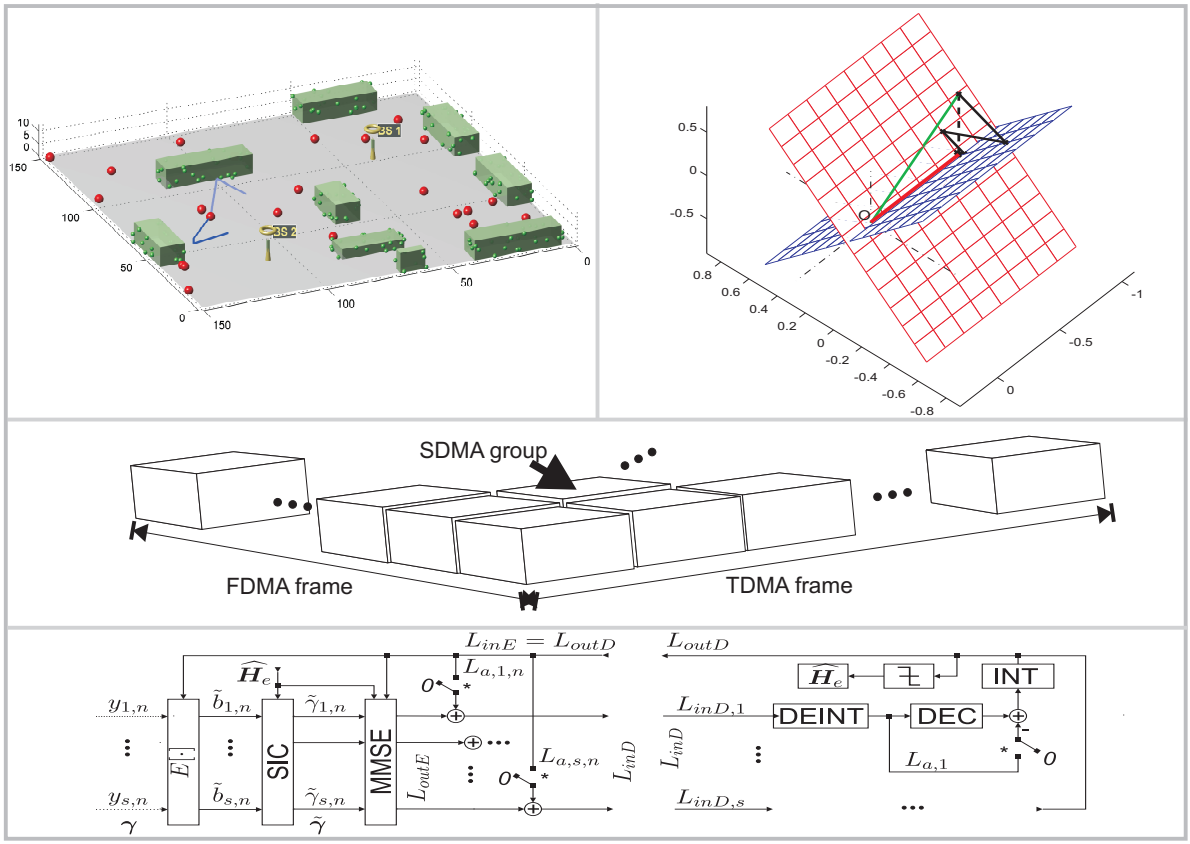
TECHNISCHE UNIVERSITÄT ILMENAU

COMMUNICATIONS RESEARCH LABORATORY

CRL



Advances in Multi-User Scheduling and Turbo Equalization for Wireless MIMO Systems

Martin Fuchs-Lautensack

Reports from the Communications Research Laboratory at Ilmenau University of Technology
Editor: Univ.-Prof. Dr.-Ing. Martin Haardt
Ilmenau, Germany, 2008

Research Reports from the Communications Research Laboratory
at Ilmenau University of Technology

Martin Fuchs-Lautensack

Advances in Multi-User Scheduling and Turbo Equalization for Wireless MIMO Systems

2009

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

ISBN 978-3-938843-43-7

Verlag ISLE, Betriebsstätte des ISLE e.V. ,
Werner-von-Siemens-Str. 16, 98693 Ilmenau

Printed in Germany

Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität Ilmenau

Advances in Multi-User Scheduling and Turbo Equalization for Wireless MIMO Systems

Martin Fuchs-Lautensack
geboren am 6.6.1977 in Ulm

Dissertation zur Erlangung des
akademischen Grades Doktor-Ingenieur (Dr.-Ing)

Anfertigung im:	Fachgebiet Nachrichtentechnik Institut für Informationstechnik Fakultät für Elektrotechnik und Informationstechnik
Gutachter:	Univ.-Prof. Dr.-Ing. Martin Haardt
	Univ.-Prof. Dr.-Ing. Albert Heuberger
	Univ.-Prof. Dr.-Ing. Eduard Jorswieck
Vorgelegt am:	21.10.2008
Verteidigt am:	19.06.2009
Verf.-Nr.:	EI 226

urn:nbn:de:gbv:ilm1-2009000114

Preface

Welcome to the second volume of the Research Reports from the Communications Research Laboratory at Ilmenau University of Technology. It is the doctoral dissertation of Martin Fuchs-Lautensack, entitled "Advances in Multi-User Scheduling and Turbo Equalization for Wireless MIMO Systems". This thesis is organized in three main parts. In Part I, Martin Fuchs-Lautensack deals with low computational complexity approaches to the problem of scheduling users to resources of an SDMA system with channel adaptive precoding. Part II extends this work to systems with multiple transmitting stations, showing various ways to implement interference avoidance scheduling with low complexity. The concept of Turbo Equalization is the main topic of Part III where optimizations to traditional approaches are developed.

Multi-user MIMO systems are a key component of wireless communication systems beyond the third generation, e.g., the fourth generation, to increase their spectral efficiency and to enable higher data rates with a reduced cost per bit as compared to third generation mobile communication systems. Specifically, linear (channel adaptive) multi-user precoding has been identified as a candidate transmit signal processing approach. Low complexity scheduling is key to fully exploiting the benefits offered by such approaches.

In this volume, Martin Fuchs-Lautensack solves a variety of challenging tasks that are highly relevant to the design of future wireless communication systems in an elegant way. These tasks are at the intersection of several scientific disciplines including transmit and receive signal processing, communications, linear algebra, and system modeling. The results of this thesis represent a significant contribution to the field of multi-user MIMO signal processing and, therefore, have an important impact on the design and realistic evaluation of multi-user MIMO systems.

Ilmenau, October 2008 Martin Haardt

Acknowledgements

My thanks go first of all to my supervisor Prof. Dr.-Ing. Martin Haardt for his guidance. With his projects he managed to open a wide and diverse field of interesting international research to his assistants.

Thank you to Prof. Dr.-Ing. Albert Heuberger and Prof. Dr.-Ing. Eduard Jorswieck for investing so much time to review this work albeit their engagement in starting up their newly created research groups. The opportunities to discuss with them helped to greatly improve the quality of the final manuscript. Dr. Karl Schran from the Digital Signal Processing group deserves my gratitude for accepting in such an unbureaucratic manner to be an examiner in the rigorosum.

A majority of the results in this thesis would not have been possible without the cooperation with Dr.-Ing. Giovanni Del Galdo. This is not only due to the fact that he developed the *IlmProp*, an astonishingly versatile channel modeling tool which is an invaluable contribution to the entire lab's work. It is mostly due to his selfless sharing of his knowledge and experience which made the cooperation feel like having a second advisor at hand.

I am also especially grateful to Dr. Veljko Stankovic. In my opinion, with his devotion to his work way beyond anything usual he was the driving force in the lab's work during the IST-WINNER projects and beyond. Ever so often we pulled each other through muddy waters by exchanging what each of us could do best and sharing our friendship. Thank you for providing the first set of throughput curves with WINNER system parameters, which resulted from a joint simulation campaign.

Thank you also to Florian Römer for the second set of throughput curves with *ProSched* scheduling and updated WINNER parameters. After Dr. Stankovic had left he picked up the speed in the WINNER work astonishingly naturally. His work and both his professional and personal advice are an immense contribution to the lab's success.

Special thanks go to Ulrike Korger and Christian Käs for joining in the effort of implementing an instance of the WINNER metropolitain area channel model while being busy with their work for projects funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

All other fellow researchers in the WINNER projects will be remembered for good discussions and for sharing their expertise.

Thank you to the colleagues Marko Hennhöfer, Marko Milojevic, Anja Grosch, Nuan Song, Bin Song, Henry Henry, Joao Paulo Lustosa Da Costa and Martin Weis as well as to all the

researchers in the Electronic Measurement Laboratory EMT, especially to Christian Schneider, for good discussions, good times and being more than just colleagues.

Dr. Mike Wolf was the driving strength behind many student tutorials tought, with his patience in explaining and his supplementary materials which helped a great deal to bring dusty topics to life. Thank you for being one of the solid rocks in difficult times, which also applies in many ways to the lab's secretary Brigitte Epler, technical assistant Wolfgang Erdtmann and former members of the lab Dr. Ralf Irmer and Dr. Wolfgang Röhr. It was a special pleasure to work together with them in both professional and personal aspects.

Thank you to all student researchers for helping in the various projects. Many of them kept on producing really good results even when time to supervise was scarce. I would like to especially mention Bernd Bandemer, whose restless implementation efforts and deep understanding of the matter were key to the Turbo Equalization results in this thesis, as well as Laurits Hamm and would like to thank both for becoming friends.

My family members made this work possible in the first place by helping me move to Ilmenau and by continuously supporting me in all possible ways. I would never have thought that a similar kind of strong company can exist again, but was disabused when the family of my wife Anke came into my life. The sheer fact of her being there for me stopped a long fall.

Ilmenau, October 2008 Martin Fuchs-Lautensack

Abstract

Multiple antennas at the transmitter and/or receiver (Multiple Input Multiple Output (MIMO) systems) have been widely recognized as a key enabler for the data rates targeted in fourth generation (4G) wireless communications systems. As discussed in the introduction, recent research projects have identified linear channel adaptive transmit processing and iterative non-linear receive processing as promising candidate signal processing techniques.

Part II of this thesis deals with multi-user scheduling for the downlink of such a system with one transmitting station: Multiple users can be served simultaneously in each resource element (time and/or frequency) of the underlying system by separating them in space with different antenna weight vectors (Space-Division Multiple Access (SDMA)). Users with spatially correlated MIMO channel matrices should not be served jointly since correlation impairs the spatial separability and leads to inefficient antenna weights. Finding an optimized user subset is especially costly since the resulting sum rate for each user subset depends on the precoding weights, which in turn depend on the user subset. One key contribution of this thesis is to decouple this problem. The rate with Zero Forcing (ZF) precoding is proposed as a scheduling metric and it is shown how ZF can be written with the help of orthogonal projection matrices. The latter enable us to estimate rates without having to compute any antenna weights by using repeated projections into single user channel subspaces. When compared to other, non-rate based low complexity schedulers, the presented rate estimates still allow to consider user rate requirements and fairness criteria. Other advantageous properties are that the receive processing does not have to be taken into account and that the method can work with both instantaneous or long term averaged channel knowledge. Several search algorithms are presented which can be used together with the proposed metric to efficiently solve user grouping or selection problems jointly for the entire system bandwidth while being able track the solution in time and frequency for further complexity reduction. As a side result, an orthogonal projection based formulation of the classical optimal ZF precoding algorithm Block Diagonalization (BD) is obtained.

Part III proposes ways how multiple transmitting stations can benefit from cooperative scheduling or joint signal processing over all involved antennas. With the help of a similar orthogonal projection based technique as above, an estimate of the interference power can be provided which results from one station serving an assigned user to other users with other stations - again without having to compute any antenna weights. In combination with the scheduling metric from Part II and a so-called virtual user concept, the cooperative scheduling problem can be solved with the help of similar search algorithms as above. It is also presented

how the approach can be extended to include any number of SDMA capable half-duplex relays in the scheduling process, provided that a central intelligence has some knowledge about all channel links. One basic approach to reduce the signalling overhead involved is discussed. Another supplementary result involves a method to estimate the sum rate of a cell enhanced with SDMA relays when no interference coordination is performed.

Part IV presents optimizations for iterative receivers or Turbo Equalizers which are based on the concept of iterating soft information between the building blocks to reach convergence. There, correlation between signals is a feature which can be exploited as a source of redundancy. Nevertheless it is shown that a combination with transmit precoding which aims at reducing correlation can be beneficial when the channel knowledge at the transmitter contains a realistic error, leading to increased correlation. Furthermore it is discussed that re-using so-called a-priori information which is conventionally discarded can greatly improve the convergence properties. A novel method for adaptive re-use of this information is developed by tracking the convergence online with the help of Extrinsic Information Transfer (EXIT) chart techniques. Finally, a method is proposed to model semi-blind channel estimation updates in an EXIT chart which is meaningful in Orthogonal Frequency Division Multiplexing (OFDM) systems.

Computer simulations are presented in all three parts to illustrate the gains of the proposed methods, e.g., based on channel models with realistic correlation properties such as the *IlmProp*, or within the context of a current 4G system proposal from the IST-WINNER projects.

Kurzfassung

Mehrantennensysteme (MIMO Systeme) werden weithin als Schlüsseltechnologie für zukünftige Mobilfunksysteme der 4. Generation gesehen, da sie die angestrebten Datenraten erst ermöglichen. Wie der Einleitung zu entnehmen ist, werden in der aktuellen Literatur sendeseitige lineare kanaladaptive Signalverarbeitungsverfahren (precoding) sowie empfangsseitige nicht-lineare Verfahren als vielversprechend erachtet.

Teil II dieser Arbeit behandelt das Thema Mehrbenutzer-Scheduling für die Abwärtsstrecke eines derartigen Systems mit einer Sendestation: In jeder (Zeit- und/oder Frequenz-) Ressource des zugrundeliegenden Übertragungssystems können gleichzeitig mehrere Teilnehmer bedient werden, indem man sie räumlich trennt mit Hilfe unterschiedlicher Antennengewichts-Vektoren (SDMA). Dabei sollten Teilnehmer mit räumlich korrelierten MIMO Kanalmatrizen nicht gleichzeitig bedient werden, da Korrelation die räumliche Trennung der Nutzer erschwert und zu ineffizienten Antennengewichten führt. Geeignete Nutzergruppen zu finden ist besonders rechenaufwendig, da die zu erwartenden Datenraten in jeder Untergruppe von den Antennengewichten abhängen, welche wiederum für jede Untergruppe verschieden sind. Ein Kernbeitrag dieser Arbeit ist es, dieses Problem zu entkoppeln. Es wird eine Scheduling-Metrik vorgeschlagen, die eine Schätzung der Rate bei Verwendung von interferenzfreiem precoding (ZF) darstellt. Es wird gezeigt, dass sich diese Ratenschätzung mit Hilfe von orthogonalen Projektionsmatrizen schreiben lässt. Letztere ermöglichen es, die erzielbaren Raten der Nutzer zu schätzen, ohne die Antennengewichte berechnen zu müssen, und zwar unter Verwendung von wiederholten Projektionen in unabhängige Unterräume der einzelnen Nutzer. Im Gegensatz zu anderen Metriken aus der Literatur, die nicht auf Raten basieren, lassen sich die vorgeschlagenen Schätzungen mit Datenratenanforderungen der Nutzer und mit Fairness Kriterien verknüpfen. Weitere Vorteile sind, dass das räumliche Empfangsverfahren nicht spezifiziert werden muss und dass die Methoden sowohl basierend auf momentanen Kanalmessungen als auch auf Kanalstatistik arbeiten können. Mehrere Suchalgorithmen werden vorgeschlagen, die zusammen mit den vorgeschlagenen Metriken eingesetzt werden können, um das Scheduling-Problem effizienter zu lösen. Die Verfahren können bei Bedarf das ganze System auf einmal bearbeiten und dabei die Lösung zeitlich und in Frequenzrichtung nachführen, um die Komplexität zusätzlich zu verringern. Als Nebenergebnis wird mit Hilfe von orthogonalen Projektionsmatrizen eine neuartige Formulierung des klassischen optimalen ZF precoding Verfahrens BD besprochen.

In Teil III werden Möglichkeiten vorgeschlagen, wie mehrere Sendestationen von kooperativem Scheduling oder kooperativer Signalverarbeitung profitieren können. Wiederum mit Hilfe

von orthogonalen Projektionen wird eine Schätzung der Interferenzleistung hergeleitet, die eine Station beim Bedienen einer ihrer Teilnehmer für die Teilnehmer einer anderen Station erzeugt. Dabei wird erneut die Berechnung von Antennengewichten umgangen. Kombiniert mit der Scheduling-Metrik aus Teil II und mit einem sogenannten Konzept virtueller Teilnehmer kann das kooperative Scheduling durch einen ähnlichen Suchalgorithmus wie oben behandelt werden. Weiterhin wird besprochen, wie halb-duplex Relays mit SDMA in den Suchalgorithmus mit eingebunden werden können, vorausgesetzt dass eine zentrale Intelligenz gemeinsame Kanalkenntnis hat. Ein einfacher Ansatz zur Reduktion der notwendigen Signalisierung wird kurz angerissen. Ein weiteres Nebenergebnis dieses Teils ist eine Simulationsmethode, mit deren Hilfe die Summenrate einer Mobilfunkzelle mit SDMA Relays ohne Interferenzkoordination geschätzt werden kann.

In Teil IV werden Optimierungen für iterative Empfänger bzw. Turbo Entzerrer vorgeschlagen. Diese Empfänger basieren auf dem Iterieren von sogenannter soft information zwischen den Teilen des Empfängers. Korrelation zwischen Signalen ist hierbei eine erwünschte Eigenschaft, die als Quelle für Redundanz ausgenutzt werden soll. Nichtsdestotrotz wird gezeigt, dass eine Kombination mit sendeseitigem Precoding sinnvoll sein kann. Letzteres zielt zwar auf die Verringerung von Interferenz und Korrelation ab, hängt aber in der Praxis von fehlerbehafteter Kanalkenntnis ab, die zu suboptimalen Antennengewichten und damit zu erhöhter Restinterferenz führt. Des Weiteren wird basierend auf EXIT chart Techniken eine Methode beschrieben, mit der a-priori Information in den Iterationen adaptiv weiterverwertet werden kann, wenn sie vorteilhaft ist. Dies kann die Konvergenzeigenschaften verbessern im Vergleich zur konventionellen Vorgehensweise, bei der a-priori Information zur Vermeidung von Fehlerfortpflanzung nicht verwendet wird. Schlussendlich wird für OFDM Systeme vorgeschlagen, wie der Einfluss von semi-blinder Kanalschätzung am Empfänger in EXIT charts einbezogen werden könnte.

Simulationsexperimente dienen in allen Teilen der Arbeit dazu, die Vorteile der vorgeschlagenen Methoden zu verifizieren. Dazu kommen Kanalmodelle mit realistischen Korrelationseigenschaften zum Einsatz, wie zum Beispiel *IlmProp* oder ein in den IST-WINNER Projekten entwickeltes messdatenbasiertes Modell sowie entsprechende Systemparameter.

Contents

List of Figures

List of Tables

Part I.

Outline and Introduction

1. Outline of this thesis

This work deals with three distinct topics in spatial processing for wireless communications systems equipped with multiple antennas at the receiver and/or transmitter, so called Multiple Input Multiple Output (MIMO) systems. MIMO techniques are widely considered one of the key enablers to higher data rates in wireless coverage techniques such as third generation (3G) Long Term Evolution (LTE) and fourth generation (4G) mobile communications systems. Good examples are the European Union 6th Framework Programme projects IST-Wireless World Initiative New Radio (WINNER) I and II [15], which are often cited in this work as one of the biggest 4G research projects.

Simply speaking, the use of multiple antennas allows a receiver to distinguish signal components from different propagation paths by their delayed arrival at the antenna elements. Such multi-path propagation is problematic in single antenna systems, requiring additional computational effort to equalize interference between delayed data symbols or to extract information contained in them. Now, with multiple antennas at hand, these components can be detected separately and used as a source of diversity. When the transmitter also has an array of antennas, multiple signal components may be multiplexed in space either blindly or by steering them into the preferred propagation directions of the channel (assuming that channel measurements are possible, also referred to as Channel State Information (CSI) adaptive precoding). Desired signals may be amplified and unwanted components suppressed. A more technical introduction to MIMO systems and precoding techniques is given in Chapter 2.

1.1. Overview of the parts of this work

The main contributions of this work are in the field of downlink user scheduling for a single transmitting station (Part II) and for multiple stations with coordination (Part III). In addition, Part IV presents various optimizations for Turbo receivers. Due to this split, the reader interested in more in-depth discussions of the three research areas including state of the art reviews is kindly referred to the respective parts, whereas a high level introduction to the main contents is given next. A list of the contributions of this work can be found in Chapter 3 at the end of this part.

Overview of Part II

Wireless communications systems with multiple antennas at the transmitter can serve multiple terminals at the same time by separating them spatially, e.g, via different transmit antenna

weight vectors (precoding). In a transmission scheme which also uses other orthogonal resources such as time and frequency, a different subset of terminals can be assigned to each resource element. If joint spatial processing is performed at the base station, the subset selection must avoid to group users with spatially correlated channels. Otherwise, a severe throughput degradation may occur due to the limited spatial separability of the users' signals, leading to inefficient precoding weights. The selection process is computationally very complex because the performance of the spatial processing depends on the user combination to be served.

Part II describes a low-complexity scheduling algorithm for wireless multi-user MIMO communication systems in which users are multiplexed via time-, frequency-, and space-division multiple access schemes, TDMA, FDMA, and SDMA, respectively. The approach presented here works with both zero-forcing and non zero-forcing SDMA precoding schemes by deciding, for each time and frequency slot, which users are to be served in order to maximize the precoding performance. Our algorithm *ProSched* is based on a novel interpretation of the precoding process using orthogonal projections which permit us to estimate the precoding results of all user combinations of interest with significantly reduced complexity. Additionally, the possible user combinations are efficiently treated with the help of a tree-based sorting algorithm (or dendrogram).

A reduced complexity scheduling approach is key in retaining the complexity advantage that so-called linear precoding techniques have over theoretically optimal but more complex non-linear ones (see Chapter 2 for more details). Furthermore, only with proper spatial scheduling, such linear techniques can approach their maximum performance.

The number of users to be served simultaneously is not a fixed parameter of our algorithm as often assumed for other schedulers present in the literature, but is also adjusted according to the channel conditions. While smaller SDMA groups allow us to transmit with higher average power per user, larger groups lead to higher multiplexing gains.

ProSched takes advantage of instantaneous channel measurements (CSI) when available, or alternatively of second order channel statistics. The individual user Quality of Service (QoS) requirements can be considered in the decision making process. The effectiveness of the algorithm is illustrated with simulations based on the *IlmProp* channel model, which features realistic correlation in space, time, and frequency, as well as within the context of the system concept developed in the European 6th Framework Programme project IST-WINNER.

Overview of Part III

In Part III it is shown how the *ProSched* approach can be extended to schedule terminals in systems with multiple interfering transmitters, be it Base Stations (BSs) or later relays. Using computer simulations it is illustrated that inter-site interference is a limiting factor and that coordination between the stations can enhance the system performance.

The *ProSched* approach allows all stations to fully exploit the benefits of channel-aware SDMA precoding without the need to pre-compute the precoding weights of all user and transmitter combinations in the system in order to predict the resulting interference. Especially in the context of Relay Enhanced Cells (RECs) this represents an advantage over existing scheduling and routing approaches which do not allow for channel-adaptive precoding at the relays. In addition, several side results are presented, as summarized in Chapter 3.

Overview of Part IV

Turbo Equalization can be, to some extend, thought of as a complementary approach to the one taken in the previous parts: Instead of avoiding signal correlation and interference with the help of transmit processing, it aims at exploiting interference as a source of redundancy at the receiver while commonly no additional computational complexity is invested at the transmitter.

In Part IV it is shown that a Turbo equalizer can even be beneficially combined with transmit precoding because when the latter is performed based on imperfect CSI it creates residual spatial interference that can be exploited in an iterative receiver. Another reason why it is attractive to combine Turbo receivers with precoding is the fact that they allow an enhanced channel estimation, which in turn is beneficial for the accuracy of the precoder. We propose a method to model such semi-blind channel estimation at the receiver in the most commonly used analysis tool for Turbo systems, the so-called Extrinsic Information Transfer (EXIT) charts. The model is meaningful for Orthogonal Frequency Division Multiplexing (OFDM) based systems and allows for a joint optimization of equalization and channel estimation.

The main contribution of this part is, however, an adaptive method to re-use a-priori information which is conventionally discarded between iterations in Turbo systems. The method is based on tracking the receiver convergence online with an EXIT chart. Without loss of generality, we illustrate the approach in the context of MIMO OFDM with Turbo equalization.

2. MIMO OFDM reviewed

The discussion in this work is centered around OFDM as a transmission system which shall be briefly reviewed for better understanding along with the basic aspects of spatial processing. For a comprehensive treatment of MIMO OFDM please refer to literature such as [16, 17]. However, many aspects of this thesis related to spatial processing apply also to other transmission systems than OFDM.

2.1. A motivation for the use of OFDM with spatial processing

Simply put, OFDM schemes are attractive because they allow us to treat spatial processing aspects independently and with reduced complexity since they take care of many severe channel conditions such as narrowband interference, frequency-selective fading due to multi-path propagation, or inter-symbol interference. To better understand why this is the case, OFDM is briefly reviewed in the following section. The straight forward incorporation of spatial processing is one of the reasons why OFDM was selected for implementation in two of the biggest research projects developing 4G systems in the European Union, the 6th Framework programme projects WINNER I and II [15]. (Note that interference components which are suppressed in OFDM may otherwise also be exploited by investing additional computational power, e.g., in an iterative Turbo receiver).

2.2. OFDM reviewed

In OFDM, N_c subcarriers are loaded with conventionally modulated complex data symbols d_n and transmitted during a symbol time T_s. The transmit signal may be written in the complex baseband (i.e., after shifting the transmit spectrum back to zero center frequency to omit the carrier) as a sum of complex exponentials

$$s(t) = \sum_{n=0}^{N_\mathrm{c}-1} d_n e^{j2\pi t \cdot n \cdot \frac{1}{T_\mathrm{s}}} . \tag{2.1}$$

The frequencies all being multiples of $\frac{1}{T_\mathrm{s}}$, orthogonality of the subcarriers is inherently achieved (assuming that $s(t)$ is instantly switched off after T_s corresponding to a rectangular window with

sinc shaped spectral amplitude density):

$$\frac{1}{T_\mathrm{s}}\int_0^{T_\mathrm{s}} \left(e^{j2\pi t n_1 \frac{1}{T_\mathrm{s}}}\right)^* e^{j2\pi t n_2 \frac{1}{T_\mathrm{s}}} dt = \frac{1}{T_\mathrm{s}}\int_0^{T_\mathrm{s}} e^{j2\pi t (n_2-n_1) \frac{1}{T_\mathrm{s}}} dt = \delta_{n_1 n_2}\,. \tag{2.2}$$

The orthogonality is expressed with the help of the Kronecker delta symbol $\delta_{ab} = \begin{cases} 1 & \text{if } b = a \\ 0 & \text{if } b \neq a \end{cases}$.

The convolution with the channel impulse response disperses $s(t)$ in time. To avoid interference with the subsequent transmit signal at the receiver, a guard time may be inserted. This guard time is commonly implemented as a cyclic extension or prefix that can be discarded at the receiver together with the overlapping components.

