_k\right) \\ &= \sum_{k=1}^{K}\log\det\left(\boldsymbol{I}_{r_k} + \sigma_n^{-2}\boldsymbol{Q}_k^H\boldsymbol{H}_k^*\boldsymbol{H}_k^T\boldsymbol{Q}_k\right) \end{aligned} \tag{3.20}$$

The equality holds for $\boldsymbol{Q}_k = \boldsymbol{U}_k^*\boldsymbol{\Phi}_k$, where the columns of $\boldsymbol{U}_k$ are the basis of the column space of $\boldsymbol{H}_k = \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^H$, and $\boldsymbol{\Phi}_k$ is diagonal power loading matrix.

The comparison of the DPC sum rate capacity bound of the BC and the previously introduced very simple (VS) BC sum rate capacity bound is shown in Figures 3.1 and 3.2. The first VS BC sum rate capacity bound is obtained using the equation (3.20) which corresponds to the capacity of the MU MIMO channel where all users are orthogonal in space. However, the influence of MUI which is neglected is too big and as a result this bound is too loose. The other option is to substitute the precoding matrices $\boldsymbol{Q}_k$ obtained by maximizing (3.20) in the expression for dual MAC in (3.18). As it can be seen from Figure 3.1, in case of low MUI, i.e., when the total number of antennas at the user terminals is less or equal to the number of antennas at the base station, the second approximate VS bound, when the interference between the users is also taken into account when calculating the system capacity, is the same as the DPC bound. In case of high MUI, i.e., when the number of antennas at the user terminals is greater than the number of antennas at the base station, the approximate VS BC sum rate capacity bound is very close to the DPC bound and at high SNRs it matches the DPC bound. The antenna configuration of the system in Figures 3.1 and 3.2 is: at the base station we have 6 antennas, and there are three users in the system. In the first figure all users are equipped with 2 antennas and in the second figure all users are equipped with 4 antennas each.

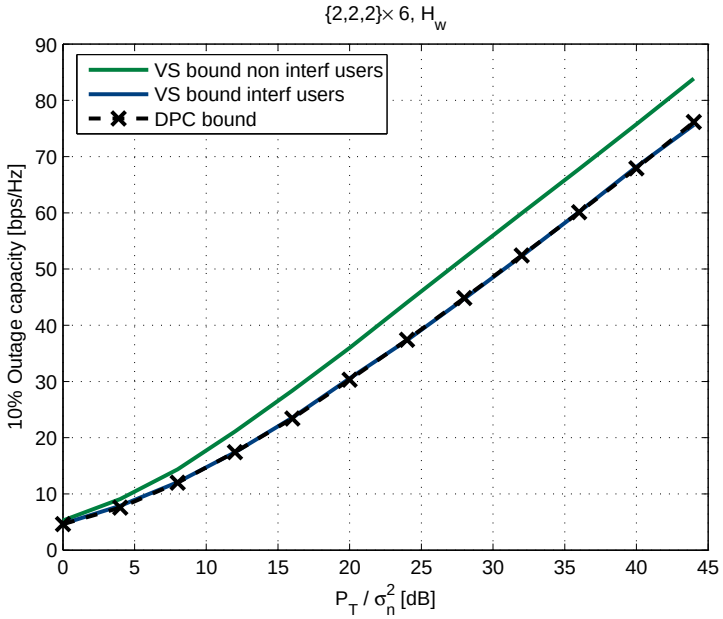

Figure 3.1: Broadcast channel upper bounds. 10 % Outage capacity. $M_R \leq M_T$ case.

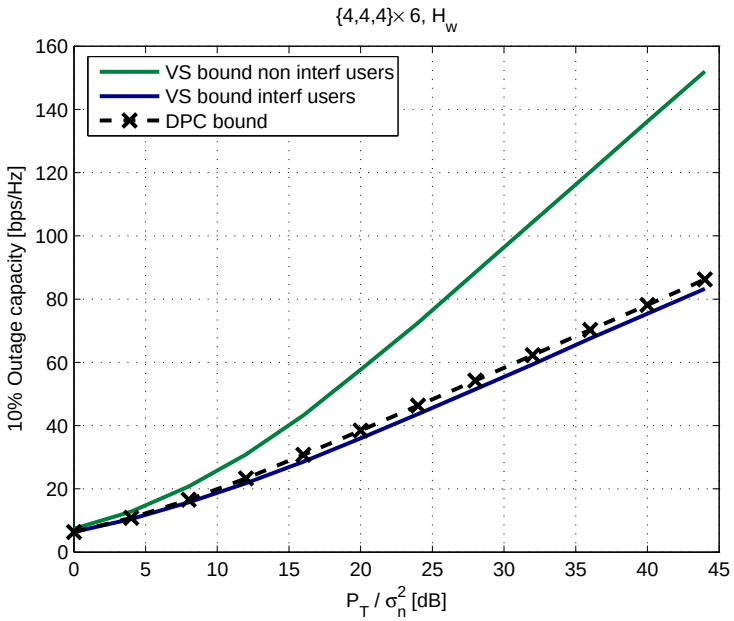

Figure 3.2: Broadcast channel upper bounds. 10 % Outage capacity. $M_R > M_T$ case.

Chapter 4

Single-user MIMO optimum precoder and decoder

The MIMO capacity gain in the single-user case is roughly $\min(M_T, M_R)$ times that of single-input single-output systems, where M_T is the number of transmit antennas and M_R is the number of receive antennas. Multiple antennas at the transmitter and receiver can provide different gains.

The diversity gain manifests itself in increasing the magnitude of the slope of the symbol error rate (SER) curve. On the other hand, the coding gain shifts the error rate curve to the left. Diversity provides the receiver with multiple (ideally independent) looks at the same transmitted signal. Each look constitutes a diversity branch. If the $M_T M_R$ links comprising the MIMO channel fade independently and the transmitted signal is suitably constructed, the receiver can combine the arriving signals such that the resultant signal exhibits considerably reduced amplitude variability in comparison to a SISO link and we get $M_T M_R{}^{th}$ order diversity. Extracting the spatial diversity gain in the absence of channel knowledge at the transmitter is possible using suitably designed transmit signals.

The array gain is the average increase in signal power at the receiver or transmitter or both. The array gain is proportional to the number of receive antennas in single-input multiple-output (SIMO) system, and to the number of transmit antennas in multiple-input single-output (MISO) system. The array gain in MIMO system is a function of the dominant eigenmode of the channel.

A system is said to have array gain A and diversity order D if the average symbol error rate is asymptotically

$$\frac{1}{A}\mathrm{SNR}^{-D}$$

The array gain shifts the symbol error rate curve, while the diversity gain determines its slope.

The spatial multiplexing gain is the increase in the transmission rate. It is proportional to the $\min(M_T, M_R)$. This gain is realized by transmitting independent data signals from the individual antennas or over singular vectors.

The spatial interference suppression gain exploits the difference between the spatial signatures of the desired signal and co-channel signals to reduce interference. Interference reduction requires knowledge of the desired signal's channel. Exact knowledge of the interferer's channel may not be necessary. Interference reduction allows aggressive frequency reuse and thereby increases multi-cell capacity.

What kind of a gain and how much of a certain gain can be exploited depends largely on the quality of the MIMO channel state information and/or channel quality information (CQI) available at the transmitter and receiver and the type of processing techniques used to precode and decode users' data. Regardless of the user terminal speed, we can assume that the instantaneous (or short-term) channel state information is always available at the receiver. The channel state information at the transmitter may be, however, hard to obtain. Type and quality of channel state information at the transmitter depends on various factors, like the speed of the user terminal, type of duplexing, feedback/feedforward overhead, etc.

One possible classification of spatial processing techniques could be done according to the availability of the channel state and/or quality information at the transmitter. We can classify all MIMO techniques as those that require some type of the channel knowledge at the transmitter and those that do not require any channel knowledge at the transmitter.

In this chapter we will give an overview of the most representative techniques and type of gains they have in single user, point-to-point (P2P), communication systems.

4.1 MIMO processing without CSI at the transmitter

In case no channel state information is available at the transmitter we extract some of the gains offered by MIMO processing by appropriately coding data symbols over space, time and/or frequency. Vertical Bell Labs Layered SpaceTime (V-BLAST) [6], where every antenna transmits its own independent substream of data, has been shown to have a good performance and simple encoding and decoding. Yet V-BLAST suffers from its inability to work with fewer receive antennas than transmit antennas. This deficiency is especially

important for modern cellular systems, where a base station typically has more antennas than the user terminals. Furthermore, because V-BLAST transmits independent data streams on its antennas it is very sensitive to deep fades from any given transmit antenna. On the other hand, there are many previously proposed STCs that have good fading resistance and simple decoding, but these codes generally have poor performance at high data rates or with many antennas. A framework for generalized designs of STBCs is given in [5] under the name linear dispersion codes (LDC). LDCs can handle any configuration of transmit and receive antennas and can reduce to both V-BLAST and many proposed space-time block codes as special cases. Frame based design of LDCs was introduced in [43].

4.1.1 Orthogonal space-time block codes

The objective of space-time block coding (STBC) is to extract the total available spatial diversity in the MIMO channel through appropriate construction of the transmitted space-time codewords. As example we consider a specific diversity coding technique, the Alamouti scheme [4].

Consider a MIMO system with 2 transmit antennas and M_R receive antennas. The Alamouti transmission technique is as follows: two different data symbols x_1 and x_2 are transmitted simultaneously from antennas 1 and 2, respectively, during the first symbol period, followed by symbols $-x_2^*$ and x_1^* that are launched from antennas 1 and 2, respectively. Note that spatial rate for the Alamouti scheme is equal to 1 since two independent data symbols are transmitted over two symbol periods.

We assume that the channel is independent identically distributed (i.i.d.) frequency-flat fading and remains constant over (at least) two consecutive symbol periods. Appropriate processing at the receiver reduces the vector channel into a scalar channel for either of the transmitted data symbols such that

$$y_i = \|\boldsymbol{H}\|_{\mathcal{F}}^2 \, x_i + \tilde{n}_i \tag{4.1}$$

where y_i is the received signal corresponding to transmitted symbol x_i and $\tilde{n}_i$ is a complex random Gaussian distributed additive noise, $\tilde{n}_i \sim \mathcal{CN}\Big(0, \|\boldsymbol{H}\|_{\mathcal{F}}^2 \, \sigma_n^2\Big)$.

Even though channel knowledge is not available to the transmitter, the Alamouti scheme extracts $2M_R{}^{th}$ order diversity. We note, however, that array gain is realized only at the receiver. The Alamouti scheme may be extended to channels with more than two

transmit antennas through orthogonal space-time block coding (OSTBC) [44], albeit at a loss in spatial rate, which is then less than 1. However, the low decoding complexity of OSTBC renders this technique highly attractive for practical applications. Another STC for more than 2 antennas at the transmitter was proposed in [45]. It results in a rate 1 orthogonal code but with reduced diversity order. The ABBA code [46] provides full diversity with rate 1 but at the price of detriment of the orthogonality of the code.

4.1.2 Open loop spatial multiplexing

Spatial multiplexing (SMUX) techniques transmit simultaneously different data streams from different antennas in order to increase the capacity of the system. Any receiver technique such as zero-forcing (ZF), minimum mean square error estimation and optimal maximum-likelihood detection (MLD) can be applied directly. MLD gives the best performance with a high computational complexity. ZF and MMSE are less expensive techniques. In case of ZF, complexity reduction comes, however, at the expense of noise enhancement which in general results in a significant performance degradation (compared to the ML decoder). The diversity order achieved by each of the individual data streams equals $M_T - M_R + 1$. The MMSE receiver balances interstream interference mitigation with noise enhancement and minimizes the total error. If we assume that the elements of $\boldsymbol{x}$ and $\boldsymbol{n}$ in (2.13) are i.i.d. random variables and

$$E\{x_k\} = 0,\ \frac{1}{2}E\left\{|x_k|^2\right\} = P_t \tag{4.2}$$

$$E\{n_k\} = 0,\ \frac{1}{2}E\left\{|n_k|^2\right\} = \sigma_n^2 \tag{4.3}$$

then the MMSE receive filter is given by

$$\boldsymbol{G} = \boldsymbol{H}^H \left(\boldsymbol{H}\boldsymbol{H}^H + \frac{\sigma_n^2}{P_t}\boldsymbol{I}_{M_R}\right)^{-1} \tag{4.4}$$

At low SNRs the MMSE receiver approximates a matched filter and is near optimal. It outperforms the ZF receiver that continues to enhance noise. At high SNRs, the MMSE receiver approaches ZF and therefore realizes the same diversity order as ZF for each data stream [9].

A higher diversity order is achieved by successive MMSE detection and interference cancellation as in V-BLAST. The key idea in a successive interference cancellation (SIC) receiver is layer peeling where the individual data streams are successively decoded and

stripped away from the received data vector layer-by-layer by selecting the stream with the highest signal-to-interference and noise ratio (SINR) at each decoding stage. Upon detection of the chosen symbol, its contribution from the received data vector is subtracted and the procedure is repeated until all symbols are detected.

4.2 Optimum design of linear precoder and decoder

When channel state information is available at both transmitter and receiver sides, channel dependent precoding and detection of data streams improves the system performance. Channel state information can be acquired at the transmitter either if a feedback channel is present or when the transmitter and receiver operate in time division duplex (TDD) so that the time-invariant MIMO channel transfer function is the same in both ways.

The precoder is a matrix with complex elements and can add redundancy to the input symbol streams to improve system performance. The precoder output is launched into the MIMO channel through M_T transmit antennas. The signal is received by M_R receive antennas and processed by the linear decoder, which is optimized for the fixed and known channel. The linear decoder also operates in the complex domain and removes any redundancy that has been introduced by the precoder.

Generalized designs of a jointly optimum linear precoder and decoder for a single user (SU) MIMO system, using a mean-squared error (MSE) criterion are given in [7] and [8]. The framework presented in these papers is general and addresses several optimization criteria like minimum MSE (MMSE), minimum bit error rate (BER), and maximum information rate.

The paradigm of linear precoding and decoding exploits the channel eigenmode decomposition in constructing the optimal precoder $\boldsymbol{F}$ and decoder $\boldsymbol{G}$. The different solutions are characterized by how the power is loaded on each channel eigenmode.

Let us start from the input-output signal model given in (2.13)

$$\boldsymbol{y} = \boldsymbol{G}\left(\boldsymbol{H}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{n}\right).$$

The linear precoding and decoding matrices are designed by minimizing a function of the MSE matrix that is given by [8]

$$\mathbf{MSE}\left(\boldsymbol{F}, \boldsymbol{G}\right) = \mathrm{E}\left\{\left(\boldsymbol{y} - \boldsymbol{x}\right)\left(\boldsymbol{y} - \boldsymbol{x}\right)^{H}\right\} \tag{4.5}$$

The optimum decoding matrix $\boldsymbol{G}_{opt}$ that minimizes the whole $\mathbf{MSE}\left(\boldsymbol{F},\boldsymbol{G}\right)$ is the MMSE (Wiener) receiver [47] which is known to minimize the $\mathrm{tr}\left(\mathbf{MSE}\left(\boldsymbol{F},\boldsymbol{G}\right)\right)$ and is given assuming (4.2) and (4.3) by

$$\boldsymbol{G}_{opt} = \boldsymbol{F}^H\boldsymbol{H}^H\left(\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{H}^H + \frac{\sigma_n^2}{P_t}\boldsymbol{I}_{M_R}\right)^{-1} \tag{4.6}$$

If we substitute $\boldsymbol{G}$ in equation (4.5) with $\boldsymbol{G}_{opt}$ from equation (4.6) we have

$$\mathbf{MSE}(\boldsymbol{F}) = \sigma_n^2\left(\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{H}^H + \frac{\sigma_n^2}{P_t}\boldsymbol{I}_{M_R}\right)^{-1} \tag{4.7}$$

Let us introduce the following singular value decomposition

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H \tag{4.8}$$

The column vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$ are the left and right singular vectors, respectively. The matrix $\boldsymbol{\Sigma}$ is a diagonal matrix with non-negative singular values $\sigma_k, k \leq \min(M_T, M_R)$, on the main diagonal arranged in decreasing order.

The optimum precoding matrix $\boldsymbol{F}$ under the transmit power constraint

$$P_t \cdot \mathrm{tr}\left(\boldsymbol{F}_{opt}\boldsymbol{F}_{opt}^H\right) \leq P_T, \tag{4.9}$$

is given by (4.3)

$$\boldsymbol{F}_{opt} = \boldsymbol{V}\boldsymbol{\Phi}, \tag{4.10}$$

where $\boldsymbol{\Phi}$ is an $M_T \times M_T$ diagonal power loading matrix, P_t is the power of one complex data symbol x_k and P_T is the maximum transmit power. The values on the main diagonal of $\boldsymbol{\Phi}$ will depend on the specific optimization criterion.

The MMSE precoder is obtained as a solution to the optimization problem

$$\boldsymbol{F}_{opt} = \arg\min_{\boldsymbol{F}} \mathrm{tr}\left(\mathbf{MSE}(\boldsymbol{F})\right),\ \text{s.t} : P_t \cdot \mathrm{tr}\left(\boldsymbol{F}_{opt}\boldsymbol{F}_{opt}^H\right) \leq P_T, \tag{4.11}$$

as in (4.10) with the elements of $\boldsymbol{\Phi}$ given by [8]

$$|\varphi_{i,i}|^2 = \left(\frac{P_T + \sum_{k=1}^{\bar{N}}\sigma_k^{-2}}{P_t\sum_{k=1}^{\bar{N}}\sigma_k^{-1}}\sigma_k^{-1} - \frac{1}{P_t\sigma_k^{-2}}\right)^+ \tag{4.12}$$

where $(x)^+ := \max(x, 0)$ and $\bar{N} < N$ is such that $|\varphi_{k,k}|^2 > 0$ for $k = 1, \ldots, \bar{N}$ and

$|\varphi_{k,k}|^2 = 0$ for all other k.

The maximum information rate precoder is obtained as a solution to the following optimization problem

$$\boldsymbol{F}_{opt} = \arg\min_{\boldsymbol{F}} \det\left(\mathbf{MSE}(\boldsymbol{F})\right), \text{ s.t} : P_t \cdot \mathrm{tr}\left(\boldsymbol{F}_{opt}\boldsymbol{F}_{opt}^H\right) \leq P_T, \tag{4.13}$$

Here, the power loading matrix $\boldsymbol{\Phi}$, given in equation (4.10), is identical to the one obtained using the water-pouring algorithm in (3.5).

The weighted MMSE design is introduced in [7] in order to take into account different QoS requirements over different spatially multiplexed data streams.

The dominant eigenmode transmission (DET) is performed by choosing $\boldsymbol{\Phi}$ as

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{4.14}$$

The dominant eigenmode transmission is motivated by the need to have maximum signal-to-noise ratio at the receiver. This leads to full $M_T M_R$ order diversity for the $\boldsymbol{H}_w$ channel. The array gain is given by $\mathrm{E}\left\{\sigma_{\max}^2\right\}$, where $\sigma_{\max}$ is the maximm singular value of the matrix $\boldsymbol{H}$. Since $\sum_{i=1}^{r} \sigma_i^2 = \|\boldsymbol{H}\|_{\mathcal{F}}^2$, where $\sigma_i, i = 1, \ldots, r$ are the singular values of $\boldsymbol{H}$ and r is the rank of $\boldsymbol{H}$, σ_{max}^2 may be upper- and lower-bounded according to

$$\frac{\|\boldsymbol{H}\|_{\mathcal{F}}^2}{r} \leq \sigma_{\max}^2 \leq \|\boldsymbol{H}\|_{\mathcal{F}}^2 \tag{4.15}$$

Using the dominant eigenmode transmission we achieve an array gain that is equal to or greater than the array gain achievable with STBCs. The dominant eigenmode transmission results in the maximum signal-to-noise ratio at the receiver which leads to minimum bit error rate. With channel state information available at the transmitter, the optimal strategy in order to achieve the maximum SNR at the receiver is to transmit on the dominant eigenmode with all power [9].

An important point to note is the optimality of modal decomposition of the channel matrix in the presence of channel state information at the transmitter and receiver, and that this strategy is optimal for a broad set of design criteria.

Chapter 5

Multi-user MIMO communications

It has been shown that time division multiple access (TDMA) systems cannot achieve a linear increase of the sum-rate capacity of multi-user (MU) MIMO systems in the number of transmit antennas [48], [49]. The solution to this problem is to serve users simultaneously using space-division multiple access (SDMA).

The information theoretic results in [14], [11], [12], [16], [17] have shown that it is necessary to use some kind of Costa's "dirty-paper" coding (DPC) or Tomlinson-Harashima precoding to reach the sum capacity of a multi-user MIMO downlink system. DPCs achieve the maximum sum rate of the system and provide the maximum diversity order. The sum rate capacity of the multi-user MIMO uplink system is achieved via an MMSE receiver with successive interference cancellation.

The sum-rate capacity of a downlink multi-user MIMO system employing DPC and an uplink multi-user MIMO system employing successive interference cancellation is at most $\min(M_T, K)$ times larger than the maximum achievable sum rate capacity of a system using TDMA [48], [49].

Motivated by the need for cheap user terminals with low power consumption, we focus on systems where the computationally demanding signal processing is performed at the base station. This means that one user will not be aware of other users sharing the same time and frequency resources and that the base station will have the task of reducing the multi-user interference. In this chapter we will address the problem of generalized designs of the precoding and decoding matrices in a multi-user MIMO communication system. The focus will be put on the multi-user MIMO downlink system, since in this case it is harder to define the cost function for system optimization and because most of the solutions for the downlink can be applied on the uplink in a straightforward way. Another

reason for this is that by using the same or similar MU MIMO processing techniques on both the uplink and the downlink we can reduce the cost of hardware at the base station. Moreover, having in mind the complexity of DPCs and their inability to combine instantaneous and long-term channel state information at the transmitter for precoding, we will give preference in our investigations to the linear precoding techniques. Beside their lower complexity than DPCs and their ability to combine instantaneous and long-term channel state information, they are also capable of reaching the achievable sum-rate capacity of the broadcast channel [23]. In the future it is very likely that the user terminals will be also equipped with more than one antenna which makes the scenario where the total number of receive antennas at the user terminals is greater than the number of antennas at the base station more likely. In this scenario, linear techniques have shown to be more effective than non-linear precoding techniques and they can also reach the sum-rate capacity of the multi-user downlink system.

As it was shown in the previous chapters, channel state information at the transmitter allows us to exploit the benefits of having multiple antennas at the base station and the user terminals to the maximum. Channel state information can be acquired at the transmitter either if a feedback channel is present or when the transmitter and receiver operate in time division duplex (TDD) so that time-invariant MIMO channel transfer function is the same in both ways. It was shown in [50] that the most practical duplexing scheme in indoor, hotspot and micro cellular scenarios is TDD. The cost of acquiring the channel state information at the transmitter is much lower in a TDD system, where it is possible to exploit the estimated uplink channel for the downlink transmission due to the reciprocity principle than in a frequency division duplex (FDD) system, where we have to rely on the feedback of the channel state information.

On the downlink the base station will use any channel state information available to mitigate or ideally completely eliminate multi-user interference through linear or non-linear (DPC or THP) precoding, which leads to significant information rate gains. The user terminal estimates the effective channel and transmits data in the next uplink frame. The effective channel is equal to the combined network channel after the precoding at the base station. However, on the uplink the base station has the possibility to use successive interference cancellation, so the effective channel on the uplink that includes the spatial processing does not have to be the same as on the downlink. Therefore, it is wise to assume that user terminals have no CSI available on the uplink which suggests the use of open-

loop MIMO techniques on the uplink. Thus, on the uplink we can identify two possible cases. In the first case, the user terminal encodes data using OSTBCs. In the second case, the users' precoding matrices on the uplink are generated at the base station and then feedforwarded to the user terminals. The diversity gains of MIMO are more desirable than spatial multiplexing gains if we take into account the limited power available at the user terminal and therefore it is enough that the base station transmits to the user terminal only the dominant singular vector of the uplink effective channel.

The framework for the generalized design presented in [8] and [7] is limited only to point-to-point communication where the transmitter and the receiver are able to perform a joint processing over all of the transmit and the receive antennas. There are a lot of results in the literature which address the optimization of multi-user MIMO downlink systems using different optimization criteria. However, there is no general solution like in the case of point-to-point communications. For the optimization of such systems it is often assumed that the users are equipped with only one antenna, [51], [22], [52], [16], [17]. Solutions that consider an arbitrary number of antennas at the user terminals often assume zero multi-user interference which imposes a constraint regarding the total number of antennas at the base station and the user terminals, [53], [25], [54], [27]. The solutions that overcome this dimensionality constraint, i.e., when the number of receive antennas is greater than the number of antennas at the base station either use only a subset of antennas or a subset of eigenmodes [25], [55], and usually require a large control overhead in order to feedback the decoding matrices to the user terminals.

In Chapter 5.4 a new approach will be introduced which will be used to derive a general framework for MU MIMO precoding and decoding design so we can target any optimization criteria with one universal algorithm like in the case of single-user MIMO processing.

5.1 Previous work on MU MIMO precoding

5.1.1 Zero forcing precoding

Since the base station has no influence on the noise at the user terminals, the most intuitive approach for precoding is a zero forcing filter (ZF) which eliminates all interference at the user terminals. ZF precoding for single antenna receivers was investigated extensively in the literature [56], [9].

Assuming single antenna terminals, the decoding matrix becomes $\boldsymbol{G} = \boldsymbol{I}_K$ and $M_R = K$. Let us define the precoding matrix $\boldsymbol{F}$ as $\boldsymbol{F} = \beta \boldsymbol{F}_a$. The precoding matrix $\boldsymbol{F}_a$ and the scaling factor β result from the following optimization

$$\boldsymbol{F}_a = \arg\min_{\boldsymbol{F}_a} \mathrm{E}\left\{ \|\boldsymbol{H}\boldsymbol{F}_a\boldsymbol{x} - \boldsymbol{x}\|_{\mathcal{F}}^2 \right\}, \text{ s.t. : } \boldsymbol{H}\boldsymbol{F}_a = \boldsymbol{I}_K \tag{5.1}$$

The parameter β is chosen such that the total transmit power is $\beta^2 \|\boldsymbol{F}_a\boldsymbol{x}\|_{\mathcal{F}}^2 \leq P_T$. We assume that the complex data symbols are i.i.d. uniformly distributed random variables and that the samples of the additive noise at the input of receive antennas are i.i.d. complex Gaussian white random variables with mean and variance as in (4.2),(4.3), respectively. The average power of the complex data symbols is set to

$$P_t = 1.$$

The solution to the optimization problem given in (5.1) is a pseudo-inverse of the combined channel matrix $\boldsymbol{H}$:

$$\boldsymbol{F}_a = \boldsymbol{H}^H \left(\boldsymbol{H}\boldsymbol{H}^H\right)^{-1}, \ \beta = \sqrt{\frac{P_T}{\|\boldsymbol{F}_a\|_{\mathcal{F}}^2}} \tag{5.2}$$

In the same way as the decoding ZF filter, the transmit ZF filter also suffers from the noise enhancement problem and requires increased transmit power. It is sub-optimal and results in a significant performance degradation. The diversity order and array gain of each stream is proportional to $M_T - M_R + 1$, [9].

5.1.2 Minimum mean-square-error precoding

The ZF precoder completely eliminates multi-user interference at the expense of noise enhancement. The minimum mean-square-error (MMSE) precoder balances the multi-user interference mitigation with noise enhancement and minimizes the total error. Unlike the ZF precoder, the MMSE precoder cannot be designed in such a straightforward way. A key to design of the MMSE precoder is to scale the transmit vector such that the total transmit power has the predefined level [51], i.e.

$$\boldsymbol{F}_a = \arg\min_{\boldsymbol{F}_a} \mathrm{E}\left\{ \left\|\beta^{-1}\boldsymbol{y} - \boldsymbol{x}\right\|_{\mathcal{F}}^2 \right\}, \text{ s.t. : } \beta^2 \|\boldsymbol{F}_a\boldsymbol{x}\|_{\mathcal{F}}^2 \leq P_T \tag{5.3}$$

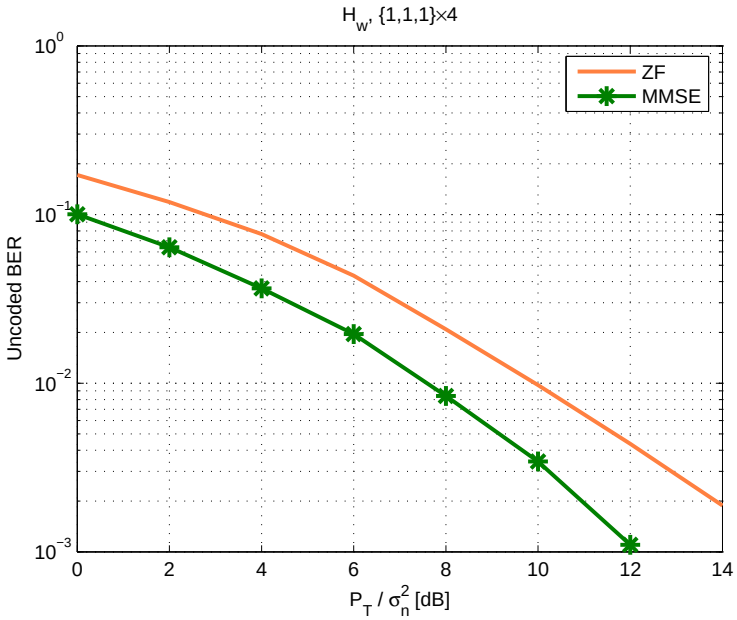

Figure 5.1: Uncoded BER performance of ZF and MMSE precoders. Flat fading, $\boldsymbol{H}_w$ channel. $K = 3$, $M_T = 4$.

The MMSE precoder is defined as

$$\boldsymbol{F}_a = \left(\boldsymbol{H}^H \boldsymbol{H} + \alpha \boldsymbol{I}_{M_T}\right)^{-1} \boldsymbol{H}^H \tag{5.4}$$

where the parameters α and β are equal to

$$\alpha = \frac{\sigma_n^2 K}{P_T} \quad \text{and} \quad \beta = \sqrt{\frac{P_T}{\|\boldsymbol{F}_a\|_{\mathcal{F}}^2}} \tag{5.5}$$

The total transmit power is P_T and the number of user terminals is K.

The MMSE precoder, in the same way as the receive spatial MMSE filter, approximates a matched filter at low SNRs and is near optimal. At high SNRs, the MMSE precoder converges to a ZF precoder and we can expect it to extract $M_T - M_R + 1$ order diversity.

In Figure 5.1 we compare the uncoded bit error performance (BER) of a ZF precoder and an MMSE precoder. A flat fading $\boldsymbol{H}_w$ channel with $M_T = 4$ antennas at the base station and $K = 3$ single antenna user terminals is assumed. As it can be seen from the previous figure, the MMSE precoder is always superior to the ZF precoder.

5.1.3 Block diagonalization

Block diagonalization (BD) is a ZF precoding technique that was first proposed in [53]. It was proposed to solve either the problem of maximizing the total system throughput under a transmit power constraint or to minimize the total transmit power for a predefined QoS level. It is restricted to channels where the number of transmit antennas M_T is not smaller than the total number of receive antennas in the network M_R. As it was shown in [25], the ZF precoding filter defined in (5.2) is suboptimal since each user is able to coordinate the processing of its own receiver outputs. Another difference is that BD, unlike ZF and MMSE precoders, allows also the option that the user terminals can also be equipped with more than one antenna.

Using BD, we can find the precoding matrix $\boldsymbol{F}$ such that all multi-user interference is zero by choosing a precoding matrix $\boldsymbol{F}_i$ such that it lies in the null space of the other users' channel matrices. Thereby, a multi-user MIMO downlink channel is decomposed into multiple parallel independent single-user MIMO channels [25], [54].