This cyclic extension implies periodicity in time inside the analysis windows and is the reason why, most commonly, equation (2.1) is implemented as an inverse Discrete Fourier Transform (DFT) of a line spectrum where each line is modulated with a different transmit symbol. When the cyclic prefix is chosen at least as long as the maximum channel delay, the channel acts on the data only in the form of independent complex scalars. This greatly facilitates the deployment of spatial processing schemes as discussed below.

2.3. Overview of MIMO OFDM and the basic system model

In a MIMO OFDM system, OFDM modulation or demodulation is employed at each antenna. On each subcarrier n, the channels between antenna pairs become scalars and can be written as a matrix $\boldsymbol{H}_n \in \mathbb{C}^{M_\mathrm{R}\times M_\mathrm{T}}$, where M_R and M_T represent the total number of receive and transmit antennas, respectively. By stacking multiple data symbols into a vector $\boldsymbol{d}_n \in \mathbb{C}^{M_\mathrm{T}\times 1}$, the system's input-output relationship may be modeled with the help of an additive noise vector $\boldsymbol{n}_n$ as

$$\boldsymbol{y}_n = \boldsymbol{H}_n \boldsymbol{d}_n + \boldsymbol{n}_n\,. \tag{2.3}$$

We assume the elements of $\boldsymbol{n}_n$ to be Zero Mean Circular Symmetric Complex Gaussian (ZMCSCG) with variance σ_n^2/N_c. The total receiver noise power per antenna in the entire band is thus σ_n^2.

The following transmission modes may be distinguished depending on the choice of symbols contained in $\boldsymbol{d_n}$:

- Spatial Multiplexing (SM): different data symbols belonging to the same stream
- SDMA: different users' data symbols (assuming that $\boldsymbol{y}_n$ contains all users' received data symbols and that $\boldsymbol{H}_n$ consists of the channels from the BS to all users stacked such that $\boldsymbol{H}_n = \begin{bmatrix} \boldsymbol{H}_{1,n}^\mathrm{T} & \cdots & \boldsymbol{H}_{N_\mathrm{u},n}^\mathrm{T} \end{bmatrix}^\mathrm{T}$)

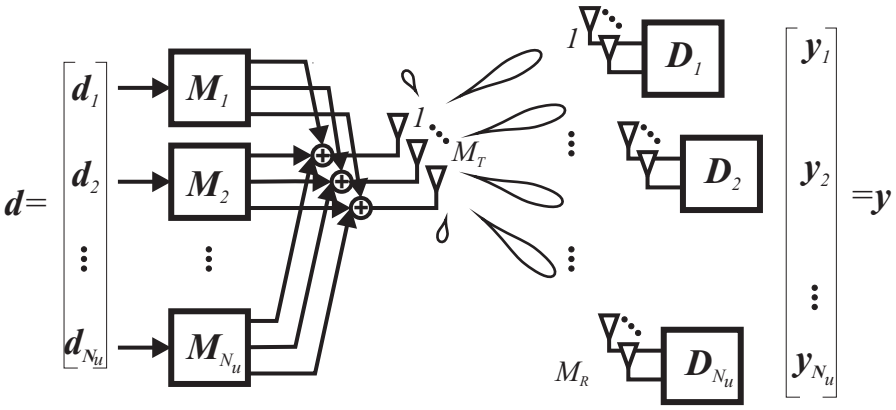


Figure 2.1.: An illustration of the multi-user MIMO downlink system model on one OFDM subcarrier.

- Diversity Mode (DM): copies of the same symbol. However, diversity gain can only be attained in combination with additional precoding or preprocessing as explained, e.g., in [16]. This is true because without precoding the resulting combined channel coefficients from all transmit antennas to each receive antenna simply become another complex random variable displaying the same variance as the channel had before.

2.4. Types of spatial processing

In the following, the subcarrier index n is often omitted for simplicity.

Most spatial processing techniques can be introduced by allowing for complex valued receive and transmit processing matrices $\boldsymbol{D}_n \in \mathbb{C}^{M_\mathrm{R} \times M_\mathrm{R}}$ and $\boldsymbol{M}_n \in \mathbb{C}^{M_\mathrm{T} \times (r_u \cdot N_\mathrm{u})}$ (also referred to as decoding and precoding matrices) such that

$$\boldsymbol{y}_n = \boldsymbol{D}_n \left(\boldsymbol{H}_n \boldsymbol{M}_n \boldsymbol{d}_n + \boldsymbol{n}_n \right) . \tag{2.4}$$

If the above is to represent the downlink with non cooperative receivers, then $\boldsymbol{D}_n \in \mathbb{C}^{M_\mathrm{R} \times M_\mathrm{R}}$ is block diagonal with the users' individually computed receive matrices on the main diagonal, i.e., it is structured as $\boldsymbol{D}_n = \begin{bmatrix} \boldsymbol{D}_{1,n} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{D}_{N_\mathrm{u},n} \end{bmatrix}$.

The decoding matrices serve to efficiently combine the antenna signals and must be tuned to estimates of the users' effective channels $\boldsymbol{HM}$ acting on the data symbols (see also Part IV). The number of spatial streams to each of the N_u users is r_u and $\boldsymbol{M}_n$ is thus $\in \mathbb{C}^{M_\mathrm{T} \times (r_u \cdot N_\mathrm{u})}$ with $\boldsymbol{d}_n \in \mathbb{C}^{(r_u \cdot N_\mathrm{u}) \times 1}$. An illustration of the downlink interpretation is given in Figure 2.1

The effect of the precoding matrix can be interpreted in different ways: In the case of so-called vector precoding, the product $\boldsymbol{Md}$ implies that a linear combination of the columns $\boldsymbol{m}_l$ of $\boldsymbol{M}$ is transmitted, namely $\boldsymbol{Md} = \boldsymbol{m}_1 d_1 + \ldots + \boldsymbol{m}_r d_r$ where $r = (r_u \cdot N_\mathrm{u})$ is the total number of multiplexed data symbols. To that extend, the columns of $\boldsymbol{M}$ can simply be understood as spanning vector subspaces. They can be taken from a (fixed or random) codebook or can be the result of various optimization criteria that may be based on measurements of the channel, so called CSI, and/or received noise and signal power (Channel Quality Indicators (CQIs)).

The precoding vectors can also be seen as beam weights or modes of the channel. In a Time Division Duplexing (TDD) system, such CSI may be conveniently obtained by transmitting pilot symbols on the uplink and assuming the downlink channel to be reciprocal, given proper calibration of the RF frontends [1]. In a Frequency Division Duplexing (FDD) system, knowledge of the downlink channel may require the transmission of pilots in the downlink to obtain an estimate at the user terminal which can then be fed back to the BS.

We may further distinguish between signal theoretic and information theoretic optimization criteria. Signal theoretic criteria are applied to the receive vector as a function of the transmit vector and include the well known Minimum Mean Squared Error (MMSE) criterion or the Zero Forcing (ZF) approach, forcing all interference between multiplexed symbols to zero. Information theoretic criteria are formulated in terms of Shannon rate or expected Bit Error Rate (BER) and include the exploitation of the three fundamental MIMO gains, array gain, spatial multiplexing gain, and diversity gain [18].

A different interpretation of $\boldsymbol{M}$ is that of matrix modulation: There, the row direction of the matrix represents time or frequency, and precoding means to distribute weighted copies of the symbols contained in $\boldsymbol{d}$ with the help of permutations such that they may be combined coherently at the receiver using $\boldsymbol{D}$ without taking the effect of the channel into account. These methods are also called Space Time or Space Frequency Codes [19, 20]. They can provide diversity but do not extract multiplexing gain as the vector precoders.

On a high level, the techniques mentioned so far may be classified as linear (or widely linear in case of Space Time codes). In addition, there exist non-linear techniques in which further operations are performed on the modulated data symbols. To accommodate non-linear techniques in the data model of equation (2.4), one can replace $\boldsymbol{d}_n$ by a function of $\boldsymbol{d}_n$ and allow for iterative receive processing.

Most of the non-linear techniques rely on Costa's principle of Writing on Dirty Paper [21] (or Dirty Paper Coding (DPC)) which states that the capacity of an interference channel can be fully exploited when the interference is known at the transmitter. With the help of CSI, the spatial interference that one data symbol is going to generate on the consecutive one is computed and pre-subtracted from the current symbol before transmission. This approach is equivalent

to the well known principle of decision feedback equalization at receivers. In [22, 23] it was derived that downlink DPC is optimal in terms of sum rate. Later it was shown that the DPC optimal sum rate region is indeed the capacity region of the MIMO multi-user downlink [24]. Well known practical DPC implementations are MMSE Tomlinson-Harashima precoding (THP) [25], Successive Minimum Mean Squared Error Tomlinson-Harashima precoding (SMMSE THP) [26], or Block Diagonalization Tomlinson-Harashima precoding (BD THP) [2] with different suboptimal approaches of computing the precoding weights. They are extensions of the THP algorithm originally developed for Single Input Single Output (SISO) inter-symbol interference channels [27]. More computationally intense algorithms for finding the optimum precoding weights in terms of rate and power constraints also exist such as discussed in [28, 29].

2.4.1. Motivation for the choice of spatial processing in this work

In accordance with the recommendations given in various publications of the WINNER project, in this work a system with linear channel adaptive precoding is considered in the parts dealing with scheduling and an iterative receiver is investigated in Part IV [1, 30, 31, 32, 33]. Channel adaptivity is considered required in this project to reach the targeted throughput values in scenarios where meaningful CSI can be measured in the form of channel coefficients or second order statistics thereof, i.e., for terminal velocities up to 50 km/h. For higher velocities, vector precoding approaches with fixed Grid of Beams (GoB) are foreseen. Downlink precoding, in general, is favored in WINNER to move as much computational complexity away from the terminals to the BS, to simplify the receiver and because SDMA and spatial multiplexing are noted as required key features to reach the targeted data rates.

Non-linear precoders have been excluded because of the following drawbacks [26]:

- In various publications they have been observed to be much more sensitive to channel estimation errors than linear precoders, caused by the fact that they make additional use of the possibly erroneous CSI in the non-linear DPC interference pre-substraction steps (as summarized, e.g., in [26]).
- They require to signal various information from the transmitter to the receiver such as the decoding order and even the precoding matrix because it is not possible to measure the equivalent channel $\boldsymbol{HM}$ with Pilot Assisted Channel Estimation (PACE) due to the non-linear preprocessing.
- The interference pre-substraction step requires an additional modulo operation at the transmitter to limit the transmit power. This raises the need for a complex sphere decoder or an approximate closest-point solution, adding to the complexity of non-linear schemes especially when the number of users is large [34].

- Less complex linear precoders such as Successive MMSE (SMMSE) [35] and Regularized Block Diagonalization (RBD) [36] have been developed and shown empirically to reach almost the same performance as the theoretically optimal DPC based precoding when the number of users is large and when appropriate scheduling is performed, see for example [37, 38, 39].

Coming back to the topics covered in this thesis, it should be noted that scheduling is a required function with all types of precoders. It is most challenging in channel adaptive precoding because the precoding matrices are different for each subset of users, as discussed in the next chapter. A scheduling approach with low complexity is of key interest when linear precoding is used to preserve the advantage linear precoders have in terms of complexity over non-linear ones and to drive their performance close to the optimal case. Scheduling for GoB is not investigated here because its fixed precoding vectors which are known at the receiver allow for less complex, straight forward scheduler implementations. There, the scheduler can directly work with the received signal to noise ratios which the users signal back to the BS.

Nevertheless we are interested in non-linear processing at the receiver (as considered in WINNER) because it offers, amongst other features, the possibility to dramatically enhance the quality of channel estimation in critical scenarios, which in turn is beneficial for the transmit precoder.

3. List of contributions

The main contributions in Parts II and III are in the field of multi-user scheduling for the downlink of wireless MIMO systems with adaptive precoding and SDMA.

1. In Part II the low-complexity approach *ProSched* is developed to schedule users to space, time, and frequency resources of a system with a single BS performing adaptive precoding. Unlike other solutions it allows the consideration of spatial correlation effects between user channels, correlation in time and frequency, optimal SDMA group size and QoS aspects without spoiling the complexity advantage linear precoders have over non-linear ones. *ProSched* has been patented in Germany and has an international patent pending [3]. The approach consists of two parts:

   a) A scheduling metric is derived which reflects the performance of one user's effective channel after MIMO transmit precoding in the presence of a set of other users that are to be served simultaneously via SDMA. The metric is an estimate of the Shannon rate with ZF precoding which can be considered an upper bound for other linear precoders. It has the following advantageous properties:

      - The ZF capacity of one user can be written with the help of an orthogonal projection into the intersection of the null spaces of all other users' channels in the same SDMA group. The projection would normally have to be recomputed for every user in every possible SDMA combination. Instead, *ProSched* approximates the intersection by a product of projection matrices into the null spaces of the single users. These matrices remain constant throughout the scheduling run, which dramatically reduces complexity. This means that the actual precoding matrices do not have to be computed while testing combinations.
      - Capacity as a metric reflects both the impairment of spatially correlated user channels on the efficiency of the precoding weights as well as the effect of the average power assigned to a user, which in turn is related to the SDMA group size. It is shown that it can be calculated based on channel matrix knowledge as well as on second order statistics channel knowledge.
      - The capacity based metric can be combined with proportional fairness or with methods taking into account QoS, e.g., in the form of user rate requirements.

   b) A best candidate search algorithm based on a search tree (or dendrogram) is developed to reduce the number of combinations to be tested. It delivers beneficial user terminal

combinations for all possible group sizes and allows in a second step a decision on the best group size based on the scheduling metric. Joint scheduling of all subcarriers is possible as well as tracking of the solution in time to exploit the gradual channel evolution for further complexity reduction.

2. As a side result, a novel formulation of the classical optimal ZF precoding algorithm Block Diagonalization (BD) is derived with the help of orthogonal projections and its equivalence to the original algorithm shown.

3. In Part III, *ProSched* is extended to coordinated scheduling for various types of MIMO systems with multiple transmitting stations performing adaptive precoding and SDMA. The extensions can be outlined as follows:

   - A low complexity estimate of the inter-site interference received by a user in such systems is derived and included into the spatial scheduling metric of Part II. The estimate uses the already available orthogonal projection matrices into the null spaces of the single users and, thus, does not represent a significant increase in computational complexity while retaining the same versatility as with the single BS metric. No further matrix decompositions are needed and, again, no precoding matrices are computed.
   - It is shown that, with the help of a so-called virtual user concept, the tree-based search algorithm can be employed to treat different types of systems including:
     - Distributed MIMO systems, also called virtual MIMO systems where all antennas of all BSs are treated jointly.
     - Multiple BSs performing interference coordination scheduling with the help of a central intelligence, with optional soft handover of users between BSs.
     - REC with interference coordination scheduling where each Relay Node (RN) can perform adaptive precoding independently.

   Several side results are also obtained:

4. A method to schedule multiple neighboring subcarriers jointly to exploit correlation for complexity reduction is shown, based on an already available covariance matrix averaging method.

5. A simulation method is developed based on the DPC bound to obtain a reference estimate of the sum rate of a multi-user REC with half duplex SDMA relays when no interference coordination is performed. There are currently no theoretical results available for such a system.

6. An approach to reduce the signaling overhead needed for coordinated scheduling is discussed based on a scaled rank one approximation.

Part IV deals with several topics in Turbo Equalization which can be, to some extend, thought of as a complementary approach to the one taken in the previous chapters: Instead of avoiding signal correlation and interference with the help of transmit processing, it aims at exploiting interference as a source of redundancy at the receiver while commonly no additional computational complexity is invested at the transmitter. The following contributions are described in this chapter:

7. An adaptive method to re-use a-priori information which is conventionally discarded between iterations in Turbo systems is developed in the context of MIMO OFDM without loss of generality. The method is based on tracking the convergence online with an EXIT chart. Simulation examples show that the iteration gain and the resulting final BER can be greatly improved compared to the cases when a-priori information is re-used non-adaptively or not at all.

8. It is shown that a Turbo equalizer can be beneficial even when combined with transmit precoding. This approach is conventionally not taken because after precoding the spatial interference is suppressed to a minimum, leaving no source of redundancy for the Turbo equalizer to work with. However, in realistic systems, CSI used for precoding is error prone or imperfect, leading to suboptimal precoding weights and residual spatial interference after precoding that may be sufficiently pronounced to justify an iterative receiver.

9. A new method is proposed to model semi-blind channel estimation through EXIT charts that is applicable to OFDM.

Part II.

The low-complexity space-time-frequency multi-user scheduling approach ProSched

4. Introduction to the basic scheduling problem

The problem of scheduling in wireless communications systems is as old as early satellite links. Already in [40] it was observed that the overall system performance may be increased by scheduling demanding services to resources with good channel quality.

A good description of the basic classes of scheduling solutions for the Single Input Single Output (SISO) case can be found in [41]. Traditionally, scheduling criteria are based on the maximization of a sum performance metric as in [42] (so-called greedy approaches) that may or may not be channel-aware, or on the Quality of Service (QoS) requirements of the services such as transmission delay [43]. Hybrid approaches between the two exist as well, leading to fairness based methods trying to equalize resource access between users.

In a Multiple Input Multiple Output (MIMO) system, the use of multiple antennas allows a Base Station (BS) to serve multiple users at once in any frequency or time slot of a traditional time or frequency division multiplexing scheme by exploiting the spatial dimension (Space-Division Multiple Access (SDMA)). The users can even be provided with more than one spatial data stream. An introduction to the basic aspects of MIMO systems can be found in Part I, along with a summary of the other parts of the work. This part deals with MIMO systems with a single BS, whereas Part III treats aspects of systems with multiple transmitting stations (BSs and/or Relay Nodes (RNs)).

As depicted in Figure 4.1, the downlink of such a system with one BS can be viewed as a rectangular grid of resources where each grid point can be a container for the data to be transmitted to a different subsets of users. A scheduler is needed to decide how many and which users to serve in any given resource element, based on a number of criteria which, in the SDMA case, must take into account several aspects of the spatial dimension as explained in the following.

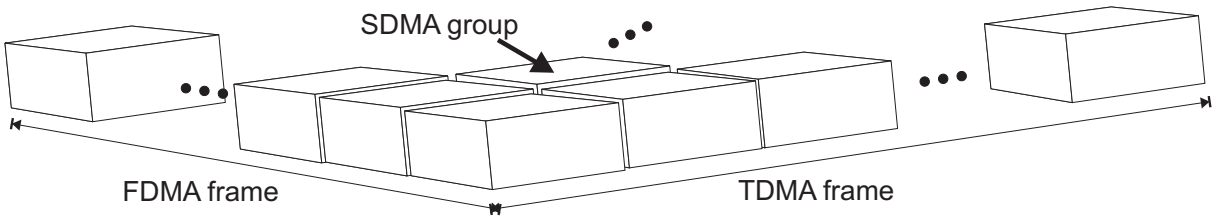


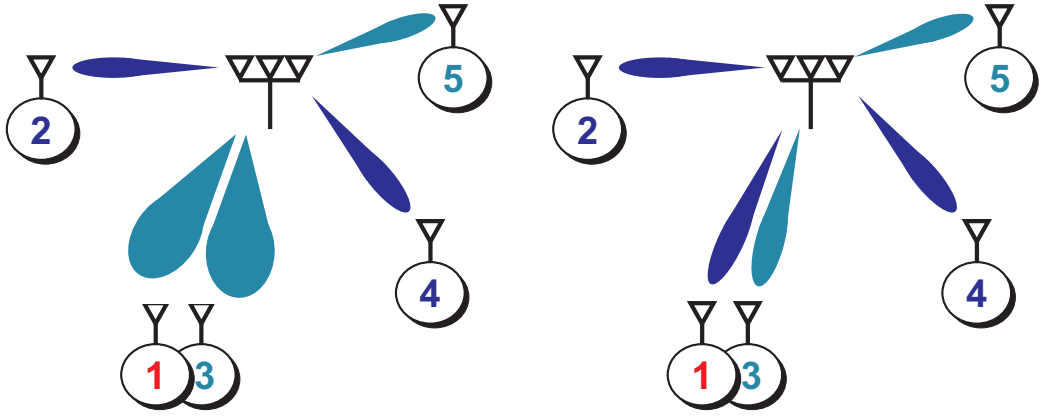
Figure 4.1.: The MIMO Orthogonal Frequency Division Multiplexing (OFDM) system can be regarded as a rectangular grid of resource elements where each resource can be used to serve a different subset of users via SDMA.

Recently, the gain offered by scheduling (often referred to as multi-user diversity) has attracted considerable interest because it has been backed by information theoretic results [44] that the system throughput can be boosted by scheduling only users with good channel quality for transmission and that linear precoding techniques can approach the maximum achievable sum rate of a system with proper scheduling if the available number of users to choose from goes to infinity [37].

One of the challenges to be solved lies in the computational complexity: The various SDMA techniques in the literature can be divided into the ones able to adapt the precoding to the spatial signatures of the downlink channels if Channel State Information (CSI) is available at the transmitter (via estimation, feedback, or prediction), and into others with fixed precoding weights. The latter case is often referred to as opportunistic beamforming because multi-user diversity may be leveraged through artificial channel fluctuations induced by randomizing the beamforming weights [45]. Yet still, scheduling for opportunistic beamforming [46, 47] is less complex than for adaptive precoding because the beam weights are fixed when the scheduling decision is to be taken. In adaptive precoding, a different subset of users results in a different set of precoding weights, which in turn might render the previous user selection suboptimal. The brute force solution to finding the optimum solution for a given linear precoding algorithm results in an exhaustive search such as in [48] which is prohibitive and may easily outweigh all complexity advantages that linear precoding techniques have over non-linear ones. Instead of considering a brute force search, complexity wise it would make more sense to employ one of the theoretically optimum Dirty Paper Coding (DPC) based algorithms such as [29] which inherently solve the scheduling problem while finding the optimum weights by assigning zero power to the unfavorable spatial modes.

In any case, scheduling is a required function in linear adaptive precoding because the channel quality degrades considerably and unpredictably if users with spatially correlated channels are served at the same time. If no scheduling is performed, the impairment due to spatially correlated users can only be alleviated by a significant increase in the number of base station antennas leading to an increased resolvability of the channel subspaces. To better understand the problem of spatial correlation a simplified example is illustrated in Figure 4.2. A BS serving five users is shown where user one is located closely to user three, causing the subspaces of the two users' channel matrices to be increasingly correlated. In the left subfigure, user one is served together with user three in the green SDMA group leading to inefficient or less well focussed beams whereas in the right figure the beams can be better focussed because the two spatially correlated users are served in different system resources.

Due to the nature of the problem, many recent works suggest to group the users on the downlink based on some measure for pairwise spatial correlation between channel matrix subspaces



(a) When user one is served via SDMA in the same group as user three, the BS has difficulties separating the channel subspaces, leading to inefficient beams.

(b) More efficient precoding weights can be found when the two collocated users are served in different groups.

Figure 4.2.: A simplistic illustration of the effect of spatially correlated user channels.

in cases where channel state information is available at the transmitter [49, 50, 41]. The author of [51] improves upon this approach by defining correlation based metrics in [52, 53] which comprise the correlation between all users' channel matrices in the system. As an alternative to correlation, users could also be grouped by their (mean) Direction Of Arrival (DOA) difference for the case with not too large angular spreads [54]. Such correlation measures, however, suffer from the drawback that there does not seem to exist an explicit relationship to the obtainable throughput after transmit precoding. As a consequence they cannot inherently take into account the influence of the transmit power allocation which may be described as follows:

For the sum performance of an SDMA group, there exists a trade-off between adding another user to the group or offering greater fractions of the available power to all users already in the group. In [49] it was illustrated that the number of users to be scheduled at the same time (i.e., the SDMA group size) has a great influence on the performance. In many cases it was observed that it can be beneficial not to fully load the base station, i.e., not to serve the maximum possible number of users at any one instant, even if the system is not overpopulated. However, this is not only caused by the fact that the base station can on average transmit with more power to the users when the SDMA group size is reduced. Also, the separability of the users' spatial signatures is increased when the number of users is low in relation to the number of transmit antennas, leading to better channels after precoding.

Other authors have already thoroughly investigated the dependency of the sum performance on the SDMA group size with rank limited precoding (see e.g., [55]) and, therefore, we do not investigate this issue here. The total number of spatial data streams which can be efficiently

provided to all users simultaneously is limited by the rank of the downlink channel matrix. For some techniques this limit expresses itself as a strict dimensionality constraint in the computation of the precoding matrix [56]. Even techniques which do not have this dimensionality constraint such as maximum Signal to Noise Ratio (SNR) beamforming [57], Regularized Block Diagonalization (RBD) or Successive MMSE (SMMSE) [38] still show a performance decrease at this rank limit.

To be able to include the influence of the power and group size in the scheduler, the main scheduling approach proposed in this work is based on low complexity estimates of the user rates with precoding, since rate figures directly reflect both the influence of spatial correlation and transmit power. Other recent works propose additional algorithm steps to adjust the group size, e.g. by removing users from the group, after the initial selection was made based on a non-capacity metric [52, 53, 51, 58, 59]. While [58, 59] rely on the computationally costly explicit computation of Shannon capacities after precoding for such a final adjustment step, [52, 53] show that the users' channel norm can be a less complex measure.

As said above, the result of the precoding depends on the user selection. The methods presented here, however, reduce the computational complexity of estimating the end performance to an effort comparable to when a user is served alone with the help of the concept of orthogonal projections as described in Chapter 5. For this reason we call our approach *ProSched*. The use of orthogonal projections for scheduling was first presented in [4] in conjunction with Zero Forcing (ZF) precoding and for a system model which may be referred to as user grouping. Other authors [58, 55] have later also developed scheduling algorithms using orthogonal projections in a different fashion. However, their proposal requires an additional pre-selection step to reduce the computational complexity, as mentioned above and discussed in more detail in Chapter 7.

The calculation of the beamforming weights for all possible user combinations is avoided with *ProSched*. To identify the combination with maximum throughput, the *ProSched* rate estimates can then be used together with a search algorithm as the one illustrated in Chapter 6 to reduce the number of user combinations to be tested. With the help of simulations and complexity estimates it is shown in Chapter 8 that *ProSched* based approaches achieve almost the same performance as an exhaustive search for the optimal user combination at a much lower computational cost. In the simulations shown here, simple linear precoders approach the theoretical sum rate limit when used together with proper spatial scheduling.

The contents of this Part of the work reflect the development of *ProSched* as presented in [5] where it was also applied to precoding techniques other than Zero Forcing (Section 5.2) and extended to a fair space-time-frequency scheduling scheme which can consider user rate requirements (Section 5.4). Furthermore it was shown how it can be based not only on instantaneous but also on long-term channel knowledge (Section 5.5).

The system model used in [5] is described in Section 4.2 and can be referred to as user selection. In Section 4.1 it is briefly explained that it has certain advantages over user grouping [4]. Nevertheless it is shown later on how *ProSched* can be used for both approaches by using different search algorithms. In addition to the above references, *ProSched* has been discussed in various technical documents such as [6, 7, 8, 9] and has been patented in Germany with an international patent pending [3].

4.1. User grouping versus user selection

In various publications such as [1, 8] a distinction is made between user selection algorithms as, e.g., [60, 5, 54], and user grouping algorithms as in [4, 37, 61]. User grouping implies to split all users into groups that are usually non-overlapping with the constraint that all groups must be served at least once whereas user selection allows for an independently chosen user subset in each resource element. In [8] a simulation comparison between grouping without and with overlapping groups shows, the latter being comparable to our notion of user selection, that strict grouping results in inferior sum performance. The reasons are not elaborated there but are rather clear:

- Grouping with non-overlapping groups may result in users being served in system resources where their choice is not optimal due to the constraint that all groups need to be served at least once. User selection on the other hand allows for an adaptation of the group to the individual channel conditions in each system resource but can still be fine tuned for fairness if a user absolutely needs to be served.
- With grouping, a wrap-around problem exists at the end of a Time Division Multiple Access (TDMA) or Frequency Division Multiple Access (FDMA) frame when the number of resource elements in the frame is not an integer multiple of the number of groups built by the scheduling algorithm. Groups need to be dropped then, resulting in users not being served.

4.2. System model and notation

Let us consider a system in which FDMA, TDMA, and SDMA are used together. A TDMA frame consists of a number of time slots, where each slot can consist of one or more OFDM symbols. In a system with a total of N_u users, the scheduler can select a different subset of users $\mathcal{G}_G(n, f)$ to be served at the same time by the SDMA scheme for every time slot n and every frequency bin f such that the precoder can deliver efficient modulation matrices. The index G denotes the size of a subgroup. To reduce complexity, the same SDMA group could be

assigned to a number of correlated neighboring subcarriers and time slots. Furthermore, it is assumed that new CSI is available at the base station at the beginning of each TDMA frame. As a consequence, the scheduling decision can remain unchanged within one frame if it solely relies on the CSI.

The users in each group are numbered consecutively for simplicity. The channels on each subcarrier are considered frequency flat. Note that in the following all variables are dependent on time and frequency and that the indices n and f are omitted for notational simplicity. Let $\boldsymbol{H}_g$ denote the $M_{\mathrm{R},g} \times M_{\mathrm{T}}$ complex channel matrix between the M_{T} transmit antennas of the base station and the $M_{\mathrm{R},g}$ receive antennas of user $g \in \{1 \ldots . G\}$ in a group of size G. Let further $\boldsymbol{M}_g$ denote a linear precoding matrix generated for the transmission to user g. To cover the most general case, the system under consideration shall have the possibility to transmit multiple data streams to each user in an SDMA group. Therefore, and because of the rank limit mentioned above, $\boldsymbol{M}_g$ is allowed to have up to $r = \operatorname{rank}\{\boldsymbol{H}_g\}$ columns. All precoding matrices are generated jointly from the information about the channels of all users in a group, as illustrated later. On each of the N_{c} subcarriers, the complex symbol vector $\boldsymbol{y}_g$ received by user g is obtained from the complex data vectors $\boldsymbol{d}_g$ as follows:

$$\boldsymbol{y}_g = \boldsymbol{H}_g \boldsymbol{M}_g \boldsymbol{d}_g + \sum_{j=1, j\neq g}^{G} \boldsymbol{H}_g \boldsymbol{M}_j \boldsymbol{d}_j + \boldsymbol{n}_g \,. \tag{4.1}$$

The vector $\boldsymbol{n}_g$ contains the additive noise at the receiver of user g. It is usually assumed that the data symbols are uncorrelated with average unit power, or $\mathrm{E}\left\{\boldsymbol{d}_g \boldsymbol{d}_g^{\mathrm{H}}\right\} = \boldsymbol{I}$. Consequently, the (instantaneous) covariance matrices of the signal vectors $\boldsymbol{t}_g = \boldsymbol{M}_g \boldsymbol{d}_g$ transmitted to the users reduce to $\boldsymbol{R}_{t_g t_g} = \boldsymbol{M}_g \boldsymbol{d}_g \boldsymbol{d}_g^{\mathrm{H}} \boldsymbol{M}_g^{\mathrm{H}} = \boldsymbol{M}_g \boldsymbol{M}_g^{\mathrm{H}}$.

Under the assumption of Gaussian signaling and infinite length codewords it is well known that the downlink sum capacity of one frequency flat subcarrier can be expressed as

$$C = \max_{\boldsymbol{R}_{t_g t_g}, \forall g} \sum_{g=1}^{G} \log_2 \frac{\det\left(\boldsymbol{R}_{n_g n_g} + \sum_{j=1, j\neq g}^{G} \boldsymbol{H}_g \boldsymbol{R}_{t_j t_j} \boldsymbol{H}_g^{\mathrm{H}} + \boldsymbol{H}_g \boldsymbol{R}_{t_g t_g} \boldsymbol{H}_g^{\mathrm{H}}\right)}{\det\left(\boldsymbol{R}_{n_g n_g} + \sum_{j=1, j\neq g}^{G} \boldsymbol{H}_g \boldsymbol{R}_{t_j t_j} \boldsymbol{H}_g^{\mathrm{H}}\right)} \tag{4.2}$$

Recall that frequency indices were omitted. The maximization is subject to a fixed total transmit power P_{T} such that $\sum_{f=1}^{N_{\mathrm{c}}} \sum_{g=1}^{G} \operatorname{trace}\left(\boldsymbol{R}_{t_g t_g}(f)\right) \leq P_{\mathrm{T}}$. Note that, since $\boldsymbol{R}_{t_g t_g} = \boldsymbol{M}_g \boldsymbol{M}_g^{\mathrm{H}}$, the squared norms of the columns of $\boldsymbol{M}_g$ represent the powers assigned to each data symbol in $\boldsymbol{d}_g$.

The derivations being based on MIMO capacity expression, *ProSched* is not restricted to a specific type of receiver. However, we restrict ourselves to the case where no joint processing between the receivers of different users can be performed.

5. Introduction of a scheduling metric

The precoding matrix for user g depends on which users are served at the same time. It is, therefore, desirable to solve the scheduling problem without relying on the knowledge of the precoding matrices to avoid the computational complexity of pre-calculating all possible precoding solutions. To overcome this problem, a metric is proposed which provides an estimate of the g-th user's received data rate while decoupling the calculation from the channel matrices of the other users served at the same time, maintaining however the influence of the other channels. Some simplifications are presented whose effectiveness is later evaluated extensively with the help of simulations. In the next chapter this metric will be used in a selection algorithm which reduces the number of combinations to be tested. First we want to look at Block Diagonalization (BD) precoding [62], which is a ZF precoding technique with theoretically optimum capacity and later apply the result also to other precoding techniques.

5.1. Metric for an optimal Zero Forcing precoder such as Block Diagonalization

The transmission to user g involves an equivalent channel $\boldsymbol{H}_g\boldsymbol{M}_g$. A precoding method which suppresses all inter-user interference (i.e., $\boldsymbol{H}_j\boldsymbol{M}_g = \boldsymbol{0}\ \forall j \neq g$) is commonly referred to as ZF. As explained in [62], this implies that the transmission to user g must take place in the intersection of the null spaces of all other users' channel matrices. One way to achieve this is to construct the columns of $\boldsymbol{M}_g$ as linear combinations of basis vectors of this joint null space. This basis can, for example, be obtained with the help of a Singular Value Decomposition (SVD) of a matrix $\tilde{\boldsymbol{H}}_g$ containing all other users' channel matrices $\tilde{\boldsymbol{H}}_g = \begin{bmatrix} \boldsymbol{H}_1^{\mathrm{T}} & \cdots & \boldsymbol{H}_{g-1}^{\mathrm{T}} & \boldsymbol{H}_{g+1}^{\mathrm{T}} & \cdots & \boldsymbol{H}_G^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ while at the same time transmitting as much power as possible into the row space (also called signal space) of user g's own channel matrix. This approach is used in [62] for the BD algorithm.

Our scheduling metric is based on a lower bound of the g-th user's capacity per frequency flat subcarrier. If σ_n^2 represents the total noise power in the entire bandwidth at one receiver, the noise power on one subcarrier is $\sigma_n^2/N_{\mathrm{c}}$. With spatially white noise, the noise covariance matrix is $\boldsymbol{R}_{n_g n_g} = \sigma_n^2/N_{\mathrm{c}}\boldsymbol{I}$. Then the capacity expression for user g becomes under the zero interference constraint $C_g = \log_2 \det\left(\boldsymbol{I} + \frac{N_{\mathrm{c}}}{\sigma_n^2}\boldsymbol{H}_g\boldsymbol{R}_{t_g t_g}\boldsymbol{H}_g^{\mathrm{H}}\right)$. Denoting $r = \operatorname{rank}\{\boldsymbol{H}_g\}$ and introducing an eigenvalue decomposition of the correlation term $\boldsymbol{H}_g\boldsymbol{M}_g\boldsymbol{M}_g^{\mathrm{H}}\boldsymbol{H}_g^{\mathrm{H}} = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^{\mathrm{H}}$, the rate with ZF

precoding can be expressed using the eigenvalues as follows:

$$\begin{aligned} C_g &= \log_2 \det\left(\boldsymbol{I} + \frac{N_\mathrm{c}}{\sigma_n^2}\boldsymbol{H}_g\boldsymbol{M}_g\boldsymbol{M}_g^\mathrm{H}\boldsymbol{H}_g^\mathrm{H}\right) \\ &= \log_2 \det\left(\boldsymbol{I} + \frac{N_\mathrm{c}}{\sigma_n^2}\boldsymbol{W}_g\boldsymbol{\Lambda}_g\boldsymbol{W}_g^\mathrm{H}\right) \\ &= \log_2 \det\left(\boldsymbol{W}_g\left(\boldsymbol{I} + \frac{N_\mathrm{c}}{\sigma_n^2}\boldsymbol{\Lambda}_g\right)\boldsymbol{W}_g^\mathrm{H}\right) \\ &= \log_2 \det\left(\boldsymbol{I} + \frac{N_\mathrm{c}}{\sigma_n^2}\boldsymbol{\Lambda}_g\right) = \log_2 \prod_{i=1}^{r}\left(1 + \frac{N_\mathrm{c}}{\sigma_n^2}\lambda_{g,i}\right) \\ C_g &= \log_2\left\{1 + \frac{N_\mathrm{c}}{\sigma_n^2}\sum_{i=1}^{r}\lambda_{g,i} + (\cdots)\right\}. \end{aligned} \tag{5.1}$$

Knowing that $\sum_{i=1}^{r}\lambda_{g,i} = \mathrm{trace}\left\{\boldsymbol{H}_g\boldsymbol{M}_g\boldsymbol{M}_g^\mathrm{H}\boldsymbol{H}_g^\mathrm{H}\right\} = \|\boldsymbol{H}_g\boldsymbol{M}_g\|_\mathrm{F}^2$ and that the other intermediate products of eigenvalues which have been skipped in Equation 5.1 are all positive numbers, we can say that the rate is lower bounded by:

$$C_g \geq \log_2\left(1 + \frac{N_\mathrm{c}}{\sigma_n^2}\|\boldsymbol{H}_g\boldsymbol{M}_g\|_\mathrm{F}^2\right). \tag{5.2}$$

Recall that the goal of our algorithm is to overcome the problem of pre-calculating the precoding matrices for all user combinations to be tested. First we want to look at the problem that, for the metric computation, ideal power allocation (such as waterpouring) cannot be used when the full precoding is not performed: To do so, we factor out a diagonal matrix $\boldsymbol{D}_g$ containing the square roots of the fractions of user g's power assigned to each of its spatial modes, where unused modes have zero entries. The remaining part of the precoding matrix has normalized columns and is denoted $\boldsymbol{N}_g$, i.e., $\boldsymbol{M}_g = \boldsymbol{N}_g\boldsymbol{D}_g$. The distribution of the eigenvalues of the equivalent channel $\boldsymbol{H}_g\boldsymbol{M}_g$ cannot be known before the precoding is performed and, therefore, the optimum distribution of the transmit power is unknown. It is, therefore, assumed that equal fractions of the total available transmit power are assigned to all subcarriers and per subcarrier on all users' spatial modes, i.e., $\boldsymbol{D}_g = \sqrt{P_\mathrm{T}/(N_\mathrm{c}\cdot G\cdot r)}\boldsymbol{I}$ (assuming also full rank channels). This reduces the capacity figure compared to an optimum power loading. The assumption works around the unknown distribution of the users' spatial modes and we can now define the following rate estimate as scheduling metric for user g in the presence of a set of other users $\mathcal{S}$, which is a lower bound for equation (5.2):

$$\eta_g^{(\mathcal{S})} = \log_2\left(1 + \frac{P_\mathrm{T}}{Gr\sigma_n^2}\|\boldsymbol{H}_g\boldsymbol{N}_g\|_\mathrm{F}^2\right) \leq C_g\,. \tag{5.3}$$

Note how N_c is canceled out due to the equal power loading assumption. Simulations indicate that this assumption is reasonable for the scheduler, even if the precoder later uses, for example, waterpouring power loading to approach the maximum sum rate. It can become problematic in near-far scenarios with significant differences in the user channel norms. However, the proposed proportional fair modification of Section 5.4 overcomes this problem. As a side-effect, the necessary equal power loading assumption allows us to skip the complexity involved with the computation of the water pouring during the scheduling process. (For a discussion on other implications of this assumption please see Section 5.3.)

If the system was designed to suppress also the interference between each user's data streams (e.g., via SVD based eigenbeamforming), the metric $\eta_g^{(\mathcal{S})}$ would equal the rate of user g [16] under the equal power assumption.

Breaking the interdependency between precoder and user selection

This metric, however, still depends on the knowledge of the precoding matrix, which for ZF depends on a basis of the common null space of all other users grouped together with user g. To combat this interdependency, we introduce the concept of orthogonal projections into the precoding: In Chapter 10.1, a new formulation of BD precoding is derived. This formulation involves a new effective channel $\boldsymbol{H}_g\tilde{\boldsymbol{P}}_g^{(0)} = \mathring{\boldsymbol{H}}_g \in \mathbb{C}^{M_{\mathrm{R}g}\times M_{\mathrm{T}}}$, which is the result of an orthogonal projection $\tilde{\boldsymbol{P}}_g^{(0)}$ into the common null space of all other users' channel matrices. The effective channel represents the g-th user's channel deprived of the part of the subspace which cannot be used for transmission since it does not lie in the other users' null space. (Note that the superscript $^{(0)}$ is added to a basis of a null space or a projection into a null space and $^{(1)}$ is added to a row space or a projection into a row space.) It is also shown that for BD precoding the norm of the equivalent channel after precoding - which is part of the metric in equation (5.3) - equals the norm of this projected channel:

$$\|\boldsymbol{H}_g\boldsymbol{N}_g\|_{\mathrm{F}}^2 = \left\|\boldsymbol{H}_g\tilde{\boldsymbol{P}}_g^{(0)}\right\|_{\mathrm{F}}^2 . \tag{5.4}$$

We then make use of a property of projectors introduced in [63]: A projection into a subspace can be approximated by repeatedly projecting into the separate subspaces whose intersection is the subspace to project into:

$$\tilde{\boldsymbol{P}}_g^{(0)} = \left(\boldsymbol{P}_1^{(0)}\cdot\ldots\cdot\boldsymbol{P}_{g-1}^{(0)}\boldsymbol{P}_{g+1}^{(0)}\cdot\ldots\cdot\boldsymbol{P}_G^{(0)}\right)^p \;,\; p\to\infty\,. \tag{5.5}$$

Note that the $\tilde{\cdot}$ symbol is used to mark a projection into all other users' space. In this repeated projection approximation the order of the projections is not significant - as long as one does not project in the same subspace multiple times in a row which is of no use because projectors

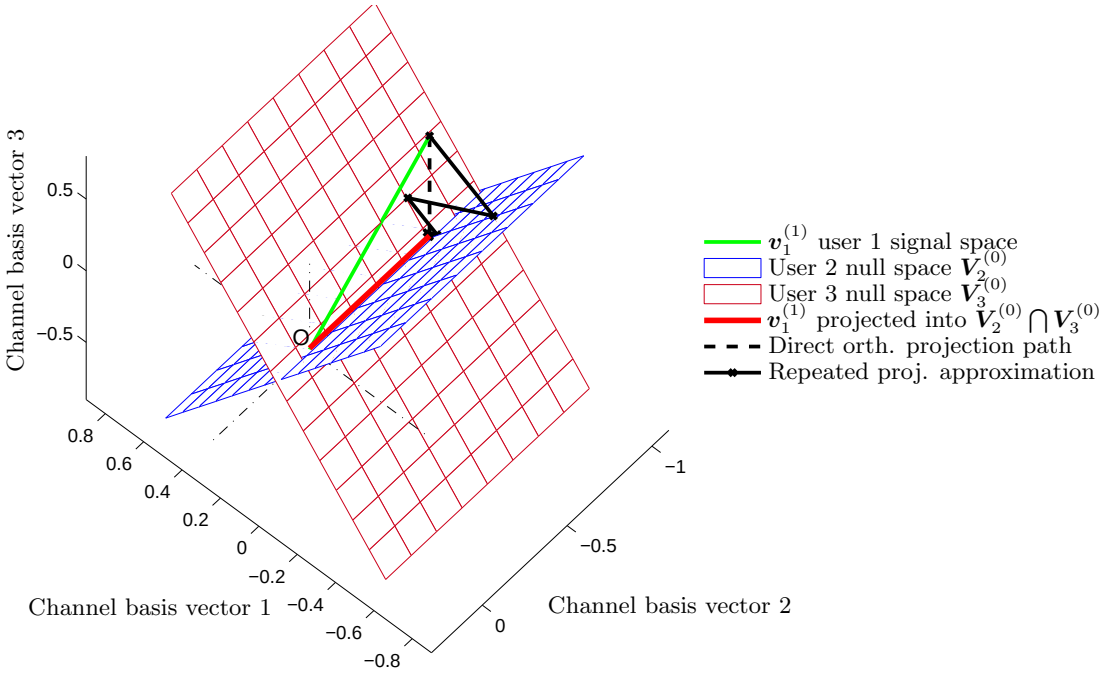


Figure 5.1.: An example for the repeated projection approximation in a real channel matrix with $M_\mathrm{T} = 3$ and 3 single antenna users: ZF precoding for user one implies a projection into the intersection of the two other users' null spaces. The zig-zag path connects the vector end points when projecting into the second and third user's null space alternatingly. This converges to the same result as a direct projection into the common null space. The latter, however, is a different one for each user combination to be tested whereas the separate null spaces stay the same.

are idempotent. By using this approximation in the calculation of the metric $\eta_g^{(\mathcal{S})}$ from equation (5.3), the scheduler no longer has to know the basis of the common null space for any subset of users to test, but can instead estimate their equivalent channels only from the knowledge of all users' null space projectors. The projection matrix on the g-th user's null space can efficiently be computed from an orthonormal basis $\boldsymbol{V}_g^{(1)}$ of its row space with $\boldsymbol{P}_g^{(0)} = \boldsymbol{I} - \boldsymbol{V}_g^{(1)}\boldsymbol{V}_g^{(1)\mathrm{H}}$. The approximation in equation (5.5) only requires G SVDs at the start of each scheduling run to obtain the bases for each of the G users' signal spaces. Furthermore, simulation results show that it is sufficient to choose the projection order p between 1 and 3.

Since the repeated projection approximation is a core feature of the *ProSched* approach, a graphical explanation is presented in Figure 5.1 for better understanding. It assumes a 3×3 channel matrix with real entries, representing a BS with $M_\mathrm{T} = 3$ and three users with one antenna each. In this case the null space of a user forms a plane and the users' signal spaces are vectors. It is shown how the signal space of user one is projected into the intersection of the

other two users' null spaces either directly or via repeatedly projecting into either other user's null space plane, reaching the same end point.

In order to reduce the complexity even further, the projectors used in the repeated projection approximation (5.5) can be obtained from rank one approximations of the signal spaces by using only their strongest mode, as proposed in [49]. If this approach is used, complexity can once more be reduced by employing a less exact rank one approximation such as the normalized column of $\boldsymbol{H}_g^{\mathrm{T}}$ with the highest norm. The latter corresponds to a Sparse Pivoted QR decomposition approximation (SPQR) [64], where the algorithm was stopped after the first basis vector was found.

As a final remark please note that the part of expression 5.3 representing the SNR on the current subcarrier, $\frac{P_{\mathrm{T}}}{Gr\sigma_n^2}\|\boldsymbol{H}_g\boldsymbol{N}_g\|_{\mathrm{F}}^2$, can also be used as a metric for scheduling, as presented in [4]. We prefer, however, to use the rate estimate because it allows for convenient ways of introducing QoS into the metric, see Section 5.4.

5.2. Metric in the case of other (non-ZF) precoding techniques

A general precoder which does not necessarily suppress inter-user interference must choose which basis to use for the transmission to the g-th user. If it chose the g-th channel's row space, it would exploit the channel as best as possible, however causing high interference to the other users in the group. On the other hand, if it chose the projection of the g-th channel into the common null space of the other users (which is the ZF solution), it would produce no interference but possibly attain a lower data rate for user g. A general precoder chooses an intermediate solution between these two extremes. For this reason, a good spatial scheduler aims to group users whose row spaces are as close as possible to the other users' common null space, thus maximizing throughput while holding back interference.

The problem can be illustrated graphically in a system with two users having one antenna each and a base station with two antennas, resulting in a combined channel with rank 2. The right part of Figure 5.2 shows a situation where user one is combined with another correlated user having a null space which is almost orthogonal to the user subspace of interest. Reducing the interference will result in a precoded channel with a small norm. The left picture shows a less correlated situation where a precoding close to the null space will leave the norm of the channel almost unchanged. (In a situation with more users, the precoded channel of a non-ZF solution can be inclined in space more or less towards any of the signal spaces and does not have to be on a plane with the joint null space.)

There exist various methods to measure similarity between subspaces. One is the correlation between subspaces which is used for scheduling in different ways, e.g., in [49, 50]. These methods

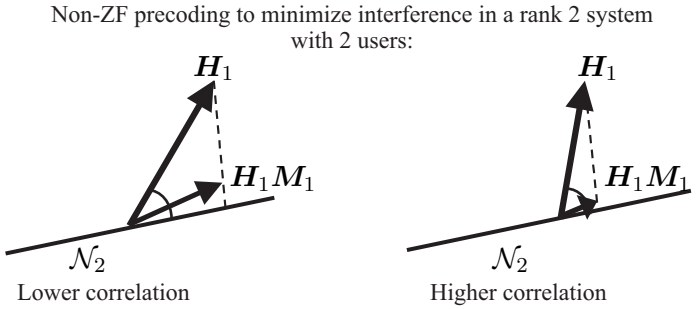


Figure 5.2.: Non-ZF precoding to minimize interference illustrated for a system with 2 single antenna users and a base station with 2 antennas: the null space $\mathcal{N}_2$ of users 2 forms a line. The precoder can either make use of user 1's channel fully or use a subspace closer to the null space to generate less interference. In a correlated situation, the latter results in a significant reduction of the norm, as can be seen on the right.

are not capable of taking into account the influence of the power assigned to the transmission to each user when estimating how many users should be served at the same time. Therefore, and because it is dual to the interference generated by user g, the ZF capacity estimate (5.3) based on the Frobenius norm of a user's channel projected completely into the null space of the other users' channel matrices is proposed as a test for user compatability also for the non-ZF case. Another explanation for the fact that the ZF metric works well enough also in other cases is as follows: Most precoders as, for example, the Minimum Mean Squared Error (MMSE) Transmit Wiener Filtering (TxWF) precoding technique (see, e.g., [25]) used in the simulations converge to the pseudo inverse of the channel at high SNRs and include, therefore, the ZF solution as a special case. (For precoders that exceed the dimensionality constraint implied in *ProSched* due to its ZF nature please see the following section).

5.3. The number of spatial modes in the proposed metric

One key feature of the proposed metric is to avoid the computation of the full precoding matrices. It was said above that, as a result, the modal distribution of the channel after precoding is unknown to the scheduler. It was proposed to work with an equal distribution of the available transmit power on all user's spatial modes. In this short section some implications besides the fact that this represents an approximation are discussed and modifications proposed.

Once a group of spatially compatible users is found, e.g., with the help of search algorithms as the ones in Chapter 6, the precoding matrices for the users can be computed including a power allocation. In contrast to the above assumption, the result could possibly involve a different number of active modes for each of the users, depending on the channel conditions

(i.e., rich scattering environment versus dominating line of sight component) and optimization criteria. It may then be worthwhile accommodating more users when some users' modes have been deactivated by the power loading step - supposed that either the chosen group size is still smaller than the maximum supportable or, alternatively, the precoding algorithm can cope with in total more receive antennas than transmit antennas at the BS.

A slight gain in sum throughput is possible from optimizing the number of active modes jointly with the scheduler compared to when the number of active modes is assumed fixed in the scheduler. In [51], for example, it was proposed to iterate a low complexity scheduler with the precoder such that an initially found user selection may be fine tuned after the precoding matrices were computed. Such an approach can also be implemented with *ProSched* when the following aspects are properly taken into account:

The *ProSched* metric as such is derived from a ZF algorithm and should, thus, suffer from the same dimensionality constraint which is that the total number of receive antennas in the selected user subset must be lower or equal the number of transmit antennas, i.e., $\sum_g = 1^G M_{\mathrm{R},g} \leq M_{\mathrm{T}}$. If this condition is violated, the combined channel matrix of all other users than the g'th, $\tilde{\boldsymbol{H}}_g$, simply doesn't have any (right) null space to project $\boldsymbol{H}_g$ into (supposed that all channel matrices are full rank).

On the other hand it is possible to accommodate more users with non-ZF precoders, supposed that the total number of active modes does not exceed the rank of the combined downlink channel. *ProSched* can still be used in this case due to the presented approximation using repeated projections since the separate projectors into single users' null spaces can always be found as long as their number of active antennas is less or equal than M_{T} (which seems reasonable to assume). However, when the ZF dimensionality constraint is violated, the repeated projection approximation is not going to result in a ZF solution. Instead, a solution with residual interference for one user will result. Such a combination of *ProSched* with SMMSE precoding and Dominant Eigenmode Transmission (DET) for all users was successfully tested in Chapter 8. This combination can schedule more users than allowed by the ZF constraint.

The use of DET in some of the simulations is based on a recommendation within the WINNER project [15] system proposal that was motivated by the underlying measurement based channel model which was found to produce low rank channels in most scenarios of interest [65, 66], see also Chapter 8.

The metric requires to be changed to take into account an unequal power distribution: DET means that only the first mode is loaded with power. From a sum capacity point of view (which is the original viewpoint in *ProSched*), this occurs only when a user's channel matrix is rank one. When going from equation (5.2) to the metric (5.3) it should be noted that, one the one hand, the power loading matrices D_g are no longer diagonal but contain only one non-zero element

instead, namely the assumed power for the dominant mode. On the other hand, this results in the Frobenius norm being taken from a matrix with only the first column being non-zero. Thus, the scalar power can be factored out again, even with D_g not being diagonal. The only necessary change is to replace P/Gr by the correct fraction of power. For example, if all users were to be assumed to be served on their dominant modes instead of all available modes, the total number of active modes should be G instead of Gr.

5.4. Fairness and QoS extension

Since the proposed scheduling metric is a rate estimate rather than a correlation only metric, it is straightforward to use it in combination with one of the many rate based QoS aware or fair algorithms proposed in the literature.

A well known notion of fairness is that of proportional fairness. Several references on proportional fair algorithms can, for instance, be found in [67]. In communications, proportional fairness is widely defined with the help of a vector $\boldsymbol{r} = [r_1, r_2, \ldots, r_{N_\mathrm{u}}]^\mathrm{T}$ containing the achievable rates $r_u,\ u \in 1..N_\mathrm{u}$ of all N_u users in the system at a certain time instance including the ones not scheduled and having rate zero. Such a rate vector is said to be proportionally fair if for any other feasible rate vector $\boldsymbol{r}^* = \left[r_1^*, r_2^*, \ldots, r_{N_\mathrm{u}}^*\right]^\mathrm{T}$ the sum of proportional changes is zero or negative: $\sum_{u=1}^{N_\mathrm{u}} \frac{r_u^* - r_u}{r_u} \leq 0,\ \forall\ \boldsymbol{r}^*$. The original equivalent definition from queueing theory [68] states that the sum of a utility function of the rates is to be maximized. Often, logarithmic utility functions are used, resulting in a maximization of the product of the rates.