If we define $\widetilde{\boldsymbol{H}}_i$ as

$$\widetilde{\boldsymbol{H}}_i = \begin{bmatrix} \boldsymbol{H}_1^T & \cdots & \boldsymbol{H}_{i-1}^T & \boldsymbol{H}_{i+1}^T & \cdots & \boldsymbol{H}_K^T \end{bmatrix}^T \in \mathbb{C}^{(M_R - M_{R_i}) \times M_T} \tag{5.6}$$

the zero MUI constraint forces $\boldsymbol{F}_i$ to lie in the null space of $\widetilde{\boldsymbol{H}}_i$. From the singular value decomposition of $\widetilde{\boldsymbol{H}}_i$ whose rank is $\widetilde{L}_i$

$$\widetilde{\boldsymbol{H}}_i = \widetilde{\boldsymbol{U}}_i \widetilde{\boldsymbol{\Sigma}}_i \begin{bmatrix} \widetilde{\boldsymbol{V}}_i^{(1)} & \widetilde{\boldsymbol{V}}_i^{(0)} \end{bmatrix}^H \tag{5.7}$$

we choose the last right $M_T - \widetilde{L}_i$ singular vectors $\widetilde{\boldsymbol{V}}_i^{(0)} \in \mathbb{C}^{M_T \times M_T - \widetilde{L}_i}$ which form an orthogonal basis for the null space of $\widetilde{\boldsymbol{H}}_i$. The effective channel of user i after eliminating the MUI is identified as $\boldsymbol{H}_i \widetilde{\boldsymbol{V}}_i^{(0)}$, whose dimension is $M_{R_i} \times \left(M_T - \widetilde{L}_i\right)$ and is equivalent to a system with $M_T - \widetilde{L}_i$ transmit antennas and M_{R_i} receive antennas. Each of these effective single-user MIMO channels has the same properties as a conventional single-user MIMO channel. Define the SVD

$$\boldsymbol{H}_i \widetilde{\boldsymbol{V}}_i^{(0)} = \boldsymbol{U}_i \boldsymbol{\Sigma}_i \begin{bmatrix} \boldsymbol{V}_i^{(1)} & \boldsymbol{V}_i^{(0)} \end{bmatrix}^H \tag{5.8}$$

and let the rank of the i^{th} user's effective channel matrix be L_i. The product of the first L_i singular vectors $\boldsymbol{V}_i^{(1)}$ and $\widetilde{\boldsymbol{V}}_i^{(0)}$ produces an orthogonal basis of dimension L_i and

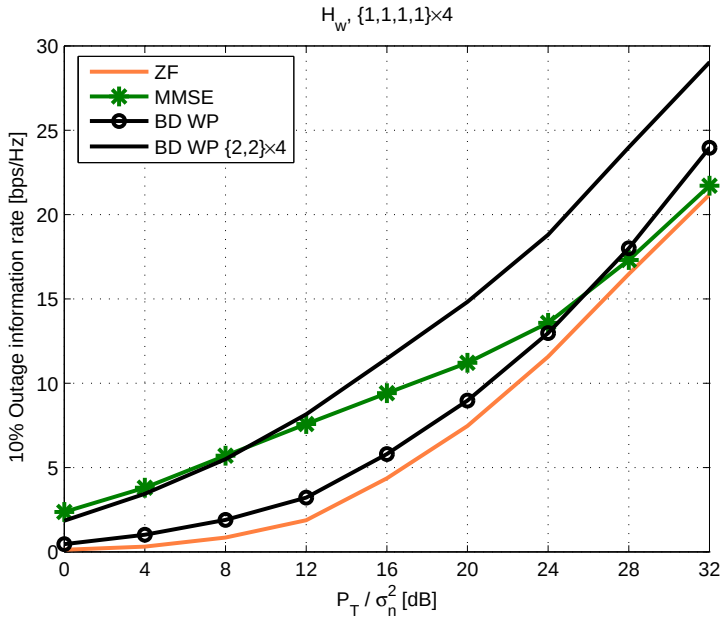

Figure 5.2: Information sum rate of BD,ZF and MMSE in a system with configuration $\{1,1,1,1\} \times 4$. BD system with configuration $\{2,2\} \times 4$.

represents the transmission vectors that maximize the information rate for user i subject to the zero MUI constraint. The demodulation matrix of the i^{th} user is chosen as $\boldsymbol{G}_i = \boldsymbol{U}_i^H$.

It can be easily verified that assuming the $\boldsymbol{H}_w$ channel, the i^{th} user can extract maximum $M_{R_i} \times \left(M_T - \widetilde{L}_i\right)$ order diversity.

In Figure 5.2 we compare the information sum rate of the systems employing BD, ZF and MMSE precoding. We assume a flat fading $\boldsymbol{H}_w$ channel with $M_T = 4$ antennas at the base station. There are $K = 4$ users in the system equipped with $M_{R_i} = 1$ antenna each. BD clearly provides higher information rate than ZF. However, at low SNRs MMSE outperforms both BD and ZF. The real advantage of BD can be seen when the users are equipped with multiple antennas too. In this case, BD outperforms ZF and MMSE at high SNRs.

5.1.4 Tomlinson-Harashima precoding

Tomlinson-Harashima precoding (THP) was first developed for SISO multipath channels, where it was used to overcome the error propagation problem of decision feedback equalization (DFE) by moving the DFE block to the transmitter side, [18], [19]. In [20] it is proposed for the spatial equalization of multi-user interference in MIMO systems.

The THP precoder can be interpreted as a one-dimensional implementation of DPCs and also requires that the channel state information is available at the transmitter. The

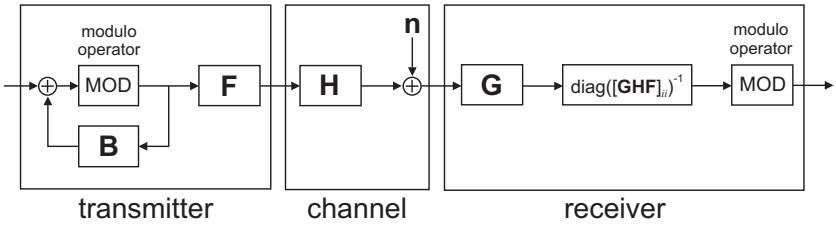

Figure 5.3: Block diagram of the THP system.

block diagram of a system employing THP at the transmitter side is shown in Figure 5.3.

At each step k of the precoding in a THP system, only the non-causal interference from the previously encoded data symbols $x_i, i < k$, is pre-subtracted from the data symbol encoded at this step x_k. It does not take into account the future side information $x_i, i > k$. By using THP at the transmit side we significantly increase the transmit power. That is why we have to introduce the modulo operator at the transmitter and the receiver in order to reduce the constellation size into certain boundaries. If we assume that the first data symbol x_1 is encoded first, than the second data symbol x_2, etc., then the THP feedback matrix $\boldsymbol{B}$ is equal to the lower triangular part of $\boldsymbol{GHF}$. The k^{th} data symbol at the output of the THP precoder is equal to

$$z_k = \mathrm{mod}\left(x_k - \sum_{i=1}^{k-1} b_{k,i} z_i\right)_\tau \tag{5.9}$$

where $\mathrm{mod}\,(\bullet)_\tau$ is the modulo operator and τ is the constant which depends on the constellation size of the used modulation alphabet. The modulo operation is defined as

$$\mathrm{mod}\,(x)_\tau = x - \left\lfloor \frac{x}{\tau} + \frac{1}{2} \right\rfloor \tau$$

where the floor operator $\lfloor \bullet \rfloor$ gives the integer number smaller than or equal to the argument. The modulo operation is done on a symbol-by-symbol basis, producing an output that is uniformly distributed between $-\tau$ and τ. This uniform distribution corresponds to a cubic shape, thus incurring a shaping loss when compared to the spherical shape of an optimal Gaussian code [20].

The exact formula for the capacity of the ZF THP temporal equalizer and the upper and lower bound for the capacity of the MMSE THP temporal equalizer were derived in [57]. The same approach was used in [58] to derive the exact capacity formula for ZF THP spatial precoding. As it was shown in these papers, the optimum information rate maximizing strategy is not Gaussian but uniformly distributed which is direct consequence

of the use of the modulo operator at the transmitter and the receiver. Following the analysis in [57], it is not hard to derive the achievable rate of the k^{th} user as

$$R_k = \log_2(2\tau) - \mathrm{h}\left(\mathrm{mod}\left(\tilde{n}_k\right)_\tau\right) \tag{5.10}$$

where $\tilde{n}_k$ is the filtered noise at the input of the k^{th} user's antenna and $\mathrm{h}(\bullet)$ denotes the differential entropy function.

5.1.5 MMSE THP precoding

In [21] THP precoding is performed at the transmitter, whereas the receiver still performs linear filtering. In [22], both feedforward and feedback filters are deployed at the transmitter which results in a significant reduction of the computational load at the receiver side.

The MMSE THP precoding filter is derived from the linear transmit MMSE precoding optimization by neglecting the contribution of the elements of the lower triangular part of $\boldsymbol{H}\boldsymbol{F}_a$ to the overall MSE:

$$\boldsymbol{F}_a = \arg\min_{\boldsymbol{F}_a} \mathrm{E}\left\{\left\|\beta^{-1}\boldsymbol{y} - \boldsymbol{x}\right\|_{\mathcal{F}}^2\right\},\ \text{s.t.}: \beta^2\left\|\boldsymbol{F}_a\boldsymbol{x}\right\|_{\mathcal{F}}^2 \leq P_T,\ \left[\boldsymbol{H}\boldsymbol{F}_a\right]_{i,j} = 0, i > j, \forall i, j \tag{5.11}$$

where $[\bullet]_{i,j}$ denotes the element in the i^{th} row and the j^{th} column of the matrix $\boldsymbol{H}\boldsymbol{F}_a$. The interference remaining after the precoding using $\boldsymbol{F}$ from these lower triangular matrix is eliminated using THP. The algorithm described in [22] is iterative and requires a certain ordering of the users. First, a precoding matrix $\boldsymbol{F}_a$ is defined column by column starting from user K. The column corresponding to the i^{th} user is obtained as the i^{th} column of the precoding matrix calculated using only the first i rows of the network channel matrix $\boldsymbol{H}$ and equation (5.4). Let us introduce the matrix

$$\widehat{\boldsymbol{H}}_i = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_{i-1} \\ \boldsymbol{H}_i \end{bmatrix}$$

which contains the first i rows of the matrix $\boldsymbol{H}$. Then the i^{th} column of the precoding

matrix $\boldsymbol{F}_a$ is given by

$$[\boldsymbol{F}_a]_{(:,i)} = \left[\left(\widehat{\boldsymbol{H}}_i^H \widehat{\boldsymbol{H}}_i + \alpha \boldsymbol{I}_{M_T} \right)^{-1} \widehat{\boldsymbol{H}}_i^H \right]_{(:,i)} \tag{5.12}$$

The parameters α and β are given in equation (5.5). After this, the users are encoded in the reverse order from the one in which their precoding matrices are generated, i.e., starting from the first user, then the second, etc. Using THP we eliminate the multi-user interference to the i^{th} user originating from the previous $i-1$ users. The ordering of the users is based on a heuristic approach, by choosing in every step of the algorithm the user with the minimum MSE.

The performance of this algorithm can be further improved by introducing the weighting which favors the stronger users. The additional gain from weighting is less than 2 dB [59]. However, the problem of precoding the users equipped with more than one antenna still remains.

5.2 BD MMSE THP precoding

A combination of BD and MMSE THP precoding was introduced in [26] that provides better performance for the users equipped with one antenna than BD and better performance for the users equipped with multiple antennas than MMSE, which results in the improvement of the overall system performance. By using BD only for multiple antenna users we effectively eliminate the interference that these users generate to single antenna users. Then by using MMSE THP only for the single antenna users we improve their and the overall system performance.

First we group the single- and the multiple-antenna users. For users with multiple receive antennas we use BD, and for single antenna users we use MMSE THP. The modulation matrices for multiple antenna users are chosen to lie in the null space of the channel matrices of the other users including the single antenna users. In this way the effective channel for the single antenna users looks as if there were no multiple antenna users. Thereby we improve the diversity of these users which influences their BER performance with different degrees of CSI at the transmitter. MMSE THP is applied only on the combined network channel corresponding to these single antenna users. The data transmitted to the multiple antenna users is also precoded using THP in order to eliminate the MUI which in this case only originates from the single antenna users.

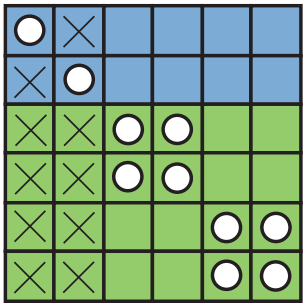

Figure 5.4: Graphical representation of the effective network channel matrix for BD MMSE THP with the configuration $\{1, 1, 2, 2\} \times 6$. Crosses represent MUI elements that will be eliminated using THP.

By using THP at the transmit side we significantly increase the transmit power. That is why we need to introduce the modulo operator at the transmitter and the receiver in order to reduce the constellation size into certain boundaries. The single antenna users multiply the received signal by $1/\beta$ from equation (5.5) and multiple antenna users with the corresponding eigenvalue in $\mathbf{\Sigma}$ so that the constellation boundaries at the receiver are the same as at the transmitter. The matrix $\mathbf{\Sigma}$ is defined as

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & & & \\ & \mathbf{\Sigma}_2 & & \\ & & \ddots & \\ & & & \mathbf{\Sigma}_K \end{bmatrix}$$

where the $\mathbf{\Sigma}_i$ are defined in equation (5.8).

The effective channel matrix for the system with the configuration $\{1, 1, 2, 2\} \times 6$, after precoding, is depicted in Figure 5.4. The first two rows correspond to the two single antenna users and the last four to the two users equipped with two antennas each. By using BD, all elements in the last four columns above and below the block matrices on the diagonal are set to zero. Crosses in the figure represent the MUI that is generated by the single-antenna users due to the MMSE processing. By using THP the elements in the first two columns below the main diagonal are set to zero. If we used BD for these users too, this interference would also be equal to zero. With the circles we denote the elements of the effective network channel matrix that contain the gains of the channel for each data stream.

The performance of this technique is better than BD or MMSE THP, however, the following precoding techniques outperform BD MMSE THP and provide higher information

rate and BER performance. The only advantage of BD MMSE THP is its simplicity.

5.3 Successive optimization THP precoding

As mentioned before, by applying BD on the combined channel matrix of all users the MU MIMO channel can be transformed into a set of parallel single-user MIMO channels. However, there is an information rate loss due to the cancellation of overlapping row subspaces of different users. In [53], the authors propose a successive optimization (SO) precoding algorithm in order to define a simplified solution to the problem of minimizing the total transmit power while achieving a predefined QoS level for each user in the network and to the near-far problem. By allowing a certain amount of interference, this algorithm reduces the data rate loss due to the subspace cancellation. It can yield better results in some situations but its performance in general is very poor and depends on the power allocation and the order in which the users' signals are pre-processed. A novel technique that combines SO and THP in order to reduce the information rate loss due to the overlapping of different users' row subspaces and to eliminate the MUI was introduced in [27]. After the precoding, the resulting effective combined channel matrix of all users is again block diagonal. This also facilitates the definition of a new ordering algorithm. Unlike in [11], [22], [51], this technique allows more than one antenna at the user terminals and has no performance loss due to the cancellation of interference between the signals transmitted to two closely spaced antennas at the same terminal.

First, we have to assume or determine a certain optimum ordering of the users, similar to V-BLAST [6] or MMSE THP [22]. Using SO, the modulation matrix for each user is designed in such a way that it lies only in the null space of the channel matrices of previous users. As a consequence, only they will generate the interference to this user. Let us define the previous $i-1$ users' combined channel matrix as

$$\widehat{\boldsymbol{H}}_i = \left[\begin{array}{cccc} \boldsymbol{H}_1^T & \boldsymbol{H}_2^T & \cdots & \boldsymbol{H}_{i-1}^T \end{array} \right]^T$$

and its corresponding SVD as

$$\widehat{\boldsymbol{H}}_i = \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{\Sigma}}_i \left[\begin{array}{cc} \widehat{\boldsymbol{V}}_i^{(1)} & \widehat{\boldsymbol{V}}_i^{(0)} \end{array} \right]^H . \tag{5.13}$$

If the rank of $\widehat{\boldsymbol{H}}_i$ is $\hat{L}_i$, then $\widehat{\boldsymbol{V}}_i^{(0)}$ contains $M_T - \hat{L}_i$ right singular vectors. As in the BD solution, we force the modulation matrix $\boldsymbol{F}_i$ to lie in the null space of $\widehat{\boldsymbol{H}}_i$ by setting

for $i = 1 : K$

$\boldsymbol{H}_i = \boldsymbol{U}_i \boldsymbol{\Sigma}_i \left[\begin{array}{cc} \boldsymbol{V}_i^{(1)} & \boldsymbol{V}_i^{(0)} \end{array} \right]^H$;

$\boldsymbol{F}_{\max,i} = \boldsymbol{V}_i^{(1)}$;

$C_{\max,i} = \log_2 \det \left(\boldsymbol{I} + \boldsymbol{R}_{n,i}^{-1} \boldsymbol{H}_i \boldsymbol{F}_{\max,i} \boldsymbol{F}_{\max,i}^H \boldsymbol{H}_i^H \right)$;

end;

$\mathrm{S} = \{1, \ldots, K\}$;

$\boldsymbol{H}_{\text{aux}} = \boldsymbol{H}$;

for $i = K : 1$

$\left[\begin{array}{cccccc} \boldsymbol{P}_1, & \ldots & \boldsymbol{P}_i, & \boldsymbol{U}_1, & \ldots & \boldsymbol{U}_i \end{array} \right] = \mathrm{BD}\left(\boldsymbol{H}_{\text{aux}}\right)$;

for $k = 1 : i$

$C_k = \log_2 \det \left(\boldsymbol{I} + \boldsymbol{R}_{n,k}^{-1} \boldsymbol{H}_k \boldsymbol{P}_k \boldsymbol{P}_k^H \boldsymbol{H}_k^H \right)$;

end;

$k_i = \arg\min_{k \in S} \left(C_{\max,k} - C_k \right)$;

$\boldsymbol{F}_i = \boldsymbol{P}_{k_i}$;

$\boldsymbol{G}_i = \boldsymbol{U}_{k_i}^H$;

$\mathrm{S} = \mathrm{S} \backslash \{k_i\}$;

$\boldsymbol{H}_{\text{aux}} = \left[\begin{array}{cccccc} \boldsymbol{H}_1^T & \ldots & \boldsymbol{H}_{k_i-1}^T & \boldsymbol{H}_{k_i+1}^T & \ldots & \boldsymbol{H}_K^T \end{array} \right]^T$;

end;

$\boldsymbol{F} = \left[\begin{array}{ccc} \boldsymbol{F}_1 & \ldots & \boldsymbol{F}_K \end{array} \right]$;

$\boldsymbol{G} = \left[\begin{array}{ccc} \boldsymbol{G}_1 & & \\ & \ddots & \\ & & \boldsymbol{G}_K \end{array} \right]$;

$\boldsymbol{B} = \text{lower triangular} \left(\boldsymbol{GHF} \cdot \mathrm{diag} \left([\boldsymbol{GHF}]_{ii}^{-1} \right) \right)$;

Table 5.1: SO THP algorithm.

$\boldsymbol{F}_i = \widehat{\boldsymbol{V}}_i^{(0)} \boldsymbol{F}_i'$ for some choice of $\boldsymbol{F}_i'$. Thereby, the i^{th} user does not see any interference from any subsequent user $(i+1, \ldots, K)$.

The combination of SO and THP (SO THP) is performed by successively calculating the BD, the reordering of users, and in the end precoding with THP. Instead of examining all $K!$ possibilities for ordering to minimize the total information rate loss in the system, we propose a heuristic simplification to minimize the information rate loss of each user in the presence of the other co-channel users separately.

The whole SO THP algorithm is summarized in Table 5.1. We use the following notation: BD () is BD as explained before, $\boldsymbol{P}_k$ is an auxiliary matrix where we store the precoding matrices generated using BD, $\mathcal{S}$ is a set of indices of the users to be processed, $\boldsymbol{G}_k$ is the k^{th} user demodulation matrix obtained by using the BD algorithm and $\boldsymbol{B}$ is the THP feedback matrix.

We first calculate the information rate that an individual user can achieve assuming there are no other users in the system. Then, we look for the user for whom the difference between its information rate when there are no other users and its BD information rate is minimum and generate the precoding matrix of this user such that it lies in the null space

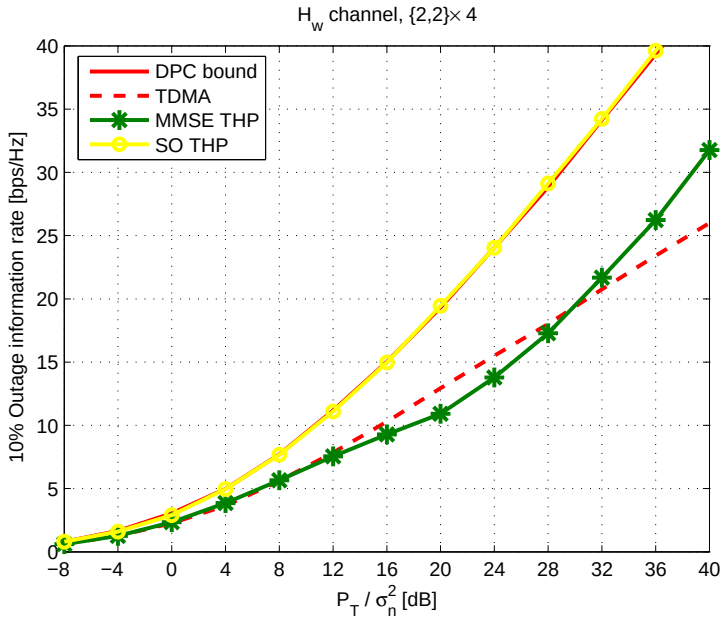

Figure 5.5: 10 % outage information rate of SO THP and MMSE THP in bps/Hz as a function of the SNR.

of the remaining users' channel matrices. In each step we find the user with the minimum information rate loss and place it as the last one. Afterwards, we form the new combined channel matrix $\boldsymbol{H}_{\text{aux}}$ without this user's channel matrix $\boldsymbol{H}_{k_i}$. We repeat these steps until the combined channel matrix is empty.

The order of the users is the reverse of the order in which their precoding matrices are generated. With this reordering of the users we achieve that the effective combined channel matrix after precoding and demodulation is lower triangular with the singular values on the main diagonal. The lower triangular feedback matrix $\boldsymbol{B}$, used in THP precoding [22], is generated from this effective combined channel matrix after the elements in each row are divided by the elements on the main diagonal, i.e., the corresponding singular values, as it can be seen from the last equation in Table 5.1.

By using THP at the transmit side we significantly increase the transmit power. That is why we have to introduce the modulo operator at the transmitter and the receiver in order to reduce the constellation size into certain boundaries. Before applying the modulo operator at the receiver we have to divide each data stream by the corresponding singular value so that the constellation boundaries at the receiver are the same as at the transmitter [22].

In Figure 5.5 we compare the 10 % outage information rate of MMSE THP in a system with configuration $\{1, 1, 1, 1\} \times 4$ and SO THP in a system with configuration $\{2, 2\} \times 4$.

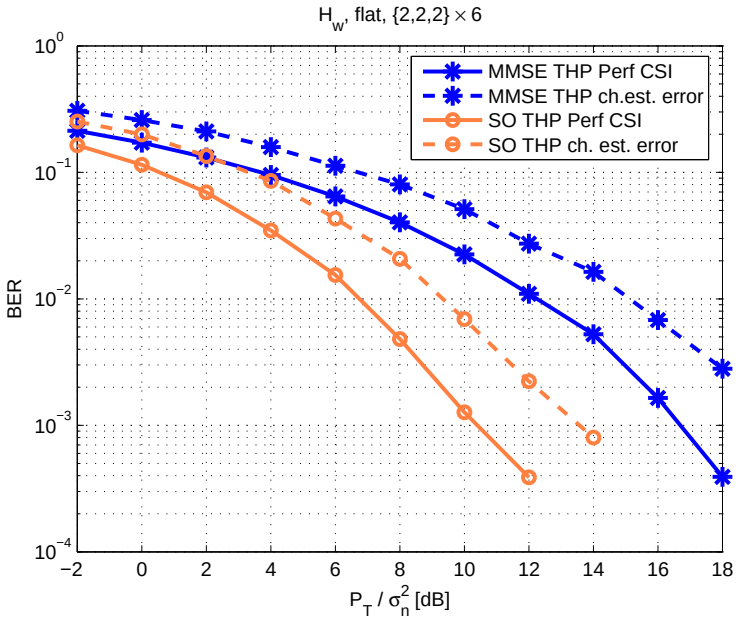

Figure 5.6: BER as a function of SNR.

We also present capacity results for a TDMA system and the DPC bound as a comparison. MMSE THP fails to reach the DPC bound and SO THP provides higher information sum rate when the users are equipped with multiple antennas. This can be also seen from the next figure, where we compare the BER performance of MMSE THP and SO THP in system with configuration $\{2, 2, 2\} \times 6$. In a MMSE THP system each data stream transmitted over different antennas is modulated using BPSK and in a SO THP system data is modulated using QAM modulation. Throught this thesis we use Gray mapping. SO THP, which is a ZF technique, performs better than MMSE THP at low SNRs and has the same diversity order as a consequence of optimum processing at the user terminals.

As we could see, SO THP is a ZF sum-rate capacity achieving technique. However, it has the same disadvantages as other ZF techniques and as BD it also has a dimensionality constraint that the total number of receive antennas at the user terminals has to be less than or equal to the number of antennas at the base station. Therefore, the applicability and gains that this and similar techniques can bring are very limited.

5.4 Generalized design of MU MIMO precoding and decoding matrices

MMSE THP precoding is optimum when all users in the system are equipped with only one antenna. MMSE balances the MUI in order to reduce the performance loss while the THP is used to eliminate the part of the MUI and improves the diversity. However, MMSE THP suffers a performance loss when it attempts to mitigate the interference between two closely spaced antennas as in the case when the user terminal is equipped with more than one receive antenna. It was shown through simulations that SO THP approaches the sum-rate capacity of the broadcast channel and can support multiple antennas at the user terminals but has the dimensionality and noise enhancement problem.

Our goal is to perform MIMO precoding in such a way that the loss due to the multi-user interference mitigation is reduced, to better use multiple antennas at the user terminals and to remove any constraint regarding the number of antennas at the user terminals.

In order to facilitate the generalized design of the precoding matrix in a MU MIMO scenario as in the SU MIMO case, we use a different approach. We separate the MUI suppression and the system performance optimization. Therefore, the precoder design is performed in two steps. In the first step we balance the MUI suppression which is achieved by reducing the overlap of the row spaces spanned by the effective channel matrices of different users and any MIMO processing gain which requires that the users use as much as possible the available subspaces. In the second step we optimize the system performance assuming parallel SU MIMO channels. Thus, the precoding matrix in equation (2.16) is factored as

$$\boldsymbol{F} = \beta \boldsymbol{F}_a \cdot \boldsymbol{F}_b, \tag{5.14}$$

where

$$\boldsymbol{F}_a = \begin{bmatrix} \boldsymbol{F}_{a_1} & \boldsymbol{F}_{a_2} & \cdots & \boldsymbol{F}_{a_K} \end{bmatrix} \in \mathbb{C}^{M_T \times M_x}, \tag{5.15}$$

and

$$\boldsymbol{F}_b = \begin{bmatrix} \boldsymbol{F}_{b_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{F}_{b_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{F}_{b_K} \end{bmatrix} \in \mathbb{C}^{M_x \times r}, \tag{5.16}$$

with $\boldsymbol{F}_{a_i} \in \mathbb{C}^{M_T \times M_{x_i}}$ and $\boldsymbol{F}_{b_i} \in \mathbb{C}^{M_{x_i} \times r_i}$, $M_{x_i} \leq r$, and $M_x = \sum_{i=1}^{K} M_{x_i}$ depending on the

specific choice of the precoding algorithm. The matrix $\boldsymbol{F}_a$ is used to suppress the MUI interference first, and then the matrix $\boldsymbol{F}_b$ is used to optimize the system performance according to a specific criterion assuming that the MU MIMO channel has been transformed into a set of parallel SU MIMO channels. Finally, the parameter β is chosen to set the total transmit power to P_T.

In the same way as for the precoding, the receive MU MIMO matrix $\boldsymbol{D}$ on the UL is factored as

$$\boldsymbol{D} = \boldsymbol{D}_b \boldsymbol{D}_a \tag{5.17}$$

where matrix

$$\boldsymbol{D}_a = \begin{bmatrix} \boldsymbol{D}_{a_1} \\ \boldsymbol{D}_{a_2} \\ \vdots \\ \boldsymbol{D}_{a_K} \end{bmatrix} \in \mathbb{C}^{M_x \times M_T}, \tag{5.18}$$

and

$$\boldsymbol{D}_b = \begin{bmatrix} \boldsymbol{D}_{b_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_{b_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{D}_{b_K} \end{bmatrix} \in \mathbb{C}^{r \times M_x}, \tag{5.19}$$

with $\boldsymbol{D}_{a_i} \in \mathbb{C}^{M_{x_i} \times M_T}$ and $\boldsymbol{D}_{b_i} \in \mathbb{C}^{r_i \times M_{x_i}}$, $M_{x_i} \leq r$, and $M_x = \sum_{i=1}^{K} M_{x_i}$ depends on the specific choice of the receive algorithm. The matrix $\boldsymbol{D}_a$ is used to suppress the MUI interference first, and then the matrix $\boldsymbol{D}_b$ is used to optimize the system performance according to a specific criterion assuming that the MU MIMO channel has been transformed into a set of parallel SU MIMO channels.

In the following sections we will introduce two classes of precoding techniques which are defined by two different cost functions. These cost functions are based on the modified MMSE optimization criterion. As a consequence these two classes of precoding techniques will have different performance.

5.5 Successive MMSE filtering

5.5.1 Successive MMSE precoding

A new algorithm was proposed in [28], that deals with the disadvantages of MMSE precoding by successively calculating the columns of the precoding matrix $\boldsymbol{F}_a$ that correspond to different receive antennas.

The successive MMSE (SMMSE) precoding filter $\boldsymbol{F}_a$ is derived from the linear transmit MMSE precoding optimization by neglecting the contribution of interference between the signals at one user's antenna array to this user's MSE. Since each user can coordinate the processing over all of its antennas, we can combine the signals at the different antennas of one user in order to extract higher diversity and array gain. The interference of other co-channel users to the signal arriving at the i^{th} user's j^{th} antenna is suppressed independently from the other antennas at the same terminal. This is done for each antenna at the same user terminal successively. Therefore, the j^{th} column of the i^{th} user's precoding matrix $\boldsymbol{F}_{a_i}$, corresponding to the i^{th} user's j^{th} receive antenna, is equal to the first column of the matrix $\boldsymbol{F}_{a_{i,j}}$ which is obtained from the following optimization

$$\boldsymbol{F}_{a_{i,j}} = \arg\min_{\boldsymbol{F}_{a_{i,j}}} \mathrm{E}\left\{ \left\| \overline{\boldsymbol{H}}_i^{(j)} \boldsymbol{F}_{a_{i,j}} \overline{\boldsymbol{z}}_i^{(j)} + \frac{\overline{\boldsymbol{n}}_i^{(j)}}{\beta} - \overline{\boldsymbol{z}}_i^{(j)} \right\|_{\mathcal{F}}^2 \right\} \tag{5.20}$$

such that $\beta^2 \left\| \boldsymbol{F}_a \boldsymbol{F}_b \boldsymbol{x} \right\|_{\mathcal{F}}^2 \leq P_T$. The matrix $\overline{\boldsymbol{H}}_i^{(j)}$ and the vectors $\overline{\boldsymbol{z}}_i^{(j)}$ and $\overline{\boldsymbol{n}}_i^{(j)}$ corresponding to the i^{th} user's, $i = 1, \ldots, K$, j^{th} receive antenna, $j = 1, \ldots, M_{R_i}$, are defined as

$$\overline{\boldsymbol{H}}_i^{(j)} = \begin{bmatrix} \boldsymbol{h}_{i,j}^T \\ \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_{i-1} \\ \boldsymbol{H}_{i+1} \\ \vdots \\ \boldsymbol{H}_K \end{bmatrix}, \ \overline{\boldsymbol{z}}_i^{(j)} = \begin{bmatrix} z_{i,j} \\ \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_{i-1} \\ \boldsymbol{z}_{i+1} \\ \vdots \\ \boldsymbol{z}_K \end{bmatrix}, \text{ and } \ \overline{\boldsymbol{n}}_i^{(j)} = \begin{bmatrix} n_{i,j} \\ \boldsymbol{n}_1 \\ \vdots \\ \boldsymbol{n}_{i-1} \\ \boldsymbol{n}_{i+1} \\ \vdots \\ \boldsymbol{n}_K \end{bmatrix}$$

where $\boldsymbol{h}_{i,j}^T$ is the j^{th} row of the i^{th} user's channel matrix $\boldsymbol{H}_i$, $z_{i,j}$ is the j^{th} element of the i^{th} user's vector $\boldsymbol{z}_i \in \mathbb{C}^{M_{R_i} \times 1}$ and $n_{i,j}$ is the noise at the input of the i^{th} user's j^{th} receive antenna. The elements of the vector $\boldsymbol{z}_i$ are zero mean, unit variance i.i.d. complex uniform random variables. The elements of the vector $\boldsymbol{n}_i$ are zero mean complex Gaussian random

variables with variance σ_n^2. Note that the vectors $\boldsymbol{z}_i = \boldsymbol{F}_{b_i}\boldsymbol{x}_i$, $i = 1, \ldots, K$, are the i^{th} user's precoded data. The statistical properties of the elements of the vector $\boldsymbol{z}_i$, in general depend on the matrix $\boldsymbol{F}_{b_i}$. However, when we generate matrices $\boldsymbol{F}_{a_i}$ we assume that the matrices $\boldsymbol{F}_{b_i}$ are unitary. This assumption is true if each user is receiving independent data streams with the same power over all of the receive antennas. In that case the statistics of the elements of the vectors $\boldsymbol{z}_i$ are the same as the statistics of the elements of the vectors $\boldsymbol{x}_i$. By assuming that the matrix $\boldsymbol{F}_{b_i}$ is unitary we assign the same priority for data transmission to all eigenmodes of the i^{th} user's effective channel $\boldsymbol{H}_i\boldsymbol{F}_{a_i}$.

The columns of the precoding matrix $\boldsymbol{F}_{a_i}$, each corresponding to one receive antenna, are calculated successively. The corresponding column of the precoding matrix $\boldsymbol{F}_{a_i}$ is equal to the first column of the following matrix:

$$\boldsymbol{F}_{a_{i,j}} = \left(\overline{\boldsymbol{H}}_i^{(j)\ H}\overline{\boldsymbol{H}}_i^{(j)} + \alpha\boldsymbol{I}_{M_T}\right)^{-1}\overline{\boldsymbol{H}}_i^{(j)\ H} \tag{5.21}$$

The parameter α is equal to $\alpha = \sigma_n^2 K/P_T$ as in the equation (5.5).

After calculating the precoding vectors for all receive antennas in this fashion, the effective combined channel matrix of all users is equal to $\boldsymbol{H}\boldsymbol{F}_a \in \mathbb{C}^{M_R \times M_R}$ after the precoding. For high SNR ratios and when $M_R \leq M_T$, this matrix is also block diagonal. We can now apply any other previously defined SU MIMO technique on the i^{th} user's effective channel matrix $\boldsymbol{H}_i\boldsymbol{F}_{a_i}$. After the precoding using the matrix $\boldsymbol{F}_{a_i}$, we first perform the singular value decomposition (SVD) and then, if we want to maximize the capacity of the system we use water-pouring on the eigenmodes of all users or if we want to extract the maximum diversity and array gain, we transmit only on the dominant eigenmode of the users' effective channels. Dominant eigenmode transmission provides maximum SNR at the receiver and minimum BER performance. The complexity of this algorithm is only slightly higher than the one of BD. By using this algorithm we efficiently improve the array and diversity gains of the system by introducing MUI and by eliminating inter-stream interference. Following the same analysis as for MMSE, we expect that at high SNRs, in case of a $\boldsymbol{H}_w$ channel, SMMSE extracts a diversity order of $M_{R_i}(M_T - M_R + M_{R_i})$.

5.5.2 SMMSE THP precoding

SMMSE deals with the problem of performance loss due to inter-stream interference cancellation of MMSE by successively calculating the columns of the precoding matrix $\boldsymbol{F}_{a_i}$ for each of the receive antennas separately. In this way it provides a higher diversity and a higher array gain. Here, the diversity is further improved by combining SMMSE and THP. By doing so we also improve the information rate of an SMMSE system and as we will see later for low SNRs approach the sum rate capacity bound of the broadcast channels.

The i^{th} user's SMMSE THP precoding matrix $\boldsymbol{F}_{a_i}$ is designed similarly to the SMMSE precoding filter under the assumption that the interference from the remaining users, $i+1,\ldots,K$ is pre-subtracted using THP. Then, the j^{th} column of the i^{th} user's precoding matrix $\boldsymbol{F}_{a_i}$, corresponding to the i^{th} user's j^{th} receive antenna, is equal to the first column of the matrix $\boldsymbol{F}_{a_{i,j}}$ which is obtained from the following optimization

$$\boldsymbol{F}_{a_{i,j}} = \arg\min_{\boldsymbol{F}_{a_{i,j}}} \mathrm{E}\left\{ \left\| \overline{\boldsymbol{H}}_i^{(j)} \boldsymbol{F}_{a_{i,j}} \overline{\boldsymbol{z}}_i^{(j)} + \frac{\overline{\boldsymbol{n}}_i^{(j)}}{\beta} - \overline{\boldsymbol{z}}_i^{(j)} \right\|_{\mathcal{F}}^2 \right\} \tag{5.22}$$

such that $\beta^2 \left\| \boldsymbol{F}_a \boldsymbol{F}_b \boldsymbol{x} \right\|_{\mathcal{F}}^2 \leq P_T$. The matrix $\overline{\boldsymbol{H}}_i^{(j)}$, and the vectors $\overline{\boldsymbol{z}}_i^{(j)}$ and $\overline{\boldsymbol{n}}_i^{(j)}$ are defined in this case as

$$\overline{\boldsymbol{H}}_i^{(j)} = \begin{bmatrix} \boldsymbol{h}_{i,j}^T \\ \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_{i-1} \end{bmatrix}, \ \overline{\boldsymbol{z}}_i^{(j)} = \begin{bmatrix} z_{i,j} \\ \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_{i-1} \end{bmatrix}, \text{ and } \ \overline{\boldsymbol{n}}_i^{(j)} = \begin{bmatrix} n_{i,j} \\ \boldsymbol{n}_1 \\ \vdots \\ \boldsymbol{n}_{i-1} \end{bmatrix}$$

where $\boldsymbol{h}_{i,j}^T$ is the j^{th} row of the i^{th} user's channel matrix $\boldsymbol{H}_i$, $z_{i,j}$ is the j^{th} element of the i^{th} user's vector $\boldsymbol{z}_i \in \mathbb{C}^{M_{R_i} \times 1}$ and $n_{i,j}$ is the noise at the input of the i^{th} user's j^{th} receive antenna. The elements of the vector $\boldsymbol{z}_i$ are zero mean, unit variance i.i.d. complex uniform random variables. The elements of the vector $\boldsymbol{n}_i$ are zero mean complex Gaussian random variables with variance σ_n^2. In the same way as in the case of SMMSE, the vectors $\boldsymbol{z}_i$, $i = 1,\ldots,K$, are auxiliary vectors used only for the design of the matrices $\boldsymbol{F}_{a_i}$.

The columns in the precoding matrix $\boldsymbol{F}_{a_i}$, each corresponding to one receive antenna, are calculated successively using SMMSE in the following way. The corresponding column of the precoding matrix $\boldsymbol{F}_{a_i}$ is equal to the first column of the following matrix:

$$\boldsymbol{F}_{a_{i,j}} = \left(\overline{\boldsymbol{H}}_i^{(j)\ H} \overline{\boldsymbol{H}}_i^{(j)} + \alpha \boldsymbol{I}_{M_T} \right)^{-1} \overline{\boldsymbol{H}}_i^{(j)\ H} \tag{5.23}$$

$\boldsymbol{H}_{\text{aux}} = \boldsymbol{H}$;
for $i = K : 1$
$\quad \left[\begin{array}{cccc} \boldsymbol{P}_1, & \ldots & \boldsymbol{P}_i, & \text{mse}_1, \quad \ldots \quad \text{mse}_i \end{array} \right] = \text{SMMSE}\left(\boldsymbol{H}_{aux}\right)$;
$\quad k_i = \arg\min_{k \in S} \text{mse}_k$;
$\quad \boldsymbol{F}_i = \boldsymbol{P}_{k_i}$;
$\quad \boldsymbol{G}_i = \boldsymbol{F}_i^H \boldsymbol{H}_i^H \left(\boldsymbol{H}_i \boldsymbol{F}_i \boldsymbol{F}_i^H \boldsymbol{H}_i^H + \sigma_n^2 \boldsymbol{I}_{M_{R_i}} \right)^{-1}$;
$\quad \text{S} = \text{S} \backslash \{k_i\}$;
$\quad \boldsymbol{H}_{\text{aux}} = \left[\begin{array}{cccccc} \boldsymbol{H}_1^T & \ldots & \boldsymbol{H}_{k_i-1}^T & \boldsymbol{H}_{k_i+1}^T & \ldots & \boldsymbol{H}_K^T \end{array} \right]^T$;
end;
$\boldsymbol{F} = \left[\begin{array}{ccc} \boldsymbol{F}_1 & \ldots & \boldsymbol{F}_K \end{array} \right]$;
$\boldsymbol{G} = \left[\begin{array}{ccc} \boldsymbol{G}_1 & & \\ & \ddots & \\ & & \boldsymbol{G}_K \end{array} \right]$;
$\boldsymbol{B} = \text{lower triangular} \left(\boldsymbol{GHF} \cdot \text{diag} \left([\boldsymbol{GHF}]_{ii}^{-1} \right) \right)$;

Table 5.2: SMMSE THP algorithm.

where $\alpha = M_R \sigma_n^2 / P_T$, P_T is the total transmit power, and σ_n^2 is the variance of a zero mean additive white Gaussian noise. The MSE corresponding to this antenna disregarding the interference from the other antennas collocated at the same user terminal is equal to:

$$\text{mse}_{i,j} = \left[\left(\overline{\boldsymbol{H}}_i^{(j)} \overline{\boldsymbol{H}}_i^{(j)\ H} + \alpha \boldsymbol{I}_{M_R - M_{M_{R_i}} + 1} \right)^{-1} \right]_{1,1} \tag{5.24}$$

where the index $1,1$ denotes the matrix element and the total per antenna MSE of the i^{th} user is:

$$\text{mse}_i = \sum_j \text{mse}_{i,j} \tag{5.25}$$

From the SVD of $\boldsymbol{H}_i \boldsymbol{F}_{a_i} = \boldsymbol{U}_i \boldsymbol{\Sigma}_i \boldsymbol{V}_i^H$ the matrix $\boldsymbol{F}_{b_i}$ is calculated as $\boldsymbol{F}_{b_i} = \boldsymbol{V}_i \boldsymbol{\Phi}_i$, where $\boldsymbol{\Phi}_i$ is the i^{th} user's power loading matrix. As in the case of SMMSE, the choice of $\boldsymbol{\Phi}_i$ will depend on the specific optimization criterion.

The combination of SMMSE and THP (SMMSE THP) is performed by successively calculating SMMSE, the reordering of users, and in the end precoding with THP. In every step we use a heuristic approach and minimize the total per antenna MSE of each user. The whole SMMSE THP algorithm is summarized in Table 5.2.

In Table 5.2, we use the following notation: $\text{SMMSE}\left(\bullet\right)$ is the SMMSE function as explained in the previous section, $\boldsymbol{P}_k$ is an auxiliary matrix where we store the precoding matrices generated using SMMSE, $\mathcal{S}$ is a set of indices of the users to be processed, $\boldsymbol{G}_i$ is the i^{th} user's demodulation matrix and $\boldsymbol{B}$ is the THP feedback matrix. In each step we find the user with the minimum total per antenna MSE and place it as the last one.

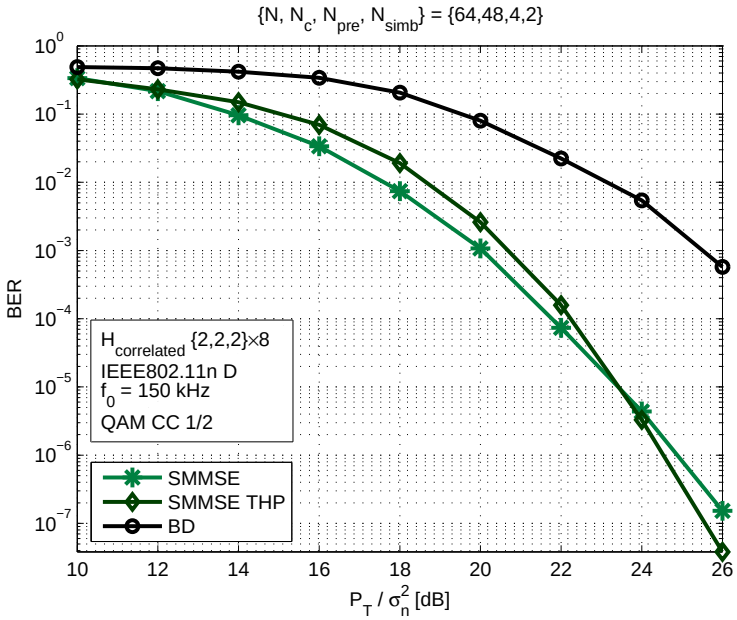

Figure 5.7: BER performance of SMMSE, SMMSE THP and BD as a function of the SNR.

Afterwards, we form the new combined channel matrix $\boldsymbol{H}_{\text{aux}}$ without this user's channel matrix $\boldsymbol{H}_{k_i}$. We repeat these steps until the combined channel matrix is empty.

We compare the performance of systems employing SMMSE, SMMSE THP and BD. To do so, we take into account a purely stochastic channel $\boldsymbol{H}_w$ and a frequency selective MIMO channel with the power delay profile as defined by IEEE802.11n - D with non-line of sight conditions [60]. We assume data transmission using an OFDM system with $N = 64$ order discrete Fourier transform (DFT), subcarrier spacing of 150 kHz and cyclic prefix $N_{\text{pre}} = 4$ samples long. The data is encoded using the convolutional code rate 1/2 $(561, 753)_{\text{oct}}$. After coding the data is mapped using QAM or 16QAM modulation. Coded and modulated symbols are transmitted using $N_c = 48$ subcarriers and $N_{\text{symb}} = 2$ OFDM symbols.

We also consider the antenna correlation at the BS and UTs. Antenna correlation is modeled in the delay domain using the Kronecker model. The channel of each user's l^{th} path component is modeled as described in Section 2.2.1.

In Figure 5.7 we compare the BER performance of SMMSE, SMMSE THP and BD. By introducing MUI, SMMSE provides a higher diversity and array gain than BD. SMMSE THP has a higher diversity gain than SMMSE and outperforms SMMSE at high SNRs. Therefore, it is an attractive solution for high SNRs and high data rates.

In Figure 5.8 we compare the information rate of SMMSE, SMMSE THP and BD.

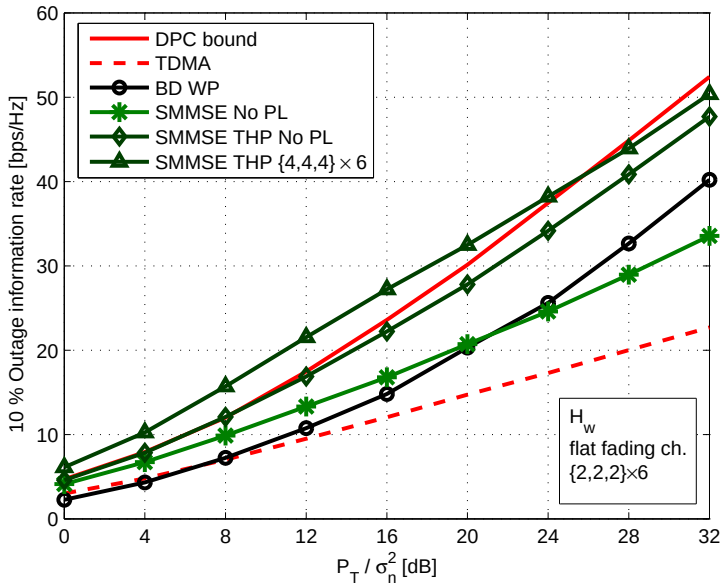

Figure 5.8: 10 % outage information rate of SMMSE, BD and SMMSE THP in bps/Hz as a function of the SNR.

SMMSE provides higher information rate than BD at low SNRs, while BD has higher information rate than SMMSE at high SNRs. By combining SMMSE and THP we approach the sum-rate capacity of the broadcast channels at the low SNRs. We also show in this figure the capacity of a SMMSE THP system with antenna configuration $\{4, 4, 4\} \times 6$. Unlike BD or SO THP, SMMSE and SMMSE THP do not have the dimensionality problem, i.e. the total number of receive antennas at the user terminals can be larger than the number of antennas at the base station. However, the number of users in the system K has to be $K \leq \text{rank}(\boldsymbol{H})$. Therefore, the SMMSE and SMMSE THP information rate improves as the number of antennas at the user terminals increases.

5.5.3 SMMSE successive interference cancellation decoding

As it was said before, in order to achieve a linear increase of the MU MIMO system capacity with the number of antennas we need to spatially multiplex users and multiple data streams to each user. A high throughput on the multi-user uplink can be achieved via an MMSE receiver with successive interference cancellation (SIC), [9]. However, as before, this introduces a loss if we try to mitigate the interference between the data streams transmitted from two closely spaced antennas located at the same user terminal. Techniques, like V-BLAST that transmit independent data streams from all or just a group of antennas, are suboptimum since they do not allow the full coordination of processing

at the antennas at the user terminals. In order to improve the system performance we can use the same approach as in SMMSE. Here, a new algorithm is introduced that deals with this problem analogous to SMMSE by successively calculating the rows of the receive matrix for each of the transmit antennas separately [30]. By applying SIC we additionally improve the diversity, similar to the SMMSE THP, but with one difference. On the uplink we do not need to use a modulo operator which gives a small advantage to SIC over THP.

The SMMSE SIC decoding filter is derived from the linear MMSE receive filter optimization by neglecting the contribution of interference between the signals from one users antenna array to this user's MSE and the influence of the previously decoded users. Let us assume that the users are ordered in such a way that the first user is decoded first, then the second one, etc. When the users' precoding matrices on the uplink $\boldsymbol{Q}_i$, $i = 1, \ldots, K$, are defined, the receive MMSE filter $\boldsymbol{D}_i$ is obtained using the following optimization

$$\boldsymbol{D}_i = \arg\min_{\boldsymbol{D}_i} \mathrm{E}\left\{ \left\| \boldsymbol{D}_i \left(\overline{\boldsymbol{H}}_i^T \overline{\boldsymbol{Q}}_i \overline{\boldsymbol{x}}_i + \boldsymbol{n} \right) - \boldsymbol{x}_i \right\|_{\mathcal{F}}^2 \right\}. \tag{5.26}$$

The matrices $\overline{\boldsymbol{H}}_i$, $\overline{\boldsymbol{Q}}_i$ and the vector $\overline{\boldsymbol{x}}_i$ are defined in this case as

$$\overline{\boldsymbol{H}}_i = \begin{bmatrix} \boldsymbol{H}_i \\ \boldsymbol{H}_{i+1} \\ \vdots \\ \boldsymbol{H}_K \end{bmatrix}, \; \overline{\boldsymbol{Q}}_i = \begin{bmatrix} \boldsymbol{Q}_i & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_{i+1} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Q}_K \end{bmatrix} \text{ and } \; \overline{\boldsymbol{x}}_i = \begin{bmatrix} \boldsymbol{x}_i \\ \boldsymbol{x}_{i+1} \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix}$$

where $\boldsymbol{H}_i$, $\boldsymbol{Q}_i$ and $\boldsymbol{x}_i$ are the i^{th} user's channel matrix, uplink precoding matrix and data vector, respectively.

In equation (5.26) we have included the processing at the user terminals in the optimization criterion. However, in a multi-user scenario, the user terminals estimate the effective channel on the downlink that includes also the processing performed at the base station. Since the processing at the base station is not necessarily the same on the uplink and the downlink, it is reasonable to assume that the user terminals do not have the exact channel state information. We can distinguish two situations. In the first case the users transmit using one of the techniques that do not require CSI at the transmitter. In the second case the base station generates the optimum precoding matrices $\boldsymbol{Q}_i$ and then feedforwards them to the user terminals. Having in mind the limited power available at the user terminals, the diversity gains are more desirable than spatial multiplexing gains.

Therefore, we assume that users transmit data either using STCs or using the dominant right singular vectors of the users' uplink effective channels. In either case the decoding matrix $\boldsymbol{D}_a$ is generated at the base station row by row under the assumption that $\boldsymbol{Q}_i = \boldsymbol{I}_{M_{R_i}}$, $\forall i$, and in the following we will always use this assumption.

In the same way as for the downlink, the interference from the co-channel users to the signal transmitted from the i^{th} user's j^{th} antenna is suppressed independently from the other collocated antennas. Therefore, the j^{th} row of the i^{th} user's decoding matrix $\boldsymbol{D}_{a_i}$, corresponding to the i^{th} user's j^{th} transmit antenna is equal to the first row of the matrix $\boldsymbol{D}_{a_{i,j}}$ which is designed such that

$$\boldsymbol{D}_{a_{i,j}} = \arg\min_{\boldsymbol{D}_{a_{i,j}}} \mathrm{E}\left\{ \left\| \boldsymbol{D}_{a_{i,j}} \left(\overline{\boldsymbol{H}}_i^{(j)\ T} \overline{\boldsymbol{z}}_i^{(j)} + \boldsymbol{n} \right) - \overline{\boldsymbol{z}}_i^{(j)} \right\|_{\mathcal{F}}^2 \right\},\ \forall i, j. \tag{5.27}$$

The matrix $\overline{\boldsymbol{H}}_i^{(j)}$ and the vector $\overline{\boldsymbol{z}}_i^{(j)}$ are defined in this case as

$$\overline{\boldsymbol{H}}_i^{(j)} = \begin{bmatrix} \boldsymbol{h}_{i,j}^T \\ \boldsymbol{H}_{i+1} \\ \vdots \\ \boldsymbol{H}_K \end{bmatrix} \text{ and } \overline{\boldsymbol{z}}_i^{(j)} = \begin{bmatrix} z_{i,j} \\ \boldsymbol{z}_{i+1} \\ \vdots \\ \boldsymbol{z}_K \end{bmatrix}$$

where $\boldsymbol{h}_{i,j}^T$ is the j^{th} row of the i^{th} user's channel matrix $\boldsymbol{H}_i$, and $z_{i,j}$ is the j^{th} element of the i^{th} user's auxiliary vector $\boldsymbol{z}_i \in \mathbb{C}^{M_{R_i} \times 1}$, where $\boldsymbol{z}_i = \boldsymbol{Q}_i \boldsymbol{x}_i$. The vector $\boldsymbol{n} \in \mathbb{C}^{M_T \times 1}$ is the noise at the input of the receive antenna array at the base station. The elements of the vector $\boldsymbol{z}_i$ are assumed for the same reason as in the case of SMMSE to be zero mean, unit variance i.i.d. complex uniform random variables. The elements of the vector $\boldsymbol{n}$ are complex Gaussian random variables with zero mean and variance σ_n^2.

The rows in the receive matrix $\boldsymbol{D}_{a_i}$, each corresponding to one transmit antenna, are calculated successively. The j^{th} row of the receive matrix $\boldsymbol{D}_{a_i}$ is equal to the first row of the following matrix:

$$\boldsymbol{D}_{a_{i,j}} = \overline{\boldsymbol{H}}_i^{(j)\ *} \left(\overline{\boldsymbol{H}}_i^{(j)\ T} \overline{\boldsymbol{H}}_i^{(j)\ *} + \sigma_n^2 \boldsymbol{I}_{M_T} \right)^{-1} \tag{5.28}$$

where σ_n^2 is the variance of the zero-mean additive white Gaussian noise at the input of one receive antenna.

In order to define the ordering of users we will use the same heuristic approach as for SMMSE THP. The mean-square error (MSE) corresponding to the j^{th} transmit antenna

$\boldsymbol{H}_{\text{aux}} = \boldsymbol{H}$;
for $i = 1 : K$
$\quad \left[\begin{array}{cccc} \boldsymbol{P}_1, & \ldots & \boldsymbol{P}_i, & \text{mse}_1, \quad \ldots \quad \text{mse}_i \end{array} \right] = \text{SMMSE}_{\text{D}}\left(\boldsymbol{H}_{aux}\right)$;
$\quad k_i = \arg\min_{k \in S} \text{mse}_k$;
$\quad \boldsymbol{D}_i = \boldsymbol{P}_{k_i}$;
$\quad \text{S} = \text{S} \backslash \{k_i\}$;
$\quad \boldsymbol{H}_{\text{aux}} = \left[\begin{array}{cccccc} \boldsymbol{H}_1^T & \ldots & \boldsymbol{H}_{k_i-1}^T & \boldsymbol{H}_{k_i+1}^T & \ldots & \boldsymbol{H}_K^T \end{array} \right]^T$;
end;
$\boldsymbol{D} = \left[\begin{array}{ccc} \boldsymbol{D}_1^T & \ldots & \boldsymbol{D}_K^T \end{array} \right]^T$;

Table 5.3: SMMSE SIC algorithm.

of the i^{th} user is equal to

$$\text{mse}_{i,j} = \sigma_n^2 \left[\left(\overline{\boldsymbol{H}}_i^{(j)\ T} \overline{\boldsymbol{H}}_i^{(j)\ *} + \sigma_n^2 \boldsymbol{I}_{M_T} \right)^{-1} \right]_{1,1} \tag{5.29}$$

Let us define the total mean square error of the i^{th} user as

$$\text{mse}_i = \sum_{j=1}^{M_{R_i}} \text{mse}_{i,j} \tag{5.30}$$

We look for the user with the minimum mse_i, demodulate its data and then subtract the reconstructed signal from the received signal. Afterwards, we form the new combined channel matrix $\overline{\boldsymbol{H}}_i^{(j)}$ without this user's channel matrix and use it in equation (5.28). We repeat these steps until the combined channel matrix $\overline{\boldsymbol{H}}_i^{(j)}$ is empty.

The whole SMMSE SIC algorithm is given in Table 5.3. We use the following notation: $\text{SMMSE}_{\text{D}}\left(\bullet\right)$ is the SMMSE decoding function as it was previously described, $\boldsymbol{P}_k$ is an auxiliary matrix where we store the decoding matrices generated using a receive SMMSE filter and $\mathcal{S}$ is a set of indices of the users to be processed. In each step we find the user with the minimum total per antenna MSE and place it as the first one. Afterwards, we form the new combined channel matrix $\boldsymbol{H}_{\text{aux}}$ without this user's channel matrix $\boldsymbol{H}_{k_i}$. We repeat these steps until the combined channel matrix is empty.

In Figure 5.9 we compare the performance of SMMSE SIC and V-BLAST in a system with the antenna configuration $\{2, 2\} \times 6$. In case of SMMSE SIC each user terminal can transmit data using either STC or on the dominant singular vector of the effective channel. In the first case, each user assumes no CSI at the transmitter and encodes the data using an Alamouti space-time code. In the second case, the base station calculates and sends back to the i^{th} user terminal the dominant right singular vector of the i^{th} user's effective channel $\boldsymbol{D}_{a_i} \boldsymbol{H}_i^T$, $\forall i$. The matrices $\boldsymbol{D}_{b_i}$ and $\boldsymbol{Q}_i$ are chosen as the dominant left and right

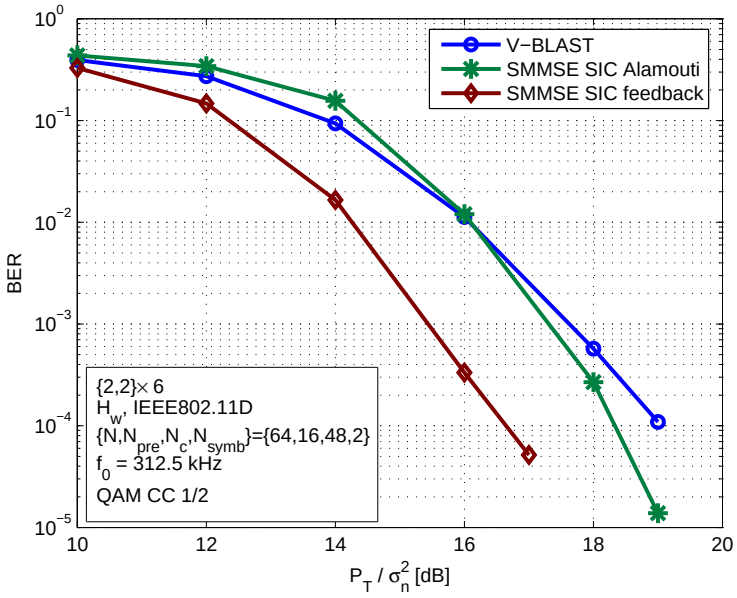

Figure 5.9: BER performance of V-BLAST and SMMSE SIC in combination with Alamouti STC and feddback of user uplink precoding vectors.

singular vectors of the matrix $\boldsymbol{D}_{a_i}\boldsymbol{H}_i^T$, respectively. When SMMSE SIC is used at the base station the data is encoded using the QAM modulation. In case of V-BLAST since we have two data streams per user, each data stream is modulated using BPSK. Information data is encoded using convolutional code rate 1/2. As we can see from the figure, SMMSE SIC in both cases provides higher diversity than V-BLAST. When we transmit data on the left dominant singular vector of each users' effective channel, in the system with this antenna configuration we obtain an additional 2 dB SNR gain.

5.6 Regularized block diagonalization

As it was shown in the previous section, SMMSE and SMMSE THP manage to provide higher array and diversity gain and in simulations reach the sum-rate capacity of the broadcast channel at the low SNRs. However, they fail to extract full array gain, diversity gain and sum-rate capacity.

In this section we will use the same approach introduced in the Section 5.4 to design the precoding matrix. The approach used for the design of the precoding matrix is general and the resulting algorithm can address several optimization criteria with an arbitrary number of antennas at the user terminals. This has been achieved by designing the precoding matrices in two steps. In the first step we minimize the overlap of the row

spaces spanned by the effective channel matrices of different users. To this end we separate the transmission to the different users using a new cost function that includes multi-user interference suppression and the avoidance of noise enhancement. In the next step, we optimize the system performance with respect to the specific optimization criterion assuming a set of parallel single-user MIMO channels. The new technique is called regularized block diagonalization (RBD) since at high signal-to-noise ratios (SNRs) and under the condition that the total number of antennas at the user terminals is less or equal to the number of antennas at the base station, the effective combined channel matrix is block diagonal. In addition to this linear technique we also present several variants that extract full gain provided by multiple antennas.

5.6.1 RBD precoding

It has been shown that the best performance for the single antenna receivers, $r = K = M_R$, is achieved by designing the precoding matrix $\boldsymbol{F}$ using the MMSE criterion [22]. Let us define the precoding matrix $\boldsymbol{F}$ as $\boldsymbol{F} = \beta \boldsymbol{F}_a$. The precoding matrix $\boldsymbol{F}_a$ and the scaling factor β result from the following optimization

$$\boldsymbol{F}_a = \min_{\boldsymbol{F}_a} E\left\{ \left\| \beta^{-1}\boldsymbol{y} - \boldsymbol{x} \right\|_{\mathcal{F}}^2 \right\} = \min_{\boldsymbol{F}_a} E\left\{ \left\| \left(\boldsymbol{H}\boldsymbol{F}_a - \boldsymbol{I}_{M_R} \right) \boldsymbol{x} \right\|_{\mathcal{F}}^2 + \frac{\left\| \boldsymbol{n} \right\|_{\mathcal{F}}^2}{\beta^2} \right\} \tag{5.31}$$

where the parameter β is used to fulfill the transmit power constraint. This can be interpreted as choosing the matrix $\boldsymbol{F}_a$ to minimize the Frobenius norm of the off-diagonal elements of the effective channel $\boldsymbol{H}\boldsymbol{F}_a$ at the high SNRs, while the elements on the main diagonal should converge to 1.