It can be shown for i.i.d. channels [69] that a scheduling algorithm fulfils these definitions if it keeps track of the long term averages of all users' rates and prioritizes the user with the highest instantaneous rate normalized to its long term average throughput - supposing that the long term throughput figures are built for each user with the help of an exponentially smoothed average.

Furthermore it was shown in [69] for i.i.d. channels that proportional fairness asymptotically schedules all users with the same fraction of time and power under the assumption that the supported data rates are linear with power, at least on a long term scale.

Following this observation, similarly to other proportional fair implementations, our proposal can be modified to include proportional fairness by normalizing the metric $\eta_g^{(\mathcal{S})}$ to a long term average, i.e., by replacing it with an $\Upsilon_g^{(\mathcal{S})} = \eta_g^{(\mathcal{S})} / \overline{\eta}_g$.

In our simulations we used a linear average of the user metrics corresponding to the final scheduling decisions from m previous time slots to construct the long term average metric $\overline{\eta}_g$: $\overline{\eta}_g(n) = \frac{1}{m} \sum_{\mu=n}^{\mu=n-m-1} \eta_g(\mu)$. With this implementation in place it becomes necessary to carry out the scheduling decision for every time slot instead of every frame, because the long term

averages change as soon as some users are served and some are not. A rectangular window is of course more memory consuming than an exponential one, but offers a more direct influence on the delay introduced. If m is small, then the time until a user is scheduled is reduced if its channel quality hits a peak. However m should not be chosen smaller than the time frame length or the rate averages converge too quickly to the current rates, especially if the number of users in the system is low and due to the fact that only one channel estimate is available per frame. Then the fairness method would have no effect.

As a welcome side effect, proportional fair scheduling helps to deal with near-far scenarios due to the normalization of the users' rates to long term averages. If a user was much further away from the base station than all other users, a non proportional rate based metric would never schedule it, since its expected rate would be much smaller than the rates of the closer users. A normalized rate, however, can anyhow yield a high value as soon as the long term average has dropped to the same order of magnitude, thus allowing the ratio to become high again and the far user to be scheduled.

To introduce a first notion of QoS, the ratio $\Upsilon_g^{(\mathcal{S})}$ can simply be multiplied with a cost factor c_g which is supposed to be chosen higher for users requiring a higher data rate to sustain their desired service. Schacht [70] introduced this idea for scheduling in third generation (3G) systems and suggests to set the cost factors based on the number of time slots a service needs to occupy in order to obtain its minimum required throughput per frame at the lowest possible transmission rate. The service with the lowest requirement is assigned cost factor one and all other services are related to it.

Alternatively, as proposed in [71], an additive cost factor can be used. Additionally, the metric can not only be normalized to its long term average, but also to a target rate T_g such that $\Upsilon_g^{(\mathcal{S})} = \eta_g^{(\mathcal{S})} / \left((\overline{\eta}_g \cdot T_g)^\kappa + c_g \right)$. The factor κ can fine tune the influence of the target rate, where $\kappa = 0$ yields back the maximum throughput scheduling metric.

5.5. Extension to second order statistics channel knowledge

Channel state information is often acquired only in one transmission direction of a system and then assumed to be reciprocal and fed back on the reverse link. If the channel is varying too rapidly, alternatively a spatial transmit covariance matrix $\boldsymbol{R}_{T,g} = \mathrm{E}\left\{ \boldsymbol{H}_g^{\mathrm{H}} \boldsymbol{H}_g \right\}$ can be tracked instead of the channel matrix $\boldsymbol{H}_g$ (on each subcarrier separately). It was shown in [72] that the eigenvectors of the transmit signal covariance matrix $\boldsymbol{R}_{t_g t_g}$ which maximizes the ergodic capacity are given by the eigenvectors of $\boldsymbol{R}_{T,g}$.

Or in other words, if $\boldsymbol{H}_g$ is not known one should transmit into the eigenspace of its (estimated) transmit covariance matrix instead.

A method to apply this principle to existing MIMO multi-user precoding schemes can be found in [73]: A pseudo channel matrix $\hat{\boldsymbol{H}}_g$ is constructed from a basis of the signal space of $\boldsymbol{R}_{T,g}$ which can, for example, be obtained with an EigenValue Decomposition (EVD)

$$\begin{aligned} \boldsymbol{R}_{T,g} &= \boldsymbol{V}_g \boldsymbol{\Sigma}_g^2 \boldsymbol{V}_g^{\mathrm{H}} \\ &= \boldsymbol{V}_g \boldsymbol{\Sigma}_g^2 \left[\boldsymbol{V}_g^{(0)} \; \boldsymbol{V}_g^{(1)} \right]^{\mathrm{H}} \qquad (5.6) \\ \text{such that} \quad & \hat{\boldsymbol{H}}_g = \boldsymbol{\Sigma}_g \boldsymbol{V}_g^{(1)^{\mathrm{H}}} . \qquad (5.7) \end{aligned}$$

where $\boldsymbol{V}_g^{(1)}$ contains the first $r = \operatorname{rank}\{\boldsymbol{H}_g\}$ columns of $\boldsymbol{V}_g$. With practical channel measurements, the matrices $\boldsymbol{H}_g$ likely are full rank. Should they not be, then padding with zero vectors is required for the $\hat{\boldsymbol{H}}_g$ to become the same size as the $\boldsymbol{H}_g$.

These pseudo channel matrices $\hat{\boldsymbol{H}}_g$ can then be used instead of the actual channel matrices in any precoding method developed for short term channel knowledge, and, which is more important in this context, in the scheduling metric proposed in the previous sections.

Note that, in equation (5.6), the elements of $\boldsymbol{\Sigma}_g^2$ are the eigenvalues of $\boldsymbol{R}_{T,g}$ but that in equation (5.7) the square root needs to be taken, i.e., $\boldsymbol{\Sigma}_g$ is used. This becomes more obvious when one considers $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ written with an economy sized SVD of $\boldsymbol{H}_g$, i.e., with only the non-zero singular values in $\boldsymbol{\Sigma}_g$ and the corresponding basis vectors in $\boldsymbol{V}_g^{(1)^{\mathrm{H}}}$, namely $\boldsymbol{H}_g = \boldsymbol{U}_g \boldsymbol{\Sigma}_g \boldsymbol{V}_g^{(1)^{\mathrm{H}}}$: Then, the covariance matrix estimate becomes $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} = \boldsymbol{V}_g^{(1)} \boldsymbol{\Sigma}_g^2 \boldsymbol{V}_g^{(1)^{\mathrm{H}}}$. Also, the missing information of $\boldsymbol{U}_g$ is not needed in practical precoding algorithms such as BD, see the algorithm description of BD in Section 10.1.

In Chapter 8 a simulation example is given to compare scheduling with instantaneous CSI to scheduling based on the above long term averaging method.

In Part III, Section 14.2, it is described that the method can also be used to jointly precode and schedule multiple correlated neighboring subcarriers of an OFDM system.

6. Scheduling algorithm

In this chapter, an algorithm is introduced to reduce the number of combinations to be tested in order to find the best users to schedule. It is based on a scheduling metric as the one introduced in the previous chapter, but could be used in conjunction with any other type of spatial metric.

6.1. User selection algorithm

The algorithm discussed in this section works independently on every time slot n and subcarrier f and searches the best subset of users out of the N_u users in the system. In Section 6.4, a modification is introduced which can treat all subcarriers jointly. As discussed before, the scheduling decisions can remain constant within one frame if they only rely on channel knowledge and not on proportional fairness. In this case, n becomes the number of the frame and the algorithm needs to be executed only at the start of each frame.

The task of the algorithm can be divided into two phases: First it produces candidate user groups $\mathcal{G}_G(n,f)$ featuring maximum sum metric for all possible group sizes G from $G=1$ to the maximum supported size of the precoder, which is limited by the rank R of the combined downlink channel matrix $\boldsymbol{H} = \left[\begin{array}{ccc} \boldsymbol{H}_1^\mathrm{T} & \cdots & \boldsymbol{H}_{N_\mathrm{u}}^\mathrm{T} \end{array} \right]^\mathrm{T}$. This is performed with the help of a best-candidate search tree, which is depicted in Figure 6.1 for $N_\mathrm{u}=5$. It is a modification of the tree-based search used in [4]. In the second phase the algorithm selects for every pair (n,f) the group with the highest sum rate out of all candidate sets $\mathcal{G}_G^{opt}(n,f)$. This final phase can also be solved with the help of our rate estimate. Alternatively, if the metric only contained a measure for spatial compatability and no reference to the transmission power, the second step could still be performed by calculating the precoding matrices and the actual capacities for the candidate sets.

For every subcarrier, the algorithm is described by the following steps (see also the example below):

- Phase 1:
  1. Start: Let group size $G=1$. Form N_u user subsets of size G. Calculate the metrics for all user subsets and identify the best one as $\mathcal{G}_1^{opt}(n,f)$.
  2. Let $G=G+1$. If G is smaller than R, add one of the remaining users at a time to $\mathcal{G}_{G-1}^{opt}(n,f)$, forming $N_\mathrm{u}-(G-1)$ new candidate sets of the new size G. Otherwise skip to Phase 2.

3. Update the metrics of the users in the new candidate sets and calculate their metric sums. Keep the one set with the highest metric sum as $\mathcal{G}_G^{opt}(n, f)$. Back to step 2.

- Phase 2:
  Out of all candidate sets $\mathcal{G}_{1\ldots R}^{opt}(n, f)$ use the one with the highest sum metric, if the metric is a rate estimate. Otherwise calculate the precoders and rates for the candidate sets to be able to identify the best one.

In the example depicted in Figure 6.1, the algorithm steps of phase 1 perform the following (note that the time and frequency indices have been skipped for notational convenience):

1. The first tree level consists of $N_\mathrm{u} = 5$ user subsets of size 1 (labeled $1 \ldots 5$). User number one is identified as having the highest metric and becomes $\mathcal{G}_1^{opt}$.
2. One user at a time is added to $\mathcal{G}_1^{opt} = \{1\}$, forming 4 new candidate sets of size 2, namely $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$ and $\{1, 5\}$. The metrics of all users in all sets are calculated as well as the metric sum of each set. In the example, the group with the highest metric sum is $\mathcal{G}_2^{opt} = \{1, 2\}$. To come to the next tree level, the groups $\{1, 2, 3\}$, $\{1, 2, 4\}$ and $\{1, 2, 5\}$ are compared and so on.

In phase 2, the metric sums of the candidate sets $\mathcal{G}_1^{opt} \ldots \mathcal{G}_5^{opt}$ (marked in gray) or their actual sum rates are compared and the best one is selected.

Essentially, the algorithm produces a tree-like structure from the bottom up. A top-down approach producing candidate user sets with decreasing size can also be thought of, i.e., by starting with all users in one group and removing the least compatible user in each step such that the resulting smaller group has the best sum performance of all possible smaller groups. In the top-down approach the word 'all' should be treated with care, simply because the system might contain more users than the precoding algorithm can handle, requiring already some extra pre-selection to find the best group for the top of the tree. In this context the top-down approach is interesting because it allows to derive a method to track the scheduling solution in time for further complexity reduction, see Section 6.3.

6.2. User grouping algorithm

With the help of a slightly different algorithm [4] the user grouping problem can be solved. We will discuss it briefly but continue afterwards only based on the user selection approach because of its advantages summarized in Section 4.1.

As with the user selection solution, the algorithm also consist of building a tree like the one depicted in Figure 6.2. Let ℓ now denote the level of the tree, so that in the exemplary tree we

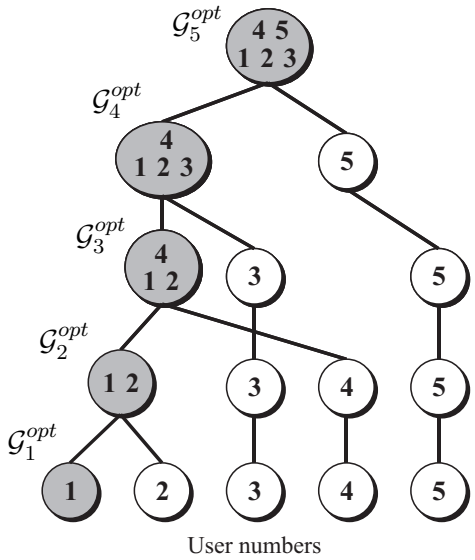


Figure 6.1.: An example for the sorting tree used in the user selection algorithm at one time instance: in a system with 5 users, candidate user sets $\mathcal{G}_G^{opt}$ of sizes $G = 1$ to 5 are produced with the help of a best candidate combining procedure and a scheduling metric reflecting the performance of the groups. In the final step the algorithm selects among the candidate sets, which can be found on the left edge of the tree.

have $\ell = 1, 2, \ldots, 5$. Each level ℓ of the tree represents an assignment of users into ℓ groups, as indicated by the connecting lines. At the lowest level, corresponding to $\ell = 5$, we have 5 groups, one for each user. For instance, for $\ell = 2$ we have two groups, the first being composed by users 1, 2, and 4. The remaining users form the second group. Again the tree can be constructed following either a bottom up or a top-down approach.

In the bottom up strategy, we start from the roots of the tree and continue by merging two groups of the previous level. At each step we have one group less. In order to decide which of the $\binom{\ell}{2}$ possible pairs of groups to join we can follow different criteria. We could choose the pair which has the best sum metric over the number of users in the group. Alternatively, we could choose the pair which penalizes its members the least in terms of their $\eta_g^{(\mathcal{S})}$ (in order to introduce a notion of QoS), or stop merging groups when a user's rate is predicted to drop below a target. Various other QoS variations of such a search tree have been developed and compared in [52].

In the example of Figure 6.2, the $\ell = 4$ level is derived by merging the first two users. Following the best sum criterion it means that $\eta_1^{(2)} + \eta_2^{(1)}$ was the maximum compared to all possible combinations of $\eta_j^{(i)} + \eta_i^{(j)}$ for $i \neq j$. Similarly, for $\ell = 3$, we merge the group composed by 1 and 2 with the group of user 4.

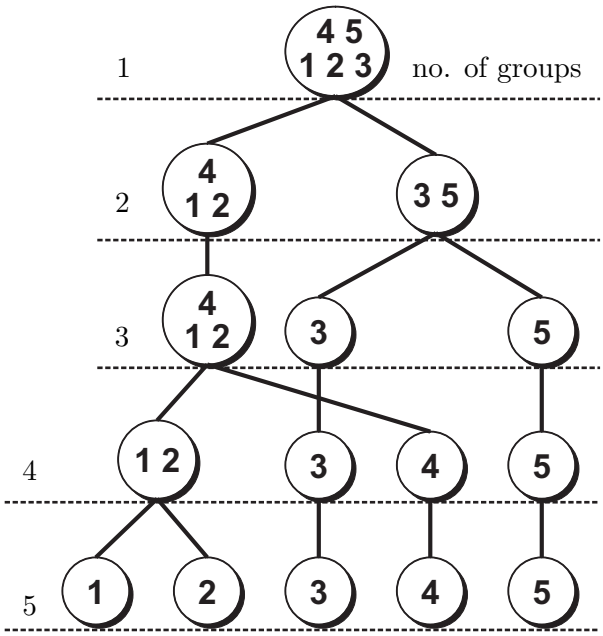


Figure 6.2.: An example for the tree-based user grouping algorithm: possible groupings of $N_\mathrm{u} = 5$ users are generated based on a performance metric. In a second phase the optimum number of groups is determined.

To proceed top-down, the tree can be started from $\ell = 1$, i.e., with all users in one group. To come to the next lower level, we split the group with the worst sum performance into two smaller groups such that it yields the best possible two new groups. We could also opt to divide the group containing the user with the worst performance.

Similarly to the user selection solution, in a second phase the problem of finding the best number of groups to be used is treated. It is often worked around in other scheduling methods such as in [50], where the number of groups is chosen to be equal to the number of highly correlated users simply based on a threshold decision. In [74] the problem is treated by imposing lower and upper bounds on the number of users per group to limit the number of combinations to be tested, thus excluding possible optimum group allocations.

Our algorithm allows us to solve this problem as follows. For example, when the metric $\eta_i^{(\cdots)}$ is an estimate of the achievable user rate, it is possible to choose the number of groups to be equal to the level ℓ that has the best predicted total system rate $\hat{C}_{\mathrm{sys},\ell} = \frac{1}{\ell} \sum_{i=1}^{N_\mathrm{u}} \eta_i^{(\cdots)}$. Here, $\frac{1}{\ell}$ accounts for the fact that ℓ system resources (e.g., TDMA time slots) have to be spent to be able to serve all groups (which is the usual constraint with user grouping). Alternatively, if the metric was chosen as an SNR estimate, we could compute an average SNR for the ℓ-th level and choose the highest level which achieves a target average SNR.

6.3. Tracking and adaptivity

In a real world situation, the scenario evolves gradually due to user movement and changes in the environment. This correlation in time can be exploited to reduce the scheduling complexity by taking new decisions based on previous values, e.g., by learning the optimal SDMA group size. To do so, our algorithm can be modified to avoid running the entire sorting tree again at every time instance. Instead, only some candidate user sets can be considered starting from the previously optimum user set $\mathcal{G}_G^{opt}(n-1,f)$ while restricting the possible update in group size to a small number, e.g., to $\{-1,0,+1\}$:

- Phase 1:
  1. A total of G new candidate sets of size $G-1$ are built from the previously optimal solution $\mathcal{G}_G^{opt}(n-1,f)$ by taking out one of the G users at a time, effectively going downwards one level in the search tree. The user metrics are then calculated and the set with the highest metric sum is kept as $\mathcal{G}_{G-1}^{opt}(n,f)$.
  2. Starting with $\mathcal{G}_{G-1}^{opt}(n,f)$, apply the algorithm from Chapter 6 twice to go up two levels only to $\mathcal{G}_{G+1}^{opt}(n,f)$, effectively updating also the solution for $\mathcal{G}_G^{opt}(n,f)$.
- Phase 2:
  The previous two steps yielded new solutions $\mathcal{G}_{(G-1),G,(G+1)}^{opt}(n,f)$. Pick the best out of these three candidate sets by selecting the one with the highest sum metric if the metric is a rate estimate. Otherwise calculate the precoders and rates for the three candidate sets to be able to identify the best one.

If a user leaves the system, it can be deleted from $\mathcal{G}_G^{opt}(n-1,f)$ before step 1. New users can simply be included in the one-by-one testing of step 2. This updating procedure fixes the number of possible combinations to be tested to three, regardless of the number of users in the system. Because of this rather significant complexity reduction, it seems reasonable to use the final precoding matrices and rates including waterpouring to select the best out of the three combinations rather than the metric. This eliminates the estimation error in this phase of the algorithm and can increase its overall performance. Additionally, one of the three sets of precoding matrices will anyway be used in the transmission and does not represent a complexity increase.

6.4. Joint 3D-Scheduling

Recall that our scheduling metric definition involved the assumption of equal power loading because of the missing knowledge about the eigenmodes. Together with the fact that all subcarriers

are treated as orthogonal, this results in independent scheduling decisions for every subcarrier. However there exists one exemption: The problem is no longer independent if the final selection out of all candidate user sets $\mathcal{G}_G^{opt}(n, f)$ is performed based on the true rate after precoding instead of using the metric. This is due to the fact that for certain precoders such as BD, joint space frequency powerloading (e.g.,waterpouring) must be used to obtain the maximum capacity, which renders the problem 3-dimensional. Also when user QoS requirements are to be considered the decisions should be taken jointly for the entire frequency band. To allow for such a joint treatment we propose the following approach:

For a multicarrier system with N_c subcarriers where the subcarriers can be regarded as orthogonal, the problem can be reduced to a virtual frequency flat system with one single carrier. To do so, a new system consisting of $N_\mathrm{u} \cdot N_\mathrm{c}$ virtual users is formed out of all N_u users' channels on all subcarriers. The best candidate search as well as the tracking algorithm can then be applied to these virtual users instead with f equal to one. However, to take into account the orthogonality of the subcarriers, the algorithm has to be modified to treat virtual users originating from different subcarriers as being not present during the calculation of the user metrics. Note that this does not result in equally sized groups per subcarrier since any candidate grouping can consist of any number of virtual users originating from any subcarrier.

In addition to the possibility of considering QoS aspects for the entire band jointly, this method of scheduling all subcarriers in one run is especially attractive when used together with the time tracking of the search tree as in Section 6.3. The tree becomes much larger in this approach and tracking it in time can reduce the combinations to be tested even further compared to when a separate tree is build on each subcarrier.

7. Complexity of ProSched user selection

In the first part of this section we illustrate the advantage in complexity *ProSched* has over an exhaustive search through all possible user combinations on a certain system resource. Later on, a comparison is drawn to another method from the literature which uses projection matrices in a different fashion.

For simplicity, we look at the effort needed to calculate the rate for a certain precoder in terms of the number of SVDs, matrix multiplications (MM) and the number of capacity calculations (CC) without considering the matrix sizes. We choose this simplification because the complexity of SVDs in terms of multiplications and sums depends on the decomposition algorithm in use and on the desired accuracy. As explained in the following, *ProSched* greatly reduces the number of required SVDs, and therefore the complexity of each SVD is, in absolute terms, not significant when comparing *ProSched* to a brute force search.

For BD precoding (for which *ProSched* was originally conceived), the effort to calculate the resulting rate for one user in a group of size G is $2 \cdot \text{SVDs} + 1 \cdot \text{MM} + 1 \cdot \text{CC}$. Note that the first SVD has to decompose matrices $\tilde{\boldsymbol{H}}$ whose size grows with the size of the groups being tested. The second SVD is a modal decomposition of the resulting smaller channel matrix, see also Chapter 10.1.

An exhaustive search through all possible combinations of N_u users in order to find the best subset for a time slot would require to test $\sum_{G=1}^{R} \binom{N_\text{u}}{G}$ combinations per subcarrier ($R = \text{rank}\{\boldsymbol{H}\}$, where $\boldsymbol{H}$ denotes the combined channel matrix as in Chapter 6).

Due to the BD precoding, the exhaustive search thus results in $\sum_{G=1}^{R} \binom{N_\text{u}}{G} \cdot 2G$ SVDs, i.e., two per user in a group. Consider for example $N_\text{u} = 10$ users (and assume that $R \geq N_\text{u}$), it tests 1023 combinations per subcarrier and requires already 10240 SVDs. As an example we would like to compare this to a variation of our algorithm where the complexity is readily computed, which is the non-tracking (i.e., full tree) algorithm with projection order $p = 1$. Due to the proposed repeated projection metric, no matter how many combinations the search algorithm has to test, it needs to perform only N_u SVDs once at the start of a search (or rank one approximations of them) to obtain the bases of the projectors on all N_u user's null spaces. During the search, any projector into the joint null space of a number of other users can simply be approximated by multiplying the other users' separate null spaces. These decompositions also provide the spatial modes of the separate users' channels. Therefore, the best candidate search tree needs in the first step only N_u MMs to find the equivalent channels and N_u CCs to identify the best user. To find the candidate set of size two, $N_\text{u} - 1$ combinations are tested. In general, to identify the best

group with G users, $(N_\mathrm{u} - G - 1)$ combinations are needed, yielding in total $\sum_{G=1}^{N_\mathrm{u}} (N_\mathrm{u} - G - 1)$ combinations. The complexity can be reduced further by using the recommended time-tracking modification, which tests only 3 group sizes. However, the number of combinations cannot be predicted for this modification, since it depends on the current position in the tree. For a group of size G, $G-1$ MMs are needed to calculate each users' projected channel, thus in total there are $\sum_{G=1}^{N_\mathrm{u}} (G-1) \cdot (N_\mathrm{u} - G - 1)$ MMs required if the whole search tree is needed. A total of $\sum_{G=1}^{N_\mathrm{u}} G \cdot (N_\mathrm{u} - G - 1)$ CCs are needed to calculate the rates of the users in the groups tested.

In the above example with 10 users, the *ProSched* search tree needs to test only 35 combinations - not considering the time tracking modification - and requires only 10 SVDs due to the repeated projection approximation rather than 10240 as in the brute force case. Note that, without the repeated projection approximation, the tree-based search alone would require $\sum_{G=1}^{N_\mathrm{u}} 2G \cdot (N_\mathrm{u} - G - 1) = 220$ SVDs of matrices with increasing size.

Other authors [58], [55] have also developed upon the concept of orthogonal projections which was first used for scheduling in [4]. They use a search algorithm which proceeds similarly in testing the combinations as the one presented here, but computes the exact capacity after precoding during the search. To keep the overall complexity low, they propose a pre-selection step to limit the number of users during the search to an initial subset with cardinality equal to the maximum number of users supportable by the precoder at the base station. This pre-selection step uses the Frobenius norm of the channel projected into the null space of all other users in a group as presented in [4], however computed with a low complexity iterative Gram-Schmidt procedure. Since they do not use the repeated projection approximation, their final user selection thus requires the full 220 SVDs in an example where the number of supportable users is also 10. If $N_\mathrm{u} < 220$ then *ProSched* is clearly less complex, because it requires N_u SVDs only, considering the fact that SVDs dominate the overall computational complexity, and because it does not require a pre-selection phase. For a very large number of users N_u the complexity reduction from the pre-selection step in [58], [59] might outweigh the alternative of using *ProSched* from the start to solve the entire problem. In other words, for a system with a large number of users, a combination of both approaches would possibly be able to achieve the lowest complexity, i.e, first a pre-selection as in [58], [59] and then a final selection with *ProSched*. However, due the iterative nature of computing the projected channels in the pre-selection step as well as due the fact that not all users are treated with the same procedure, it could not be used to provide fairness as *ProSched* can (see Section 5.4 for *ProSched* with fairness).

8. Simulations

Two sets of simulation results are presented in the following. The first one is intended to investigate the relative performance difference between various versions of the proposed algorithm and some reference schemes. With this comparison in mind and also due to limitations in the channel model and simulation setup we provide these results in the form of normalized sum rates.

To give an idea of the performance impact of scheduling in a practical systems we review simulation results obtained within the Wireless World Initiative New Radio (WINNER) project [15] in cooperation with fellow experts, see also the acknowledgements.

8.1. Results based on the IlmProp channel model

To show the potential gain from proper spatial scheduling in simulations, the channel model must be able to reproduce spatial correlation between users. Here we use the geometry-based channel model *IlmProp* [75, 76]. It features realistic correlation in space, time, and frequency as well as realistic far-field antenna radiation patterns. The importance of a full-featured correlation model can be seen from the earlier results presented in [4]. There, an example was shown based on a random channel matrix with identically distributed independent entries where it was visible that hardly any gain can be achieved due to scheduling when there is no correlation in the channel matrix.

Simulation setup

The scenario in use spans an area of 150 m × 120 m and buildings of up to 8 m height are modeled. Up to $N_\mathrm{u} = 18$ users move randomly with speeds of up to 70 km/h, as depicted in Figure 8.1. The users change their heading and speed by a limited amount after a random time interval. The BS mounts a Uniform Circular Array (UCA) with 12 antennas while each mobile has 2 omnidirectional antennas, $\lambda/2$ spaced. The system operates at 2 GHz. The simulation results are shown in Figures 8.2a to 8.4b in form of complementary cumulative distribution functions of the total system throughput for a fixed SNR of 20 dB as well as 90 % outage rate curves. The SNR is defined as total transmit power over total receiver noise variance. The time variant frequency selective channels have been computed for 24 frequency bins, spanning a bandwidth of 1.2 MHz. The coherence bandwidth was estimated to be 8 bins at 0.7 of the maximum correlation.