Let us rewrite here the matrix $\widetilde{\boldsymbol{H}}_i$ from equation (5.6)

$$\widetilde{\boldsymbol{H}}_i = \left[\begin{array}{ccccccc} \boldsymbol{H}_1^T & \cdots & \boldsymbol{H}_{i-1}^T & \boldsymbol{H}_{i+1}^T & \cdots & \boldsymbol{H}_K^T \end{array} \right]^T \in \mathbb{C}^{(M_R - M_{R_i}) \times M_T} \tag{5.32}$$

First we address the design of the matrix $\boldsymbol{F}_a$. The effective combined channel matrix of all users after the precoding is equal to

$$\boldsymbol{H}\boldsymbol{F}_a = \left[\begin{array}{cccc} \boldsymbol{H}_1\boldsymbol{F}_{a_1} & \boldsymbol{H}_1\boldsymbol{F}_{a_2} & \cdots & \boldsymbol{H}_1\boldsymbol{F}_{a_K} \\ \boldsymbol{H}_2\boldsymbol{F}_{a_1} & \boldsymbol{H}_2\boldsymbol{F}_{a_2} & \cdots & \boldsymbol{H}_2\boldsymbol{F}_{a_K} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{H}_K\boldsymbol{F}_{a_1} & \boldsymbol{H}_K\boldsymbol{F}_{a_2} & \cdots & \boldsymbol{H}_K\boldsymbol{F}_{a_K} \end{array} \right] \tag{5.33}$$

where the i^{th} user's effective channel is given by $\boldsymbol{H}_i\boldsymbol{F}_{a_i}$ and the interference that this user generates to the other users is determined by $\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}$.

The Frobenius norm of the matrix $\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}$ is related to the level of the overlap of the row subspaces of the effective channels of different users $\boldsymbol{H}_i\boldsymbol{F}_{a_i}, i = 1, \ldots, K$. Analogous to (5.31), the matrix $\boldsymbol{F}_a$ is chosen such that the off-diagonal block matrices converge to zero as the SNR increases. The specific optimization of the block matrices on the main diagonal is addressed in the next step where we design the matrix $\boldsymbol{F}_b$. We rewrite the optimization criterion in (5.31) as

$$\boldsymbol{F}_a = \min_{\boldsymbol{F}_a} E\left\{\sum_{i=1}^{K}\left\|\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}\right\|_{\mathcal{F}}^2 + \frac{\|\boldsymbol{n}\|_{\mathcal{F}}^2}{\beta^2}\right\} \tag{5.34}$$

where vector $\boldsymbol{n} \in \mathbb{C}^{M_R\times 1}$ contains the samples of a zero mean additive white Gaussian noise at the input of the receive antennas and β is chosen to fulfill the transmit power constraint $\beta^2\left\|\boldsymbol{F}_a\boldsymbol{F}_b\boldsymbol{x}\right\|_{\mathcal{F}}^2 \leq P_T$. We assume that the noise at the input of different receive antennas is uncorrelated with the same variance σ_n^2. Therefore, the matrices $\boldsymbol{F}_{a_i}$ are obtained by minimizing the Frobenius norm of the effective channels between different users and the Frobenius norm of the scaled noise vector. The parameter β is a function of $\boldsymbol{F}_a$ and $\boldsymbol{F}_b$, and it influences the receive SNR, [22].

The power of the transmission of the i^{th} user into the subspace spanned by the rows of the other users' channel matrices is $\left\|\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}\right\|_{\mathcal{F}}^2$. The matrix $\boldsymbol{F}_a$ is designed to minimize the power of this interference plus noise. The optimization crtierion in equation (5.34) can be rewritten similarly as

$$\begin{aligned}
\boldsymbol{F}_a = &\ \min_{\boldsymbol{F}_a} E\left\{\sum_{i=1}^{K}\mathrm{tr}\left(\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\widetilde{\boldsymbol{H}}_i^H\right) + \frac{\|\boldsymbol{n}\|_{\mathcal{F}}^2}{\beta^2}\right\} \\
= &\ \min_{\boldsymbol{F}_a}\sum_{i=1}^{K}\left(\mathrm{tr}\left(\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\widetilde{\boldsymbol{H}}_i^H\right) + E\left\{\frac{\|\boldsymbol{n}\|_{\mathcal{F}}^2}{\beta^2}\right\}\right) \\
= &\ \min_{\boldsymbol{F}_a}\sum_{i=1}^{K}\left(\mathrm{tr}\left(\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\widetilde{\boldsymbol{H}}_i^H\right) + \frac{M_R\sigma_n^2}{\beta^2}\right)
\end{aligned} \tag{5.35}$$

The parameter β is chosen to set the total transmit power to P_T

$$\beta^2 E\left\{\sum_{i=1}^{K}\left\|\boldsymbol{F}_{a_i}\boldsymbol{F}_{b_i}\boldsymbol{x}_i\right\|_{\mathcal{F}}^2\right\} \leq P_T \tag{5.36}$$

where the vector $\boldsymbol{x} = \left[\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_K^T\right]^T \in \mathbb{C}^{r\times 1}$ is a collection of all data vectors $\boldsymbol{x}_i$ trans-

mitted to the users. We can write

$$\beta^2 E\left\{\sum_{i=1}^{K}\|\boldsymbol{F}_{a_i}\boldsymbol{F}_{b_i}\boldsymbol{x}_i\|_{\mathcal{F}}^2\right\}=\beta^2 E\left\{\sum_{i=1}^{K}\mathrm{tr}\left(\boldsymbol{F}_{a_i}\boldsymbol{F}_{b_i}\boldsymbol{x}_i\boldsymbol{x}_i^H\boldsymbol{F}_{b_i}^H\boldsymbol{F}_{a_i}^H\right)\right\} \tag{5.37}$$

If we assume that the users' data are uncorrelated, uniformly distributed, with zero mean and unit power, $E\left\{\boldsymbol{x}_i\boldsymbol{x}_i^H\right\}=\boldsymbol{I}_{r_i}$ and that $\boldsymbol{F}_{b_i}$ is unitary, which is justified if we assume that we use all eigenmodes of the effective channel $\boldsymbol{H}_i\boldsymbol{F}_{a_i}$ for transmission with an equal power distribution, we have

$$\begin{aligned}\beta^2 E\left\{\sum_{i=1}^{K}\mathrm{tr}\left(\boldsymbol{F}_{a_i}\boldsymbol{F}_{b_i}\boldsymbol{x}_i\boldsymbol{x}_i^H\boldsymbol{F}_{b_i}^H\boldsymbol{F}_{a_i}^H\right)\right\} = &\quad \beta^2\textstyle\sum_{i=1}^{K}\mathrm{tr}\left(\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\right)\leq P_T\\ \Rightarrow &\quad \beta^2=\tfrac{P_T}{\sum_{i=1}^{K}\mathrm{tr}\left(\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\right)}.\end{aligned} \tag{5.38}$$

From the previous equation we see that the matrix $\boldsymbol{F}_a$ is a function of the matrix $\boldsymbol{F}_b$, and vice versa, the matrix $\boldsymbol{F}_b$ is a function of the matrix $\boldsymbol{F}_a$. In order to design the matrix $\boldsymbol{F}_a$ independently from the matrix $\boldsymbol{F}_b$ we have initially assumed that the matrix $\boldsymbol{F}_b$ is unitary. This assumtion means that when we design the matrix $\boldsymbol{F}_a$ we assume that each user uses all singular vectors of the effective channel $\boldsymbol{H}_i\boldsymbol{F}_{a_i}$ for data transmission with equal priority, i.e., he uses all of his available subspace. If later, a user does not use all of the available subspace, the performance can be improved by iterating the closed form solution for RBD as it will be explained later.

By substituting β^2 from (5.38) in (5.35), we get

$$\begin{aligned}&\min_{\boldsymbol{F}_a}\sum_{i=1}^{K}\left(\mathrm{tr}\left(\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\widetilde{\boldsymbol{H}}_i^H\right)+\frac{M_R\sigma_n^2}{P_T}\mathrm{tr}\left(\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\right)\right)\\ &=\min_{\boldsymbol{F}_a}\sum_{i=1}^{K}\mathrm{tr}\left(\widetilde{\boldsymbol{H}}_i^H\widetilde{\boldsymbol{H}}_i\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H+\frac{M_R\sigma_n^2}{P_T}\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\right)\\ &=\min_{\boldsymbol{F}_a}\sum_{i=1}^{K}\mathrm{tr}\left(\left(\widetilde{\boldsymbol{H}}_i^H\widetilde{\boldsymbol{H}}_i+\frac{M_R\sigma_n^2}{P_T}\boldsymbol{I}_{M_T}\right)\boldsymbol{F}_{a_i}\boldsymbol{F}_{a_i}^H\right)\\ &=\min_{\boldsymbol{F}_a}\sum_{i=1}^{K}\mathrm{tr}\left(\boldsymbol{F}_{a_i}^H\left(\widetilde{\boldsymbol{H}}_i^H\widetilde{\boldsymbol{H}}_i+\frac{M_R\sigma_n^2}{P_T}\boldsymbol{I}_{M_T}\right)\boldsymbol{F}_{a_i}\right).\end{aligned} \tag{5.39}$$

If the SVD of $\widetilde{\boldsymbol{H}}_i$ is given by

$$\widetilde{\boldsymbol{H}}_i=\widetilde{\boldsymbol{U}}_i\widetilde{\boldsymbol{\Sigma}}_i\widetilde{\boldsymbol{V}}_i^H\in\mathbb{C}^{(M_R-M_{R_i})\times M_T} \tag{5.40}$$

then

$$\min_{\boldsymbol{F}_a} \sum_{i=1}^{K} \mathrm{tr}\left(\boldsymbol{F}_{a_i}^H \left(\widetilde{\boldsymbol{H}}_i^H \widetilde{\boldsymbol{H}}_i + \frac{M_R \sigma_n^2}{P_T} \boldsymbol{I}_{M_T} \right) \boldsymbol{F}_{a_i} \right)$$

$$= \min_{\boldsymbol{F}_a} \sum_{i=1}^{K} \mathrm{tr}\left(\boldsymbol{F}_{a_i}^H \widetilde{\boldsymbol{V}}_i \left(\widetilde{\boldsymbol{\Sigma}}_i^T \widetilde{\boldsymbol{\Sigma}}_i + \frac{M_R \sigma_n^2}{P_T} \boldsymbol{I}_{M_T} \right) \widetilde{\boldsymbol{V}}_i^H \boldsymbol{F}_{a_i} \right) \tag{5.41}$$

Using the results from [61], the expression in (5.41) is minimized by choosing $\boldsymbol{F}_{a_i}$ as

$$\boldsymbol{F}_{a_i} = \boldsymbol{M}_{a_i} \boldsymbol{\Phi}_{a_i} \tag{5.42}$$

and

$$\boldsymbol{M}_{a_i} = \widetilde{\boldsymbol{V}}_i \in \mathbb{C}^{M_T \times M_T}, \tag{5.43}$$

Then (5.41) reduces to

$$\min_{\boldsymbol{\Phi}_{a_i}} \sum_{i=1}^{K} \mathrm{tr}\left(\left(\widetilde{\boldsymbol{\Sigma}}_i^T \widetilde{\boldsymbol{\Sigma}}_i + \frac{M_R \sigma_n^2}{P_T} \boldsymbol{I}_{M_T} \right) \boldsymbol{\Phi}_{a_i}^2 \right) \tag{5.44}$$

Since the elements on the main diagonal of $\left(\widetilde{\boldsymbol{\Sigma}}_i^T \widetilde{\boldsymbol{\Sigma}}_i + \frac{M_R \sigma_n^2}{P_T} \boldsymbol{I}_{M_T} \right)$ are greater than zero, the matrices $\boldsymbol{\Phi}_{a_i}$ have to be positive definite if we would like to find a nontrivial solution. Using the results from [8] (Appendix F), the solution to (5.44) is

$$\boldsymbol{\Phi}_{a_i} = \left(\widetilde{\boldsymbol{\Sigma}}_i^T \widetilde{\boldsymbol{\Sigma}}_i + \frac{M_R \sigma_n^2}{P_T} \boldsymbol{I}_{M_T} \right)^{-1/2} \tag{5.45}$$

so that the minimum and maximum eigenvalues of (5.44) coincide. Note that there is no additional constraint regarding the values of $\boldsymbol{\Phi}_{a_i}$ except that it is positive definite.

From equations (5.42),(5.43) and (5.45) we can see that the cost function in equation (5.34) is minimized if each user transmits in the space spanned by the combined matrix of all other users with the power that is inversely proportional to the singular values of the combined channel matrix of these users $\widetilde{\boldsymbol{H}}_i$. As a result, at high SNRs and $M_R \leq M_T$ each user transmits only in the null space of all other users as in BD.

The effective combined channel matrix of all users after the first step of the precoding is equal to $\boldsymbol{H}\boldsymbol{F}_a \in \mathbb{C}^{M_R \times K M_T}$. For high SNRs and $M_R \leq M_T$, this matrix will also be block diagonal. We can now apply any other previously defined SU MIMO technique on the i^{th} user's effective channel matrix $\boldsymbol{H}_i \boldsymbol{F}_{a_i}$.

After we have suppressed MUI using $\boldsymbol{F}_a$ we optimize the system performance by op-

timizing each SU MIMO channel separately under the total transmit power constraint. Now we use the results presented for the generalized design of SU MIMO precoding and decoding matrices [8]. The matrix $\boldsymbol{F}_{b_i}$ has the form $\boldsymbol{F}_{b_i} = \boldsymbol{M}_{b_i} \cdot \boldsymbol{\Phi}_{b_i}$ where $\boldsymbol{M}_{b_i} \in \mathbb{C}^{M_T \times M_T}$ is a unitary matrix and $\boldsymbol{\Phi}_{b_i} \in \mathbb{R}^{M_T \times M_T}$ is a diagonal power loading matrix, with elements on the main diagonal greater or equal to zero, [8]. If the BS transmits r_i data streams to the i^{th} user, then only r_i elements on the diagonal of $\boldsymbol{\Phi}_{b_i}$ will be greater than zero. The optimum $\boldsymbol{M}_{b_i}$ is obtained from the SVD of the i^{th} user's effective channel

$$\boldsymbol{H}_i \boldsymbol{F}_{a_i} = \boldsymbol{U}_i \boldsymbol{\Sigma}_i \boldsymbol{V}_i^H \tag{5.46}$$

as $\boldsymbol{M}_{b_i} = \boldsymbol{V}_i$ [8], [9]. The choice of the power loading matrix $\boldsymbol{\Phi}_{b_i}$ depends on the optimization criteria.

Power loading

Let us define the matrix $\boldsymbol{\Sigma}_e$ as

$$\boldsymbol{\Sigma}_e = \begin{bmatrix} \boldsymbol{\Sigma}_1^{(r_1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2^{(r_2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_K^{(r_K)} \end{bmatrix} \in \mathbb{R}^{r \times r}, \tag{5.47}$$

where $\boldsymbol{\Sigma}_i^{(r_i)}$ is a diagonal matrix containing the largest r_i values of $\boldsymbol{\Sigma}_i$ which is obtained from the SVD given in equation (5.46). By applying the water-pouring (WP) algorithm on this matrix we can maximize the system information rate.

The maximum signal-to-noise ratio (SNR) at the user terminals is obtained by transmitting only over the strongest eigenmode, i.e., $r_i = 1, i = 1, \ldots, K$.

To minimize the average BER in the system we introduce a new power loading algorithm. Here we rely on the fact that the overall system BER performance is limited by the performance of the weakest user, i.e., the one with the highest BER. Therefore, we assign the power to users in such a way as to balance the SNR over all users, average the BER performance and improve the system BER. Thus, we first introduce a power loading matrix for which we use the term "MMSE" power loading (PL):

$$\boldsymbol{\Phi}_{\text{MMSE}} = \left(\boldsymbol{\Sigma}_e^T \boldsymbol{\Sigma}_e + \alpha \boldsymbol{I}_r\right)^{-1} \boldsymbol{\Sigma}_e^T = \left(\boldsymbol{\Sigma}_e^2 + \alpha \boldsymbol{I}_r\right)^{-1} \boldsymbol{\Sigma}_e \in \mathbb{R}^{r \times r} \tag{5.48}$$

where $\alpha = \sigma_n^2 M_R / P_T$. Additive noise is the dominant source of interference at the low SNRs. Thus, no power loading (No PL)

$$\boldsymbol{\Phi}_{b_i} = \begin{bmatrix} \boldsymbol{I}_{r_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{M_T \times M_T},$$

reduces the BER, while at high SNRs, MMSE power loading yields a lower BER. For this reason, we introduce an improved diversity power loading which adapts from no power loading at low SNRs to MMSE power loading at high SNRs. Therefore, we define the improved diversity (impD) precoding matrix as

$$\boldsymbol{\Phi}_{\text{impD}} = \left(\boldsymbol{\Sigma}_e^2 + \alpha \boldsymbol{I}_r\right)^{-1} \boldsymbol{\Sigma}_e \left(\boldsymbol{I}_r - \left(\boldsymbol{\Sigma}_e^2 + \alpha \boldsymbol{I}_r\right)^{-1} \alpha\right) + \left(\boldsymbol{\Sigma}_e^2 + \alpha \boldsymbol{I}_r\right)^{-1} \alpha. \tag{5.49}$$

These power loading matrices are used to redistribute the total transmit power over the eigenmodes. After we have generated the matrices $\boldsymbol{F}_a$ and $\boldsymbol{F}_b$, the parameter β is used to set the total transmit power to P_T, i.e., $\beta^2 = P_T / \left\|\boldsymbol{F}_a \boldsymbol{F}_b \boldsymbol{x}\right\|^2$.

5.6.2 RSO THP precoding

In this section we introduce a combination of RBD and THP which we call regularized successive optimization THP (RSO THP). We combine RBD and THP in order to reach the maximum sum rate capacity of the Gaussian broadcast channel.

We assume that a certain optimum/suboptimum ordering of users is done beforehand. Let us define the previous $i-1$ users' combined channel matrix as

$$\widehat{\boldsymbol{H}}_i = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_{i-1} \end{bmatrix}$$

and its corresponding SVD as

$$\widehat{\boldsymbol{H}}_i = \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{\Sigma}}_i \widehat{\boldsymbol{V}}_i^H. \tag{5.50}$$

The optimization of the RSO THP precoding filter is written as:

$$\boldsymbol{F}_a = \min_{\boldsymbol{F}_a} E\left\{ \sum_{i=1}^{K} \left\| \widehat{\boldsymbol{H}}_i \boldsymbol{F}_{a_i} \right\|_{\mathcal{F}}^2 + \frac{\|\boldsymbol{n}\|_{\mathcal{F}}^2}{\beta^2} \right\} \tag{5.51}$$

where $\boldsymbol{n}$ is the vector of the additive noise at the input of all receive antennas and the

parameter β is chosen to fulfill the transmit power constraint (5.36).

The combination of RBD and THP is performed by successively calculating RBD, then the reordering of users, and in the end precoding with THP. Instead of examining all $K!$ possibilities for ordering to maximize the total information rate in the system, we use the same heuristic simplification as in SO THP, where we minimize the information rate difference when we precode all users jointly and when the users are served separately. Using RBD, the modulation matrix for each user is designed in such a way that it lies only in the row space of the effective channel matrices of previous users. As a consequence, only they will generate the interference to this user.

Using RBD, the precoding matrix $\boldsymbol{F}_{a_i}$ of the i^{th} user is calculated as $\boldsymbol{F}_{a_i} = \boldsymbol{M}_{a_i}\boldsymbol{\Phi}_{a_i}$, where

$$\boldsymbol{M}_{a_i} = \widehat{\boldsymbol{V}}_i \quad \text{and} \quad \boldsymbol{\Phi}_{a_i} = \left(\widehat{\boldsymbol{\Sigma}}_i^T\widehat{\boldsymbol{\Sigma}}_i + \frac{M_R\sigma_n^2}{P_T}\boldsymbol{I}_{M_T}\right)^{-1/2}. \tag{5.52}$$

The whole RSO THP algorithm is summarized in Table 5.4. We use the following notation: RBD $(\bullet)$ is RBD as previously explained, $\boldsymbol{P}_k$ is an auxiliary matrix where we store the precoding matrices generated using RBD, $\mathcal{S}$ is a set of indices of the users to be processed, $\boldsymbol{G}_i$ is the i^{th} user's demodulation matrix and $\boldsymbol{B}$ is the THP feedback matrix. Note that the matrices $\boldsymbol{U}_i$ are given in (5.46). In short, we first calculate the information rate that an individual user can achieve assuming there are no other users in the system. Then, we look for the user with the minimum difference between its information rate when it is served alone and its information rate when it is served jointly with other users and generate the precoding matrix of this user. In each step we find the user with the minimum information rate loss and place it as the last one. Afterwards, we form the new combined channel matrix $\boldsymbol{H}_{\text{aux}}$ without this user's channel matrix $\boldsymbol{H}_{k_i}$. We repeat these steps until the combined channel matrix is empty. The order of the users is the reverse of the order in which their precoding matrices are generated. The strictly lower triangular feedback matrix $\boldsymbol{B}$, used in THP precoding, is generated from the effective combined channel matrix after the elements in each row are divided by the elements on the main diagonal, as it can be seen from the last equation in Table 5.4.

The individual users' channel matrices and demodulation matrices are grouped in the matrices $\boldsymbol{H}$ and $\boldsymbol{G}$. The feedback matrix $\boldsymbol{B}$, generated in the last step of the RSO THP algorithm, is now used to precode the users' data streams starting with the data stream of the first user whose precoding matrix $\boldsymbol{F}_1$ was generated as the last one.

Again, by using THP at the transmit side we significantly increase the transmit power

for $i = 1 : K$

$\boldsymbol{H}_i = \boldsymbol{W}_i \boldsymbol{\Sigma}_i \left[\begin{array}{cc} \boldsymbol{V}_i^{(1)} & \boldsymbol{V}_i^{(0)} \end{array} \right]^H$;

$\boldsymbol{F}_{\max,i} = \boldsymbol{V}_i^{(1)}$;

$C_{\max,i} = \log_2 \det \left(\boldsymbol{I} + \sigma_n^{-2} \boldsymbol{H}_i \boldsymbol{F}_{\max,i} \boldsymbol{F}_{\max,i}^H \boldsymbol{H}_i^H \right)$;

end;

$S = \{1, \ldots, K\}$;

$\boldsymbol{H}_{\text{aux}} = \boldsymbol{H}$;

for $i = K : 1$

$\left[\begin{array}{cccccc} \boldsymbol{P}_1, & \ldots & \boldsymbol{P}_i, & \boldsymbol{U}_1, & \ldots & \boldsymbol{U}_i \end{array} \right] = \text{RBD}\left(\boldsymbol{H}_{\text{aux}}\right)$;

for $k = 1 : i$

$C_k = \log_2 \det \left(\boldsymbol{I} + \sigma_n^{-2} \boldsymbol{H}_k \boldsymbol{P}_k \boldsymbol{P}_k^H \boldsymbol{H}_k^H \right)$;

end;

$k_i = \arg\min_{k \in S} \left(C_{\max,k} - C_k \right)$;

$\boldsymbol{F}_i = \boldsymbol{P}_{k_i}$;

$\boldsymbol{G}_i = \boldsymbol{U}_{k_i}^H$;

$S = S \backslash \{k_i\}$;

$\boldsymbol{H}_{\text{aux}} = \left[\begin{array}{cccccc} \boldsymbol{H}_1^T & \ldots & \boldsymbol{H}_{k_i-1}^T & \boldsymbol{H}_{k_i+1}^T & \ldots & \boldsymbol{H}_K^T \end{array} \right]^T$;

end;

$\boldsymbol{F} = \left[\begin{array}{ccc} \boldsymbol{F}_1 & \ldots & \boldsymbol{F}_K \end{array} \right]$;

$\boldsymbol{G} = \left[\begin{array}{ccc} \boldsymbol{G}_1 & & \\ & \ddots & \\ & & \boldsymbol{G}_K \end{array} \right]$;

$\boldsymbol{B} = \text{lower triangular} \left(\boldsymbol{GHF} \cdot \text{diag} \left([\boldsymbol{GHF}]_{ii}^{-1} \right) \right)$;

Table 5.4: Regularized successive optimization THP (RSO THP) algorithm.

and therefore a modulo operation is introduced at the transmitter and the receiver in order to limit the constellation size.

5.6.3 Iterative RBD

The performance of RBD can be further improved by exploiting the row subspace that remains unused after the precoding. In RBD, the precoding matrix for each user is generated under the assumption that the other user terminals use all the available right singular vectors for transmission. However, after the power loading some singular values remain unused and they can be exploited to improve the system performance.

We can identify two cases. In the first case, if the number of the transmitted data streams of the i^{th} user r_i is less than the rank of the i^{th} user's channel matrix, then other users could also transmit in this unused subspace without causing additional interference. In the second case, when $M_T \leq M_R$ and $K \leq M_T$, users must leave a part of their own subspaces unused in order to reduce the overall MUI.

Iterative RBD is defined as a solution to the following optimization problem:

$$\boldsymbol{F}_a^{(l)} = \min_{\boldsymbol{F}_a^{(l)}} E\left\{\sum_{i=1}^{K}\left\|\widetilde{\boldsymbol{H}}_i^{(l)}\boldsymbol{F}_{a_i}^{(l)}\right\|_{\mathcal{F}}^2 + \frac{\|\boldsymbol{n}\|_{\mathcal{F}}^2}{\beta^2}\right\} \tag{5.53}$$

where $\boldsymbol{F}_a^{(l)}$ is the precoding matrix obained in the l^{th} iteration. The modified combined channel matrix of interfering users in the l^{th} iteration is defined as

$$\widetilde{\boldsymbol{H}}_i^{(l)} = \begin{bmatrix} \boldsymbol{H}_1^{(l)} \\ \vdots \\ \boldsymbol{H}_{i-1}^{(l)} \\ \boldsymbol{H}_{i+1}^{(l)} \\ \vdots \\ \boldsymbol{H}_K^{(l)} \end{bmatrix} \in \mathbb{C}^{(r-r_i)\times M_T} \tag{5.54}$$

where $r \leq M_T$ is the total number of the spatially multiplexed data streams and r_i is the number of data stremas transmitted to the i^{th} user. The i^{th} user's effective channel matrix in the l^{th} iteration is equal to:

$$\boldsymbol{H}_i^{(l)} = \boldsymbol{U}_i^{(r_i)\ (l-1)\ H}\boldsymbol{H}_i \tag{5.55}$$

where $\boldsymbol{U}_i^{(r_i)\ (l-1)}$ contains the first r_i vectors of $\boldsymbol{U}_i^{(l-1)}$ which is obtained from the following SVD

$$\boldsymbol{H}_i\boldsymbol{F}_{a_i}^{(l-1)} = \boldsymbol{U}_i^{(l-1)}\boldsymbol{\Sigma}_i^{(l-1)}\boldsymbol{V}_i^{(l-1)\ H}. \tag{5.56}$$

The first r_i vectors of $\boldsymbol{U}_i^{(l-1)}$ correspond to the r_i strongest singular values of $\boldsymbol{H}_i\boldsymbol{F}_{a_i}^{(l-1)}$. Following the analysis from Section 5.6, the solution to the optimization problem given in (5.53) is equal to

$$\boldsymbol{F}_{a_i}^{(l)} = \widetilde{\boldsymbol{V}}_i^{(l)}\left(\widetilde{\boldsymbol{\Sigma}}_i^{(l)\ T}\widetilde{\boldsymbol{\Sigma}}_i^{(l)} + \frac{M_R\sigma_n^2}{P_T}\boldsymbol{I}_{M_T}\right)^{-1/2} \tag{5.57}$$

where M_R is the total number of receive antennas at the user terminals, P_T is the transmit power, σ_n^2 is the variance of the additive noise at the input of the receive antennas, and matrices $\widetilde{\boldsymbol{V}}_i^{(l)}$ and $\widetilde{\boldsymbol{\Sigma}}_i^{(l)}$ are obtained from the following EVD:

$$\widetilde{\boldsymbol{H}}_i^{(l)\ H}\widetilde{\boldsymbol{H}}_i^{(l)} = \widetilde{\boldsymbol{V}}_i^{(l)}\widetilde{\boldsymbol{\Sigma}}_i^{(l)\ T}\widetilde{\boldsymbol{\Sigma}}_i^{(l)}\widetilde{\boldsymbol{V}}_i^{(l)\ H}. \tag{5.58}$$

The whole IRBD algorithm is sumarized in Table 5.5. We assume a random ordering

$\boldsymbol{U}_i^{(0)} = \boldsymbol{I}_{M_{R_i}}, \forall i;$
for $l = 1 : N_{\text{iterat}}$
  for $i = 1 : K$
    $\boldsymbol{H}_i^{(l)} = \boldsymbol{U}_i^{(l-1)\ H} \boldsymbol{H}_i;$
  end;
  for $i = 1 : K$
    $\widetilde{\boldsymbol{H}}_i^{(l)} = \left[\begin{array}{cccccc} \boldsymbol{H}_1^{(l)\ T} & \ldots & \boldsymbol{H}_{i-1}^{(l)\ T} & \boldsymbol{H}_{i+1}^{(l)\ T} & \ldots & \boldsymbol{H}_K^{(l)\ T} \end{array} \right]^T;$
    $\widetilde{\boldsymbol{H}}_i^{(l)\ H} \widetilde{\boldsymbol{H}}_i^{(l)} = \widetilde{\boldsymbol{V}}_i^{(l)} \widetilde{\boldsymbol{\Sigma}}_i^{(l)\ T} \widetilde{\boldsymbol{\Sigma}}_i^{(l)} \widetilde{\boldsymbol{V}}_i^{(l)\ H};$
    $\boldsymbol{F}_{a_i}^{(l)} = \widetilde{\boldsymbol{V}}_i^{(l)} \left(\widetilde{\boldsymbol{\Sigma}}_i^{(l)\ T} \widetilde{\boldsymbol{\Sigma}}_i^{(l)} + \frac{M_R \sigma_n^2}{P_T} \boldsymbol{I}_{M_T} \right)^{-1/2};$
    $\boldsymbol{H}_i^{(l)} \boldsymbol{F}_{a_i}^{(l)} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^H;$
    $\boldsymbol{U}_i^{(l)} = \boldsymbol{U}_{(:,1:r_i)};$
  end;
end;
$\boldsymbol{F}_a^{(N_{\text{iterat}})} = \left[\begin{array}{ccc} \boldsymbol{F}_{a_1}^{(N_{\text{iterat}})} & \ldots & \boldsymbol{F}_{a_K}^{(N_{\text{iterat}})} \end{array} \right];$

Table 5.5: Iterative RBD (IRBD) algorithm.

of the users since simulation results have shown that the performance of the algorithm does not depend on the user ordering. In iterative RBD (IRBD), the precoding matrices $\boldsymbol{F}_{a_i}$ are calculated by repeatedly performing RBD. The difference to RBD is that after calculating $\boldsymbol{F}_{a_i}$ for the i^{th} user, for all other users we use $\boldsymbol{U}_i^{(r_i)\ (l-1)\ H} \boldsymbol{H}_i$ instead of the channel matrix $\boldsymbol{H}_i$ to generate the matrix $\widetilde{\boldsymbol{H}}_j^{(l)}, j \neq i$, where $\widetilde{\boldsymbol{H}}_j^{(l)}$ is defined in equation (5.54) and $\boldsymbol{U}_i^{(r_i)\ (l-1)}$ is a matrix containing the first r_i singular vectors of the matrix $\boldsymbol{U}_i^{(l-1)}$ which is given in equation (5.56).

Similar iterative solutions were previously introduced as, for example, coordinated beamforming [25]. However, the difference compared to our proposal is that the authors assume cooperation between UTs and the BS which requires that a large portion of the system throughput is used for the transmission of demodulation matrices from the UTs back to the BS. In our case, the matrices $\boldsymbol{U}_i$ are calculated using only the CSI available at the BS. The optimum receive MIMO processing at the UTs is still MMSE spatial filtering of the effective channel $\boldsymbol{H}_i \boldsymbol{F}_i$, where $\boldsymbol{F}_i = \beta \boldsymbol{F}_{a_i} \boldsymbol{F}_{b_i}$.

5.6.4 Joint processing in space, time and frequency

Joint multi dimensional RBD (JRBD) exploits the other two dimensions, time and frequency, to reduce the overlap of the row subspaces of users' channel matrices. Let us assume that $M_R \leq M_T$. At high SNRs, RBD will in this case reduce to regular BD as described in [25]. If the rank of $\widetilde{\boldsymbol{H}}_i$ is $\widetilde{L}_i$, the dimension of the i^{th} user's effective channel is $M_{R_i} \times \left(M_T - \widetilde{L}_i \right)$ and is equivalent to a system with $M_T - \widetilde{L}_i$ transmit antennas and

M_{R_i} receive antennas. The maximum diversity order that this system can provide is then $M_{R_i}\left(M_T - \widetilde{L}_i\right)$. By increasing the number of transmit antennas we could reduce the diversity loss. However, an increase of the number of antennas is very expensive so we have to consider other solutions.

This problem can be avoided by combining the channel matrices in equation (2.14) from several OFDM symbols and subcarriers. However, if we use channel matrices from adjacent subcarriers and consecutive OFDM symbols this grouping will not result in any performance improvement since the channel matrices are correlated. This is equivalent to the scaling of the vector basis. Another option is to use channel matrices from subcarriers that are separated in frequency by more than the coherence bandwidth and in time by more than the coherence time interval. The third option is to perform subspace perturbation on the adjacent subcarriers and consecutive OFDM symbols. The subspace perturbation is performed at the UTs by multiplying the channel matrices on different subcarriers and OFDM symbols with pseudo-random sequence. For example, we precode jointly the signals over S_P adjacent subcarriers in one OFDM symbol by using

$$\check{\boldsymbol{H}}_i = \left[\begin{array}{ccc} c_1 \boldsymbol{H}_i^T(1) & \cdots & c_{S_P} \boldsymbol{H}_i^T(S_P) \end{array} \right]^T \in \mathbb{C}^{M_R \times S_P M_T} \tag{5.59}$$

instead of $\boldsymbol{H}_i$ in (2.14). $\boldsymbol{H}_i(k)$ is the i^{th} user's channel matrix on the k^{th} subcarrier, and $\left[\begin{array}{ccc} c_1 & \cdots & c_{S_P} \end{array} \right]$ is the pseudo-random sequence of length S_P. After substituting $\check{\boldsymbol{H}}_i$ for $\boldsymbol{H}_i$ in (2.14) the precoding is performed in the same manner as described previously.

The difference between JRBD and multi-carrier CDMA (MC-CDMA) is that in MC-CDMA, MIMO processing is performed on every subcarrier separately and then the data is spread over a group of adjacent subcarriers. In JRBD every channel matrix is multiplied with one chip of a pseudo random sequence, and then these matrices are grouped into one big channel matrix. This matrix is used for precoding.

5.6.5 Simulation results on RBD family of precoding algorithms

In this section we compare the performance of systems employing the precoding techniques introduced in this section to SMMSE, SMMSE THP, BD, SO THP and TDMA. To this end we simulate a purely stochastic spatially white channel $\boldsymbol{H}_w$ and the second is a frequency selective MIMO channel with a power delay profile as defined by IEEE802.11n - D with non-line of sight conditions [60]. The elements of the channel matrices on each subcarrier are zero mean, unit variance complex Gaussian variables. We assume data transmission

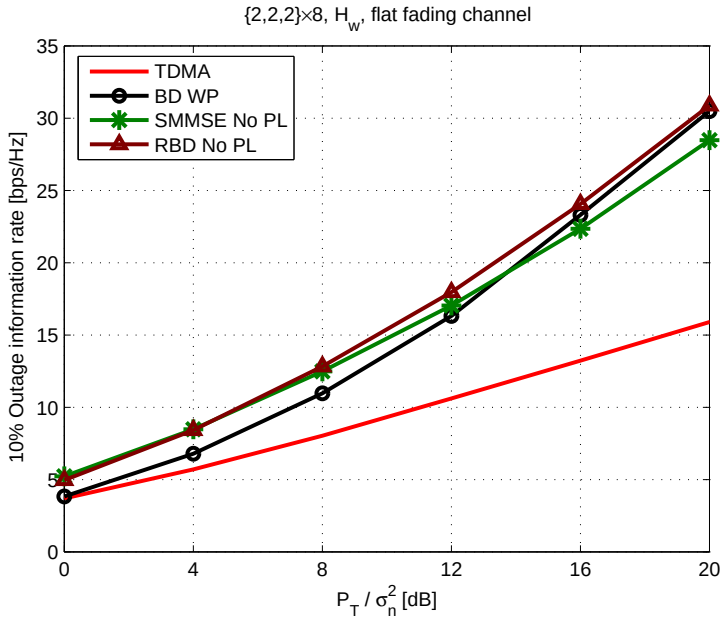

Figure 5.10: 10 % outage information rate of BD, SMMSE and RBD as a function of SNR.

using an OFDM system with DFT size $N = 64$, a subcarrier spacing of 150 kHz and a cyclic prefix that is $N_{\text{pre}} = 4$ samples long. The data is encoded using the convolutional code rate 1/2 $(561, 753)_{\text{oct}}$. After coding the data is mapped using QAM or 16 QAM modulation. Coded and modulated symbols are transmitted using $N_c = 48$ subcarriers and $N_{\text{symb}} = 2$ OFDM symbols.

In the second channel model we also consider antenna correlation at the BS and UTs. Antenna correlation is modeled in the delay domain using the Kronecker model defined in Section 2.2.1.

In Figure 5.10 we show the 10 % outage information rate as a function of the ratio of the total transmit power P_T and the power of additive white Gaussian noise at the input of every antenna, σ_n^2. The information rate is calculated using the results on the information rate of MIMO broadcast channels in [14]. We also present capacity results for a TDMA system as a comparison. As this figure shows, SMMSE provides higher information rate than BD at low SNRs while BD is better at high SNRs. RBD practically adapts to different levels of noise, allowing more MUI at low SNRs and canceling MUI at high SNRs to improve the information rate of the system. When $M_T > M_R$ RSO THP approaches the maximum achievable sum rate capacity of the MU MIMO downlink system as it can be seen from Figure 5.11. RBD and IRBD have a loss in this scenario of around 5 bps/Hz.

If the subspaces of the users' channel matrices significantly overlap, e.g., if $M_R > M_T$,

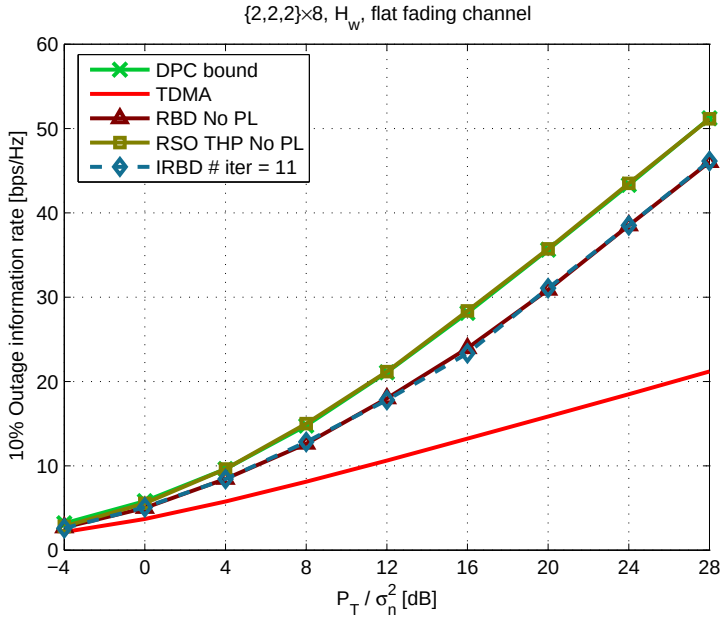

Figure 5.11: 10 % outage information rate of RBD, RSO THP and IRBD as a function of SNR. $M_R \leq M_T$.

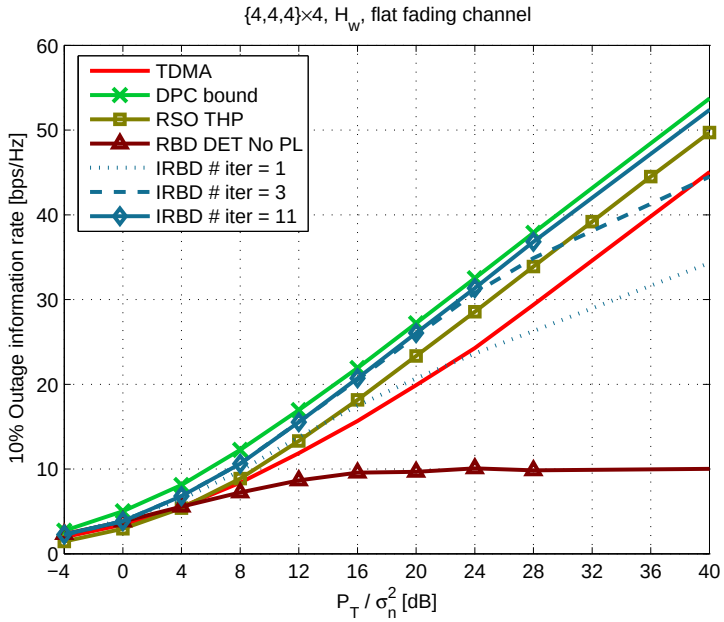

Figure 5.12: 10 % outage information rate of RBD, RSO THP and IRBD as a function of SNR. $M_R > M_T$.

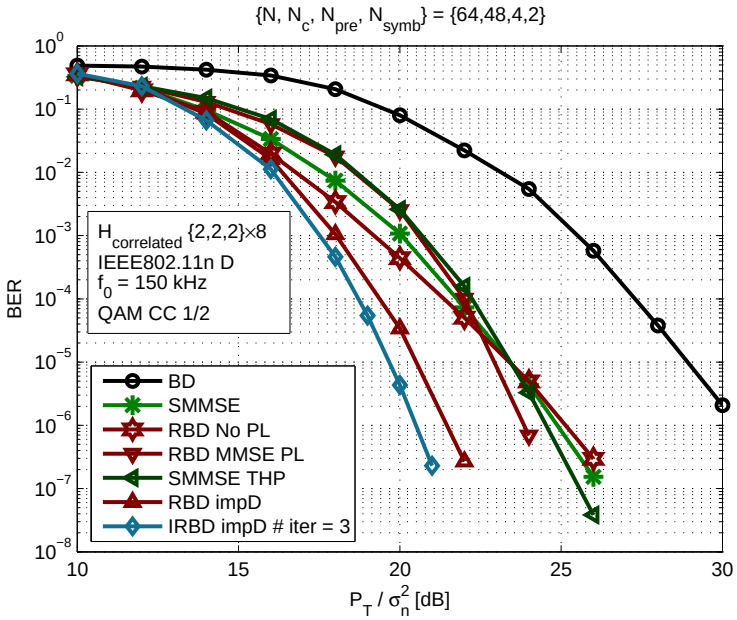

Figure 5.13: BER performance comparison of RBD with different power loading algorithms with BD, SMMSE and SMMSE THP.

MUI will substantially degrade the system information sum rate. In a system with $K = 3$ users each equipped with $M_{R_i} = 4$ antennas and with $M_T = 4$ antennas at the BS, RSO THP fails to reach the DPC bound. However, IRBD with dominant eigenmode transmission approaches this bound as the number of iterations increases. From Figure 5.12 we see that after 11 iterations, IRBD reaches the DPC bound.

In Figure 5.13 we compare the BER performance of BD, SMMSE, SMMSE THP, RBD and IRBD. By introducing MUI, SMMSE outperforms BD, and its diversity order can be further improved by combining it with THP. The performance of RBD depends on the power loading algorithm being used. No power loading produces better results at low SNRs and MMSE power loading is more advantageous at high SNRs. By combining these two power loading strategies, RBD, which is a linear precoding technique, provides an SNR gain of more than 3 dB over SMMSE THP which is a non-linear precoding technique. IRBD with only 3 iterations provides a further improvement and a higher diversity than RBD.

In Figure 5.14 we show the BER performance of RBD, IRBD, JRBD, MC-CDMA and the BER curve for a similar "genie aided" system where the users are assumed perfectly orthogonal in order to show the diversity inherent in this type of system. In MC-CDMA spatial processing is done on each subcarrier separately. Unlike in Figure 5.13, IRBD has a better performance than RBD only at very high SNRs. JRBD outperforms both RBD

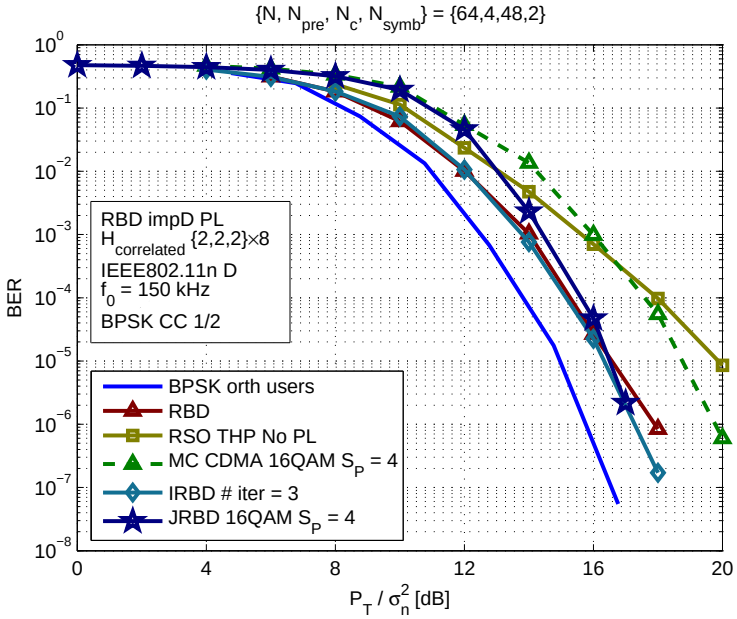

Figure 5.14: BER performance of RBD, IRBD, JRBD and MU MIMO system with orthogonal users in configuration $\{2, 2, 2\} \times 8$.

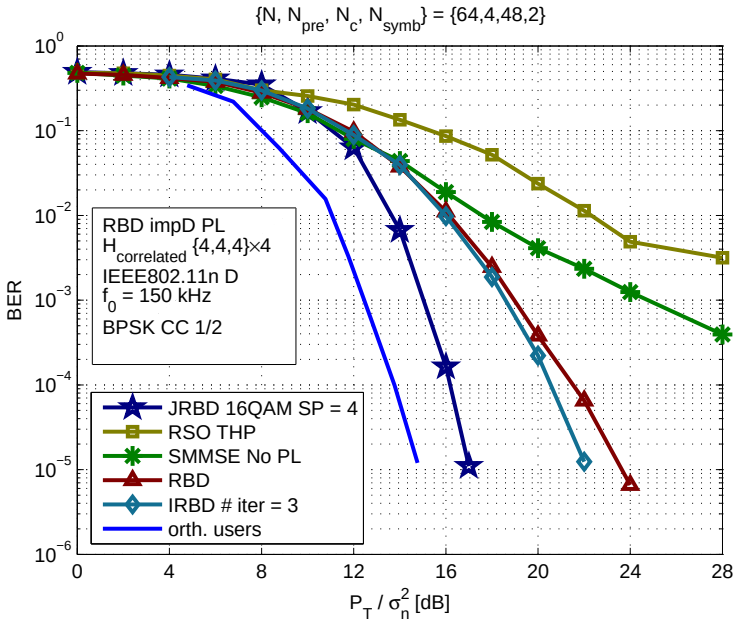

Figure 5.15: BER performance of RBD, IRBD, JRBD, RSO THP and MU MIMO system with orthogonal users in configuration $\{4, 4, 4\} \times 4$.

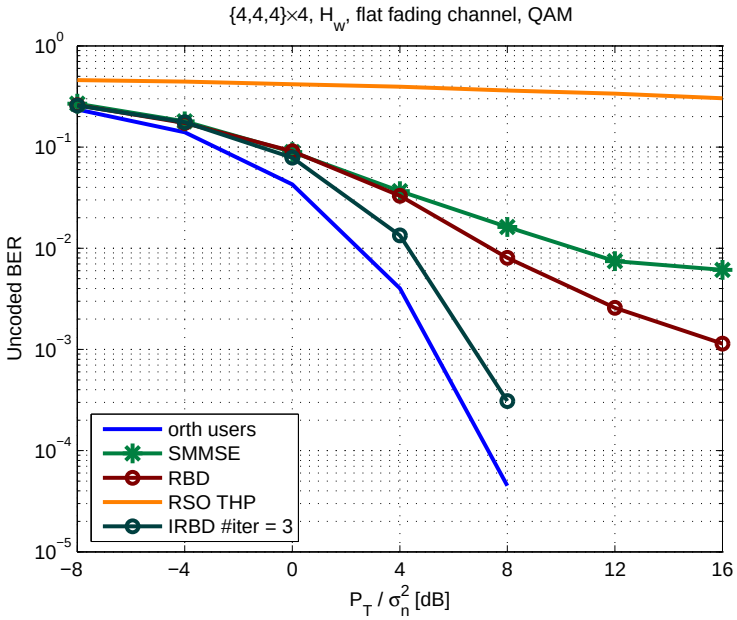

Figure 5.16: Uncoded BER performance of SMMSE, RBD, RSO THP, IRBD and MU MIMO system with orthogonal users in configuration $\{4, 4, 4\} \times 4$.

and IRBD at very high SNRs. However, both JRBD and IRBD manage to provide the full diversity of the system. The difference between these precoding techniques becomes more visible in case of high MUI as shown in Figure 5.15. In this case both SMMSE and RSO THP experience an error floor. By using IRBD we provide an additional gain of 2 dB in comparison with RBD, and by further increasing the computational complexity using JRBD we manage to reduce the loss compared to the "genie aided" system to only 2 dB.

In Figure 5.16 we investigate the asymptotic performance at high SNRs of SMMSE, RBD, RSO THP and IRBD relative to a "genie aided" system. The antenna configuration of the system is $\{4, 4, 4\} \times 4$. As we can see, SMMSE, RBD and RSO THP experience an error floor at high SNRs due to the high MUI. However, IRBD performs very good in this case and extracts full antenna diversity gain inherent in the system.

5.7 Precoding using instantaneous and long-term channel state information

It has been shown in [2] that the capacity of a MIMO system increases linearly with the minimum out of the number of receive and transmit antennas. Fading correlation reduces the system capacity, especially when there is no channel knowledge available at the transmitter [35]. If it is impossible to acquire instantaneous CSI at the base station, the spatial channel correlation can nevertheless be used to effectively eliminate or reduce the multi-user interference. In this section we introduce the equivalent channel which facilitates the use of MU MIMO precoding techniques when there is either instantaneous CSI or long-term CSI available at the base station. This will help us apply the same linear precoding techniques requiring instantaneous channel state information at the transmitter also if we only have the information on the transmit correlation matrices. But instead of using the exact channel knowledge we will use the equivalent channel, which will be introduced later.

If we assume that the channel is varying too rapidly to track its mean, the information regarding the relative geometry of the propagation paths is captured by a colored spatial correlation matrix. This problem is also addressed in [25] where the authors assume that each user's channel matrix can be represented as $\boldsymbol{H}_i = \boldsymbol{A}_i\boldsymbol{B}_i$ where the transmitter has the information about the matrix $\boldsymbol{B}_i \in \mathbb{C}^{r_i \times M_T}$ but not $\boldsymbol{A}_i \in \mathbb{C}^{M_{R_i} \times r_i}$, $r_i \leq M_{R_i}$. The multi-user interference in the system can be set to zero by performing BD on the matrices $\boldsymbol{B}_i$. This solution corresponds to beamforming based on the long-term beams with the additional constraint that the i^{th} user's long-term beams are in the null space of all other users' long-term beams.

Let us introduce the i^{th} user's average correlation matrix $\boldsymbol{R}_i^{(k)}$ over the k^{th} chunk, where the chunk represents the smallest time-frequency resource allocation unit. The dimensions of a chunk in a OFDM-MIMO system are defined by the number of consecutive OFDM symbols in the time dimension and the number of adjacent subcarriers in the frequency direction. Thus, the i^{th} user's average correlation matrix over the k^{th} chunk $\boldsymbol{R}_i^{(k)}$ is defined as

$$\boldsymbol{R}_i^{(k)} = \frac{1}{N_{\text{chunk}}} \sum_j \boldsymbol{H}_i^{(k,j)\ H} \boldsymbol{H}_i^{(k,j)} \tag{5.60}$$

and its singular value decomposition as

$$\boldsymbol{R}_i^{(k)} = \boldsymbol{V}_i^{(k)} \boldsymbol{\Lambda}_i^{(k)} \boldsymbol{V}_i^{(k)\ H} \tag{5.61}$$

where $\boldsymbol{H}_i^{(k,j)}$ is the i^{th} user's channel matrix on the j^{th} symbol of the k^{th} chunk and N_{chunk} is the number of complex symbols in one chunk.

The equivalent channel is defined based on the matrix $\boldsymbol{R}_i^{(k)}$ as follows:

$$\widehat{\boldsymbol{H}}_i^{(k)} = \boldsymbol{\Lambda}_i^{1/2} \boldsymbol{V}_i^{(k)\ H} \tag{5.62}$$

It can easily be shown that the solution proposed in [25] is similar to the solution obtained by applying BD on the matrices $\widehat{\boldsymbol{H}}_i^{(k)}$. Unlike in [25] where the authors address only the cancellation of the multi-user interference without the optimization of the isolated single-user performance, the approach based on (5.62) completely defines the modulation matrices. It was shown that when only the channel correlation matrix is available at the transmitter, the optimum strategy is to transmit on the long-term eigenmodes of the matrix $\boldsymbol{R}_i^{(k)}$ [9].

We will use the matrix $\widehat{\boldsymbol{H}}_i^{(k)}$ defined in (5.62) as a long-term equivalent channel and perform the precoding on this matrix as if it represented the actual channel. The matrices $\widehat{\boldsymbol{H}}_i^{(k)}, i = 1, 2, \ldots, K$, contain all the information about the long-term subspace of each user available at the transmitter, in this case the base station. To illustrate this fact we look at zero-forcing (ZF) precoding that is defined as the pseudo-inverse of the channel.

The pseudo-inverse of the matrix $\widehat{\boldsymbol{H}}_i^{(k)}$ is equal to

$$\left(\widehat{\boldsymbol{H}}_i^{(k)\ H} \widehat{\boldsymbol{H}}_i^{(k)}\right)^{-1} \widehat{\boldsymbol{H}}_i^{(k)\ H} = \left(\boldsymbol{V}_i^{(k)} \boldsymbol{\Lambda}_i^{(k)} \boldsymbol{V}_i^{(k)\ H}\right)^{-1} \boldsymbol{V}_i^{(k)} \boldsymbol{\Lambda}_i^{(k)\ 1/2} = \boldsymbol{V}_i \boldsymbol{\Lambda}_i^{(k)\ -1/2}.$$

From the previous equation we can see that the pseudo-inverse of the matrix defined in (5.62) also results in a transmission on the scaled long-term beams $\boldsymbol{V}_i^{(k)}$ of the channel. This means that we could apply the same linear precoding techniques requiring perfect channel state information at the transmitter also if we only have the information on the transmit correlation matrices, but instead of using the exact channel knowledge we will use the matrix in (5.62). By using the equivalent channel $\widehat{\boldsymbol{H}}_i^{(k)}$ we facilitate an easier adaptation from instantaneous channel state information to the long-term channel state information. Since we use the equivalent channel in the same way as the instantaneous channel knowledge we allow the combination of these two types of channel knowledge for

precoding.

Next, we will compare the 10 % outage information rate of IRBD, SMMSE and BD when we use the equivalent channel for precoding. We assume flat fading channel which is modeled as in equation (2.9). There are three users in the system, each equipped with two antennas and there are six antennas at the base station. All users have the same speed, equal to 70 km/h. As a comparison we show the 10 % outage information rate of these precoding techniques when we assume perfect channel knowledge at the base station. We also introduce the assumption that the communication is done using TDD. Each frame consists of a downlink transmission interval followed by an uplink transmission interval. With downlink:uplink asymmetry 1:1, the TDD slot consists of 15 downlink OFDM symbols followed by 15 uplink OFDM symbols. The equivalent channels are calculated by averaging the correlation matrices over 15 OFDM symbols. The delay from the time slot when the equivalent channels are calculated to the time slot when they are used for the precoding is equal to one frame, i.e., 30 OFDM symbols. This is denoted with (15+30+1). As we can see from the Figure 5.17, the performance of these techniques when we use the equivalent channel for precoding is close to the curves obtained assuming perfect channel knowledge at the base station. It can be seen that BD is very sensitive to channel correlation and that IRBD provides the best performance among these three. At high SNRs IRBD also has an information rate loss. However, this point when the information rate of IRBD using long-term channel state information breaks away from the information rate of IRBD based on the perfect channel state information can be further shifted to the higher SNR values by increasing the number of iterations.

How well IRBD performs with only long-term channel state information at the transmitter can also be seen from Figure 5.18. As it can be seen from the figure, the performance of IRBD is the same as in the case when the precoding is done using instantaneous CSI. In this case we again have three users in the system, each equipped with two antennas and with six antennas at the base station. All users have the same speed of 50 km/h. The channel is frequency selective and it is modeled as described in Section 2.2.1. The power delay profile is defined by IEEE802.11n - D with non-line of sight conditions [60]. We assume data transmission using an OFDM system with DFT size $N = 64$, a subcarrier spacing of 312.5 kHz and a cyclic prefix that is $N_{\text{pre}} = 16$ samples long. The data is encoded using the convolutional code rate 1/2 $(561, 753)_{\text{oct}}$. After coding the data is mapped using QAM modulation. Coded and modulated symbols are transmitted using $N_c = 48$

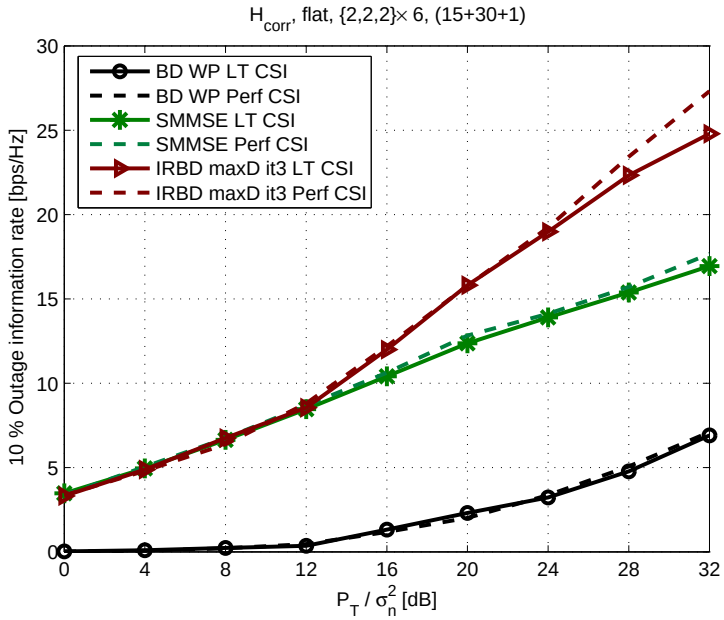

Figure 5.17: 10 % outage information rate of IRBD, SMMSE and BD with long-term CSI at the transmitter as a function of the SNR. Each user is equipped with two antennas.

subcarriers and $N_{\text{symb}} = 2$ OFDM symbols. Note that until now we did not perform grouping of spatially low correlated users. Again, even without spatial scheduling IRBD is able to achieve the same performance as with perfect knowledge of the instantaneous CSI.

The previous results are obtained using the average correlation matrix on one chunk. However, if we assume that the averaging window is long enough, then this average correlation matrix is the estimate of the transmit correlation matrix given in equation (2.7) in Section 2.2.1. The performance of BD and SMMSE in the system with configuration $\{4, 4\}\times 4$ can be seen in Figure 5.19. We assume flat fading channel modeled as in equation (2.9). Unlike in the case when we have perfect CSI available at the transmitter, in the case where long-term CSI is available BD is not limited by the number of receive antennas. This is the consequence of the fact that we perform the precoding on an equivalent channel. When the channel is rank deficient like it is the case in the wide area scenario, the dimension of the equivalent channel is different from the actual and the dimensionality restriction is met if the rank of the transmit correlation matrix is less than the number of receive antennas. We should note here that the performance of these precoding techniques does not change significantly compared to the case when there is perfect CSI available at the transmitter, which is a consequence of the smart scheduling assumption [62], [63]. If designed well a scheduling algorithm selects the users that are well separated in space. In

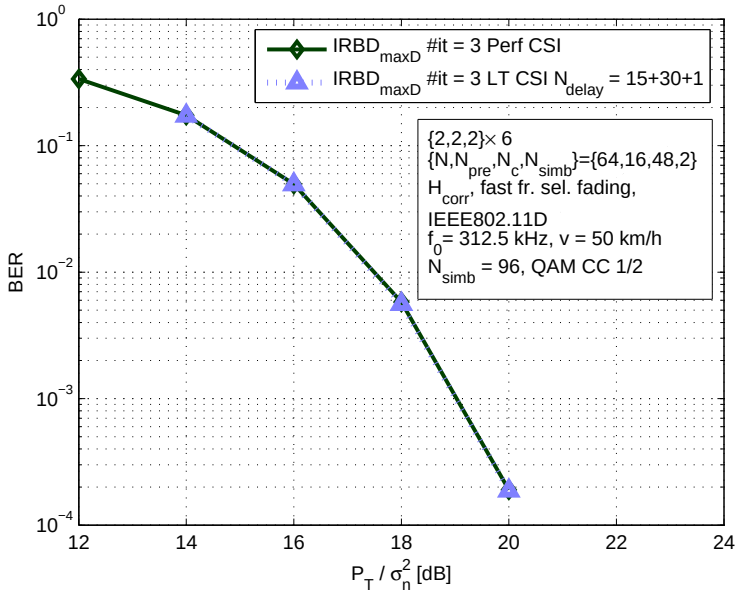

Figure 5.18: BER performance of IRBD with long-term CSI at the transmitter as a function of the SNR. System configuration is $\{2, 2, 2\} \times 6$.

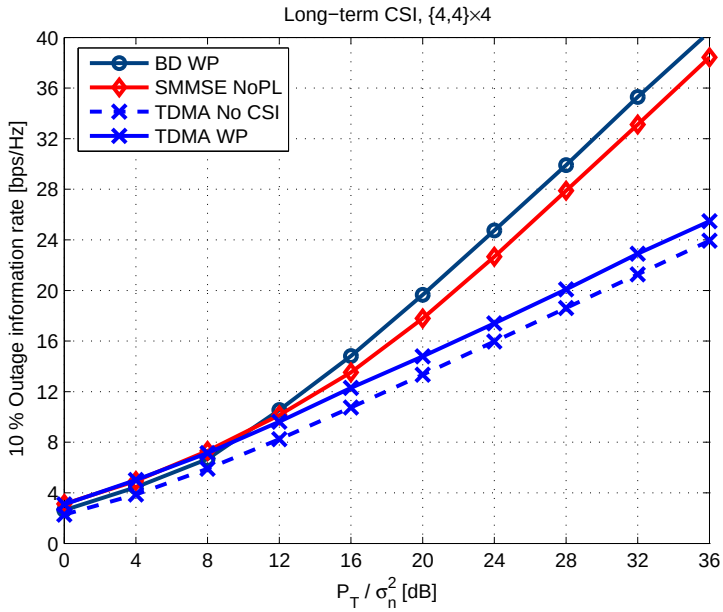

Figure 5.19: 10 % outage information rate with long-term CSI at the transmitter as a function of the receive SNR. Each user is equipped with four antennas.

this case even the knowledge about the transmit correlation of the channel allows us to completely eliminate multi-user interference at high SNRs. The more antennas there are at the base station, the greater are the benefits of multi-user MIMO precoding.

5.8 Calibration and transceiver front-end impairments

In order to reduce the overhead required to acquire the channel state information at the base station or access point we rely on the reciprocity principle so we could use the channel measurements on the uplink to perform the spatial processing on the downlink. Various degrees of channel state information at the base station can help us to significantly improve system performance especially in multi-user scenarios. The major advantage of channel state information at the transmitter based on the channel reciprocity assumption, compared to introducing a feedback link, is that it is a low complexity solution.

When the channel is estimated on the uplink, the downlink channel matrix is just the transpose of the uplink matrix, assuming the channel is reciprocal. However, the channel is actually made up of the propagation channel, the antennas and the transceiver radio-frequency (RF) and baseband circuits at both sides of the link. The transceiver circuits are usually not reciprocal, so since the channel estimation and spatial processing are performed in the baseband, it is necessary to calibrate the amplitude and phase errors between the branches of the antenna array due to individual differences in the RF circuits of the receivers and the transmitters. The main sources of imbalance in the channel reciprocity are analog-to-digital converter (ADC), baseband filter, in phase/quadrature (IQ) imbalance, phase noise, and amplifiers.