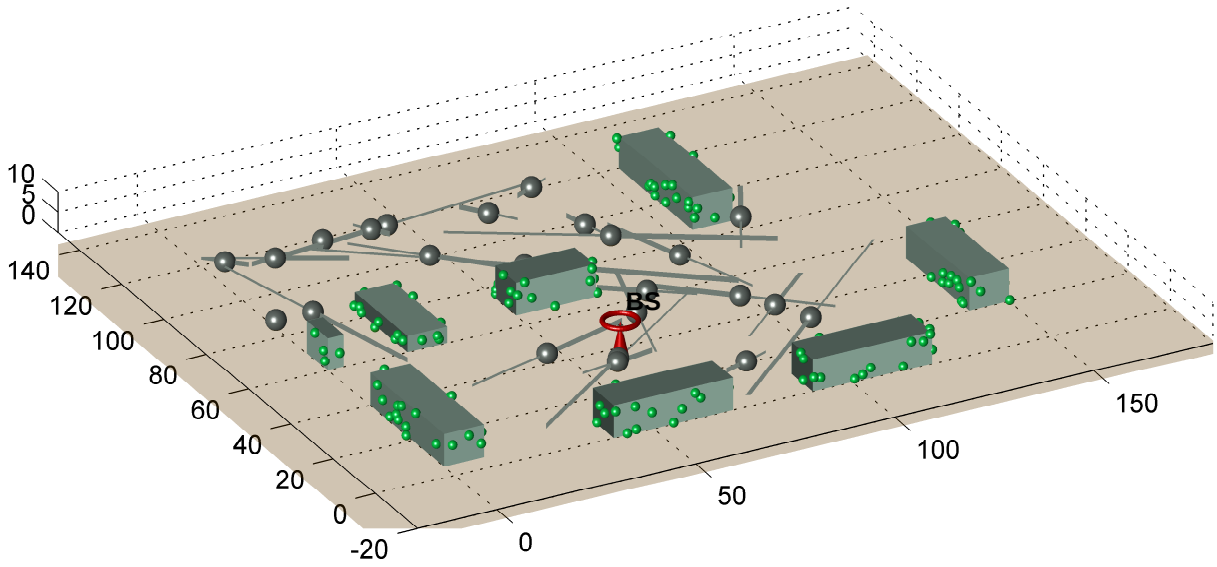


Figure 8.1.: An illustration of the geometry used to generate the channel with the *IlmProp* model for the single BS simulations.

An OFDM symbol duration of 20 μs is assumed without considering the length of the guard period and a TDMA frame consists of 50 OFDM symbols. As precoding schemes we choose one ZF scheme, Block Diagonalization [62], see also Section 10.1 (in Figures 8.2a, 8.3 and 8.4a), and one non-ZF scheme, namely MMSE TxWF [25](in Figures 8.4b and 8.2b). Ideal waterpouring power loading is used to calculate the resulting rates [16] (but not within the scheduling metric of course).

Since the basic algorithm itself is not affected by the OFDM system parameters, we focus on showing relative gains between the algorithm variations in this first set of simulations. To do so, all plots have been normalized to the maximum sum rate value attained in the TDMA case shown in the first figure for BD precoding.

Figures 8.2a to 8.3 correspond to simulations performed on a single frequency bin, thus assuming a narrowband system operating on a frequency flat channel. Consequently the (normalized) sum rates are expressed in bps/Hz. To illustrate the effectiveness of our algorithm it is necessary to start with single subcarrier simulations for one SNR value only, because we want to use as a reference the maximum rate achievable by exhaustively searching through all possible user subsets, which is computationally too demanding to perform it on multiple subcarriers. To facilitate the relative comparison of the normalized rates we also show the maximum achievable sum rate under a sum power constraint (or DPC bound, which equals the downlink capacity region [24]). To obtain these figures we have used the iterative algorithm from [29] which solves the problem for the dual uplink. (Note that, at the time when the results were obtained, only

frequency flat versions of such algorithms were readily available in the literature.)

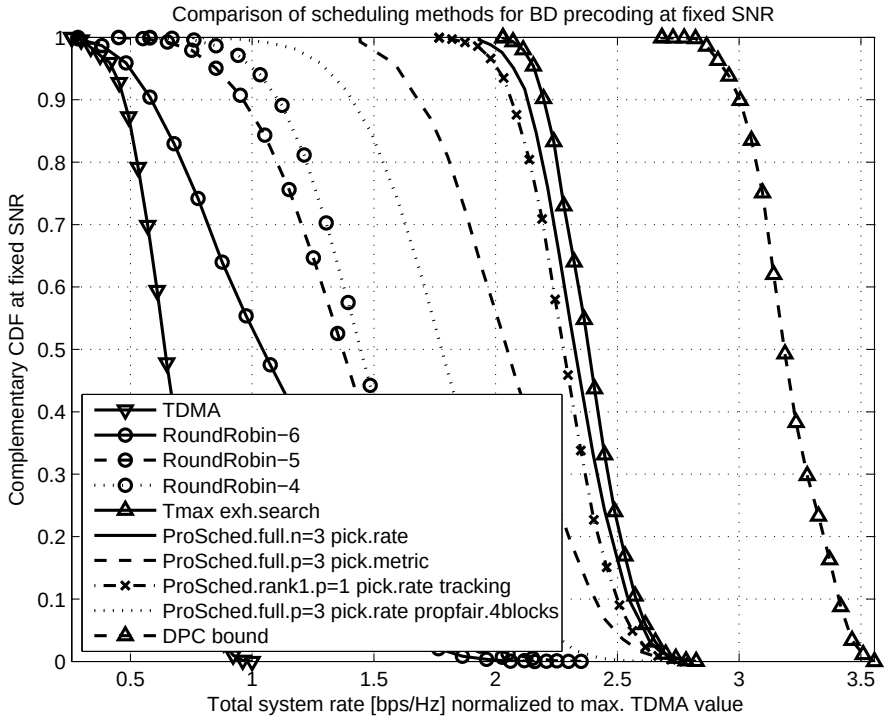
The 90 % outage curves in Figures 8.4a and 8.4b, on the other hand, are computed for the whole 1.2 MHz band, and the (normalized) rates are thus expressed in bits after multiplying the capacity with the number of subcarriers divided by the duration of an OFDM symbol. The proposed algorithm is also compared to pure TDMA and to SDMA using a so-called Round Robin (RR) scheduler. The RR scheme re-schedules every time slot by cycling through the N_{u} available users. The number after the letters RR denotes how many users are to be scheduled at every time slot. The 12-element array at the base station can spatially multiplex up to 6 users, with two antennas each. For instance, an RR-5 scheme would schedule the following users out of $N_{\mathrm{u}} = 12$ for successive time snapshots: $\{1, 2, 3, 4, 5\}$, $\{6, 7, 8, 9, 10\}$, $\{11, 12, 1, 2, 3\}$, etc. Note that in the case of the frequency selective channel (Figures 8.4a and 8.4b), the RR solution is applied to all subcarriers. Although the system supports up to RR-6 we compute RR-5 and RR-4 as well because in smaller SDMA groups, greater fractions of the available power can be assigned to the group members.

Discussion

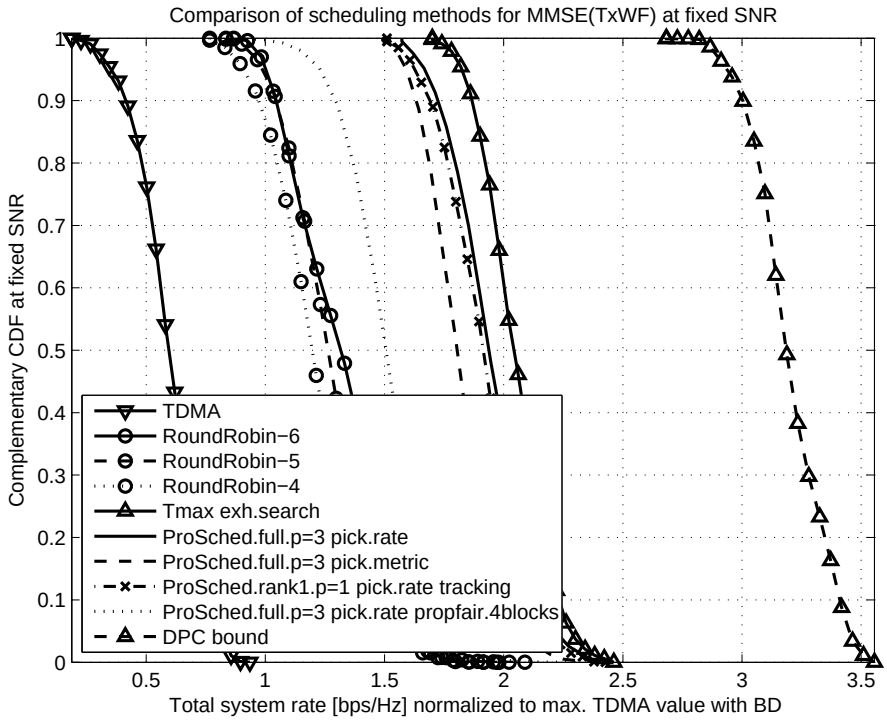
In Figures 8.2a and 8.2b we show the performance of the proposed algorithm in different modifications: The variation of the projection based scheduling algorithm (*ProSched*) displaying the best performance uses the repeated projection metric calculated with full rank basis matrices and projection order $p = 3$ to select the candidate user sets while the final set is then selected by the exact rate (denoted as `ProSched.full.p=3 pick.rate`). Its performance is comparable to that of an exhaustive search through all combinations. Of course it would reduce complexity to use the scheduling metric to perform the final subset selection step instead of the true rate. However, it can be seen that switching to `pick.metric` the performance decreases noticeably. This suggests that the final selection step is especially sensitive to estimation errors in the metric. Instead, the time tracking modification should be used to reduce the number of final candidate sets to three and the selection should be performed based on the true rate. Thus, the *ProSched* version which offers the best trade-off between complexity and performance uses basis matrices of rank one only and order 1 repeated projections together with the tracking algorithm and selection of the final set based on the true rate (`ProSched.rank1.p=1 pick.rate tracking`).

In any case, *ProSched* outperforms a non-intelligent RR scheduler in terms of sum rate. As a side result it can be seen once again from the RR results that the sum rate depends on the group size, which was already observed in [49]. However, one must not conclude from these graphs that decreasing the SDMA group size monotonically leads to an increased sum rate: A further reduction in group size resulted in a decrease in performance (not shown), as can be imagined based on the TDMA case.

The effect of the proportional fairness extension proposed in Section 5.4 is also shown in both



(a) Frequency flat case, BD precoding, $N_u = 18$ users



(b) Frequency flat case, MMSE(TxWF) precoding, $N_u = 18$ users

Figure 8.2.: Performance of various versions of *ProSched* with different complexity compared to the references Dirty Paper Coding bound, exhaustive search, Round Robin and TDMA

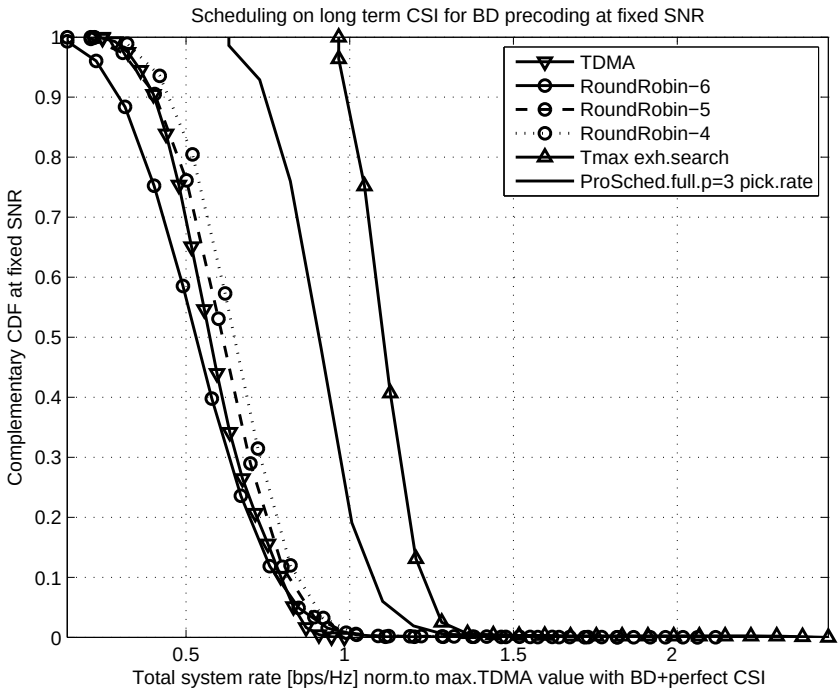
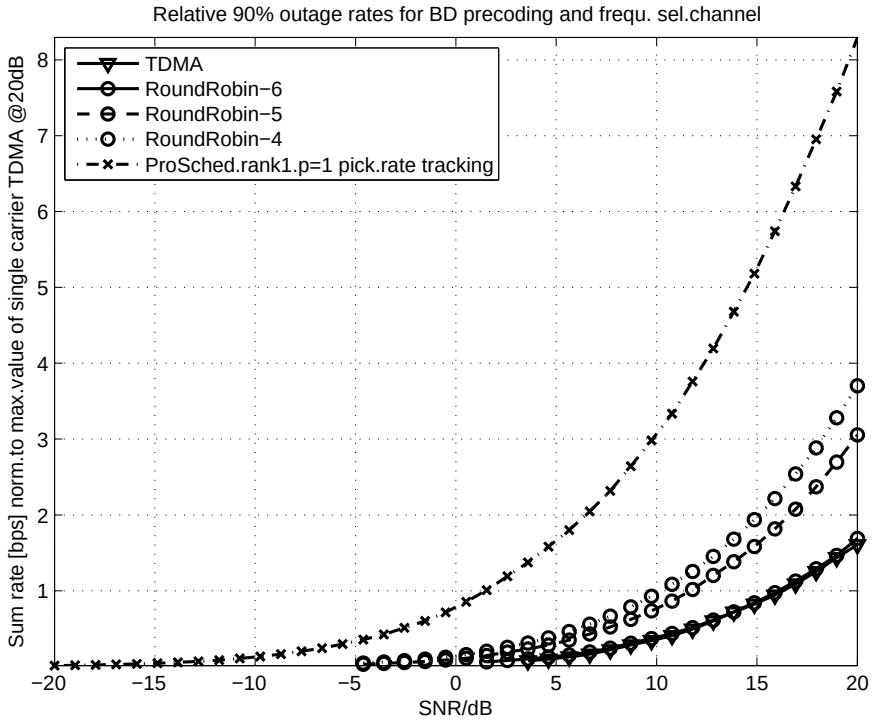


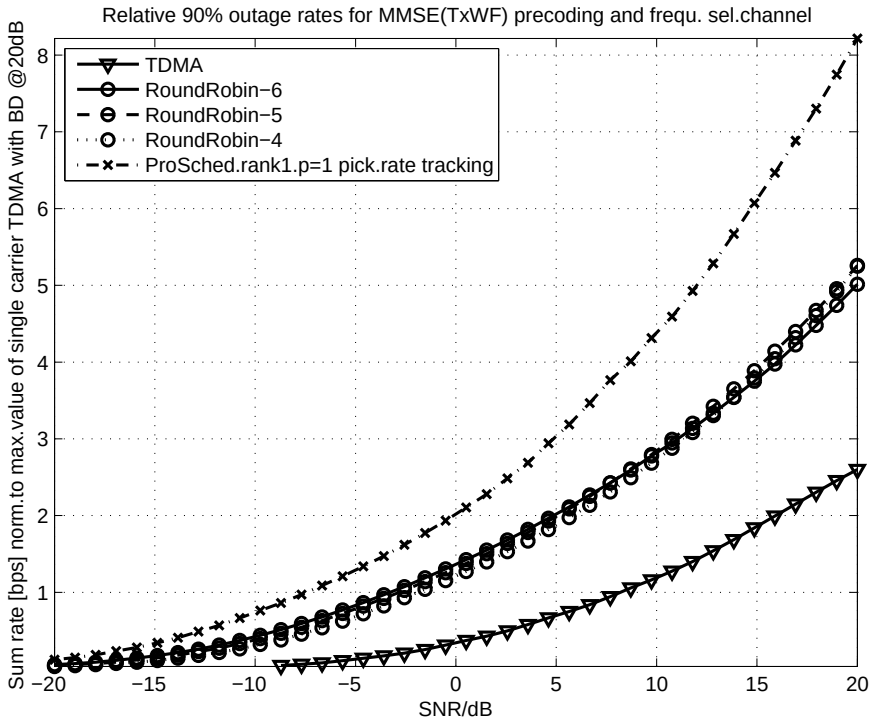
Figure 8.3.: Scheduling performed on averaged covariance matrix knowledge with an averaging window of five frames, frequency flat case, BD precoding, $N_u = 18$ users

figures (see the curves labeled `propfair.4blocks`). Following the definition of proportional fairness, which is based on a sum of logarithmic rates, it should express itself in an increase of the product of all users' rates for every time slot. For the frequency flat case at 20 dB where the long term metrics were tracked over 4 frames it was able to increase the rate products at every time instance on average by a factor of 2.3 for the optimum algorithm and MMSE(TxWF). In [41] it is explained that fairness also depends on the underlying precoding scheme and should, therefore, not be measured by one number only. Instead, it is proposed to analyze a system by studying plots of rate mean versus standard deviation, where fairness would result in a high mean and low standard deviation. An extensive study of fairness issues is beyond the scope of this work but in the next section more results with fairness are shown which illustrate the behavior of proportional fairness when compared to maximum throughput scheduling in terms of average user rates.

In Figure 8.3 we show some results for scheduling based on covariance matrices as long-term channel knowledge averaged over five frames using a rectangular window. From the RR curves it can be seen that the impact of the group size has decreased. Also, the performance gap between the exhaustive search and RR has decreased. Our proposed algorithm still outperforms the RR scheme significantly.



(a) BD precoding



(b) MMSE(TxWF) precoding

Figure 8.4.: Frequency selective case, 24 OFDM subcarriers, $N_u = 12$ users: 90% outage of the total system rate for a low complexity version of the scheduling algorithm using the Joint 3D-Scheduling extension from Section 6.4 compared to TDMA and RR.

In Figures 8.4a and 8.4b we show 90 % outage rates in the frequency selective case using the Joint 3D-Scheduling extension from Section 6.4. For complexity reasons, only the variation of the scheduling algorithm offering the best complexity versus performance trade-off was simulated (`ProSched.rank1.p=1 pick.rate tracking`), because the relative performance compared to more complex versions can be judged from the figures for the frequency flat case. The conclusions remain the same as in the frequency flat case, i.e., that already the low complexity version greatly outperforms the RR scheduler. Furthermore, it can be seen that a gain from spatial scheduling is possible in the entire simulated SNR range, whereas the biggest improvement is possible in the medium to high SNR range.

8.2. Results in the context of the WINNER 4G system proposal

The first two Figures 8.5 discussed in this section use the channel model developed within the first half of the WINNER project [65] whereas the second example is based on the WINNER II channel model [66] in a version as published in [77]. The fundamental difference between the *IlmProp* used in the previous Section and the WINNER model is that, in the latter physical effects are rendered as realizations of stochastic processes. The process parameters have been obtained from extensive measurement campaigns whereas in the *IlmProp* model only a limited tuning to existing path loss models is possible. To that extend, the absolute throughput values obtained with the WINNER model can be considered more significant. On the other hand, the abilities to produce settings with realistic correlation is far more limited in the WINNER model compared to the *IlmProp* due to the fact that each link is rendered in its very own independent space and correlation is introduced only by the antenna arrays.

The only similarity of the two approaches lies in the fact that the WINNER model is also based on a geometry in space which, however, cannot evolve continuously in time. Because of this aspect, running a proportional fair algorithm implementation in time direction would not produce meaningful results. The proportional fair *ProSched* algorithm was, thus, changed to perform averaging along frequency at each time instance instead of along the time axis and the three-dimensional (3D) extension was, consequently, not used.

Simulation setup

The results were obtained with the help of a so-called link-to-system level interface [78, 30]. Its purpose is to reproduce realistic throughput figures on a system scale with multiple users without the need to simulate link level coding and modulation procedures. This is achieved by computing SNRs after precoding, converting them into mutual information which can then be averaged over several system resources and converted back and fed into a look-up table to obtain throughput estimates.

The fact that realistic Modulation and Coding Schemes (MCSs) are emulated with this method limits the maximum attainable bit rate per frequency resource. To take this into account in the scheduler, *ProSched* can be modified by capping the projection based rate estimates to the very same maximum. Such modifications intended to match a Shannon rate curve to a practical MCS are often referred to as Shannon fitting [79]. Only in the second example such a capping was implemented in the scheduling metric, namely to 4.8bps/Hzw to reflect the maximum spectral efficiency attainable with 64 Quadrature Amplitude Modulation (QAM) modulation and rate 4/5 coding. Nevertheless, the scheduling algorithm worked well enough in the first case which suggests that in this scenario a capping is not crucial during the scheduling - probably because the rate estimates hardly reach the limit imposed by the MCS.

It should also be noted that each user employs a realistic linear MMSE receiver computed from an estimate of the equivalent channel including the precoding matrix, whereas the scheduler continues to work with rate based estimates which do not include the receiver, but can be considered an upper bound that is more or less approached by different receivers. Yet still, the scheduling decision is accurate enough.

The scenario used in both cases is the Indoor Local Area Office. An isolated cell is simulated, consisting of 40 light wall rooms of size 10 m × 10 m, separated into three rows by two corridors (see, e.g., [30]).

For the results in Figures 8.5 the BS is placed in the middle corridor and mounts a Uniform Linear Array (ULA) with six antennas. This reflects a system proposal of an earlier project phase, as described in the earlier report [80]. A total of six stationary users are present, having two cross-polarized antennas each. The precoding algorithm is RBD.

In the simulation setup used for Figures 8.6 the BS has a total of 32 antennas in the form of four remote radio heads, according to later recommendations summarized in [30]. In Part III, such a placement is referred to as distributed antennas and some of the advantages are mentioned. The total number of users is increased to 50 with a uniform distribution. Both SMMSE [35] and RBD [36] precoding have been simulated.

Scheduling and precoding is carried out once per resource element or chunk as it is called in WINNER. It uses a pseudo channel matrix as CSI which is obtained with the help of a covariance matrix averaging method discussed in Section 14.2, based on estimates obtained with a sparse uplink pilot grid. In the second set of simulations, a chunk is 8 subcarriers wide and 15 OFDM symbols long [30] which is half as wide as in the older simulations based on an earlier version of [80].

Several impairments have been included using models described in the above references: Calibration errors between uplink and downlink transmitter and receiver RF front ends are modeled according to the self-calibration approach of [81]. They result in a degradation of CSI that has

been measured with uplink pilots when it is re-used for the downlink precoding at the BS. Both the estimates of the effective channels $\boldsymbol{HM}$ needed for tuning the users' spatial receive filters as well as the uplink estimate contain residual errors and the required downlink pilot overhead is accounted for in the throughput.

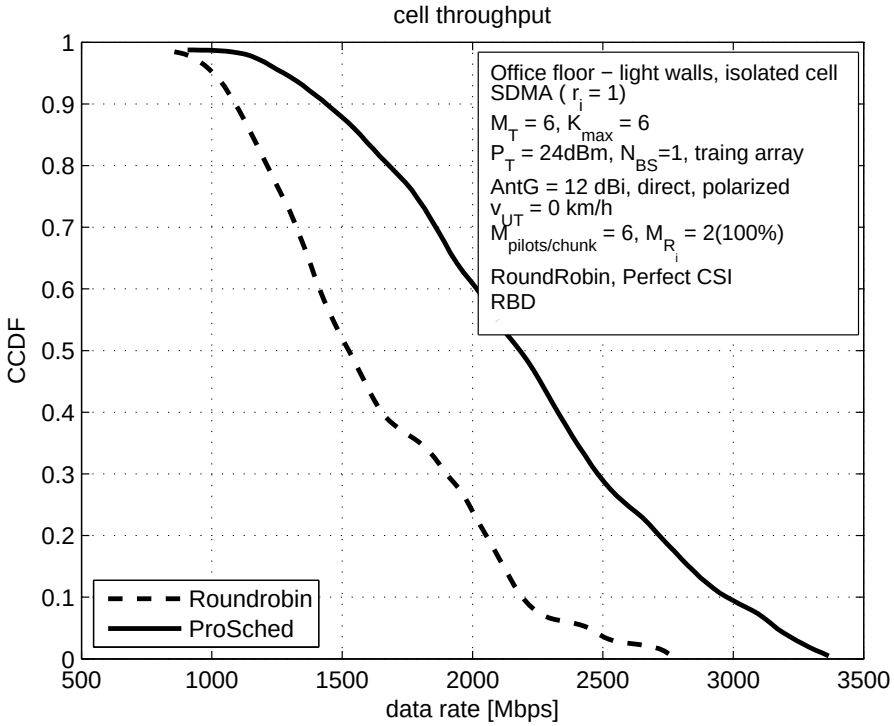
Discussion

Figure 8.5a shows the Complementary Cumulative Distribution Function (CCDF) of the estimated total cell throughput for *ProSched* maximum throughput scheduling versus RR scheduling. The *ProSched* algorithm version used is the one recommended in the previous section. An increase in throughput of up to 1.5 is visible at the median and 1.35 at 90 % outage. The increase is less significant than in the *IlmProp* simulations because of the limited correlation modeling capabilities of the WINNER model and also because of the additional impairments simulated.

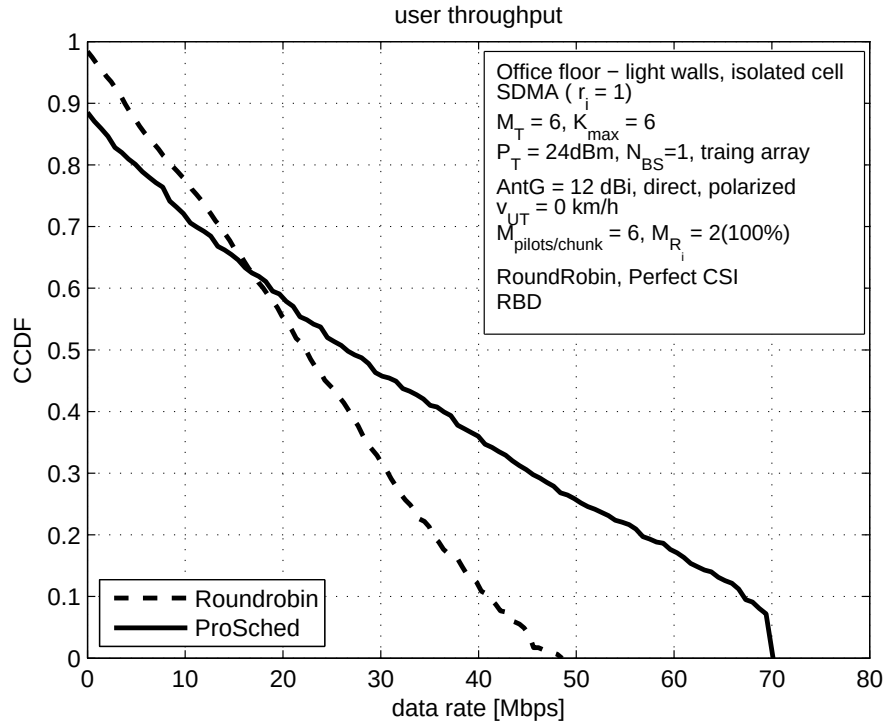
A CCDF of the average user rate is shown in Figure 8.5b. Here, it is especially noteworthy how maximum throughput scheduling shifts the histograms towards higher rates at the expense of users not being served. This expresses itself in a decrease of probability for the lower rates when compared to RR.