We consider an MU OFDM-MIMO system, where M_T transmit antennas are located at the base station and M_{R_i} receive antennas are located at the i^{th} user's terminal, $i = 1, \ldots, K$. All users downlink over-the-air propagation channel matrices on one subcarrier are grouped in one combined network channel matrix given in equation (2.14). The uplink overall channel between the base station and the user terminals that is estimated at the base station on one subcarrier is equal to:

$$\boldsymbol{H}^{UL} = \boldsymbol{\Psi}_{T_x,UT} \boldsymbol{H} \boldsymbol{\Psi}_{R_x,BS} \in \mathbb{C}^{M_R \times M_T} \tag{5.63}$$

and the overall channel on the downlink estimated at the user terminals equals

$$\boldsymbol{H}^{DL} = \boldsymbol{\Psi}_{R_x,UT}\boldsymbol{H}\boldsymbol{\Psi}_{T_x,BS} \in \mathbb{C}^{M_R \times M_T} \tag{5.64}$$

The matrices $\boldsymbol{\Psi}_{T_x,UT} \in \mathbb{C}^{M_R \times M_R}$, $\boldsymbol{\Psi}_{R_x,UT} \in \mathbb{C}^{M_R \times M_R}$, $\boldsymbol{\Psi}_{T_x,BS} \in \mathbb{C}^{M_T \times M_T}$, and $\boldsymbol{\Psi}_{R_x,BS} \in \mathbb{C}^{M_T \times M_T}$ are complex perturbation matrices containing the user terminals and the base station transmit and receive front-end frequency response on the specific subcarrier, respectively. These matrices are diagonal, which implies that they can be inverted and that their products are commutative. The elements of these matrices are slowly varying in time.

The combined overall network channel matrix on the downlink can be rewritten as

$$\begin{aligned} \boldsymbol{H}^{DL} = &\quad \boldsymbol{\Psi}_{R_x,UT}\boldsymbol{\Psi}^{-1}_{T_x,UT}\boldsymbol{\Psi}_{T_x,UT}\boldsymbol{H}\boldsymbol{\Psi}_{R_x,BS}\boldsymbol{\Psi}^{-1}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS} \\ = &\quad \boldsymbol{\Psi}_{R_x,UT}\boldsymbol{\Psi}^{-1}_{T_x,UT}\boldsymbol{H}^{UL}\boldsymbol{\Psi}^{-1}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS} \end{aligned} \tag{5.65}$$

where we assume that based on the measurements at the base station we have the perfect knowledge of the uplink channel $\boldsymbol{H}^{UL}$. The two matrices $\boldsymbol{\Psi}_{R_x,UT}\boldsymbol{\Psi}^{-1}_{T_x,UT}$ and $\boldsymbol{\Psi}^{-1}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS}$ represent calibration errors at the user terminals and the base station, respectively. The calibration matrices are defined as

$$\begin{aligned} \boldsymbol{K}_{BS} = &\quad \boldsymbol{\Psi}^{-1}_{T_x,BS}\boldsymbol{\Psi}_{R_x,BS} \\ \boldsymbol{K}_{UT} = &\quad \boldsymbol{\Psi}_{T_x,UT}\boldsymbol{\Psi}^{-1}_{R_x,UT} \end{aligned} \tag{5.66}$$

Using these calibration matrices, the reciprocity of the over-the-air channel can be exploited to achieve the reciprocity of the overall channel including RF front-end effects on both sides and the free space channel.

Several approaches can be used for calibration:

- By estimating the effective channel at the receiver, which includes also the contribution of spatial processing at the transmitter, the system is more robust to channel and calibration errors than in the case when the receiver estimates the exact channel.
- RF approaches: each of the RF imperfections is compensated specifically, or the front-ends are designed to be reciprocal. One can re-use the IQ mixer, the low-noise amplifier (LNA), and power amplifier (PA) in the TDD receive and transmit mode, since the terminal never transmits and receives simultaneously. By using an RF

transfer switch, the direction of the active components, being the LNA and PA, can be reversed. Moreover, the IQ mixer can be reused, both as a modulator and a demodulator. Provided that no imperfection occurs, in particular that in- and output impedances of the LNA and PA are respectively the same and that the IQ mixer has no imbalance, the RF front-end can be regarded as inherently reciprocal and it needs no calibration.

- A self calibration approach in which the transmitter and the receiver estimate and compensate independently the distortions;

- A global approach proposed by Qualcomm [64], in which the knowledge of the channel in both directions is used to pre-compensate the distortions (whatever their origin). The BS observes a MIMO pilot from the UT and derives an estimate of the channel between the UT and the BS. In response, the BS transmits a MIMO pilot, which is observed by the UT and used to derive an estimate of the channel between the BS and the UT. Then, the UT transmits the quantized estimated channel to the BS. The BS can now determine the diagonal calibration matrices. Finally, the BS sends one diagonal calibration matrix to the UT, then each end of the link has its own calibration matrix.

We will focus on the self calibration approach since it does not require any communication protocol modifications as the global approach and it is less complex than the RF approach. One possible approach to perform self calibration at the transceiver (user terminal or base station) is proposed in [65]. As an example we will consider the calibration at the base station. Before calibration, the carrier frequency and all transceiver parameters that have an effect on the amplitude or phase response of the transmit and/or receive chain are set. Once the parameters are set, the frequency responses are assumed static. The calibration is achieved in two steps: measurement of $\boldsymbol{\Psi}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS}$ and measurement of $\boldsymbol{\Psi}_{R_x,BS}$.

In order to measure the product $\boldsymbol{\Psi}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS}$, the signal from the output of the i^{th} transmitter is fed to the input of the i^{th} receive branch. In each antenna branch, a suitable known signal $\boldsymbol{s}$ is generated L times. The j^{th} frequency domain received signal, $j = 1, \ldots, N$, is

$$\boldsymbol{r}_j = \boldsymbol{\Psi}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS}\boldsymbol{s} + \boldsymbol{n}_j \tag{5.67}$$

where $\boldsymbol{s} = \begin{bmatrix} s_1 & \cdots & s_{M_T} \end{bmatrix}^T$, M_T is the number of antennas at the base station and $\boldsymbol{n}_j$

is a noise vector. The signal $\boldsymbol{s}$ should have a low peak to average power ratio. The product $\boldsymbol{\Psi}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS}$ is estimated by averaging the L values of $\boldsymbol{r}_j$.

$$\boldsymbol{a} = \frac{1}{L}\sum_{j=1}^{L} \boldsymbol{r}_j./\boldsymbol{s} \cong \mathrm{diag}\left(\boldsymbol{\Psi}_{R_x,BS}\boldsymbol{\Psi}_{T_x,BS}\right) \tag{5.68}$$

where ./ stands for element-wise division, and $\mathrm{diag}\left(\bullet\right)$ is the function that extracts the elements on the main diagonal.

In the second step the receiver is isolated from both the transmitter and the antenna. At the input of each receive branch we apply the same reference noise source whose power exceeds the thermal noise power by 20 dB or more. The signal is sampled and measured at baseband in the receiver of all antenna branches simultaneously. The received frequency domain signal is

$$\boldsymbol{r}_j = \boldsymbol{\Psi}_{R_x,BS}\left(n_{\mathrm{ref}}\boldsymbol{I} + \boldsymbol{n}_j\right) \tag{5.69}$$

where n_{ref} is the reference noise. Since n_{ref} is a multiplicative term, $\boldsymbol{\Psi}_{R_x,BS}$ cannot be extracted directly, and the signal in the first antenna branch $r_{j,1}$ is taken as a reference. After averaging over L measurements we will have

$$\boldsymbol{c} = \frac{1}{L}\sum_{j=1}^{L} \frac{1}{r_{j,1}}\boldsymbol{r}_j \cong \frac{1}{\psi_{R_x,BS,1}}\mathrm{diag}\left(\boldsymbol{\Psi}_{R_x,BS}\right) \tag{5.70}$$

which is a vector containing the frequency responses of the receiver branches with a complex error coefficient, common to all antenna branches. The estimation of $\boldsymbol{K}_{BS}^{-1}$ is obtained as

$$\boldsymbol{a}./\boldsymbol{c}^2 = \left(\psi_{R_x,BS,1}\right)^2 \mathrm{diag}\left(\boldsymbol{\Psi}_{R_x,BS}^{-1}\boldsymbol{\Psi}_{T_x,BS}\right) = \left(\psi_{R_x,BS,1}\right)^2 \mathrm{diag}\left(\boldsymbol{K}_{BS}^{-1}\right) \tag{5.71}$$

It is the relative difference in the UL/DL blocks that is important when considering the channel reciprocity assumption. For instance, if the UL filter differs from the DL filter, then so will the channels. In fact, even if both filters correspond to the same product device, they will never have exactly the same characteristics due to the production spread of the device used.

It is very important to take into account RF impairments in order to quantify their impact on the performance of multi-user MIMO algorithms. In our investigations we will use the model of RF impairments that was introduced in [65].

We assume that the self-calibration is performed only at the base station in order to

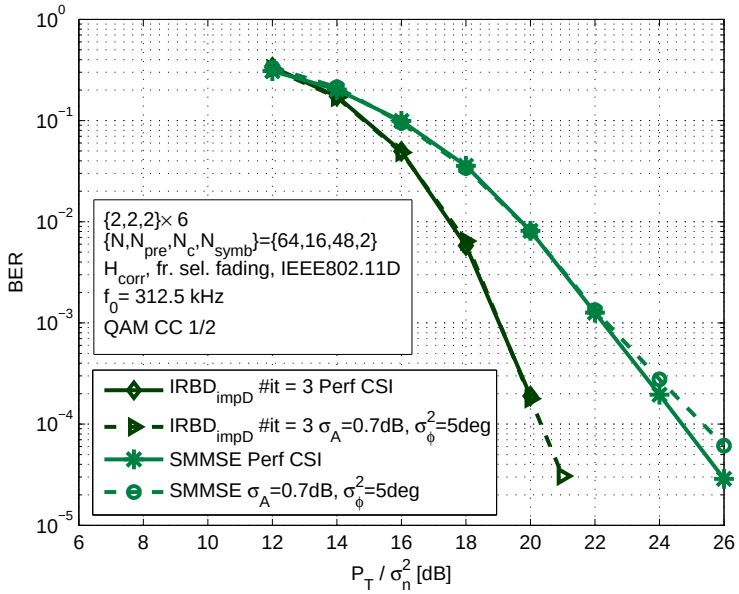

Figure 5.20: Influence of the calibration errors on the performance of IRBD and SMMSE. The antenna system configuration is $\{3, 3, 3\} \times 6$.

keep the user terminals as simple as possible. Thus, the precoding at the base station is done under the assumption that the matrix $\mathbf{\Psi}_{R_x,UT}\mathbf{\Psi}_{T_x,UT}^{-1}$ is an identity matrix and that the calibration is performed only at this end of the link. We assume that the mobile stations have perfect channel knowledge of the effective channel after the precoding.

In the simulations, the amplitude and phase of each element of matrices $\mathbf{\Psi}_{R_x,UT}\mathbf{\Psi}_{T_x,UT}^{-1}$ and $\mathbf{\Psi}_{R_x,BS}^{-1}\mathbf{\Psi}_{T_x,BS}$ are modelled as a Gaussian random variable with mean 1 and variance 0.0839 (0.7 dB at 1 sigma) and a uniform random variable on $[-5°, 5°]$, respectively.

In Figure 5.20 we show the influence of calibration errors on the performance of IRBD and SMMSE. We assume data transmission using an OFDM system with DFT size $N = 64$, the chunk size is $N_c = 48$ subcarriers $\times$ $N_{symb} = 2$ OFDM symbols, the cyclic prefix is $N_{pre} = 16$ samples long, and the subcarrier spacing is 312.5 kHz. The propagation channel is frequency selective, spatially correlated with a power delay profile as defined by the IEEE 802.11n D channel model. Spatial correlation is modeled using the Kronecker model as defined in Section 2.2.1. Information is coded using the convolutional code $(561, 753)_{\text{oct}}$ and modulated using QAM. P_T denotes the total transmit power at the base station, and σ_n^2 is the AWGN variance at the input of each receive antenna.

As it can be seen from Figure 5.20 when the calibration errors are modeled using [65], the performance of IRBD is close to perfect. However, in case of SMMSE at high SNRs we have a smaller diversity in the presence of calibration errors. These results show that the

base station does not need the information about the calibration matrices at the mobile stations in order to perform the precoding. The base station and the mobiles can perform the calibration independently.

5.9 Channel estimation errors

An important issue in a multi-user MIMO transmission is the impact of imperfect channel state information on the general system performance.

The delay between the moment when the channel state information is estimated to the moment when it is used for multi-user MIMO processing or scheduling and adaptive coding and modulation might cause a serious problem from the adaptation point of view. When the channel is fast varying in the time domain, it is important that the CSI is estimated and used within a time period that is shorter than the coherence time of the channel.

For channel estimation purposes in OFDM MIMO systems, two types of pilots could be used. Pilot patterns may be generated in the frequency domain by using a scattered pilot grid. Alternatively, pilot patterns may be generated in the time domain, in the form of short training blocks time-multiplexed with data blocks.

Channel estimation by interpolation in time and frequency based on a scattered pilot grid is considered to be an efficient solution for an OFDM-based radio interface.

For multi-antenna transmission, a combination of dedicated pilots per flow, common pilots per cell/sector, common pilots per antenna and common pilots per beam are required. Especially in case of dedicated pilots, purely pilot aided techniques may have severe limitations. Conventional channel estimation by interpolation may then require a pilot boost and/or a significant degree of over-sampling. Advanced solutions, such as iterative channel estimation, aim to make a pilot boost redundant, at the expense of increased complexity.

Adaptive transmission requires channel prediction for use in the resource allocation. Channel prediction can be based on common pilot symbols that are also used for other purposes. It should utilize the channel correlation in time and frequency.

In [66], [67] a simple and easy to implement model for channel estimation errors is proposed. The estimation error is approximated by white Gaussian noise. This model is well suited for pilot aided channel estimation schemes. For decision directed techniques this is a reasonable approximation for high SNRs, where decision errors are negligible.

The estimation error can be separated into an error caused by the noise and an interpolation or lag error. In low SNR regions the MSE is dominated by the error due to the noise. Hence, the MSE linearly decreases with the SNR. The ratio $G_{\mathrm{CSI}} = 1/(\mathrm{SNR} \cdot \mathrm{MSE})$ is the channel estimator gain. It is seen that the estimator gain is in the range from 5 to 10 dB, [67]. The interpolation error, on the other hand, is independent of the SNR. Therefore, the MSE curve experiences an error floor at high SNRs, which is around -30 dB. Unfortunately, both the noise and the interpolation error are dependent on the subcarrier index. In particular, near the beginning and the end of the frequency range, edge effects result in an increased estimation error.

The influence of the channel estimation errors at the base station on the performance of multi-user MIMO precoding techniques is investigated using a parametric channel estimation error model. The performance of an estimator can be approximated by modeling the channel estimation error as an additional noise source with variance σ_e^2 and a mean corresponding to the bias of the estimator. A good estimator should have a small bias, so in most cases it is justified to assume that the conditional mean of the channel estimation error is zero. We assume that the channel estimation error is a zero mean Gaussian random variable with variance equal to the MSE. The MSE is proportional to the SNR and is modeled as

$$\sigma_e^2 = \begin{cases} \sigma_n^2/(P_t\ G_{\mathrm{CSI}}), & \sigma_n^2/(P_t\ G_{\mathrm{CSI}}) > -30 \text{ dB} \\ -30 \text{ dB}, & \sigma_n^2/(P_t\ G_{\mathrm{CSI}}) \leq -30 \text{ dB} \end{cases} \tag{5.72}$$

where G_{CSI} defines the estimator gain. Hence, the MSE linearly decreases with the SNR. The estimator gain is set to 10 dB, but may vary depending on the number of users, antennas and on the preamble design.

The MIMO channel is modeled as

$$\boldsymbol{H}_{\mathrm{est}} = \boldsymbol{H} + \boldsymbol{E} \tag{5.73}$$

where $\boldsymbol{H}$ denotes the combined network channel matrix, $\boldsymbol{H}_{\mathrm{est}}$ is the channel estimate, and $\boldsymbol{E}$ is a channel estimation error matrix. Each element of the matrix $\boldsymbol{E}$ is modeled as a zero mean Gaussian random variable with variance σ_e^2.

The influence of the channel estimation errors on the BER performance of IRBD and SMMSE is shown in Figure 5.21. As it can be seen, the channel estimation errors cause an SNR loss of about 3 dB for both precoding techniques.

The influence of the channel estimation errors is much higher at the user terminals than

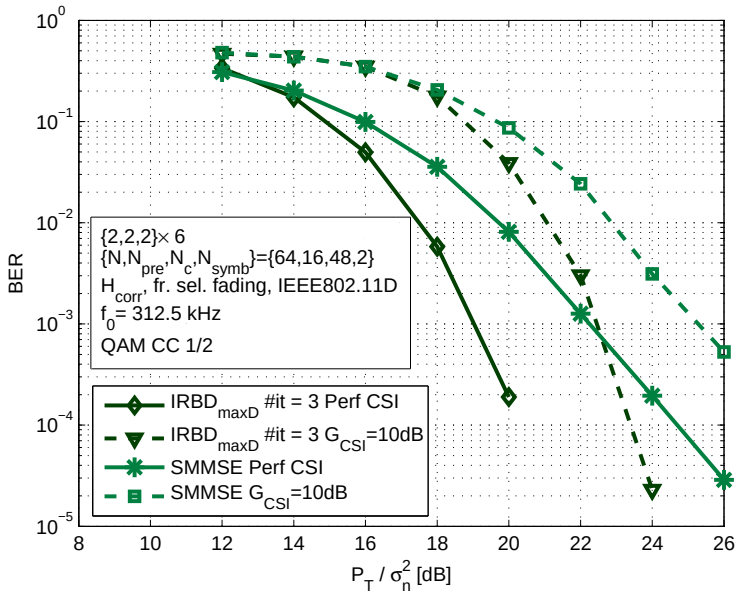

Figure 5.21: Influence of the channel estimation errors on the performance of IRBD and SMMSE. Antenna system configuration is $\{3, 3, 3\} \times 6$.

at the base station due to the processing and power limits. In order to reduce the influence of the channel estimation errors on the performance of the precoding techniques we can either invest in more processing power to implement more efficient channel estimation techniques or increase the pilot overhead. By using a denser pilot grid at the bandwidth edges we can reduce the MSE [67]. One more straightforward way of reducing the channel estimation errors is to increase the transmit power or antenna gain. Values of the transmit power and the antenna gain are set by the government regulations so there we do not have too many options. Therefore, a logical conclusion is that if we cannot improve the estimator performance by increasing the processing power, then the other option is to either increase the number of antennas or to reduce the number of spatial data streams. As we will see in the chapter on system level investigations, the last approach gives very good results.

5.10 Implementation and complexity of MU MIMO processing techniques

Clearly spatial processing has an impact on system and terminal complexity, which are important criteria for system design. Some algorithms are very powerful, yet complex that might be impractical for current systems. We can investigate the baseband and the radio frequency (RF) complexity of a system independently.

RF complexity arises from the number of antennas, the number of separate RF chains, and whether the signal paths of the different RF chains need to be calibrated for phase coherency. One beneficial side effect from using multiple antennas is that the total transmitted power is distributed over several RF chains, relaxing the requirements in terms of peak transmitted power for the high power amplifiers, thus enabling the use of cheaper high power amplifiers (HPA). As Dirty RF effects, i.e., RF non-linearities (such as phase noise, I/Q imbalance, etc.), can be expected to worsen as systems are designed for higher carrier frequencies, lower supply voltages and higher SNR regions, such effects should be taken into account when designing multi-antenna systems.

The baseband complexity is related to the relative energy and cycle count required for one run of the algorithm. The results presented in this section constitute only a rough estimation of the baseband complexity of different precoding and decoding techniques. The details of the general methodology for complexity assessment are described in [68]. The focus in the following assessments is on power consumption and delay only.

It is reasonable to classify the operations required for a run of a transceiver algorithm according to their complexity as follows: Simple arithmetic operations (add, abs, shift, max, etc), multiplications, divisions, square roots, and non-linear functions in general. The cost related to these functions, in terms of energy and cycle count are shown in Table 5.6, [68].

	Simple operation	Multiplication	Division	Square root	Non-linear function
Relative energy cost	1	10	40	50	60
Relative cycle count	1	1	4	5	6

Table 5.6: Relative energy and processing time costs

The cost in terms of energy and cycle count will be equal whenever we require only simple arithmetic operations for the execution of the algorithm. Processors can usually be designed to have the same cycle delay for addition and multiplication, at the expense of large area and energy consumption. Divisions and other non-linear functions are imple-

for $i = 1 : K$

$\widetilde{\boldsymbol{H}}_i = [\ \boldsymbol{H}_1^T \ \ldots \ \boldsymbol{H}_{i-1}^T \ \boldsymbol{H}_{i+1}^T \ \ldots \ \boldsymbol{H}_K^T\ ]^T$;

(1) $\boldsymbol{A} = \widetilde{\boldsymbol{H}}_i^H \widetilde{\boldsymbol{H}}_i$;

(2) $\boldsymbol{A} = \boldsymbol{A} + \frac{\sigma_n^2 M_R}{P_T} \boldsymbol{I}_{M_T}$;

for $j = 1 : M_{R_i}$

(3) $\boldsymbol{B} = \boldsymbol{h}_{i,j}^* \boldsymbol{h}_{i,j}^T$;

(4) $\boldsymbol{P} = (\boldsymbol{A} + \boldsymbol{B})^{-1} [\ \underset{\widetilde{\boldsymbol{H}}_i}{\boldsymbol{h}_{i,j}^T}\ ]$;

$\boldsymbol{F}_a = [\ \boldsymbol{F}_a \quad \boldsymbol{P}(:,1)\ ]$;

end;

end;

Table 5.7: SMMSE algorithm.

mented via iterative interpolations, explaining the higher cost for these operations. Energy consumption is mostly of interest for the user terminal, since power consumption is usually not a limiting factor at the base station.

As an example we will show the relative energy and cycle count for SMMSE and IRBD. SMMSE and IRBD consist of two steps. First, the multi-user interference is suppressed using $\boldsymbol{F}_a$ and then in the second step the independent processing of the users' effective channel matrices is performed. In this section we consider only the first step when we generate the matrix $\boldsymbol{F}_a$ and evaluate its complexity. The SMMSE algorithm is summarized in Table 5.7 and IRBD algorithm is given in Table 5.10.

The number of cycles and the relative energy count for operations (1) and (2) are given in Table 5.7 and for operations (3) and (4) are given in Tables 5.8 and 5.9, respectively.

	Matrix multiplication (1) +addition (2)
Real ADD	$M_T^2(2(2(M_R - M_{R_i}) - 1) + M_T$
Real MUL	$4M_T^2(M_R - M_{R_i}) + M_T$
Real DIV	–
Total cycle count	$M_T^2(8(M_R - M_{R_i}) - 2) + 2M_T$
Total energy consumption	$M_T^2(44(M_R - M_{R_i}) - 2) + 11M_T$

Table 5.8: Relative energy and processing time costs for steps (1) and (2) of SMMSE algorithm given in Table 5.7.

	Outer product (3)+matrix inversion (4)
Real ADD	$\frac{M_T^3}{2} + \frac{3M_T^2}{2} + \frac{3M_T}{4} + (M_R - M_{R_i} + 1)(M_T^2 - M_T) + 2M_T^2$
Real MUL	$\frac{M_T^3}{2} + \frac{M_T}{4} + (M_R - M_{R_i} + 1)(M_T^2 - M_T) + 4M_T^2$
Real DIV	$2M_T + (M_R - M_{R_i} + 1)M_T$
Total cycle count	$M_T^3 + 7.5M_T^2 + 9M_T + 2(M_R - M_{R_i} + 1)(M_T^2 + M_T)$
Total energy consumption	$5.5M_T^3 + 43.5M_T^2 + 83.25M_T + (M_R - M_{R_i} + 1)(11M_T^2 + 29M_T)$

Table 5.9: Relative energy and processing time costs for steps (3) and (4) of SMMSE algorithm given in Table 5.7.

for $i = 1$: no iterations
  for $i = 1 : K$
    $\widetilde{\boldsymbol{H}}_i = \left[\begin{array}{cccccc} \boldsymbol{H}_1^T & \cdots & \boldsymbol{H}_{i-1}^T & \boldsymbol{H}_{i+1}^T & \cdots & \boldsymbol{H}_K^T \end{array} \right]^T$;
(1) $\widetilde{\boldsymbol{H}}_i = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}}^H$;
(2) $\boldsymbol{S} = \text{sqrt}(\widetilde{\boldsymbol{\Sigma}}^T\widetilde{\boldsymbol{\Sigma}} + \frac{\sigma_n^2 M_R}{P_T}\boldsymbol{I}_{M_T})$;
(3) $\boldsymbol{F}_{a_i} = \widetilde{\boldsymbol{V}}\boldsymbol{S}$;
(4) $\boldsymbol{H}_i\boldsymbol{F}_{a_i} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H$;
(5) $\boldsymbol{H}_i = \boldsymbol{U}^{(r_i)\ H}\boldsymbol{H}_i$;
  end;
end;

Table 5.10: IRBD algorithm.

	SVD
Real ADD	$18(M_R - M_{R_i})M_T^2 + 4M_T^3$
Real MUL	$24(M_R - M_{R_i})M_T^2 + \frac{16}{3}M_T^3$
Total cycle count	$42(M_R - M_{R_i})M_T^2 + \frac{28}{3}M_T^3$
Total energy consumption	$258(M_R - M_{R_i})M_T^2 + \frac{172}{3}M_T^3$

Table 5.11: Relative energy and processing time costs for step (1) in the first iteration of IRBD algorithm given in Table 5.10.

	SVD
Real ADD	$18(r - r_i)M_T^2 + 4M_T^3$
Real MUL	$24(r - r_i)M_T^2 + \frac{16}{3}M_T^3$
Total cycle count	$42(r - r_i)M_T^2 + \frac{28}{3}M_T^3$
Total energy consumption	$258(r - r_i)M_T^2 + \frac{172}{3}M_T^3$

Table 5.12: Relative energy and processing time costs for step (1) in iterations after the first one of IRBD algorithm given in Table 5.10.

The IRBD algorithm is summarized in Table 5.10. Note that the dimensions of the matrices $\widetilde{\boldsymbol{H}}_i$ and $\boldsymbol{H}_i$ are $(M_R - M_{R_i}) \times M_T$ and $M_{R_i} \times M_T$ only in the first iteration. In the following iterations their dimensions are $(r - r_i) \times M_T$ and $r_i \times M_T$, respectively, where $r \leq \text{rank}(\boldsymbol{H})$ is the total number of data streams and r_i is the number of data streams transmitted to the i^{th} user. Therefore, the baseband complexity of the first iteration will be the highest. In order to estimate the number of cyclic counts and relative power consumption of SVD operations in steps (1) and (4) of the algorithm, we will use the results reported in [69].

The complexity of the SVD in step (1) in the first iteration is given in Table 5.11. In every following iterations, the complexity of the SVD in step (1) is given in Table 5.12. The complexity of steps (2), (3), (4) and (5) is given in Tables 5.14, 5.15 and 5.16.

From the tables given in this section we can see that the complexity of IRBD is higher than the complexity of SMMSE. The higher complexity of IRBD is the price that we have

	Scaling matrix $\boldsymbol{S}$
Real ADD	$\text{rank}(\widetilde{\boldsymbol{H}}_i)$
Real MUL	$\text{rank}(\widetilde{\boldsymbol{H}}_i)$
Real SQRT	M_T
Total cycle count	$2\text{rank}(\widetilde{\boldsymbol{H}}_i) + 5M_T$
Total energy consumption	$11\text{rank}(\widetilde{\boldsymbol{H}}_i) + 50M_T$

Table 5.13: Relative energy and processing time costs for step (2) of IRBD algorithm given in Table 5.10.

	Matrix multiplication
Real MUL	$2M_T^2$
Total cycle count	$2M_T^2$
Total energy consumption	$20M_T^2$

Table 5.14: Relative energy and processing time costs for step (3) of IRBD algorithm given in Table 5.10.

	SVD
Real ADD	$18M_{R_i}M_T^2 + 4M_T^3$
Real MUL	$24M_{R_i}M_T^2 + \frac{16}{3}M_T^3$
Total cycle count	$42M_{R_i}M_T^2 + \frac{28}{3}M_T^3$
Total energy consumption	$258M_{R_i}M_T^2 + \frac{172}{3}M_T^3$

Table 5.15: Relative energy and processing time costs for step (4) of IRBD algorithm given in Table 5.10.

	Matrix multiplication
Real ADD	$r_iM_T(4M_{R_i} - 2)$
Real MUL	$4r_iM_{R_i}M_T$
Total cycle count	$r_iM_T(8M_{R_i} - 2)$
Total energy consumption	$r_iM_T(44M_{R_i} - 2)$

Table 5.16: Relative energy and processing time costs for step (5) of IRBD algorithm given in Table 5.10.

to pay for a much better performance of IRBD than SMMSE. In order to achieve the same or a similar performance of SMMSE and IRBD we have to deploy more antennas at the base station while in case of IRBD we can use less antennas which reduces the difference between these two techniques. This trade off will depend on a specific deployment scenario and the target performance. The gain that IRBD provides compared to SMMSE is so significant that it justifies the higher baseband cost.

Chapter 6

System level performance investigations

For a comprehensive assessment of multi-antenna techniques, it is mandatory to consider the performance at system level, since many effects of spatial processing, like multi-user precoding, the impact of spatially-colored interference, and the benefits of interference management techniques are not tractable at the link level. In this chapter, we will investigate the performance of MU MIMO techniques in a system that was proposed within the WINNER project [67].

Major requirements for the next generation of wireless systems include among others high performance, robustness and adaptability to a wide range of scenarios and terminal classes. WINNER (Wireless World Initiative New Radio) is one of the most ambitious international research projects aiming at the development of a ubiquitous radio system concept providing wireless access for a wide range of services and applications across all environments from indoor/hotspot to wide area suburban and rural with one single scalable and adaptive system concept for all envisaged radio environments. The system concept proposed in [67] is a packet-oriented user-centric always-best concept.

Advanced multi-antenna solutions are an integral part of the future wireless communication systems. The WINNER overall system requirements mandate a single ubiquitous radio access system concept that is able to adapt to a comprehensive range of mobile communication scenarios with scalability in complexity. Under the constraint of low deployment effort and cost, it aims at complete coverage, while at the same time providing significant performance enhancements compared to legacy systems and their evolutions. Some of the goals of WINNER for the future wireless systems are:

- Improved spectral efficiency and increased user peak data rate
- Increased range or coverage in a cost-efficient manner
- Enhanced interference management
- Adaptivity to scenario and channel conditions.

Three different types of scenarios have been identified within the WINNER project: wide area, metropolitan area and local area. These three scenarios are illustrated in Figure 6.1.

In general there is a trade-off between different types of gains that multiple-antenna processing can provide. A specific processing at the transmitter and/or receiver is needed in order to leverage them, and the exact link gains depend critically on the properties of the radio channel and the amount of channel knowledge available at the receiver and transmitter.

Channel knowledge is typically described with two sorts of measures; channel state information and channel quality indicators (CQI). The term CSI usually refers to knowledge of the complex valued radio channel, while CQI, on the other hand, is rather a real valued measure of the quality of the channel, for example an SINR after receiver processing that may be used to adapt the code rate, modulation order, and spreading at the transmitter. The amount of channel knowledge dictates which methods are applicable and the potential benefits of spatial processing techniques.

For wide-area scenarios only long-term channel state information (CSI) seems to be reasonable for spatial processing in the majority of cases, most favorably combined with short-term channel quality information (CQI) for link adaptation. In all other scenarios, we assume that the instantaneous CSI is available, either due to reduced mobility or even fixed point-to-(multi)point connections.

TDD supports a high degree of adaptivity to actual propagation conditions with reasonable signaling overhead and is thus a key technique to reach the targets of future wireless communication systems. More important is the fact that using the reciprocity principle, in a TDD system it is possible to use the estimates of the uplink channel to perform precoding on the downlink. This significantly reduces the cost of acquiring the CSI at the base station and allows the use of more advanced MU MIMO precoding techniques for spatial processing.

MU MIMO precoding and decoding facilitates the simultaneous transmission of multiple data streams (SMUX) to multiple users (SDMA) which results in a significant through-

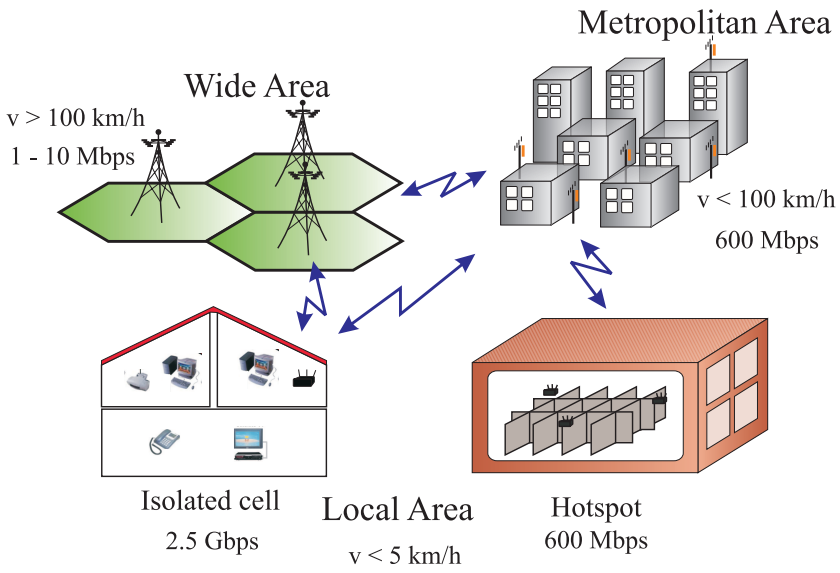

Figure 6.1: WINNER scenarios.

put improvement. In the previous chapter we have introduced several linear and non-linear precoding techniques.

Non-linear precoding techniques provide higher diversity than linear techniques at high SNRs. However, the point where non-linear precoding techniques become better than linear depends on the specific antenna configuration of the system, e.g., the number of antennas at the base station and the number of user terminals and antennas at the user terminals. Linear precoding techniques can achieve the sum-rate capacity bound of the broadcast channel when the number of users in the system is large and appropriate spatial scheduling of users is performed or when the total number of antennas at the user terminals is greater than the number of antennas at the base station. This was illustrated with the example of IRBD in Section 5.6. Furthermore, linear precoding techniques allow the combination of instantaneous CSI for some users and long-term CSI for others, unlike non-linear precoding techniques which require the exact CSI in order to be able to pre-subtract the non-causal interference. Together with a lower computational complexity this renders linear precoding techniques more favorable for practical implementation than non-linear precoding techniques. Therefore, we will focus only on the system level performance of linear precoding techniques. Based on the link level investigations the most promising technique is IRBD. This linear precoding techniques can be used on the downlink and with minor modifications also on the uplink which will have a big impact on the lower complexity of the base station hardware.

Scenarios that will be considered describe the relevant characteristics of a selection of

environments where a system will be operated. Our focus will be on an indoor/hotspot scenario where the MU MIMO processing is expected to provide the maximum throughput gains due to the low user mobility and availability of the channel-state information at the base station. We will also investigate a micro cellular scenario where MU MIMO processing will be used with the combination of instantaneous and long-term CSI. In [70] it is proposed that in the indoor hotspot scenario the TDD physical layer mode and 100 MHz bandwidth at 5 GHz should be used, and in the urban micro-cellular scenario the TDD physical layer mode and 100 MHz bandwidth at 3.95 GHz should be used. Micro cellular deployment envisage smaller cells for more dense usage in a typical Manhattan deployment whereas indoor/hotspot considers isolated cells or a couple of isolated cells for home or small office coverage.

Metropolitan area scenarios consider large urban environments and should provide contiguous outdoor coverage especially in city centers of large and medium size cities. It should support high user density, high system throughput, and mobility up to reasonable velocities in urban environments, e.g., 50 km/h. In urban scenarios base stations and relays are placed clearly below the rooftop level. Spatial schemes in the metropolitan area have to cope with these challenging radio propagation conditions, and must be able to meet the high throughput requirements. The limited mobility makes availability of instantaneous CSI at the transmitter possible [70].

Local area scenarios cover few cells, isolated sites, and peer-to-peer communication in an indoor/hotspot scenario. They are characterized by high data rates and a high traffic demand. Low mobility (up to 5 km/h) allows the estimation of short-term CSI at the transmitter based on the uplink measurements and the reciprocity principle.

For in-home scenarios the WINNER base station controls a single cell, encountering limited interference from other systems as it can be assumed that the house represent a well protected environment. In hot-spot scenarios, the coverage area is larger than in the home deployment. Several WINNER base stations may be required and thus the assumption of limited interference may not be valid any more. However, if an appropriate wire-line infrastructure already exists, it is possible to interconnect the base stations with a central unit in order to allow the full cooperation between them. The cooperation between distributed antenna arrays or base stations provides a significant additional throughput gain.

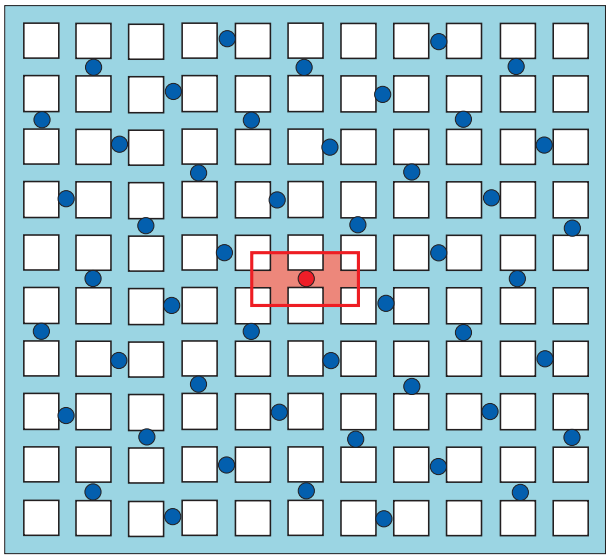

Figure 6.2: Manhattan grid. The buildings are represented with squares. The dots mark the positions of the base stations. The cell of interest is denoted in the center of the grid.

6.1 Simulation setup

The environment specific characteristics that will be used later for system level investigations are given in the Table 6.1, [70]. The channel is modeled using the parameters reported in [34].

	Metropolitan area	Local area
Environment characteristics	Two-dimensional regular grid of buildings ("Manhattan grid")	One floor of a building with regular grid of rooms and corridors, three dimensional
User distribution model	Number of users is a variable parameter All users are uniformly distributed in the streets	Number of users is a variable parameter 90% of users are uniformly distributed in rooms and 10% of users are uniformly distributed in corridors

Table 6.1: Environment specific parameters.

The considered deployment scenario in the metropolitan area is a two-dimensional regular grid of buildings, the so-called Manhattan grid, where the users are located in the streets only. In the local area we consider the three-dimensional deployment scenario which is one floor of a building with rectangular grid of rooms and corridors. These two deployment scenarios are illustrated in the Figure 6.2 and Figure 6.3, respectively. In our simulations we use WINNER channel model A1 for indoor scenario and WINNER channel model B1 for metropolitan area, [34], [71].

In the Manhattan grid, the base stations are placed in the center of the streets, in

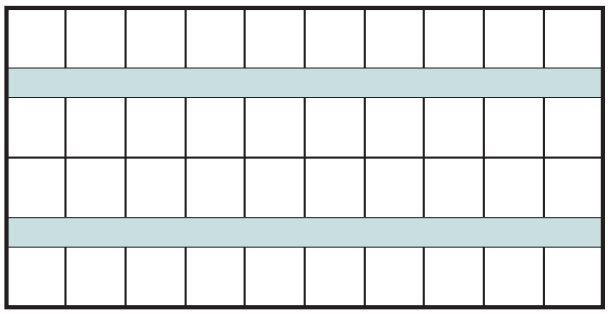

Figure 6.3: Small office scenario. There are two corridors with 10 offices on each side of the corridors.

the center of blocks. The streets are 30 m wide, and the blocks are 200×200 m. Each base station covers one cell, and is equipped with a cross-polarized 8-element uniform linear array (ULA) with omni-radiating elements. The UTs are equipped with two cross-polarized antennas.

In the small office scenario we consider one floor 50×100 m, room size 10×10 m and with 2 corridors of size 100×5×3 m. The base station and the UTs are equipped with linearly polarized ULAs with omni-radiating elements.

The physical channel structure proposed within WINNER divides the available time-frequency resources into chunks, each chunk consisting of a set of (orthogonal) waveforms. The chunks are considered two-dimensional, and for the downlink, each chunk consists of a number of subcarriers and a number of consecutive OFDM symbols. When multiple antennas are introduced, the spatial dimension is added to the chunks. The chunks may thus be viewed as three dimensional, and the third dimension will be referred to as a layer. The chunk durations and frame durations are short, to ensure a low transmission delay over the radio interface (below 1 ms). The chunks are organized into frames, and for the TDD mode each frame contains a downlink transmission interval followed by an uplink transmission interval, referred to as slots. A super-frame consists of a preamble, which is transmitted in a commonly available frequency band, followed by 8 frames.

A number of different users are scheduled and concurrently use the available resources. In the most general case, the multiple access can use the time, frequency, code, and space domain.

In our simulations we will adopt a simplified traffic model based on the full queue assumption without detailed modeling of packets. The investigations will be based on snapshots without short-term evolution of the individual and uncorrelated drops of large-scale parameters (like user location, channel realization, etc). Users have fixed positions,

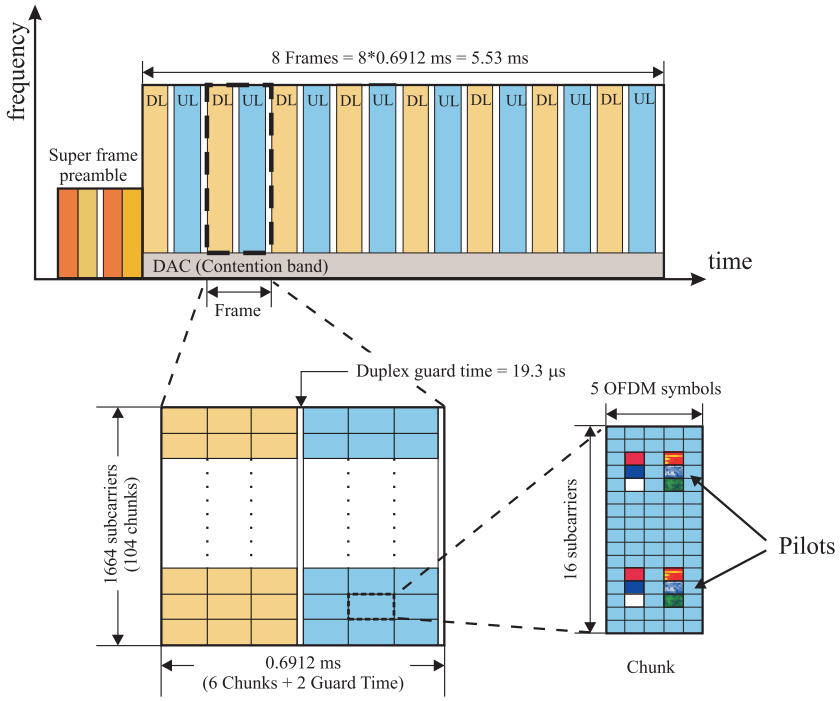

Figure 6.4: WINNER super frame, frame and chunk configuration.

and the users' channels vary only due to the fast fading in one snapshot by considering Doppler spread.

Conceptually the full buffer model is most closely related to traffic like ftp. For a given number of users per cell, the full queue traffic model offers maximum multi-user diversity.

As a performance measure we use user and cell throughput.

The user throughput is defined as the ratio of correctly received information bits on one link to one user to the total simulation time (i.e., the time that elapses in the real system) for this link. Statistics are collected from all links within the evaluation area (e.g., center cell or original cells in case of a simulator using the wrap-around technique). The cumulative distribution function (CDF) of these user throughput values is provided and comparisons are based on the corresponding percentiles.

The cell throughput is defined as the aggregate number of correctly received information bits within one cell per simulation time step. Samples are taken at each time step from all cells within the evaluation area. From these samples the CDF of the cell throughput is calculated. Cells are defined as the parts of a site with a fixed allocation of antenna resources. A site itself is the location of the base station hardware.

Frame and deployment specific parameters are listed in the Tables 6.2, 6.3 and 6.4, [70].

	Metropolitan area	Local area
duplexing (asymmetry)	TDD (1:1)	TDD (1:1)
carrier frequency f_c	3.95 GHz	5 GHz
system bandwidth	100 MHz	100 MHz
BS max. transmit power	37 dBm	24 dBm
number of antennas at BS	8	8, 16, 24
BS antenna configuration	ULA cross polarized	ULA linearly polarized
antenna element spacing	0.5 λ	0.5 λ
azimuth antenna element pattern	omni directional	omni directional
BS receiver noise figure	5dB	5dB
UT max transmit power	24 dBm	21 dBm
number of antennas at UT	2	2
UT antenna configuration	ULA cross polarized	ULA linearly polarized
antenna element spacing	0.5 λ	0.5 λ
azimuth antenna element pattern	omni directional	omni directional
UT receiver noise figure	7 dB	7 dB

Table 6.2: Deployment specific parameters.

DFT size	2048
Subcarriers used	1664
Subcarrier spacing f_0	48828.125 Hz
Useful symbol duration T_0	20.48 μs
Gurad interval T_{pre}	1.28 μs (128 samples)
Total symbol duration T_s	21.76 μs

Table 6.3: OFDM parameters.

The evaluation of the associated control overhead of each technique in the downlink and uplink and its impact on the user and cell throughput is an important part of the assessment process. We will assume that 12 symbols in each chunk are used for channel estimation and 8 symbols are used for transmission of the control information. We use 2 pilot symbols out of 12 in each chunk to estimate the channel from one transmit antenna. This implies that by using 12 pilot symbols per chunk we can estimate the channel from 6 transmitting antennas in total.

Channel estimation errors at the receiver cause a degradation in the decoding performance of the current data. Robustness needs in particular to be investigated for spatial processing techniques using instantaneous channel knowledge at the transmitter, like most precoding techniques. To obtain a first insight into this topic, a simple Gaussian error model is used to generate the erroneous channel estimates. Additionally, we use the outdated CSI which is estimated in the previous uplink transmission to perform the precoding.

Overall frame length	0.6912 ms
Number of OFDM symbols per frame	30
Chunk dimension in symbols × subcarriers	5 × 16
Pilot and signalling overhead per chunk (in number of resource elements)	20
Number of chunks per frame (in time and frequency direction)	6 × 104
Duplex guard time	2 × 19.2 μs

Table 6.4: Environment specific parameters.

The relative performance loss compared to the ideal case is used to characterize robustness with respect to channel estimation and CSI feedback errors. We also include the influence of calibration errors at the base station and the user terminals on the performance of the system.

6.2 Antenna multiplexing and complexity reduction

The main issue with the spatial processing is how to estimate the information required for their implementation. In case of precoding it is the problem of estimating the exact channel at the BS and the exact and/or effective channel at the mobile terminals. In order to do this we need to multiplex the sufficient number of pilots to estimate the channel from all of the transmitting antennas or spatial streams. In the current WINNER proposal [67], [70] for the metropolitan and local area scenario, TDD duplexing with the asymmetry 1:1 is used and here we rely on the reciprocity assumption in order to acquire the CSI at the BS necessary to perform the multi-user MIMO precoding.

The proposed 12 pilots in the adaptive mode per one chunk are not enough to multiplex the number of users necessary to exploit the multi-user diversity. Therefore, here we rely on the assumption that the channel is slowly varying, almost constant over one frame. Since there are 3 time slots in the frame, we can multiplex in every time slot pilots from the different antennas regardless of whether these antennas are located at the same terminal or not. We assume that one user transmits using only one fourth of the available bandwidth, and in this way we can multiplex pilots from a total of 72 different antennas if we multiplex 12 pilot symbols per chunk. Assuming that in the system there are 20% users equipped with one antenna, 40% equipped with 2 and 40% equipped with 4 antennas, we can multiplex on the average $72/(1 \cdot 0.2 + 2 \cdot 0.4 + 4 \cdot 0.4) = 27.7$ users.

The MU MIMO precoding should be performed on every subcarrier in every OFDM symbol. However, this results in high computational load. In order to reduce the complexity we will not perform the precoding on every subcarrier and symbol in the chunk but only once per chunk by using the equivalent channel as defined in Section 5.7. Moreover, if we multiplex pilots from different antennas as described in the previous paragraph we will use the same channel estimate, i.e., the precoding matrix, for all three time slots which further reduces the complexity. This means that in every frame the precoding is performed only 104 times. The same approach can be used on the uplink also.

6.3 Simulation results

In this section we present the system level simulation results of a WINNER system downlink with IRBD precoding.

The number of users in the system is variable and it depends on the pilot overhead assumed. We assume 12 pilot tones per chunk which facilitates the estimation of the channel from 6 antennas. The frequency band is divided in 4 contention bands. If we have 12 pilot tones per chunk, we can estimate the channel from 6 antennas per chunk, i.e., from 18 antennas per contention band in total (3 chunks $\times$ 6 antennas). If we assume that every user is equipped with 2 antennas this means that we can estimate the channel from $18/2 = 9$ users per contention band. It is assumed that all users on one floor are assigned a unique, orthogonal pilot pattern for the channel estimation.

The maximum number of spatially multiplexed users $K_{\max}$ is fixed. This results in a 3-D resource that will be allocated to the users prior to transmission. Users are scheduled for transmission using the round-robin algorithm. More advanced adaptation, scheduling and resource allocation algorithms that take into account the spatial dimension and SINR which results in a better performance.

First we simulate assuming perfect CSI at the base station and user terminals. In order to investigate more realistic system performance we introduce channel estimation errors and calibration errors that are modeled as described in Sections 5.9 and 5.8, respectively. We assume an interpolation error floor of 30 dB and a channel estimator gain of 10 dB.

We compare the cell and user throughput with the SISO system. In case of SISO communication, we assume that there are 6 pilot symbols per chunk used to estimate the signals from 3 users.

The solid curves are simulated assuming perfect CSI at the base station and the user

terminals. The dashed curves are simulated assuming channel estimation errors and calibration errors at the base station and the user terminals.

6.3.1 Isolated cell indoor scenario

In this section the complementary cumulative distribution functions (CCDF) of the cell and the user throughput when there is one base station in the middle of one floor will be presented. We compare the system performance when there are 8, 16 and 24 antennas. The base station is in the center of the floor. The base station transmits one data stream per user, ($r_i = 1$).

In case when there are 8 antennas at the base station the maximum number of spatial layers is set to $K_{\max} = 5$. Note the high sensitivity of the system to the channel estimation errors and RF impairments.

In Figures 6.7 and 6.8 we present system level results when there are 16 antennas at the base station. In Figures 6.9 and 6.10 we present system level results when there are 24 antennas at the base station. The maximum number of spatial streams is fixed to 6 and 9 when there are 16 and 24 antennas at the base station, respectively. By deploying more antennas at the base station the sensitivity of the system to the calibration and channel estimation errors has been reduced.

In Figures 6.11 and 6.12 we investigate what happens when users have more than two antennas. In this case we have assumed that 60% of the users are equipped with two antennas and that 40% of the users are equipped with four antennas. By deploying more antennas at the user terminals we significantly improve the user throughput. However, since the number of the antennas from which we can estimate the signals is limited, the number of users and consequently the number of data streams will be lower. This reflects on the cell throughput which is slightly reduced. Note, that by using more antennas at the receiver we also reduce the sensitivity of the system to the calibration and channel estimation errors.

In Tables 6.5 and 6.6 we summarize the 90%, 50% and 10% cell and user throughput gains relative to the SISO system.

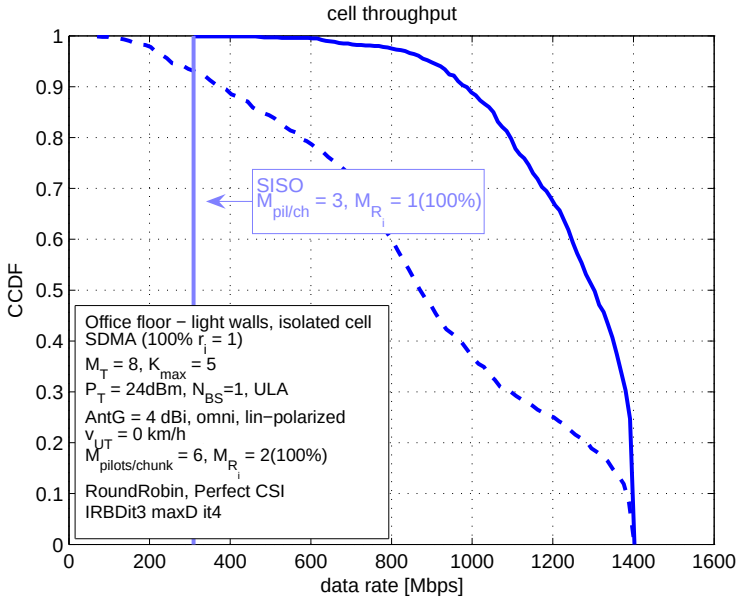

Figure 6.5: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 8$. Isolated cell.

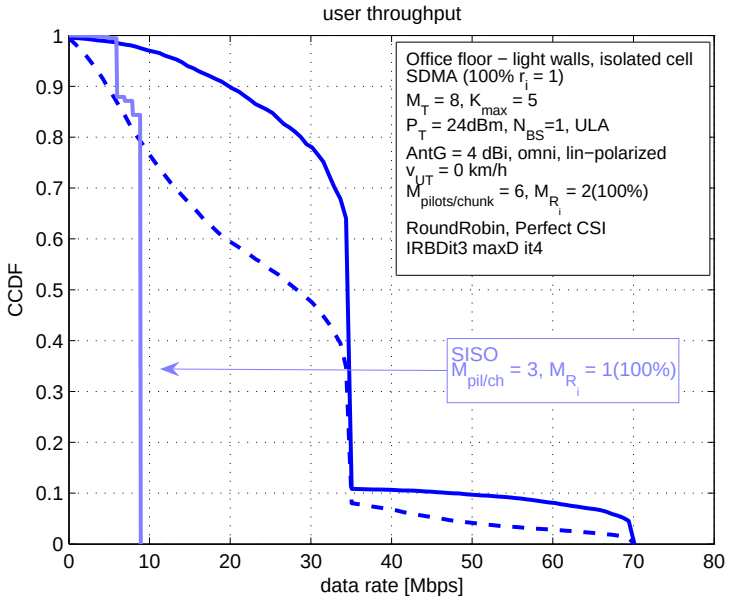

Figure 6.6: CCDF of user throughput. IRBD precoding. Number of antennas at the base station $M_T = 8$. Isolated cell.

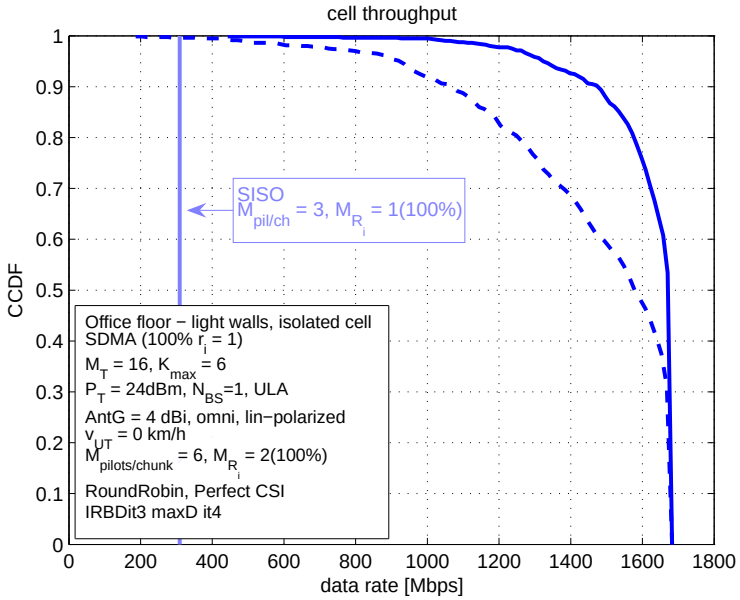

Figure 6.7: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 16$. Isolated cell.

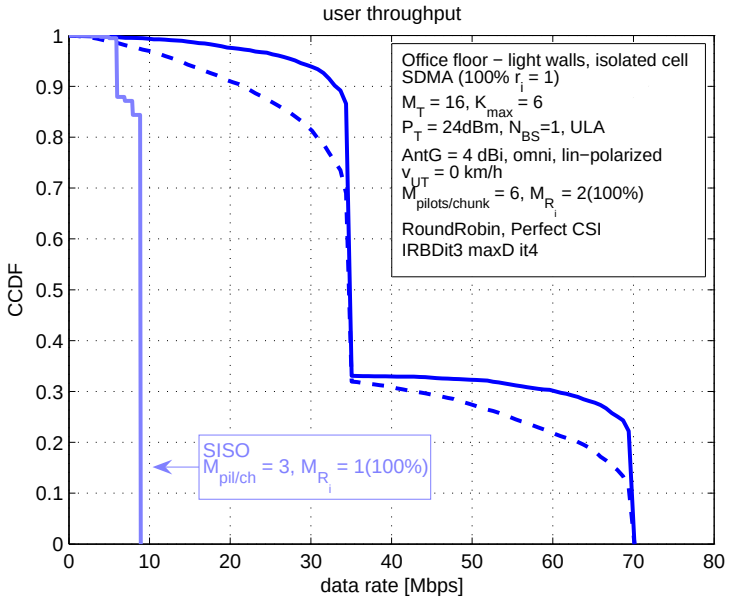

Figure 6.8: CCDF of user throughput. IRBD precoding. Number of antennas at the base station $M_T = 16$. Isolated cell.

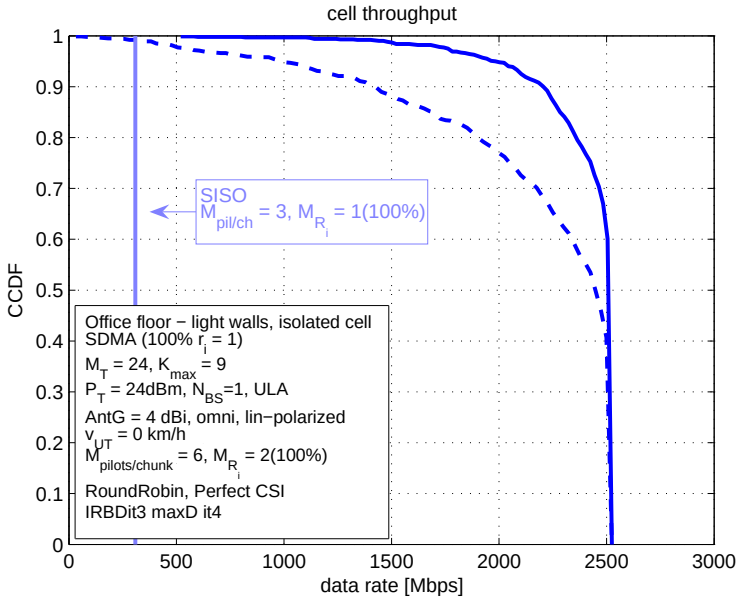

Figure 6.9: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 24$. Isolated cell.

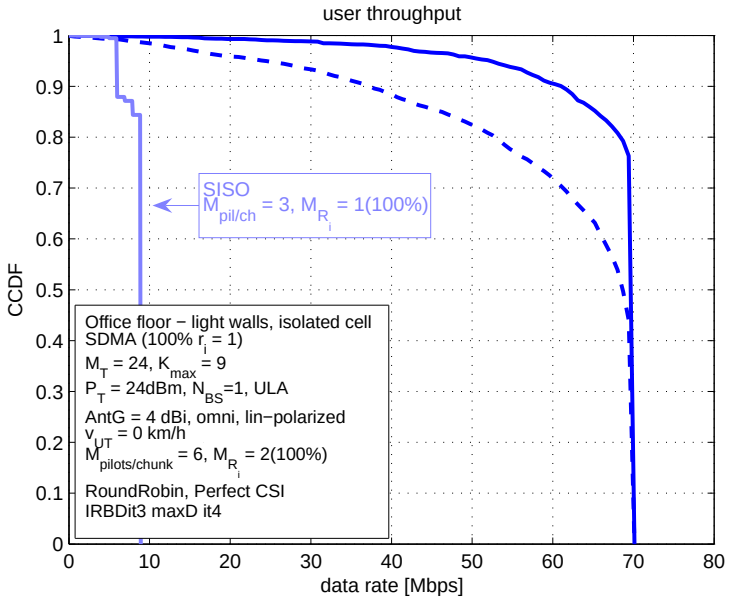

Figure 6.10: CCDF of user throughput. IRBD precoding. Number of antennas at the base station $M_T = 24$. Isolated cell.

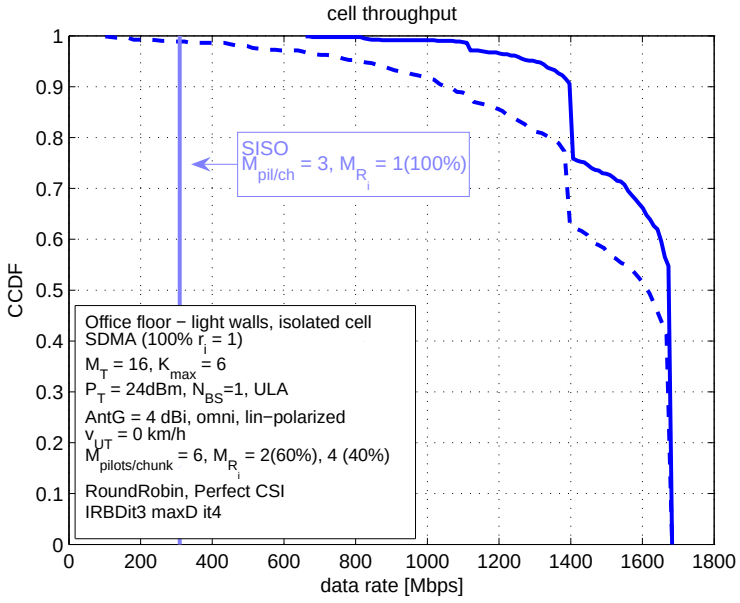

Figure 6.11: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 16$. Isolated cell.

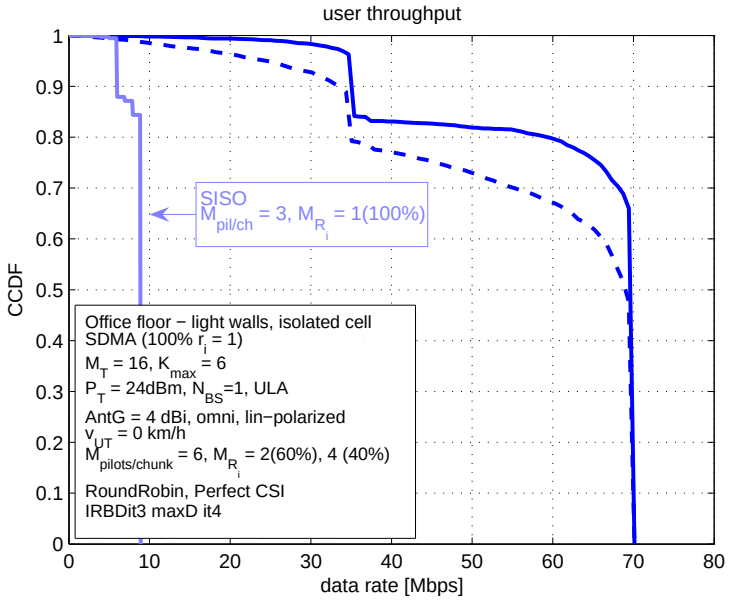

Figure 6.12: CCDF of user throughput. IRBD precoding. Number of antennas at the base station $M_T = 16$. Isolated cell.