In the next set of curves in Figure 8.6, the effect of *ProSched* with proportional fairness is compared to that of the maximum throughput scheduling version for both RBD and SMMSE precoding.

Especially in the user rate CCDF (Figure 8.5b) it can be seen again how proportional fairness increases the probability of users being served at the expense of peak data rate. In the cell throughput (Figure 8.5a), a large gap between TDMA and SDMA with proper scheduling is visible.

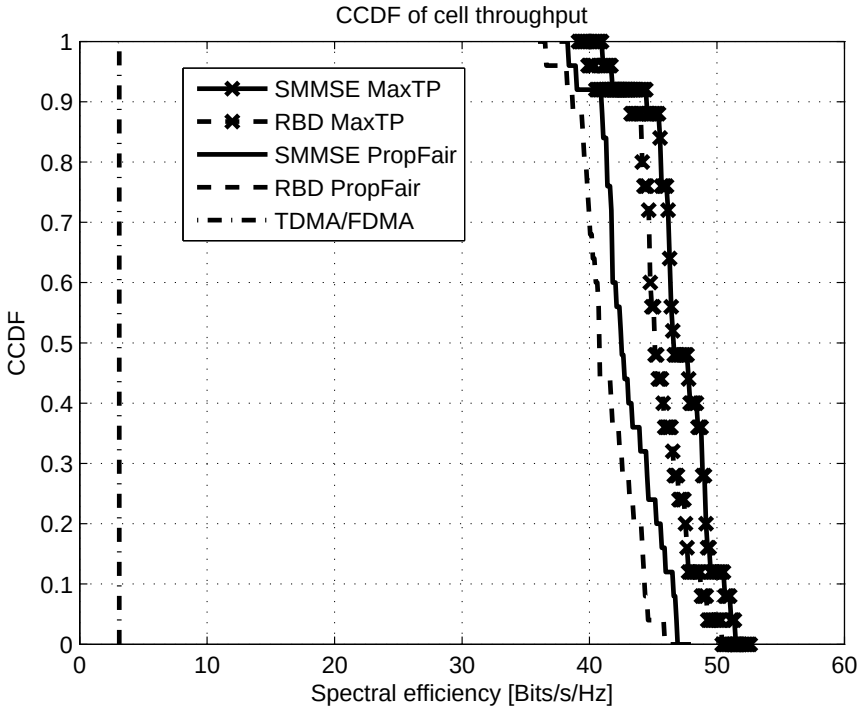


(a) CCDF of total cell throughput

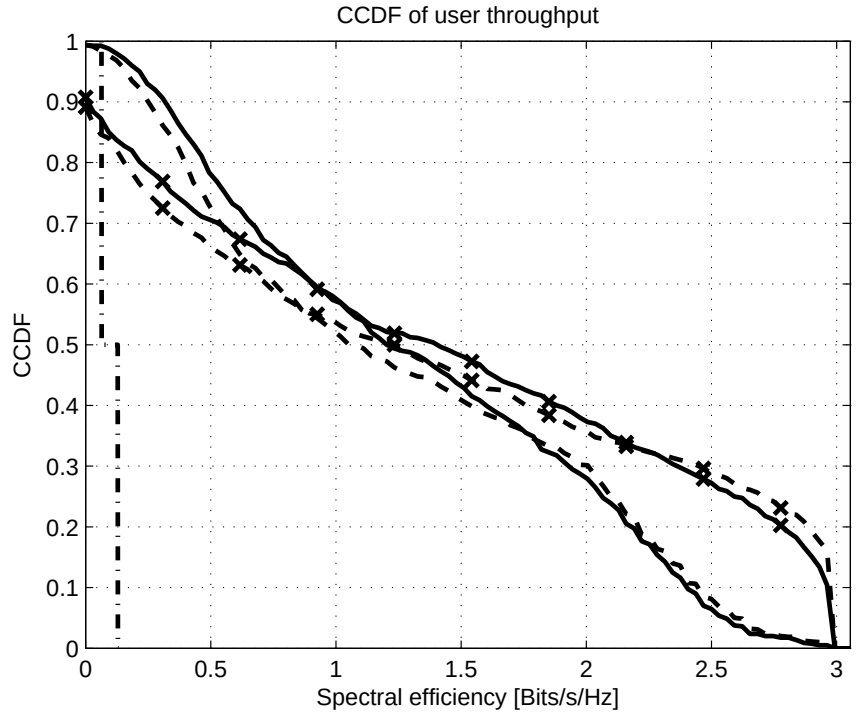


(b) CCDF of average user rate

Figure 8.5.: WINNER project system level performance of *ProSched* versus RR scheduling with RBD precoding. These figures were first presented in [8] (see acknowledgements section).



(a) CCDF of total cell throughput



(b) CCDF of average user rate

Figure 8.6.: WINNER II project system level performance of *ProSched* maximum throughput scheduling versus *ProSched* with proportional fairness. These results were first presented in an internal report of the WINNER II project (see acknowledgements section).

9. Conclusions

In the downlink of MIMO systems with SDMA, scheduling is absolutely required to prevent the huge performance losses due to spatially correlated users and to fully exploit the gains offered by multi-user diversity. In the simulations presented here, the system performance gap between non-intelligent SDMA with Round Robin scheduling and the presented novel reduced complexity scheduling algorithm *ProSched* is in some situation as high as a factor of 3.5 at 90 % outage. *ProSched* is based on a novel way of interpreting the process of precoding with the help of orthogonal projectors, on an efficient search algorithm, and on efficient approximation techniques. It reduces the effort of the precoder calculations to an amount comparable to a situation where every user was alone. Thus it renders scheduling practical for realistic systems with a large number of terminals while offering a similar performance as an exhaustive search through all possible user combinations. The latter would require the unfeasible calculation of all precoding matrices for all possible user combinations on all subcarriers. Even for closed form scheduling solutions such a high complexity is expected, because the result of all precoder combinations would still be needed in order to know all channel quality indicators.

The presented algorithm can treat space, time, and frequency jointly. It not only considers the spatial correlation but inherently also the degree of freedom in choosing the right SDMA group size. The solution is not restricted to a specific type of spatial precoder and can be applied to systems with long-term channel knowledge based on second order statistics only. We have also discussed how to include proportional fairness and QoS into the algorithm.

Uplink discussion

One may tend to argue that a user selection which was found to be beneficial for the downlink should also be beneficial for the uplink, based on the results from [22, 82] which show that the achievable DPC downlink rate region equals the uplink capacity region with non-cooperating transmitters. These results allow to find the DPC rate region by first finding the optimum transmit weights and power allocations for the dual uplink where the rates are tractable convex functions of the individual transmit correlation matrices. During a conversion back to the dual downlink [28], users with zero power remain inactive, or in other words the same user selection is applied. In practical systems, however, a sum power constraint, which is a key point in the above duality argumentation, usually does not exist explicitly in the uplink. To find the optimum uplink transmit covariance matrices when successive decoding is used, instead of a sum power constrained optimal solution [29] an algorithm with individual user power constraints is to be applied [83]. Both approaches have in common that each user's transmit covariance matrix is

optimized iteratively while considering all other users' signals as noise. A water filling power allocation is performed on the resulting modes after each iteration.

An implicit uplink sum power constraint exists due to the fact that only a limited number of spatial streams or users can be efficiently spatially separated at the receiver. To that extend, re-using the selection found by downlink *ProSched* might be possible, provided that the sum power constraint is replaced by individual uplink power constraints.

To be able to re-use the downlink *ProSched* solution it would be desirable from the above theoretical point of view to derive how the orthogonal projection based ZF downlink precoding approximations relate to the respective optimal uplink transmit covariance matrices (or transmit antenna weights) - presuming that one wants to deploy precoding on the uplink in a practical system.

If precoding on the uplink is to be used, the recent solutions in [84] are of interest. The authors derived that adding a greedy user exclusion step to the iterative waterfilling is near-optimal and still offers a significant complexity reduction compared to the optimal solution of [83], especially when the precoding algorithm is constrained to rank one transmissions (i.e., beamforming).

From the computational complexity point of view, however, it might be desirable to avoid any costly channel adaptive transmit processing at mobile stations. This view has also been adopted in the WINNER project [31, 32, 33]. In this case, other previously published downlink scheduling approaches which do not inherently take into account power, such as approaches based on DOA and correlation measures, have been applied successfully to the uplink, e.g., in [85]. They do, however, not offer the inherent possibility to take into account QoS constraints. With this in mind, the recent solutions in [86] are more interesting because they are based on rates. With the help of a receive side correlation channel model the authors provide computationally efficient Taylor series approximations of the ergodic sum capacity which can then be used for scheduling.

To adopt *ProSched* in the case without transmit precoding seems to be more tractable, since a ZF receiver at the BS could again be used as a boundary condition for other receivers, thus offering a possibility of applying a projection based notation of the sum rate.

10. Derivations

10.1. The Block Diagonalization Algorithm expressed with an Orthogonal Projection

A formulation for BD precoding is shown in this section based on the concept of orthogonal projections as first derived in [5]. It is in contrast to the original formulation of BD in [62] which uses coordinate transformations, as explained in the following.

The channel of user $g \in G$ is denoted $\boldsymbol{H}_g \in \mathbb{C}^{M_{R,g} \times M_T}$ as before. It is assumed that the dimensionality constraint discussed in [62] for the original BD precoding algorithm is satisfied, i.e., $M_\mathrm{T} \geq M_\mathrm{R} = \sum_{g=1}^{G} M_{R,g}$ for the G users in one SDMA group. A diagonal power loading matrix $\boldsymbol{D}_g$ is factored out from the sought precoding matrix. It contains the fractions of the total transmit power assigned to each of user g's spatial modes. The remaining part has normalized columns and is denoted as $\boldsymbol{N}_g$, i.e., $\boldsymbol{M}_g = \boldsymbol{N}_g \boldsymbol{D}_g$.

In the original algorithm, the precoding matrix $\boldsymbol{N}_g^{\mathrm{BD}}$ is obtained in two steps as follows:

1. Let $\tilde{\boldsymbol{H}}_g = \begin{bmatrix} \boldsymbol{H}_1^\mathrm{T} & \cdots & \boldsymbol{H}_{g-1}^\mathrm{T} & \boldsymbol{H}_{g+1}^\mathrm{T} & \cdots & \boldsymbol{H}_G^\mathrm{T} \end{bmatrix}^\mathrm{T}$ contain the channels of all users except the gth. Compute an SVD

$$\tilde{\boldsymbol{H}}_g = \tilde{\boldsymbol{U}}_g \tilde{\boldsymbol{\Sigma}}_g \tilde{\boldsymbol{V}}_g^\mathrm{H} = \tilde{\boldsymbol{U}}_g \tilde{\boldsymbol{\Sigma}}_g \begin{bmatrix} \tilde{\boldsymbol{V}}_g^{(1)} & \tilde{\boldsymbol{V}}_g^{(0)} \end{bmatrix}^\mathrm{H} \tag{10.1}$$

   where $\tilde{\boldsymbol{V}}_g^{(1)}$ contains the first $\tilde{r} = \operatorname{rank}\left\{\tilde{\boldsymbol{H}}_g\right\}$ columns of $\tilde{\boldsymbol{V}}_g$. Then $\tilde{\boldsymbol{V}}_g^{(0)} \in \mathbb{C}^{M_\mathrm{T} \times (M_\mathrm{T} - \tilde{r})}$ is a basis of the intersection of all other users' null spaces.

2. Introduce an effective channel $\boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)}$. Its SVD is:

$$\boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} = \boldsymbol{U}'_g \boldsymbol{\Sigma}'_g \boldsymbol{V}_g'^\mathrm{H} = \boldsymbol{U}'_g \boldsymbol{\Sigma}'_g \begin{bmatrix} \boldsymbol{V}_g'^{(1)} & \boldsymbol{V}_g'^{(0)} \end{bmatrix}^\mathrm{H}, \tag{10.2}$$

   where $\boldsymbol{V}'_g \in \mathbb{C}^{(M_\mathrm{T} - \tilde{r}) \times (M_\mathrm{T} - \tilde{r})}$ and where $\boldsymbol{V}_g'^{(1)} \in \mathbb{C}^{(M_\mathrm{T} - \tilde{r}) \times r}$ contains the first $r = \operatorname{rank}\{\boldsymbol{H}_g\} = \operatorname{rank}\left\{\boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)}\right\}$ columns corresponding to the non-zero singular values.

   Choose the normalized part of the precoding matrix as $\boldsymbol{N}_g^{\mathrm{BD}} = \tilde{\boldsymbol{V}}_g^{(0)} \boldsymbol{V}_g'^{(1)} \in \mathbb{C}^{M_\mathrm{T} \times r}$.

Alternatively, the precoding matrix can be obtained with the help of an orthogonal projection and the following algorithm where step one stays the same as for BD:

2. Introduce an effective channel $\mathring{\boldsymbol{H}}_g = \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)\mathrm{H}} = \boldsymbol{H}_g \tilde{\boldsymbol{P}}_g^{(0)} \in \mathbb{C}^{M_{R,g} \times M_\mathrm{T}}$ where $\tilde{\boldsymbol{P}}_g^{(0)} \in \mathbb{C}^{M_\mathrm{T} \times M_\mathrm{T}}$ is an orthogonal projection into the intersection of the null spaces of all other users' channel matrices. Its SVD is then:

$$\boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)\mathrm{H}} = \boldsymbol{U}_g'' \boldsymbol{\Sigma}_g'' \boldsymbol{V}_g''^{\mathrm{H}} = \boldsymbol{U}_g'' \boldsymbol{\Sigma}_g'' \left[\begin{array}{c} \boldsymbol{V}_g''^{(1)} \boldsymbol{V}_g''^{(0)} \end{array} \right]^{\mathrm{H}}, \tag{10.3}$$

where $\boldsymbol{V}_g''^{(1)} \in \mathbb{C}^{M_\mathrm{T} \times r}$ contains again the first r columns.

Choose $\boldsymbol{N}_g^{\mathrm{BDP}} = \boldsymbol{V}_g''^{(1)}$, resulting in the same dimensionality as for BD.

In the following it is shown that the precoding matrices obtained by the two algorithms are identical without the powerloading part. As a consequence, the power loading matrix will also be identical and can, therefore, be skipped in the explanations:

The effective channel $\mathring{\boldsymbol{H}}_g$ introduced in step two of the BD with projection (BDP) is the result of the projection of user g's channel into the intersection of the null spaces of all other users' channel matrices. Using the notation introduced in the algorithm descriptions, it can now be represented by two separate SVDs. The reduced form SVDs are used here, where the $\boldsymbol{V}$ matrices only contain the first r columns associated with the non-zero singular values. These two SVDs already include the two modulation matrices obtained by the different algorithms:

$$\begin{aligned} \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \cdot \tilde{\boldsymbol{V}}_g^{(0)\mathrm{H}} &= \\ \boldsymbol{U}_g' \boldsymbol{\Sigma}_g' \boldsymbol{V}_g'^{(1)\mathrm{H}} \cdot \tilde{\boldsymbol{V}}_g^{(0)\mathrm{H}} &= \boldsymbol{U}_g'' \boldsymbol{\Sigma}_g'' \boldsymbol{V}_g''^{(1)\mathrm{H}} = \mathring{\boldsymbol{H}}_g \\ = \boldsymbol{U}_g' \boldsymbol{\Sigma}_g' \boldsymbol{N}_g^{\mathrm{BD}^\mathrm{H}} &= \boldsymbol{U}_g' \boldsymbol{\Sigma}_g' \boldsymbol{N}_g^{\mathrm{BDP}^\mathrm{H}} \end{aligned} \tag{10.4}$$

Lines two and three show the equality of the two precoding matrices except for a possible ambiguity in the calculation of an SVD. It holds because of the following: Multiplying $\tilde{\boldsymbol{V}}_g^{(0)\mathrm{H}}$ from the right to $\boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)}$ yields a result in the same column space. Therefore, $\boldsymbol{U}_g'$ remains unchanged. It represents the column space of $\mathring{\boldsymbol{H}}_g$ and so does $\boldsymbol{U}_g''$. Furthermore, $\boldsymbol{N}_g^{\mathrm{BD}^\mathrm{H}}$ is orthonormal, because it is the product of two orthonormal matrices. It is, therefore, a candidate for the basis of the row space of $\mathring{\boldsymbol{H}}_g$ and so is $\boldsymbol{V}_g''^{(1)\mathrm{H}}$. Both SVDs represent the same matrix $\mathring{\boldsymbol{H}}_g$ leading to the final conclusion that $\boldsymbol{\Sigma}_g' = \boldsymbol{\Sigma}_g''$.

Norm of the equivalent channel

For BD precoding, the Frobenius norm squared of user g's precoded equivalent channel equals the norm of its channel projected into the joint null space of all other users channel matrices

$$\|\boldsymbol{H}_g \boldsymbol{M}_g\|_\mathrm{F}^2 = \left\| \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)\mathrm{H}} \right\|_\mathrm{F}^2 = \left\| \boldsymbol{H}_g \tilde{\boldsymbol{P}}_g^{(0)} \right\|_\mathrm{F}^2 . \tag{10.5}$$

It can be shown using the precoding matrix from the original BD formulation without the power loading matrix, $\boldsymbol{N}_g^{\mathrm{BD}} = \tilde{\boldsymbol{V}}_g^{(0)} \boldsymbol{V}_g'^{(1)}$ (which is equivalent to the precoding matrix from the new formulation using projections). It holds that

$$\begin{aligned} \left\| \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \boldsymbol{V}_g'^{(1)} \right\|_{\mathrm{F}}^2 &= \mathrm{trace} \left\{ \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \boldsymbol{V}'^{(1)} \boldsymbol{V}'^{(1)^{\mathrm{H}}} \tilde{\boldsymbol{V}}^{(0)^{\mathrm{H}}} \boldsymbol{H}_g^{\mathrm{H}} \right\} \\ &= \mathrm{trace} \left\{ \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}^{(0)^{\mathrm{H}}} \boldsymbol{H}_g^{\mathrm{H}} \right\} \\ &= \left\| \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \right\|_{\mathrm{F}}^2 \end{aligned} \tag{10.6}$$

which comes from the fact that $\boldsymbol{V}'^{(1)} \boldsymbol{V}'^{(1)^{\mathrm{H}}}$ projects $\boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)}$ into its very own row space of which $\boldsymbol{V}'^{(1)}$ is a basis (see equation (10.2)) and has thus no effect.

Furthermore, $\left\| \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \right\|_{\mathrm{F}}^2 = \left\| \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)^{\mathrm{H}}} \right\|_{\mathrm{F}}^2$ because

$$\left\| \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)^{\mathrm{H}}} \right\|_{\mathrm{F}}^2 = \mathrm{trace} \left\{ \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)^{\mathrm{H}}} \left[\boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)^{\mathrm{H}}} \right]^{\mathrm{H}} \right\} \tag{10.7}$$

$$= \mathrm{trace} \left\{ \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)^{\mathrm{H}}} \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)^{\mathrm{H}}} \boldsymbol{H}_g^{\mathrm{H}} \right\} \tag{10.8}$$

$$= \mathrm{trace} \left\{ \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \tilde{\boldsymbol{V}}_g^{(0)^{\mathrm{H}}} \boldsymbol{H}_g^{\mathrm{H}} \right\} \tag{10.9}$$

$$= \left\| \boldsymbol{H}_g \tilde{\boldsymbol{V}}_g^{(0)} \right\|_{\mathrm{F}}^2 . \tag{10.10}$$

Part III.

ProSched for systems with multiple transmitting stations

11. Introduction

We consider again the downlink of a Multiple Input Multiple Output (MIMO) Orthogonal Frequency Division Multiplexing (OFDM) mobile communications system. An introduction to MIMO OFDM can be found in Part I. Here we deal with systems with multiple transmitters in which traffic of adjacent transmitting stations is not (or not exclusively) separated by other means such as orthogonal spreading codes or a fixed bandwidth partitioning.

The transmitting stations are assumed to be connected via a backbone network or other fast link rendering them able to coordinate their traffic or have it organized by a central controller. Scheduling solutions for three different coordination strategies will be developed in this part of the thesis, whereas MIMO scheduling for a single transmitting station is treated in II.

The first coordination strategy is often called distributed MIMO [87, 9] and has recently attracted considerable interest. It implies that all spatial processing is done jointly as if all Base Station (BS) antennas were located at one device and that, consequently, there does not exist any inter-site interference between the combined BS. Some more theoretical background on distributed antenna systems is reviewed in Section 12.2.

In the second strategy, the BSs only coordinate the scheduling and not the spatial processing. New extensions to the low complexity scheduling approach *ProSched* from Part II are derived to be able to estimate the interference originating from other base stations during the scheduling process. If coordination is possible between adjacent sectors or cells, such a low complexity interference estimation approach is an alternative to deploying orthogonal cells or orthogonal sectors and could help increase frequency and code reuse on the downlink.

The third strategy is that of Relay Enhanced Cells (RECs) where again coordination is assumed while data can be transported over multiple hops. It is treated in a separate Chapter 16 where also more background information on MIMO RECs is given.

Apart from the interference problem, the implications of the spatial domain on the terminal scheduling process are the same as in Part II: Terminals with highly spatially correlated channels must not be assigned to the same Space-Division Multiple Access (SDMA) group. High spatial correlation severely degrades the possible throughput of any SDMA scheme, because it impairs the spatial separability of the terminals, leading to less efficient precoding matrices and to increased interference. Additionally, the size of a group has a significant impact on the overall performance, because the SDMA throughput gain comes at the expense of serving each user with a smaller fraction of the available transmit power.

The *ProSched* approach [4, 5] considers all of the above issues by using a low complexity

estimate of a user's channel capacity as scheduling metric. It is applicable in the case where channel knowledge is used for SDMA at the transmitter, i.e., no additional channel quality feedback is required. The result of the precoding depends on the user selection, but *ProSched* reduces the effort of estimating it by using the concept of orthogonal projections to an effort comparable to when a user is served alone. Additionally, the complexity of testing all possible terminal combinations for each SDMA group is avoided with the help of a tree-based search algorithm.

The multiple transmitter extensions consist of two modifications originally presented in [10]. The first one is to extend the per-user scheduling metric by an estimate of the total received inter-cell interference power at each terminal, see Chapter 12. This estimate is obtained using the already available orthogonal projection matrices and requires only matrix multiplications but no additional matrix decompositions. The second part of the multi-BS extension is a virtual user concept allowing the *ProSched* tree based search algorithm to be employed (Chapter 13). Modifications of the tree search algorithm are shown to enable *ProSched* to be used with different types of MIMO systems including distributed MIMO systems, multiple BSs performing joint interference avoidance scheduling (with optional soft handover of user between BS), and REC with coordination. It is shown in simulations that interference avoidance coordination may yield a substantial gain in these settings.

In addition to the *ProSched* complexity reduction methods of Part II we show in Section 14.2 how neighboring OFDM subcarriers can share the same spatial processing solution with the help of a channel covariance matrix averaging method.

Several side results are obtained such as a reference simulation method to estimate the sum rate of a REC with half duplex relays performing adaptive precoding and SDMA. An approach to reduce the signaling overhead with *ProSched* is developed.

11.1. System model

As in Part II, the system resources time and frequency are assumed to be partitioned in a number of orthogonal resource elements. In each element, a different subset of the N_{u} terminals in the system can be served. This corresponds to the user selection approach which is considered to be advantageous over user grouping (see Section 4.1). When no Quality of Service (QoS) modifications are active, the scheduling relies solely on Channel State Information (CSI), which is assumed to be available once per time frame. Consequently, the scheduling solution will be the same for every time slot in a frame in this case.

To start with, the well known system model for the single BS is repeated. Later on, a BS index will be used in some cases. For each subcarrier and OFDM symbol, the complex data symbols

to be transmitted to user number $g \in \mathbb{N},\ 1 \leq g \leq G$ in a group of scheduled terminals $\mathcal{G}(n, f)$ of size G are contained in the column vector $\boldsymbol{d}_g$. The channels on each subcarrier are assumed to be frequency flat. Time and frequency indices will furtheron be skipped for notational simplicity. The complex received data vector for user g is then

$$\boldsymbol{y}_g = \boldsymbol{H}_g \boldsymbol{M}_g \boldsymbol{d}_g + \sum_{j=1, j \neq g}^{G} \boldsymbol{H}_g \boldsymbol{M}_j \boldsymbol{d}_j + \boldsymbol{n}_g \,, \tag{11.1}$$

where $\boldsymbol{H}_g$ denotes the $M_{R,g} \times M_{\mathrm{T}}$ complex channel matrix between the M_{T} transmit antennas of the base station and the $M_{R,g}$ receive antennas of user g. Let further $\boldsymbol{M}_g$ denote a linear precoding matrix generated for the transmission to user g. To cover the most general case, the system under consideration shall have the possibility to transmit multiple data streams to each user in an SDMA group. Therefore, $\boldsymbol{M}_g$ is allowed to have up to $r = \operatorname{rank}\{\boldsymbol{H}_g\}$ columns. The elements of vector $\boldsymbol{n}_g$ contain the additive noise only but do not contain interference components, since the purpose of the presented metric extension will be to estimate such inter-site interference power. The total noise power at each receiving antenna of user g in the entire band is σ_n^2.

12. Extended ProSched metric with interference estimate

12.1. The single BS case reviewed

The precoding matrices for all users to be served by a BS are generated jointly. Therefore, the result of the precoding depends on the terminal combination. The first goal of the *ProSched* approach was, thus, to avoid the pre-calculation of the precoding matrices for every user combination of interest during the search for the best user combinations. Instead, an estimate η_g of the g-th user's capacity in the group $\mathcal{G}$ was proposed as scheduling metric for every subcarrier (or resource element in frequency) at every time instant at which the scheduler is executed.

Originally, *ProSched* was derived for Block Diagonalization (BD) [62] as a precoding technique, which forces the interference between the transmissions to the users to be zero. Under this Zero Forcing (ZF) constraint, serving users with correlated channel matrices at the same time will lead to a reduction of the channel norm after precoding. ZF implies that user g's precoding matrix $\boldsymbol{M}_g \in \mathbb{C}^{r \times r}$ must lie in the joint null space of all other users' channel matrices so that the sum term in equation (11.1) is zero. In the case where the precoding allows interference (non-ZF case), spatially correlated channels also lead to a reduced channel norm and additionally to an increase in interference. The amount of interference produced is related to the channel quality reduction in the ZF case and, therefore, we use the same metric for scheduling, as also discussed in more detail in Section 5.2.

To come to the metric, in Section 10.1 an equivalent formulation of the BD algorithm was given which obtains the same ZF precoding solution as [62]. It is based on an orthogonal projection matrix $\tilde{\boldsymbol{P}}_g^{(0)}$ which projects $\boldsymbol{H}_g$ into the null space of a matrix $\tilde{\boldsymbol{H}}_g$ containing all the channel matrices of all other users in the group except the g-th, i.e.,

$$\tilde{\boldsymbol{H}}_g = \left[\begin{array}{cccccc} \boldsymbol{H}_1^{\mathrm{T}} & \cdots & \boldsymbol{H}_{g-1}^{\mathrm{T}} & \boldsymbol{H}_{g+1}^{\mathrm{T}} & \cdots & \boldsymbol{H}_G^{\mathrm{T}} \end{array} \right]^{\mathrm{T}}.$$

The resulting matrix $\boldsymbol{H}_g \tilde{\boldsymbol{P}}_g^{(0)} = \mathring{\boldsymbol{H}}_g \in \mathbb{C}^{M_{R_g} \times M_{\mathrm{T}}}$ fulfils the ZF condition and can, in a second step, be used together with other MIMO precoding techniques such as Singular Value Decomposition (SVD) based precoding (or eigenmode precoding) to decompose it into spatial modes. As before, the superscript $^{(0)}$ is added to a basis of a null space or a projection into a null space and $^{(1)}$ is added to a row space or a projection into a row space. The scheduling

metric was defined in equation (5.3) as

$$\eta_g = \log_2 \left(1 + \frac{P_\mathrm{T}}{G r_g \sigma_n^2} \left\| \boldsymbol{H}_g \tilde{\boldsymbol{P}}_g^{(0)} \right\|_\mathrm{F}^2 \right) . \tag{12.1}$$

It is an estimate of the link rate for user g and represents a lower bound on the ZF capacity obtained by the BD algorithm. This metric involves the assumption that equal fractions of the available transmit power P_T are assigned to all subcarriers as well as all r_g spatial modes of a user in a group of size G. This is necessary because the calculation of the full precoding matrix is avoided for the scheduling and thus the optimum power allocation is unknown (see Chapter 5).