Number of antennas	X% Outage cell throughput/SISO throughput		
	90%	50%	10%
8 ($M_{R_i} = 2$)	3.2036 (1.2515)	4.2335 (2.8545)	4.5449 (4.4970)
16 ($M_{R_i} = 2$)	4.8503 (3.4832)	5.4027 (5.1208)	5.5170 (5.5170)
24 ($M_{R_i} = 2$)	7.1798 (4.5955)	8.1685 (8.1724)	8.2135 (8.2135)
16 ($M_{R_i} = 2, (60\%)$) ($M_{R_i} = 4, (40\%)$)	4.5302 (3.4161)	5.4161 (5.2416)	5.5170 (5.5170)

Table 6.5: Outage cell throughput relative to SISO system throughput. Small office scenario. Isolated cell.

Number of antennas	X% Outage cell throughput/SISO throughput		
	90%	50%	10%
8 ($M_{R_i} = 2$)	3.3385 (0.8154)	3.8660 (3.1414)	5.3505 (3.9895)
16 ($M_{R_i} = 2$)	5.7302 (3.6154)	3.9072 (3.9072)	7.9684 (7.7835)
24 ($M_{R_i} = 2$)	10.169 (6.2308)	7.9684 (7.6804)	8.0105 (8.0105)
16 ($M_{R_i} = 2, (60\%)$) ($M_{R_i} = 4, (40\%)$)	5.8615 (5.5846)	7.9474 (7.7010)	7.8454 (7.8454)

Table 6.6: Outage user throughput relative to SISO system throughput. Small office scenario. Isolated cell.

6.3.2 Cellular large office scenario

In this section we investigate the system level performance of IRBD in a small office scenario when there are $N_{BS} = 4$ base stations in the system. The position and the orientation of the antenna arrays at the base stations are shown in Figure 6.13. We assume that the base stations are connected to the central unit via a high-speed wireline network. The channel estimates are sent to the central unit where the spatial processing is performed. The precoding matrices are then sent back to the base stations. We assume that the base stations are perfectly synchronized.

In Figures 6.14 and 6.15 we show the CCDFs of the cell and the user throughput when there are $M_T = 4$ antennas at each base station. When there is perfect CSI available in the system, by placing smaller antenna arrays that can cooperate, 90% cell throughput is greater than 90% cell throughput of a system where all antennas are placed in the center of the floor, which is shown in Figure 6.7, by more than 1 Gbps. In presence of calibration and channel estimation errors 90% cell throughput is 500 Mbps greater.

In Figures 6.16 and 6.17 we show the CCDFs of the cell and the user throughput when there are $M_T = 6$ antennas per base station. In this case, when there is perfect CSI available at the BS we combine SDMA and SMUX by allowing approximately 50% of the users to transmit $r_i = 2$ data streams. As a consequence the peak cell throughput is close to 4 Gbps and the peak user throughput is close to 140 Mbps. When we include

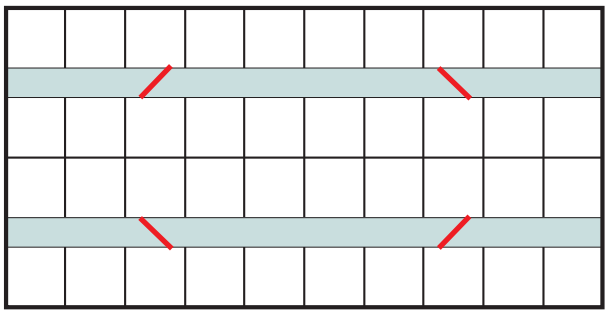

Figure 6.13: Position of the antenna arrays in a small office scenario.

both calibration and channel estimation errors it is not possible to transmit more than one data stream per user. However, by using the distributed antenna arrays that can fully cooperate, the cell and user throughput in presence of channel estimation errors and calibration errors are almost the same as in the case when we have one big antenna array in the center of the floor and perfect CSI which is shown in Figure 6.9. Combination of SDMA and SMUX provides higher peak cell and user throughputs. However, when we take into account hardware impairments SDMA is more favorable since it is more resistant to the calibration and channel estimation errors due to the higher diversity.

The 90%, 50% and 10% outage cell and user throughput are given in Tables 6.7 and 6.8. As a result of base station cooperation, we have larger cell and user throughput and a reduced sensitivity to the calibration and channel estimation errors.

Number of antennas per BS	X% Outage cell throughput/SISO throughput		
	90%	50%	10%
4	7.5169 (4.5955)	8.3563 (7.0230)	8.4023 (8.1954)
6	6.6418 (6.4925)	10.723 (8.1692)	12.283 (8.1642)

Table 6.7: Outage cell throughput relative to SISO system throughput. Small office scenario. Distributed MIMO.

Number of antennas per BS	X% Outage cell throughput/SISO throughput		
	90%	50%	10%
4	10.630 (5.8000)	7.9684 (6.7113)	7.8454 (7.8842)
6	6.0476 (5.7937)	7.9474 (7.8842)	8.0105 (8.0105)

Table 6.8: Outage user throughput relative to SISO system throughput. Small office scenario. Distributed MIMO.

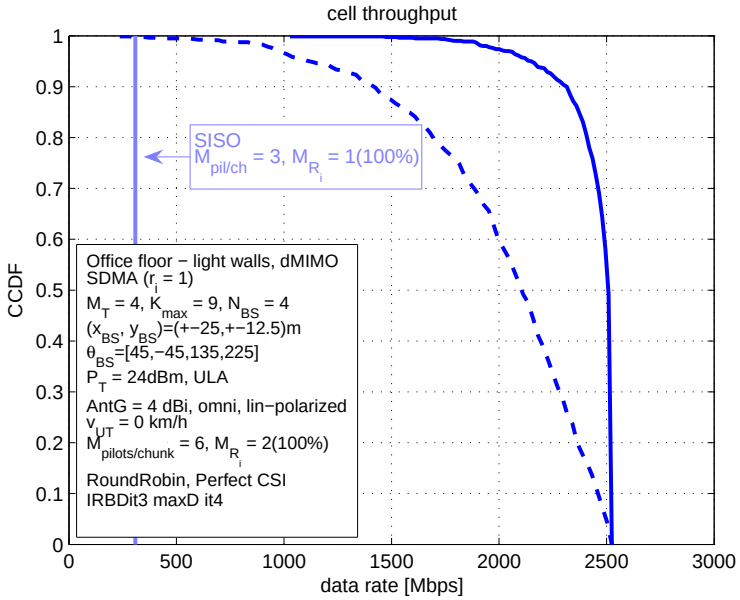

Figure 6.14: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 4$. Number of base stations $N_{BS} = 4$.

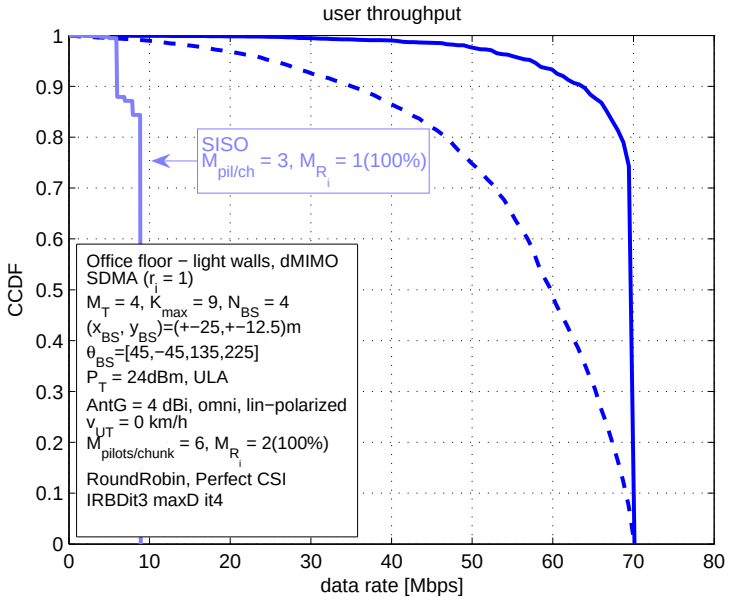

Figure 6.15: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 4$. Number of base stations $N_{BS} = 4$.

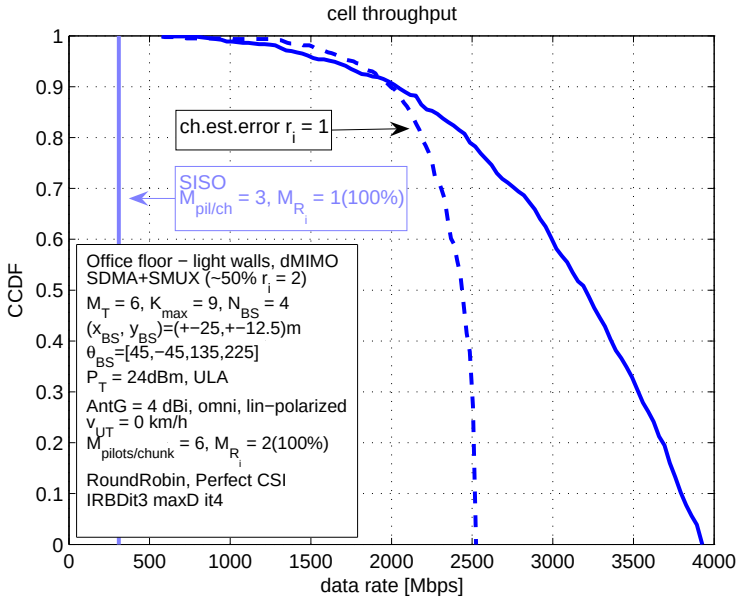

Figure 6.16: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 6$. Number of base stations $N_{BS} = 4$.

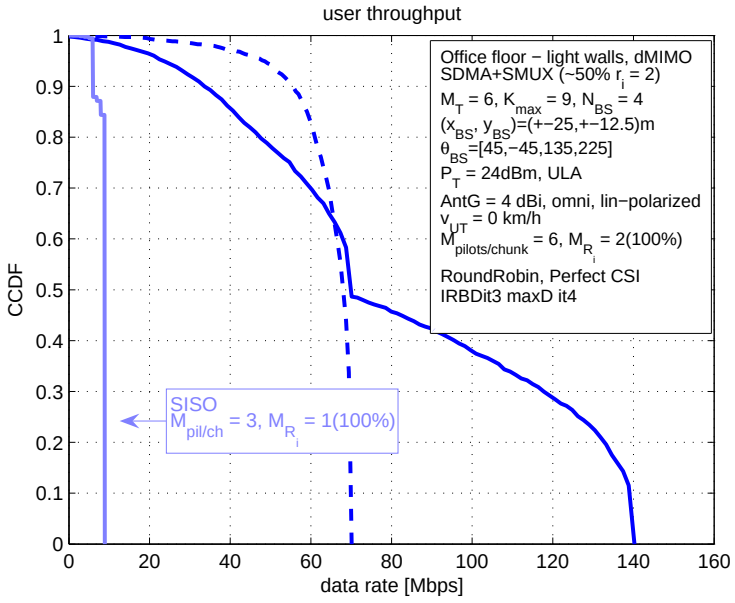

Figure 6.17: CCDF of user throughput. IRBD precoding. Number of antennas at the base station $M_T = 6$. Number of base stations $N_{BS} = 4$.

6.3.3 Manhattan scenario

In this section we show the results of the system level simulations of a MU MIMO cellular system employing IRBD for downlink precoding. We compare the cell and the user throughput of this system to the cell and the user throughput of a SISO system. All users have the same speed. User speed is fixed during one simulation run and it is equal to 3 km/h and 50 km/h.

As it can be seen from Figures 6.18 and 6.19, when users are slowly moving, MU MIMO processing provides around 60 % higher average cell and user throughput. However, when users are moving fast, 90 % cell and user throughput are even worse than in case of SISO transmission. This is the consequence of the precoding based on the long-term CSI. We have assumed perfect channel estimation and calibration. Similar to the isolated/hotspot scenario, the performance of the MU MIMO precoding in a cellular scenario when users are fast moving should improve if we deploy more antennas at the base station. By using more antennas at the base station it becomes easier to separate users' left signal subspaces which results in an improved performance.

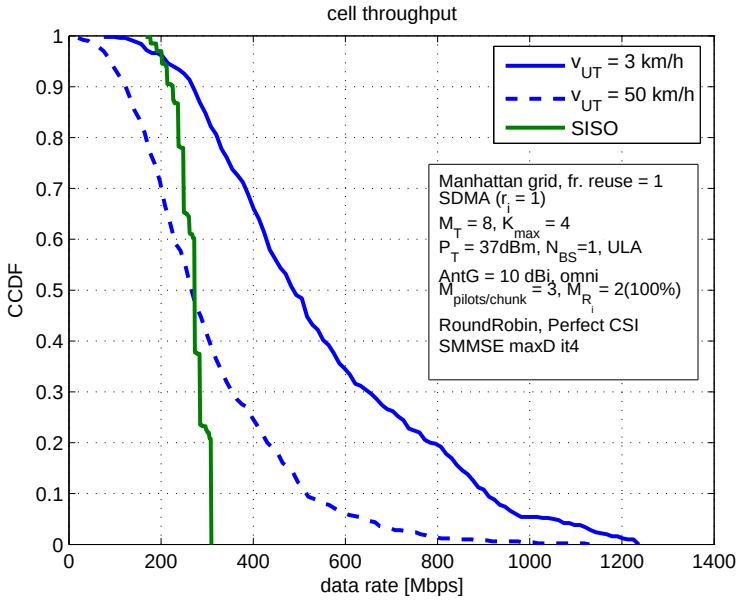

Figure 6.18: CCDF of cell throughput. IRBD precoding. Number of antennas at the base station $M_T = 8$. Manhattan scenario.

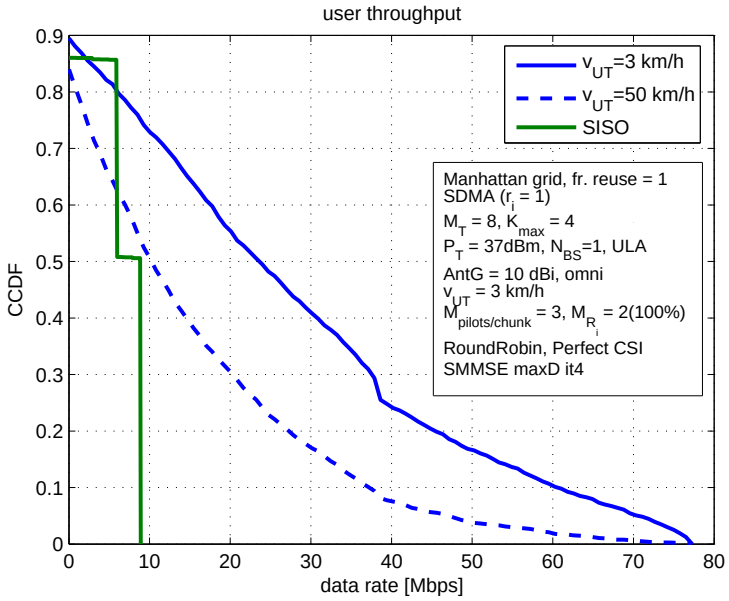

Figure 6.19: CCDF of user throughput. IRBD precoding. Number of antennas at the base station $M_T = 8$. Manhattan scenario.

Chapter 7

Conclusions

In this thesis we have addressed the problem of designing multi-user MIMO processing techniques. The goal was to define one technique that can target several optimization criteria, a technique that will be able to adapt to different qualities of channel state information and that can combine instantaneous and long-term channel state information.

Information theoretic results have shown that the linear increase of capacity in a multi-user MIMO system is possible to obtain only by spatially multiplexing users and by sending multiple data streams to each user. It is known that the maximum sum-rate capacity in a multi-user uplink system is achieved by MMSE decoding and successive interference cancellation. In order to achieve high data rates foreseen by the information theoretic investigations in a multi-user downlink system, it is necessary to use "dirty paper" codes (DPCs). A simple example for such a code is Tomplinson-Harashima precoding (THP). Except THP, DPCs are in general very complex and almost impossible to implement. On the other hand, it was shown in the literature that it is possible to approach the sum-rate capacity of a multi-user downlink system either by combining linear precoding techniques with THP or by using smart resource allocation and user scheduling. We have also shown through simulations that linear techniques can approach the DPC sum-rate capacity bound when the total number of antennas at the user terminals is greater than the number of antennas at the base station.

As it was previously reported in the literature, when user terminals are equipped with one antenna, MMSE in combination with successive interference cancellation is optimum on the uplink, while MMSE precoding in combination with THP is optimum on the downlink. However, MMSE suffers from a performance loss when users are equipped with more than one antenna. Block diagonalization (BD), on the other hand, is a zero forcing tech-

nique that can serve users with an arbitrary number of antennas under the condition that the total number of antennas at the user terminals is less than or equal to the number of antennas at the base station.

The first new technique that we have introduced in this thesis is BD MMSE THP which combines BD for users that are equipped with multiple antennas and MMSE THP for users equipped with a single antenna. This technique is very simple and provides better performance in a mixed antenna scenario compared to MMSE THP or BD. SO THP is a technique that combines BD and THP. In simulations it approaches the sum-rate capacity of the multi-user MIMO downlink system. However, both BD MMSE THP and SO THP still suffer from the dimensionality problem that the total number of antennas at the user terminals has to be less than or equal to the number of antennas at the base station.

The design of the precoding matrix introduced in this thesis is general and can target several optimization criteria like maximum information rate, maximum receive SNR, minimum BER, etc. We have achieved this by separating the problems of MUI suppression and optimization. In the first step, we minimize the overlap of the row spaces spanned by the effective channels of different users. In the next step, we optimize the system performance with respect to the specific optimization criteria assuming that we have a set of parallel single user MIMO channels.

The two MU MIMO processing techniques that result from the two different MSE criteria are successive MMSE (SMMSE) and regularized block diagonalization (RBD). The only limitation in this case is that the total number of users in the system has to be less than or equal to the rank of the combined network channel matrix of all users.

Successive MMSE precoding and decoding techniques are designed by using the modified MSE cost function. As a result, by using SMMSE we do not suppress the interference between two or more antennas at the same user terminal but combine their signal which results in a higher array and diversity gain. In combination with THP, on the downlink we are able to approach the maximum sum rate capacity at low SNRs. SMMSE in combination with SIC on the uplink provides higher diversity gain than V-BLAST.

Regularized block diagonalization is balancing the overlap of the row spaces spanned by the effective channel matrices of different users and MIMO gains by minimizing the Frobenius norm of the effective channels between different users and the Frobenius norm of the scaled noise vector. We separate the transmission to the different users using a new cost function that includes multi-user interference suppression and the avoidance of noise

enhancement. The new technique is called regularized block diagonalization since at high signal-to-noise ratios and under the condition that the total number of antennas at the user terminals is less than or equal to the number of antennas at the base station, the effective combined channel matrix is block diagonal. In addition to this linear technique we also present several variants. The first variant of this technique is a non-linear alternative where we combine RBD with THP (RSO THP). RSO THP approaches the maximum sum-rate capacity in simulations for low multi-user interference, i.e., when the total number of antennas at the user terminals is less than or equal to the number of antennas at the base station. The other way to improve the system performance is to iterate the closed form solution or perform joint processing over a group of multi-user MIMO channels in different frequency and time slots. Performance gains that result from iterative (IRBD) or joint processing (JRBD) are especially visible when the total number of antennas at the user terminals is greater than the number of antennas at the base station which leads to high multi-user interference. By iterating the closed form solution with appropriate power loading we are able to extract the full diversity in the system and empirically approach the maximum sum-rate capacity in case of high multi-user interference. Joint processing of MIMO channels yields maximum diversity regardless of the level of multi-user interference.

As these techniques rely on the fact that there is either instantaneous or long-term channel state information (CSI) available at the base station to perform precoding and decoding, it is very important to investigate the influence of the transceiver front-end imperfections and channel estimation errors on their performance. The CSI at the transmitter can be acquired either through feedback of the channel coefficients or by using the estimates of the channel transfer function and the reciprocity principle. In a TDD system the reciprocity principle allows us to use the estimates of the channel on the uplink to perform precoding on the downlink. This significantly reduces required feedback needed to acquire the channel state information at the transmitter. However, in this case the problem of the influence of the radio front-end characteristics on the performance of the system arises. This is a consequence of the fact that the transfer functions of the transmit RF chain and the receive RF chain in general are not identical. In order to cope with the RF front-end impairments and to meet the conditions so that reciprocity principle holds, it is necessary to either perform calibration or use reciprocal transceivers.

We have shown in our simulations that channel estimation errors result in an SNR loss and that by using the self-calibration methods reported in the literature only at the

base station, the influence of the calibration errors on the system performance is almost negligible.

Another important issue is the complexity of the multi-user MIMO precoding and decoding techniques. RF complexity is directly related to the number of antennas, the number of separate RF chains, and whether the signal paths of different RF chains need to be calibrated for phase coherency. The baseband complexity is related to the relative energy and cycle count required for one run of the multi-user MIMO processing algorithm.

Non-linear precoding techniques provide higher diversity than linear techniques at the high SNRs. However, the point where non-linear precoding techniques become better than linear techniques depends on the specific antenna configuration of the system, e.g., the number of antennas at the base station and the number of user terminals and antennas at the user terminals. Linear precoding techniques can approach the sum-rate capacity bound of the broadcast channel when the number of users in the system is large and appropriate spatial scheduling of users is performed or when the total number of antennas at the user terminals is greater than the number of antennas at the base station. Furthermore, linear precoding techniques can adapt from instantaneous CSI to the long-term CSI and allow the combination of instantaneous and long-term CSI unlike non-linear precoding techniques which require the exact CSI in order to be able to pre-subtract the non-causal interference. Together with a lower computational complexity this renders linear precoding techniques more favorable for practical implementation than non-linear precoding techniques.

System level investigations have shown that MU MIMO precoding techniques provide several times higher data rates than SISO systems with only slightly increased pilot and control overhead. The biggest problem is the influence of user mobility, and calibration and channel estimation errors on their performance. A straightforward way to reduce the sensitivity of these techniques to real-life impairments is to deploy more antennas at the base station or to jointly process the signal from a group of spatially distributed antenna arrays. Distributed antenna arrays are capable of providing very large spectral efficiencies with reduced sensitivity to the real-life impairments at the cost of increased complexity.

Finally, we can conclude that multi-user MIMO processing is capable of providing high spectral efficiencies and thus will be an important part of the next generation wireless systems. However, special care must be taken during the system design, in order to account for the deployment scenario characteristics, hardware imperfections and limitations that might reduce their gain.

Bibliography

[1] J. Winters, "On the capacity of radio communication systems with diversity in a rayleigh fading environment," *IEEE JSAC*, vol. 5, pp. 871–878, June 1987.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *ATT Bell Technical Memorandum*, 1995.

[3] G.J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas," *Bell Labs Tech. J.*, 1996.

[4] S.M. Alamouti, "A simple transmit diversity scheme for wireless communications," *IEEE JSAC*, vol. 16, pp. 1451–1458, October 1998.

[5] B. Hassibi and B.M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. on Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[6] P.W. Wolniansky, G.J Foschini, G.D. Golden, and R.A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE 98*, September 1998.

[7] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. on Comm.*, vol. 49, no. 12, pp. 2198–2206, December 2001.

[8] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[9] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[10] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for vector multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory, (ISIT)*, June 2001.

[11] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Trans. on Inf. Theory.*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[12] M.H.M. Costa, "Writing on dirty paper," *IEEE Trans. on Inf.Theory*, vol. 29, no. 12, pp. 439–441, May 1983.

[13] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Trans. on Inf. Theory*, vol. IT-24, pp. 374–377, May 1978.

[14] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. of the IEEE International Conference on Communications (ICC), New York, NY*, April 2002.

[15] P. Viswanath and N.C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. on Inf. Theory*, vol. 49, pp. 1912–1921, August 2003.

[16] C. Peel, B. Hochwald, and L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication," in *Proc. of the 41st Allerton Conference on Communication, Control, and Computing*, October 2003.

[17] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," in *Proc. 5th International ITG Conference on Source and Channel Coding (SCC), Erlangen, Germany*, January 2004, pp. 403–408.

[18] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics Letters*, vol. 7, no. 5/6, pp. 138–139, March 1971.

[19] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans on Comn.*, vol. 20, no. 4, pp. 774–780, August 1972.

[20] G. Cinis and J. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, November 2000, vol. 2, pp. 1627–1637.

[21] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," in *4. ITG Conference on Source and Channel Coding, Berlin*, January 2002, pp. 139–147.

[22] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. 5th International ITG Conference on Source and Channel Coding (ITG SCC'04)*, January 2004, pp. 387–394.

[23] T. Yoo and A. Goldsmith, "On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming," *IEEE JSAC*, vol. 24, no. 3, pp. 528–541, March 2006.

[24] N. Jindal, S. Jafar, S. Vishwanath, and A.J. Goldsmith, "Sum power iterative water-filling for multi-antenna gaussian broadcast channels," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, November 2002, pp. 3–6.

[25] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in Multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, February 2004.

[26] V. Stankovic, M. Haardt, and M. Fuchs, "Combination of block diagonalization and THP transmit filtering for downlink beamforming in multi-user MIMO systems," in *Proc. European Conference on Wireless Technology (ECWT 2004), Amsterdam, The Netherlands*, October 2004.

[27] V. Stankovic and M. Haardt, "Successive optimization Tomlinson-Harashima precoding (SO THP) for multi-user MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP), Philadelphia, PA, USA*, March 2005.

[28] V. Stankovic and M. Haardt, "Multi-user MIMO downlink precoding for users with multiple antennas," in *Proc. of the 12-th Meeting of the Wireless World Research Forum (WWRF), Toronto, ON, Canada*, November 2004.

[29] V. Stankovic and M. Haardt, "Novel linear and non-linear multi-user MIMO downlink precoding with improved diversity and capacity," in *Proc. of the 16-th Meeting of the Wireless World Research Forum (WWRF), Shanghai, China*, April 2006.

[30] V. Stankovic and M. Haardt, "Improved diversity on the uplink of multi-user MIMO systems," in *Proc. European Conference on Wireless Technology (ECWT 2005), Paris, France*, October 2005.

[31] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *submitted to IEEE Trans. on Wireless Com.*, 2006.

[32] V. Stankovic and M. Haardt, "Multi-user MIMO downlink beamforming over correlated MIMO channels," in *Proc. International ITG/IEEE Workshop on Smart Antennas, Duisburg, Germany*, April 2005.

[33] R. Nabar, *Performance analysis and transmit optimization for general MIMO channels*, Ph.D. thesis, Stanford University, 2003.

[34] D.S. Baum et al., "D5.4 Final Report on Link Level and System Level Channel Models," Tech. Rep. IST-2003-507581 WINNER, IST WINNER, September 2005.

[35] M. T. Ivrlac and J. A. Nossek, "Correlated fading in MIMO-systems - Blessing or Curse?," in *Proc. of the 39st Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA*, October 2001.

[36] B. Suard, G. Xu, H. Liu, and T. Kailath, "Uplink channel capacity of Space-Division-Multiple-Access schemes," *IEEE Trans. on Inf. Theory*, vol. 44, pp. 1468–1476, July 1998.

[37] S. Vishwanath, N. Jindal, and A.J. Goldsmith, "Duality, achievable rates, and sum-rate capacity of MIMO broadcast channels," *IEEE Trans. on Inf. Theory*, vol. 49, pp. 2658–2668, October 2003.

[38] D. Tse and S. Hanly, "Multiaccess fading channels Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. on Inf. Theory*, vol. 44, pp. 2796–2815, November 1998.

[39] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, and A. Goldsmith, "Capacity bounds for gaussian vector broadcast channels," in *Proc. DIMACS workshop on signal processing for wirelesss transmission*, Piscataway, NJ, October 2002.

[40] A. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE JSAC*, vol. 21, pp. 684–702, June 2003.

[41] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. of the IEEE international conference on communications ICC*, New York, NY, April 2002.

[42] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for gaussian vector multiple access channels," *IEEE Trans. on Inf. Theory*, vol. 50, pp. 145152, January 2004.

[43] R.W. Heath Jr. and A.J. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 10, pp. 2429–2441, October 2002.

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

[45] L.M.A. Jalloul, K. Rohani, K. Kuchi, and J. Chen, "Performance analysis of CDMA transmit diversity methods," in *Proc. IEEE VTC*, 1999.

[46] O. Tirkkonen, A. Boariu, and A. Hottinen, "Minimal non-orthogonality rate 1 space-time block code for 3+ Tx," in *Proc. of the International Symposioum on Spread Spectrum Techniques and Applications*, September 2000.

[47] S. Haykin, *Adaptive filter theory*, Prentice hall, 2002.

[48] N. Jindal and A. Goldsmith, "Dirty paper coding vs. TDMA for MIMO broadcast channels," in *Proc. of the IEEE international conference on communications ICC*, June 2004, p. 682686.

[49] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. on Inf. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[50] K. Wesolowski et al., "D2.6 Assessment of Multiple Access Technologies," Tech. Rep. IST-2003-507581 WINNER, IST WINNER, October 2004.

[51] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener filter for the downlink of TDD DS-CDMA systems," in *Proc. IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'02)*, September 2002, vol. 1, pp. 9–13.

[52] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. on Vehicular Tech.*, vol. 53, no. 4, pp. 18–28, January 2004.

[53] Q. Spencer and M. Haardt, "Capacity and downlink transmission algorithms for a multi-user MIMO channel," in *Proc. 36th Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, IEEE Computer Society Press*, November 2002.

[54] L. U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 20–24, January 2004.

[55] P. Tejera, W. Utschick, G. Bauch, and J. Nossek, "A novel decomposition technique for multi user MIMO," in *Proc. 6th International ITG Conference on Source and Channel Coding (ITG SCC'05)*, April 2005.

[56] M. Joham et al., "Linear transmit processing in MIMO communications systems," *IEEE Trans. on Sig. Proc.*, vol. 53, no. 8, pp. 2700–2712, August 2005.

[57] R.D. Wesel and J.M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Trans. on Inf. Theory*, vol. 44, no. 2, pp. 824–831, March 1998.

[58] J. Jiang, R.M. Buehrer, and W.H. Tranter, "Spatial T-H precoding for packet data systems with scheduling," in *Proc. Vehicular Technology Conference, VTC-Fall*, October 2003.

[59] A. Mezghani, M. Joham, R. Hunger, and W. Utschick, "Transceiver Design for Multi-User MIMO Systems," in *Proc. 7th International ITG Conference on Source and Channel Coding (ITG SCC'06)*, January 2006.

[60] "IEEE P802.11 Wireless LANs, TGn Channel Models," Tech. Rep. IEEE 802.11-03/940r2, IEEE, January 2004.

[61] S. Zhou and G.B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. on Inf. Theory*, vol. 50, pp. 2599– 2613, October 2002.

[62] G. Del Galdo and M. Haardt, "Comparison of zero-forcing methods for downlink spatial multiplexing in realistic multi-user MIMO channels," in *Proc. IEEE Vehicular Technology Conference 2004-Spring, Milan, Italy*, May 2004.

[63] M. Fuchs, G. Del Galdo, and M. Haardt, "A novel tree-based scheduling algorithm for the downlink of multi-user MIMO systems with ZF beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP), Philadelphia, PA, USA*, March 2005.

[64] J. Ketchum et al., "System description and operating principles for high throughput enhancemenats to 802.11," Tech. Rep. IEEE 802.11-04/0870r1, IEEE, 2004.

[65] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation," in *Proc. RAWCON*, August 2003.

[66] G. Auer, "Modeling of ofdm channel estimation errors," in *Proc. 10th International OFDM Workshop, Hamburg, Germany*, August 2005.

[67] J. Axnas et al., "D2.10 Final report on identified RI key technologies, system concept, and their assessment," Tech. Rep. IST-2003-507581 WINNER, IST WINNER, December 2005.

[68] M. Dottling et al., "D2.7 Assessment of Advanced Beamforming and MIMO Technologies," Tech. Rep. IST-2003-507581 WINNER, IST WINNER, February 2005.

[69] T.F. Chan, "An improved algorithm for computing the singular value decomposition," *ACM Trans. on Math. Software*, vol. 8, no. 1, pp. 72–83, March 1982.

[70] G. Vivier et al., "D6.13.1 WINNER 2 Test scenarios and calibration cases issue 1," Tech. Rep. IST-4-027756 WINNER 2, IST WINNER, June 2006.

[71] J. Salo et al., "MATLAB implementation of the WINNER Phase I Channel Model ver1.5," December 2005.