The main advantage of using this metric was that the projection can be approximated by repeatedly applying projections into the separate users' null spaces instead (equation (5.5)):

$$\tilde{\boldsymbol{P}}_g^{(0)} = \left(\boldsymbol{P}_1^{(0)} \cdot \ldots \cdot \boldsymbol{P}_{g-1}^{(0)} \boldsymbol{P}_{g+1}^{(0)} \cdot \ldots \cdot \boldsymbol{P}_G^{(0)} \right)^p \ , \ p \to \infty \, . \tag{12.2}$$

These separate projectors can be conveniently computed from a basis $\boldsymbol{V}_u^{(1)}$ for the row space of a user u's channel matrix as $\boldsymbol{P}_u^{(0)} = \boldsymbol{I} - \boldsymbol{V}_u^{(1)} \boldsymbol{V}_u^{(1)\mathrm{H}}$, where the basis can be obtained via an SVD of $\boldsymbol{H}_u$. As a result, *ProSched* does not require decompositions of $\tilde{\boldsymbol{H}}_g$ for every user combination to be tested. Instead it requires only N_u small SVDs at the beginning to obtain the bases of all N_u users' row spaces to calculate the re-usable projectors $\boldsymbol{P}_u^{(0)}$. A projector $\tilde{\boldsymbol{P}}_g^{(0)}$ for any user combination is then obtained by simply multiplying the respective projectors $\boldsymbol{P}_u^{(0)}$. The number of multiplications needed during the scheduling process is negligible, because $p = 1$ was shown to be already sufficiently accurate to achieve a good grouping of the users and also because rank one approximations of the $\boldsymbol{V}_u$ can be used.

12.2. Extension to multiple Base Stations

If the system has B BSs, we distinguish them by an additional subscript $b \in \mathbb{N}$, $1 \leq b \leq B$. Furthermore, let $u \in \mathbb{N}$, $1 \leq u \leq N_\mathrm{u}$, now be a "global" user number, whereas g was used above as a "local" user number in a group $\mathcal{G}$ of size G. The channel between user u and BS b is furtheron given additional indices such as in $\boldsymbol{H}_{u,b}$.

12.2.1. The distributed MIMO case

If the distributed MIMO approach is taken, all BSs can be thought of as combined to form one virtual BS having a channel matrix $\mathcal{H}$ consisting of all BSs' channel matrices stacked together such that $\mathcal{H} = \left[\begin{bmatrix} \boldsymbol{H}_{1,1}^\mathrm{T} & \cdots & \boldsymbol{H}_{N_\mathrm{u},1}^\mathrm{T} \end{bmatrix}^\mathrm{T} \cdots \begin{bmatrix} \boldsymbol{H}_{1,B}^\mathrm{T} & \cdots & \boldsymbol{H}_{N_\mathrm{u},B}^\mathrm{T} \end{bmatrix}^\mathrm{T} \right]$. With this definition of $\mathcal{H}$ it seems to be intuitive that the same spatio-temporal transmit processing algorithms should be

applicable to both co-located and distributed antennas. The rows of $\mathcal{H}$ corresponding to user g-th antennas can be used in the calculation of the metric in equation (5.3) as if only one BS was present in the system.

In practice, the usage of distributed antenna systems is limited by aspects such as signaling delay between the antennas, CSI processing delay in the case of dislocated processing as well as possible separate power constraints for each antenna site. Only recently it has been shown that both types of systems are indeed equivalent in many theoretical aspects, while still neglecting the practical aspects: In [88] it was derived that any spatially correlated MIMO system employing an Orthogonal Space-Time Block Code (OSTBC) can be transformed into a distributed OSTBC system having an equivalent average symbol error probability. In [89], this proof was extended to the equivalence of distributed and co-located MIMO systems in the sense that systems with equivalent capacity distributions exist - under the assumption of flat Rayleigh fading channels which follow the Kronecker correlation model. The authors of the latter reference conclude that the same design rules for optimality of spatial processing schemes such as, e.g., sufficient antenna spacing, are applicable to both system types. This is illustrated with the help of transmit power allocation schemes as an example. One should however not conclude the optimality of either one of the two approaches, which can be highly scenario dependent as discussed in the simulations of Chapter 15.

12.2.2. Metric for multiple BSs with coordination

If only the scheduling is coordinated among the BSs but no joint spatial processing is performed, any BS sends interference to the users which are not assigned to it and are served by other BSs (inter-cell interference). To estimate this interference, we work again with the assumption of a ZF algorithm like BD at the transmitter (for the same reasons as in the single BS, see the review in Section 12.1).

As before, the precoding matrix $\boldsymbol{M}_g$ for user number g in a group of size G is split into a matrix with normalized columns $\boldsymbol{N}_g$ and a diagonal power loading matrix $\boldsymbol{D}_g$ such that $\boldsymbol{M}_g = \boldsymbol{N}_g \boldsymbol{D}_g$. The diagonal elements of $\boldsymbol{D}_g$ contain the square roots of the fractions of the total transmit power assigned to each spatial mode. This involves the usual assumption of uncorrelated data symbols with average unit power, i.e., $\mathrm{E}\left\{\boldsymbol{d}_g \boldsymbol{d}_g^{\mathrm{H}}\right\} = \boldsymbol{I}$.

The new formulation for BD precoding derived in Section 10.1 is used in the following. There, $\boldsymbol{N}_g = \mathring{\boldsymbol{V}}_g^{(1)}$ where $\mathring{\boldsymbol{V}}_g^{(1)}$ is a basis of the row space of the matrix $\mathring{\boldsymbol{H}}_g$, representing user g's channel matrix projected into the joint null space of all other users' channel matrices - the basis being obtainable, e.g., via SVD. It represents the spatial modes of $\mathring{\boldsymbol{H}}_g$.

To calculate the interference estimate for one user of interest with global number u, let the BS serving it be assigned number $b = 1$ for simplicity. Without loss of generality, we consider

only the hard handover scenario because a plethora of system design possibilities exist in the soft handover case as discussed briefly in Section 13.2. Let there be another BS $b > 1$ which is not serving u. Its transmission to a user with local user number $1 \leq g \leq G_b$ out of its assigned group $\mathcal{G}_b$ creates interference for user u with the following receive covariance matrix

$$\boldsymbol{R}_{u,b,g} = \boldsymbol{H}_{u,b}\mathring{\boldsymbol{V}}_g^{(1)}\boldsymbol{D}_g\boldsymbol{D}_g^{\mathrm{H}}\mathring{\boldsymbol{V}}_g^{(1)\mathrm{H}}\boldsymbol{H}_{u,b}^{\mathrm{H}}\,, \tag{12.3}$$

Recall from above that $\boldsymbol{H}_{u,b}$ is the channel between user u and BS b.

Since our algorithm works with low complexity estimates and avoids the calculation of the full precoding matrix, the basis $\mathring{\boldsymbol{V}}_g^{(1)}$ of the spatial modes is not known to the scheduler. As a consequence, it is impossible to know the optimal power allocation on the modes. As in the single BS metric, the assumption of loading equal power on all subcarriers and all modes of all users in a group of size G is used in the following, i.e., $\boldsymbol{D}_g = \sqrt{P_{\mathrm{T}}/(N_{\mathrm{c}} \cdot G \cdot r_g)}\boldsymbol{I}$. Thus, $\boldsymbol{R}_{u,b}$ becomes

$$\boldsymbol{R}_{u,b,g} = \frac{P_{\mathrm{T}}}{G \cdot r_g}\boldsymbol{H}_{u,b}\mathring{\boldsymbol{V}}_g^{(1)}\mathring{\boldsymbol{V}}_g^{(1)\mathrm{H}}\boldsymbol{H}_{u,b}^{\mathrm{H}} = \frac{P_{\mathrm{T}}}{G \cdot r_g}\boldsymbol{H}_{u,b}\mathring{\boldsymbol{P}}_g^{(1)}\mathring{\boldsymbol{P}}_g^{(1)}\boldsymbol{H}_{u,b}^{\mathrm{H}}\,. \tag{12.4}$$

The term $\mathring{\boldsymbol{V}}_g^{(1)}\mathring{\boldsymbol{V}}_g^{(1)\mathrm{H}} = \mathring{\boldsymbol{P}}_g^{(1)}$ is an orthogonal projection matrix into the row space of user g's zero forced channel. Since projectors are idempotent, it is possible to insert $\mathring{\boldsymbol{P}}_g^{(1)}$ a second time, which will be convenient for the following derivations.

Generally, the interference is highly directional due to the spatial processing and differs at each receive antenna of a user. However, our single BS scheduling metric uses a scalar norm of the projected channel and, therefore, only a scalar interference power term can be considered as an increment of the noise power term σ_n^2. It was confirmed during simulations that our scalar interference estimate is accurate enough for scheduling. Let us define $\sigma_{i,u,b,g}^2 = \mathrm{trace}\{\boldsymbol{R}_{u,b,g}\}$ which represents the total estimated inter-cell interference power received by user u generated by BS b while serving user g out of its assigned group. Then the following holds:

$$\begin{aligned}\sigma_{i,u,b,g}^2 &= \mathrm{trace}\left\{\frac{P_{\mathrm{T}}}{N_{\mathrm{c}} \cdot G \cdot r_g}\boldsymbol{H}_{u,b}\mathring{\boldsymbol{P}}_g^{(1)}\mathring{\boldsymbol{P}}_g^{(1)}\boldsymbol{H}_{u,b}^{\mathrm{H}}\right\} \\ &= \frac{P_{\mathrm{T}}}{N_{\mathrm{c}} \cdot G \cdot r_g}\left\|\boldsymbol{H}_{u,b}\mathring{\boldsymbol{P}}_g^{(1)}\right\|_{\mathrm{F}}^2\,.\end{aligned} \tag{12.5}$$

The Frobenius norm term expresses the similarity between the subspace into which the BS transmits and the subspace of the user receiving interference. We can now exploit the fact that the row space of a matrix is an orthogonal complement of its null space:

$$\sigma_{i,u,b,g}^2 = \frac{P_{\mathrm{T}}}{N_{\mathrm{c}} \cdot G \cdot r_g}\left(\|\boldsymbol{H}_{u,b}\|_{\mathrm{F}}^2 - \left\|\boldsymbol{H}_{u,b}\mathring{\boldsymbol{P}}_g^{(0)}\right\|_{\mathrm{F}}^2\right), \tag{12.6}$$

which is the equivalent of Pythagoras' theorem for multidimensional subspaces. Introduce a matrix $\tilde{\boldsymbol{V}}_g^{(0)}$ whose columns span the null space of the matrix $\tilde{\boldsymbol{H}}_g$ containing all the channel matrices of all other users in the group except the g-th. Since $\mathring{\boldsymbol{P}}_g^{(1)}$ projects into a subspace spaned by $\tilde{\boldsymbol{V}}_g^{(0)}$, we can replace $\boldsymbol{H}_{u,b}$ by $\boldsymbol{H}_{u,b}\tilde{\boldsymbol{P}}_g^{(0)}$ such that

$$\sigma^2_{i,u,b,g} = \frac{P_\mathrm{T}}{N_\mathrm{c} \cdot G \cdot r_g} \left(\left\| \boldsymbol{H}_{u,b} \tilde{\boldsymbol{P}}_g^{(0)} \right\|_\mathrm{F}^2 - \left\| \boldsymbol{H}_{u,b} \tilde{\boldsymbol{P}}_g^{(0)} \mathring{\boldsymbol{P}}_g^{(0)} \right\|_\mathrm{F}^2 \right). \tag{12.7}$$

The projection matrix obtained by the product $\tilde{\boldsymbol{P}}_g^{(0)} \mathring{\boldsymbol{P}}_g^{(0)}$ is equal to the projector into the subspace obtained by the intersection of the column spaces spanned by $\tilde{\boldsymbol{V}}_g^{(0)}$ and $\boldsymbol{V}_g^{(0)}$. Therefore, exploiting the repeated projection approximation shown in equation (5.5), we can write

$$\tilde{\boldsymbol{P}}_g^{(0)} \mathring{\boldsymbol{P}}_g^{(0)} = \left(\tilde{\boldsymbol{P}}_g^{(0)} \boldsymbol{P}_g^{(0)} \right)^p \ , \ p \to \infty. \tag{12.8}$$

This approximation allows us to write the total estimated inter-cell interference power as

$$\sigma^2_{i,u,b,g} = \frac{P_\mathrm{T}}{G \cdot r_g} \left(\left\| \boldsymbol{H}_{u,b} \tilde{\boldsymbol{P}}_g^{(0)} \right\|_\mathrm{F}^2 - \left\| \boldsymbol{H}_{u,b} \left(\tilde{\boldsymbol{P}}_g^{(0)} \boldsymbol{P}_g^{(0)} \right)^p \right\|_\mathrm{F}^2 \right). \tag{12.9}$$

This notation is particularly convenient because both $\tilde{\boldsymbol{P}}_g^{(0)}$ and $\boldsymbol{P}_g^{(0)}$ need to be computed anyhow for the scheduling metric η_g so that no additional SVDs are required. Again the number of repeated projections p determines the accuracy of the estimate. In the simulations it is set to the same value as for the approximation of $\tilde{\boldsymbol{P}}_g^{(0)}$.

Consequently, on the current subcarrier, the estimated total inter-cell interference power received by a user with global number u is the sum of all interference generated by all other BS while serving their assigned users:

$$\sigma^2_{i,u} = \sum_{b=2}^{B} \sum_{k \in \mathcal{G}_b} \sigma^2_{i,u,b,k} \, . \tag{12.10}$$

The scheduling metric for user g in a group $\mathcal{G}_b$ with inter-cell interference then becomes

$$\eta_g = \log_2 \left(1 + \frac{P_\mathrm{T}}{N_\mathrm{c} G r_g (\sigma_n^2 / N_\mathrm{c} + \sigma^2_{i,g})} \left\| \boldsymbol{H}_g \tilde{\boldsymbol{P}}_g^{(0)} \right\|_\mathrm{F}^2 \right) . \tag{12.11}$$

(Note that, when computing $\sigma^2_{i,g}$ according to equation (12.10), the group user index g needs to be replaced by its corresponding global user number u.)

Since the interference situation depends on the user allocation at all considered transmitting stations, a method is required to traverse this large combinatorial space.

13. Extended tree-based sorting algorithm

13.1. The single BS solution applied to distributed MIMO

To reduce the number of user combinations in a search, a best candidate algorithm in the form of a dendrogram (or search tree) was already proposed in Part II. It can work with the *ProSched* metric or any other which represents an estimate of a user's link quality. In the algorithm's first phase, candidate user groups in all possible sizes from one to the maximum size supported by the precoder are identified (typically up to the rank of the combined downlink channel matrix). To do so, in the case of only one (virtual) BS, the first phase starts with single user groups containing terminals $1 \dots N_\mathrm{u}$ which form the first level of the tree as explained previously in Figure 6.1 for $N_\mathrm{u} = 5$ users. The terminal with the best metric is identified and is assigned number one. Next, the pairing with the highest metric sum of user one with any of the remaining users is searched for and stored as candidate grouping with size two. The algorithm continues in this way until candidate groups for all possible sizes have been found.

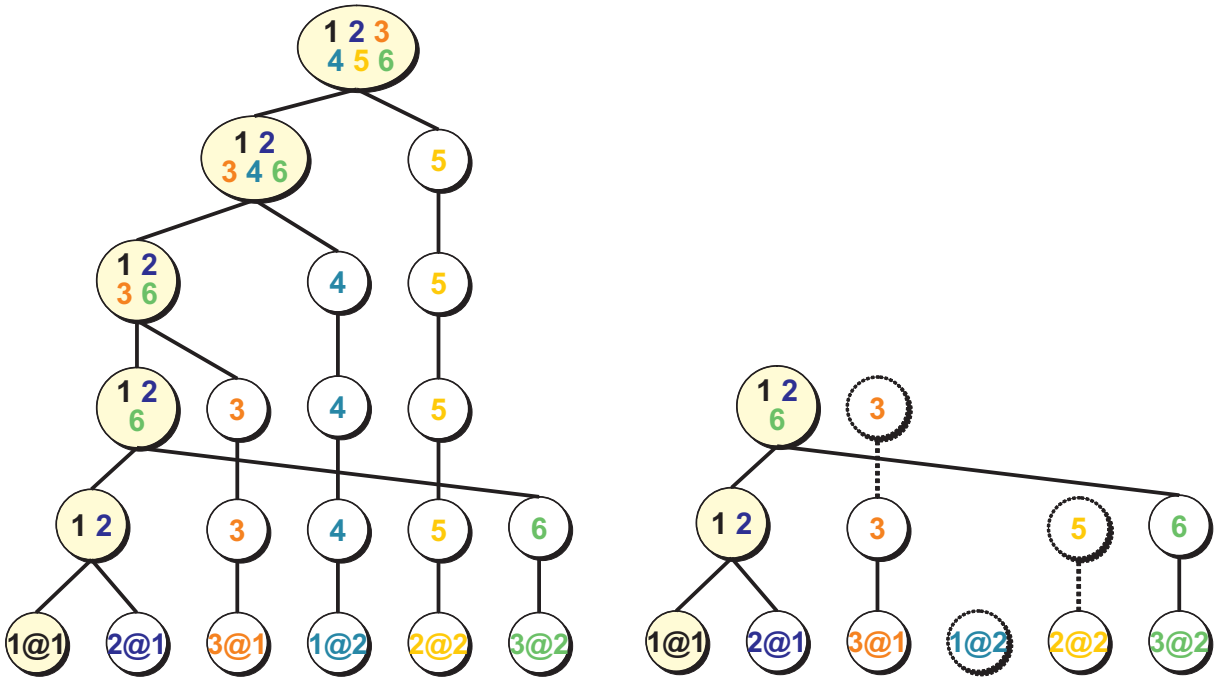
In a second phase, the group with the highest metric sum out of the candidate groups on the left edge of the tree is selected as the scheduling solution for the current time slot and frequency resource. The final selection can also be performed by calculating the exact Shannon rates for the few candidate sets. This eliminates the estimation error in this step of the algorithm and can increase its performance.

13.2. Multiple BSs with coordination

The above search algorithm can be extended to the case of multiple BSs with coordinated scheduling. To do so, the scheduling metric must contain the influence of the interference as in the *ProSched* metric from equation (12.11).

The tree is started with $N_\mathrm{u} \cdot B$ "virtual" users representing all possible combinations of users and BSs. By applying the tree as it is on those virtual users, a soft handover-like strategy is obtained, i.e., more than one BS can be allowed to serve a terminal. To achieve a hard handover, once a physical terminal has been assigned to a candidate grouping solution, all of its other virtual representations must be deleted from the current tree level before the next level is calculated.

In the soft handover case, the calculation of the rate estimate must take into account how multiple BSs are supposed to transmit to a single user as well as its spatial receive processing.



(a) Soft handover: a user can be served by multiple BSs, but the receiver strategy has to be taken into account in the metrics.

(b) Hard handover: all instances of a terminal already assigned to a base station are removed (dotted outline) before the next level is generated.

Figure 13.1.: Examples for the tree search with soft and hard handover with $B = 2$ BSs and $K = 4$ users: The 6 virtual user numbers represent all combinations of users and BSs. As before in the single BS case, the subsets detected as optimal for each possible group size appear on the left edge

The options to do so are manifold and we are not attempting a comparison of them. To simplify, it is assumed that all spatial modes are used by a BS to transmit to a user. Furthermore, all BSs serving a single user simultaneously are assumed to share a system resource via an additional multiple access scheme (Time Division Multiple Access (TDMA) or Code Division Multiple Access (CDMA)) and, thus, the metric for each link is divided by the number of BSs serving a user. Alternatively, each BS could for example only use the dominant spatial mode, allowing the user to employ a VBLAST like receiver on the stacked channel matrix $\mathcal{H}$ to detect multiple serving BSs at the same time. It is also to be decided whether diversity or spatial multiplexing gain should be the goal in a soft handover situation, but this is beyond the scope of this work.

14. Algorithm Extensions

14.1. Fairness, QoS, and long term CSI

Our scheduling metric is a rate estimate. Therefore, it can be combined with other existing approaches for QoS and fairness which are based on the calculation of user rates. In Section 5.4, examples are given how proportional fairness [69] in the form of methods given in [67] can be implemented and user rate requirements or user service priorities can be considered as in [70] or [71].

Also a method for scheduling on long term channel knowledge was investigated in Section 5.5 based on averaged spatial covariance matrices as in [73]. The basis of the user row spaces are obtained from eigendecompositions of such covariance matrices for use in both the precoding and scheduling. In the next section it is shown how this method can also be used to reduce scheduling complexity by exploiting channel correlation in frequency and time direction.

14.2. Complexity reduction in frequency and time

It is known that a wireless channel transfer function displays correlation between the bins in the discrete frequency domain. The complexity of spatial processing can, therefore, be reduced by performing it only once for a cluster of correlated frequency bins and then replicating the solution on all C bins within the cluster at the expense of accuracy. The method mentioned above for precoding/scheduling on long term CSI [73] was modified in [90] for usage with multi-user precoding in the frequency domain. Here we describe how our scheduling algorithm can be used with it:

Within a cluster number c of size C, the scheduling is performed only once with a pseudo channel matrix $\hat{\boldsymbol{H}}_{u,b}(c)$ for each user u instead of using the measured channel matrix $\boldsymbol{H}_{u,b}(f)$ on every subcarrier. To do so, first an estimated transmit covariance matrix $\boldsymbol{R}_{T,u,b}(c)$ is calculated for each cluster c and each user with the help of averaging:

$$\boldsymbol{R}_{T,u,b}(c) = \frac{1}{C} \sum_{f=1}^{C} \boldsymbol{H}_{u,b}^{\mathrm{H}}(f) \boldsymbol{H}_{u,b}(f)\,. \tag{14.1}$$

Then the pseudo channel matrix $\hat{\boldsymbol{H}}_{u,b}$ is constructed from a basis of the signal space of $\boldsymbol{R}_{T,u,b}(c)$

which can, for example, be obtained with an EigenValue Decomposition (EVD)

$$\begin{aligned} \boldsymbol{R}_{T,u,b}(c) &= \boldsymbol{V}_{u,b}(c)\boldsymbol{\Sigma}^2_{u,b}(c)\boldsymbol{V}^{\mathrm{H}}_{u,b}(c) \\ &= \boldsymbol{V}^{(1)}_{u,b}\boldsymbol{\Sigma}^2_{u,b}\left[\boldsymbol{V}^{(1)}_{u,b}\ \boldsymbol{V}^{(1)}_{u,b}\right]^{\mathrm{H}} \qquad (14.2) \\ \text{such that} \quad & \hat{\boldsymbol{H}}_{u,b}(c) = \boldsymbol{\Sigma}_{u,b}\boldsymbol{V}^{(1)^{\mathrm{H}}}_{u,b} \qquad (14.3) \end{aligned}$$

where the matrices $\boldsymbol{V}^{(1)^{\mathrm{H}}}_{u,b}$ contain the first $r = \operatorname{rank}\{\boldsymbol{H}_{u,b}\}$ columns of the $\boldsymbol{V}_{u,b}$. With practical channel measurements, the matrices $\boldsymbol{H}_{u,b}$ likely are full rank. Should they not be, then padding with zero vectors is required for the $\hat{\boldsymbol{H}}_{u,b}$ to become the same size as the $\boldsymbol{H}_{u,b}$. (Note that more explanations on the absence of a $\boldsymbol{U}$ matrix in the pseudo channel matrix and on the square factor can be found in Part II, Section 5.5.)

This method is also usable in the distributed MIMO case by applying it to the respective rows of $\mathcal{H}$.

Two more complexity reduction methods were given in Sections 6.3, 6.4 that are also applicable to the multi-BS scheduler: It is possible to track the solution found by the tree based algorithm in time. Instead of building the entire tree at every instance of the scheduler, only a number of levels above and below the previously found solution are calculated. This is possible with the help of an equivalent tree algorithm which proceeds top-down. In our simulations we calculate only one level above the previous solution and then proceed down two levels to update also the past solution.

This tracking is especially attractive when all users on all subcarriers are treated in one big tree (referred to as Joint 3D-Scheduling before). To do so, all users' channels from all subcarriers (or clusters) are considered as one frequency flat system and the tree is started with $N_{\mathrm{u}} \cdot N_{\mathrm{c}} \cdot B$ virtual users (N_{c} being the number of subcarriers or clusters). The calculation of the metrics has to take into account the orthogonality of virtual users originating from different subcarriers. In our simulations we still update only three levels when using this virtual system together with the tracking in time, although the tree levels span many more combinations than without the virtual system.

15. Simulations with multiple BSs

For the simulations, a two-BS MIMO scenario with up to $N_u = 24$ users moving randomly at speeds of up to 70 km/h is used as shown in Figure 15.1. The BSs have 12 element Uniform Circular Arrays (UCAs) while each user has two omnidirectional antennas, $\lambda/2$ spaced. The model is generated with the geometry based channel model *IlmProp* [75] featuring realistic correlation between users in space, time, and frequency. This is crucial in assessing the performance of a spatial scheduling scheme. As discussed in Chapter 8 already, the *IlmProp* is best suited for creating test scenarios but has some limitations when it comes to producing throughput values in absolute terms. Therefore, we show only normalized performance estimates and relative gains between algorithms.

To illustrate the effectiveness of our algorithm (which exists in different variations), Figure 15.2 shows a comparison with the performance of user selections found by exhaustively searching through all possible user subsets as well as to a trivial Round Robin (RR) scheduler with different maximum group sizes as explained at the end of this Section. Due to the computational complexity of an exhaustive search this is only shown in the scenario when the two BS coordinate the scheduling but precode separately and only for the frequency flat case at a frequency of 2 GHz, a total of 10 users and for a single Signal to Noise Ratio (SNR) value. Consequently, the results are given as Complementary Cumulative Distribution Functions (CCDFs) of the total system rate, normalized to the maximum rate value of the reference TDMA curve. (SNR is defined here as total transmit power over total noise variance at one receiver.)

An OFDM symbol duration of 20 μs is assumed (without considering the length of the guard period) and a TDMA frame consisting of 50 OFDM symbols. We show the case in which the two BSs perform only the scheduling jointly and do BD precoding and water pouring powerloading separately. Ideal power loading is used in the calculation of the resulting Shannon rates (but not within the scheduling metric), assuming that in a real world implementation a power and bit loading algorithm with the goal of increasing throughput such as in [91] can be used.

Our algorithm reaches a performance close to that of the exhaustive search in all displayed variations. The following abbreviations are used in the legend: `ProSched.full.p=1` stands for our projection based algorithm using full basis matrices and an order of $p = 1$ for the repeated projection approximations. `ProSched.rank1.p=1` indicates that reduced rank basis matrices are used in the calculation of the projection matrices. In the tree based algorithm, the final solution can be selected with the help of the scheduling metric (`pick=metric`) or based on the true rate (`pick=rate`). The suffix `tracking` stands for the complexity reduction method in time

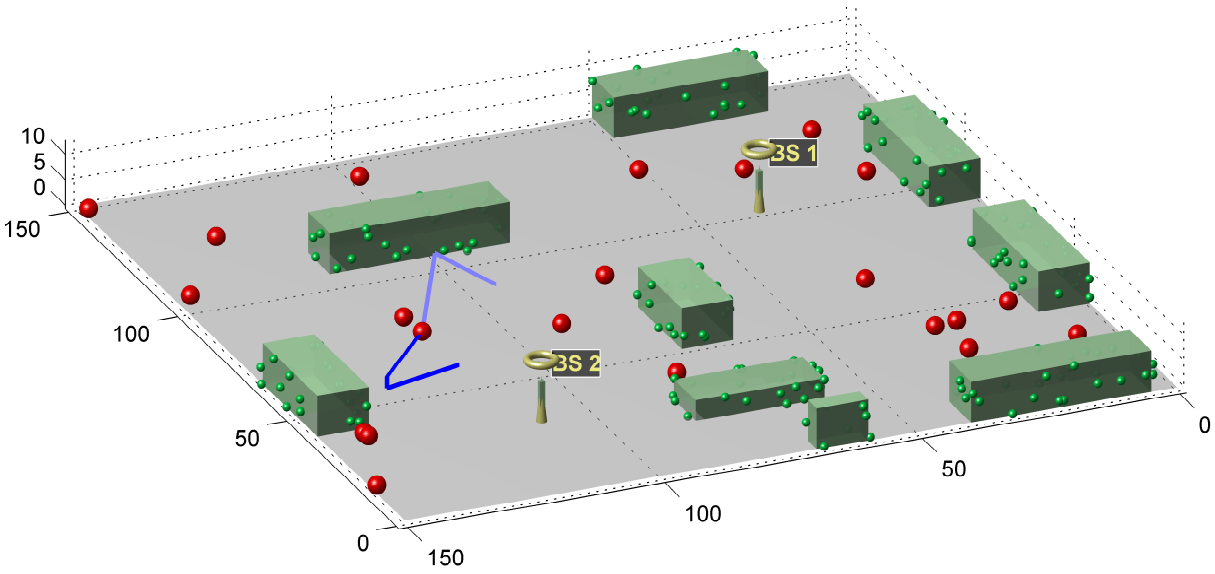


Figure 15.1.: The *IlmProp* scenario used for the two-BSs simulations: up to $N_u = 24$ mobile stations (red) move randomly in an area with two base stations.

by tracking of the search tree.

A first conclusion is that an essentially random scheduler like RR is not recommended. The risk of scheduling a user with blocked propagation paths to any BS even increases when bigger groups are chosen and, therefore, even TDMA performs better in our simulations than SDMA with RR. The next conclusion is that the second step of the tree based algorithm, i.e., the selection among the candidate user subsets, is especially sensitive to the estimation error in our metric. Therefore, it is recommended to use the calculated rate for this step and on the other hand reduce the complexity again with the help of the time tracking algorithm (which reduces the number of candidate sets to choose from to three only).

Next, we emphasize the usability of *ProSched* in both the distributed MIMO approach with one virtual BS as well as the case with coordinated scheduling but separate spatial processing. In Figure 15.3, simulation results for the frequency selective case with 24 frequency bins spanning a bandwidth of 1.2 MHz and 24 users are shown in the form of 90% outage curves as a function of SNR. The scheduling algorithm in use is the one which displayed the best trade-off in complexity and performance in the frequency flat case, namely `ProSched.rank1.p=1.pick=rate.tracking`. This time MMSE(TxWF) precoding [25] is used which does not force the interference between spatial streams to be zero.

We observe that RR performs surprisingly well in the distributed or virtual MIMO case for low SNRs. We attribute this observation to the scenario displaying many situations in which users have obstructed paths to any BS. Distributed antennas can help to overcome

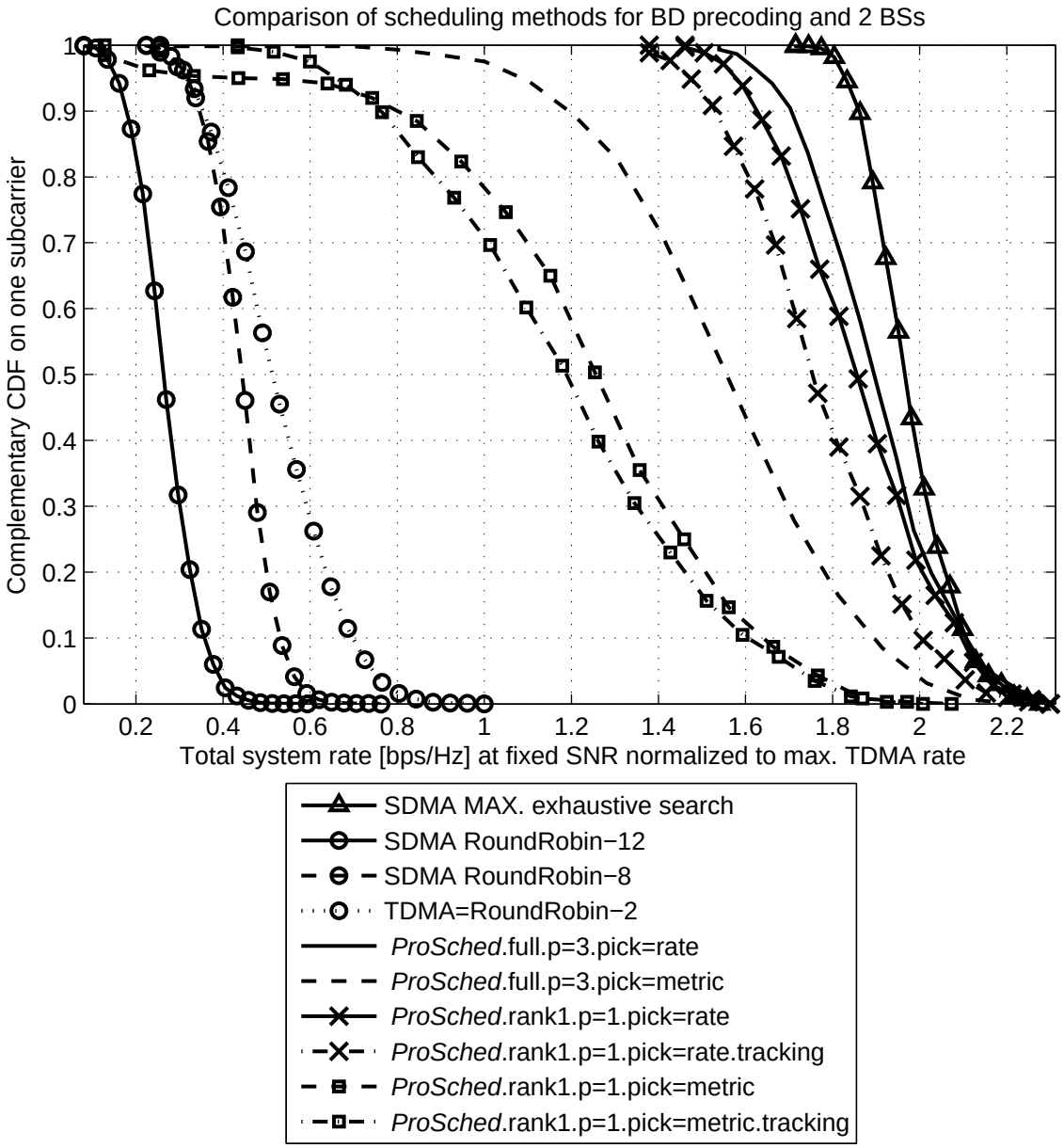


Figure 15.2.: Variations of *ProSched* coordinated scheduling with different complexity compared to an exhaustive search over all possible combinations in the given scenario with two BSs and 10 mobile stations. The curves belonging to the *ProSched* methods which offer the best performance vs. complexity trade-off are marked with an 'x'.

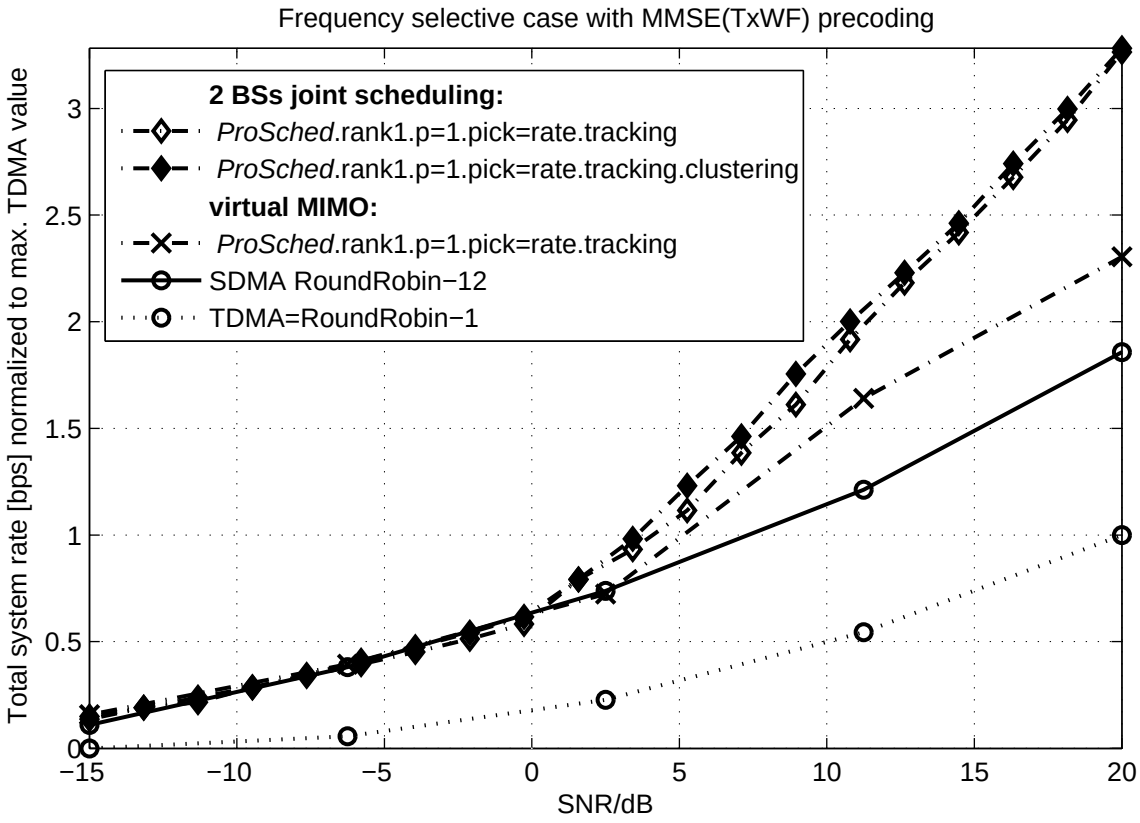


Figure 15.3.: Comparison of the virtual MIMO approach to two separate BSs performing joint scheduling. Frequency selective case with 24 mobile stations and MMSE(TxWF) precoding.

shadowing and blocked paths, especially when their placement has been optimized. In our case the antenna positions are the same in both cases, which may well be a reason why two separate coordinated BSs outperform the virtual MIMO approach. However, such results are highly scenario dependent. In channel realizations with a lot of shadowing, the distributed antennas often performed better.

The short distances between users and BSs lead to a lot of inter-site interference on the one hand, but good link budgets on the other. This often causes users to be assigned to both BSs, albeit in our system model the two BSs have to alternate in a time slot when serving the same user. (Different strategies exist to enable such soft handover, as discussed in Chapter 13). It is more beneficial for the users to be served by two times two spatial modes in half the time rather than by only two modes with slightly higher quality in the virtual MIMO case.

To show the efficiency of our frequency averaging method from Section 14.2, a curve was simulated for the two BS case, where the bandwidth was divided into four clusters of six bins. In the legend it is marked with the suffix `clustering`. The performance drop due to the reduced accuracy is negligible. Note that the averaging was used only for the scheduling and not for the precoding to assess its implications on the scheduler only.

Definition of the Round Robin scheme for multiple BS:
The RR scheme re-schedules every time slot. For the frequency selective case, the same solution is applied to all subcarriers. In the legends, the number given after Round Robin denotes how many users are to be scheduled at every time slot. In the distributed MIMO case the virtual base station having a total of 24 antenna elements can spatially multiplex up to 12 users with two antennas each when two data streams are sent to each user. The RR-X scheme selects X users, by cycling through the N_u available. For instance, a RR-5 scheme would schedule the following users out of $N_\mathrm{u} = 12$ for successive time snapshots: $\{1, 2, 3, 4, 5\}$, $\{6, 7, 8, 9, 10\}$, $\{11, 12, 1, 2, 3\}$, etc. In the case of two separate BSs, each BS can serve at most six users. To obtain the Round Robin solution, we split the one BS case into two parts, effectively cycling the user assignment also through the two BSs.

16. ProSched for RECs

The option of relaying data over fixed or mobile stations to enhance service quality is considered a key feature of a flexible fourth generation (4G) communications system in projects such as Wireless World Initiative New Radio (WINNER) [11, 12]. For MIMO systems without SDMA, solutions for the scheduling and resource partitioning problem have previously been well established for both non-adaptive and adaptive precoding [92, 93, 94]. When the possibility to use SDMA shall be given to the relays, however, the very same problems of spatial correlation and interference arise as described above in the case of multiple BSs, in addition to the data routing problem.

In the following it is described that *ProSched* can be used to solve these problems when coordination is possible at least on a long term scale. The benefits of *ProSched* translate to the REC setting, namely that during any testing of user assignments, the precoding solutions at each transmitting station do not have to be re-computed but can be estimated with the help of fixed orthogonal projection matrices. The same applies to the interference among the transmitters which is different for each possible user assignment. It can also be estimated without knowing the final precoding solution. Additionally, *ProSched* may work with rank one approximations of the users' long term channels, offering a possibility to reduce the required pilot overhead for channel estimation.

ProSched for RECs has been developed and tested so far for a maximum of two hops [12]. It is based on two ideas: The first is to express the Shannon rate of a multi-hop link type under consideration with the help of rate estimates as in our scheduling metric with interference of equation (12.11). The second part consists of an adapted search algorithm.

16.1. Estimating the multi-hop link rate

In [94], mutual information bounds are given for various types of point-to-point relay links. On a high level we may distinguish them between amplify and forward relays and decode and forward relays (and intermediate solutions). Since we are interested in the general case where the Relay Nodes (RNs) can perform adaptive precoding and SDMA, we are dealing with decode and forward only. Also it is common to assume that the same set of antennas is used for both transmission and reception, resulting in half duplex RNs - not that half duplex is necessarily the best strategy in all cases. One could think of using two sub-arrays especially in scenarios with well separated users such as in street canions.

For a single user link over a total of h_t hops with half duplex RNs having individual power constraints, the rate limit $C_{\mathrm{RN},u}$ is

$$C_{\mathrm{RN},u} = \frac{1}{h_\mathrm{t}} \min_h C_{(h)} \tag{16.1}$$

where the $C_{(h)}$ shall stand for the rate limits of the separate hops with number h. The division by h_t takes into account the loss due to the use of h_t time slots (or other resource elements). In the full duplex case this division is not needed. The expression has a simple interpretation: An RN can only re-transmit as much information as it has received before through its feeder link, but only if the following link is strong enough. With *ProSched*, the hop link rates $C_{(h)}$ can be estimated in the same way as was shown previously in equation (12.11) for an end user rate.

16.2. Testing user combinations

When SDMA is used, the $C_{(h)}$ are not unique because each one is a function of all other modes that are sent from the same transmitter and so is the resulting link capacity $C_{\mathrm{RN},u}$. To test combinations, one could interpret each possible realization of $C_{\mathrm{RN},u}$ as a virtual user and insert those into the tree based search algorithm. However, a better solution to traverse these combinations is likely to exist for the general case which is yet to be developed fully.

Therefore, a temporary solution for the two hop case is given in the following. It exploits the fact that in WINNER a Time Division Duplexing (TDD) system is proposed for the scenarios in which relaying is envisioned. When the relays are half duplex, they can be treated as user terminals in the time slots when they receive. And when they transmit, they become additional base stations for the algorithm. To take into account that the relays are part of a multi-hop link, a key element of the proposal is to model a buffer at the RNs: RNs can only transmit as much as they have received before. To simplify the implementation, each RN's buffer is implemented as one number rather than storing a vector for each user attached to it. The buffer is kept in bits/sec/Hz because no exact time reference is needed for the scheduler in a TDD system with equal time slot duration. When an RN is scheduled for reception, its estimated achievable rate is added to its buffer. In this way, the buffers are continuous as long as the scheduler is running.

The rate estimates η of the users to be served by a certain RN are then limited to values generated out of the single RN buffer value as follows. It is assumed that the buffer for transmission to each user has been loaded optimally based on the achievable rates of the attached users in the current time slot. This is reasonable if maximization of the sum rate is targeted. In real systems, this knowledge is of course not available a-priori and represents a simplification which is justified because the channel changes only gradually. In other words, the situation in the time

slot in which the buffers would have been filled can be assumed to be similar to the situation when transmission takes place. To generate the user specific buffer levels out of the single value buffer of a RN, the RN buffer value is distributed via a standard water pouring algorithm power allocation algorithm [16] on the users. To do so, the users' achievable rates when served from a certain transmitter are treated in the water pouring algorithm as the squared coefficients of the channels to be loaded and the RN buffer number is treated as the power to be distributed.

16.3. Reduction of signaling overhead for interference coordination

A method to reduce the required signaling overhead with *ProSched* interference coordination is given in this section.

The genie performing the coordination needs to know the channel matrices between all combinations of nodes in the system to be able to estimate interference. However, it can be safely assumed that no additional overhead may be needed to signal measurements of the channels between any RN and the BSs. These channels can be obtained with a similar Pilot Assisted Channel Estimation (PACE) procedure as for the users in the single BS system: The RNs could send uplink pilots without precoding and the BSs could then perform channel estimation. If the system uses TDD, this estimate could then be used for precoding on the reverse link if the RF frontends have been calibrated such that the channel is reciprocal. Otherwise, feedback is required. The stationarity of the RNs can be exploited to reduce the pilot frequency.

The only overhead stems from the necessity to signal the channels between any RN and all users back to the central intelligence. Assuming that the BSs are connected by some sort of backbone network, it may be sufficient for an RN to only signal to the nearest BS. The RNs having multiple antennas, spatial multiplexing can be used for this purpose.

Nevertheless, the properties of *ProSched* can be used to further reduce the signaling overhead. It was already discussed before that the projection matrices needed can be computed from rank one bases approximations of the users' subspaces. But so far the projectors were multiplied always to the full channel matrix of a user g when computing its metric $\eta_g^{(\dots)}$ while being served by BS b in group $\mathcal{G}_{\lfloor}$. In Figure 16.3 a simulation result is shown where only the first basis vectors of the user signal spaces $\boldsymbol{V}_g$, $\boldsymbol{v}^{(1)}$ scaled with the channel norm was used instead of the full channel matrices. In equation (12.11) the channel $\boldsymbol{H}_g$ of the current user seen from BS b is thus to be replaced by

$$\|\boldsymbol{H}_g\|_{\mathrm{F}} \cdot \begin{bmatrix} \boldsymbol{v}^{(1)^{\mathrm{H}}} \\ 0 \dots 0 \\ \vdots \\ 0 \dots 0 \end{bmatrix} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{T}}} . \tag{16.2}$$

Note that this is not a best rank one approximation. It is simply an attempt to preserve the total channel Frobenius norm which is important because *ProSched* also considers the transmit power. The all zero vectors are needed to reproduce the original matrix dimensions of $\boldsymbol{H}_g$. The interference channels $\boldsymbol{H}_{u,b}$ between a user u and BS b, needed in equations (12.9,12.10) are replaced accordingly. This effectively halves the number of complex coefficients to be signaled. The simulation discussed further below suggest that a reasonable scheduling gain can still be achieved. Due to the precoder being limited to dominant eigenmode transmission, this reduced channel knowledge can also be used for precoding in the simulated case. To allow for higher rank transmission, the network could request more accurate CSI only for the found scheduling solution later on. Appropriate vector quantization methods may be used to reduce the overhead even further.

16.4. Simulations and Discussion

In this section a proof of concept investigation is discussed in which the proposed approach of low-complexity centralized scheduling with interference coordination is compared to a reference approach without interference coordination (Section 16.4.1).

The scenario under consideration is a Metropolitan area deployment based on the guidelines given in [30], see Figure 16.1. It was rendered using the WINNER II channel model [66] in a version as published in [77]. The RNs are placed in both horizontal and vertical streets, thus causing a high occurrence probability of Line Of Sight (LOS) conditions in the channel. They have omnidirectional antennas arranged in Uniform Linear Arrays (ULAs) perpendicular to the streets. For interference avoidance, the operation of the RNs in at least two groups using different resources is recommended in WINNER [95], as indicated by the different colors of the circles. In the simulations this recommendation is considered as an optional alternative to using all RNs at the same time.

The system parameters are kept close to those of [30], except for a restriction due to the reference method used (see below) and for a change in base station antenna element elevation gain, which is set 8 dBi instead of 14 dBi, corresponding approximately to an assumed beam width of 60 degrees in elevation which seems reasonable due to the street canyons. Only one SNR operating point needs to be considered since the transmit and noise powers are fixed in [30]. Relays are half-duplex, resulting in the TDD frame structure given in [95]. Only two hops are considered and indoor users are not present in this simulation.

All performance figures are given per square meter to allow independence of the chosen deployment section size. The total number of users in the system is obtained from a uniform user density per square meter which was set to $10^{-5}/\mathrm{m}^2$. The deployment section having a surface

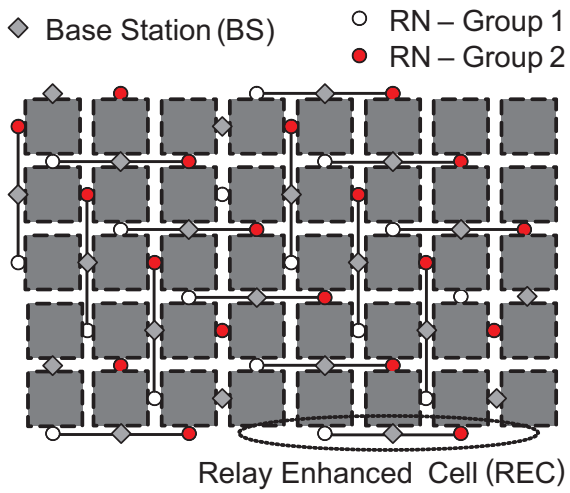


Figure 16.1.: Sketch of WINNER microcellular cell layout with Relay Nodes which was implemented except for the transmitters on the edges. Source:[95]

of 2 km^2 contains 20 users. The rate figures were obtained using the Shannon formula, bounded to 4.8 bits per channel use to take into account the fact that the spectral efficiency is limited by the Modulation and Coding Scheme (MCS) to 64 QAM with rate 4/5 coding in the WINNER system (part of a so-called Shannon fitting procedure [79]).

16.4.1. A reference method to obtain the sum rate of a REC without coordination

In this section a simulation method is described to obtain an estimate of the total system rate of a REC when no interference coordination is performed. The approach is an updated version of the one published in [8, 13] as a method for comparing different relay deployment strategies. To obtain an estimate of the maximum achievable rate without the need to choose a certain space-time processing, the method is based on the theoretically achievable maximum sum rate under sum power constraint when channel knowledge is available at the transmitter (Dirty Paper Coding (DPC) bound).

The simulation method was conceived because, to our best knowledge, other available results are not applicable to the system of interest and do not easily apply to a measurement based channel model. For example, in [96] a simulation method with similar goals can be found, however, for amplify and forward RNs and based on specific MIMO techniques that do not exploit CSI at the transmitter. It would have been desirable to base the method on any theoretical capacity bounds for MIMO RECs which have only recently attracted a growing research interest rather than construct an approach around a single base station capacity limit. However, results such as in [97, 98] are valid for a single user only and rely on the key assumption that the RNs

are operating in full duplex mode and have two sets of antennas, one for transmission and one for reception, which is not the case in the WINNER system. These works present upper and lower capacity bounds for the Gaussian and Rayleigh channel, with more tight lower bounds in [98]. General capacity scaling laws for half duplex relays are given in [99], however again only for a single user in a system with an asymptotically high number of relays.

Our method being bases on the DPC bound, the chosen assessment criterion is capacity improvement (which is also the goal in WINNER, see for example [95]) as opposed to a possible coverage extension. The DPC bound rates are computed using the frequency flat, iterative uplink algorithm of [29]. As a consequence, the method is limited to one subcarrier and the possibility of using Orthogonal Frequency Division Multiple Access (OFDMA) (which is given in the basic design of [95]) cannot be investigated. This is due to the fact that in the literature no algorithm to treat also the space frequency power loading problem at the same time was available when this method was published. A later result in [100] should allow to translate the simulation approach to multiple subcarriers.

A simplified frame structure is used such that one instance of the DPC algorithm is run per drop of the channel. (In the WINNER channel model, a drop is one fixed set of user positions, see also Chapter 8).

To assign the users to the RNs and/or BSs, a genie-aided scheduler is used as described below. It does not perform coordination and can, thus, be suitably compared to *ProSched* with coordination. In the genie-aided scheduler the RNs also have a data buffer as in the *ProSched* implementation proposal at the beginning of this section.

The simulation steps for the reference performance are as follows:

1. Compute the DPC bound rates for all users when served by each one of the BSs or RNs separately, assuming independent single cell systems with one transmitter only. In the odd time slots RNs do not transmit but are also users (receivers). In the even time slots, the RNs act as BSs.
2. Genie-aided scheduler knowing all achievable rates: Decides on the assignment of users to RNs and BSs based on the achievable DPC rates from step 1 (no interference considered in this step, suboptimal).
3. Recompute DPC covariance matrices for the newly assigned groups (second run of DPC algorithm required).
4. Perform uplink-downlink conversion of the newly computed covariance matrices as in [28].
5. Compute downlink rates for the entire system WITH interference (all transmitters) using the downlink DPC covariance matrices from step 4 and taking into account a buffer level at the RNs: The role of the RNs depends on the time slot number. When the RNs receive, they fill up their buffer. When they transmit, the achievable rates of the users assigned

to RNs is limited by a user specific buffer level of the serving RN (see the description of *ProSched* for RECs on how the buffer is implemented).

16.4.2. Discussion

The basic *ProSched* algorithm exists in variations with different complexities. The version simulated here uses rank one reduced bases of the users' subspaces to compute their null space projection matrices, which was originally meant for complexity reduction. In addition it allows to reduce considerably the overhead data to be transmitted to the central intelligence, see also Section 16.3. The final subset selection is performed on the actual estimated Shannon rate rather than on the metric, which was identified as the *ProSched* version with best performance and complexity trade-off in the previous part of this work.

To reduce simulation time, the time tracking version of the algorithm was used. Since the WINNER channel model is not time-continuous, the tracking was reset at the start of each new drop, i.e., after a certain number of statistically generated channel samples with the same user positions.

The precoding scheme used together with *ProSched* is Successive MMSE (SMMSE) with Dominant Eigenmode Transmission (DET) [38], which is one of the preferred schemes in WINNER [30]. The performance is compared to different flavors of the reference method of Section 16.4.1. The first one is the method as given, featuring no coordination but optimal DPC precoding. The second one uses the same simulation steps but with SMMSE instead of DPC. It was added to show that the optimum precoder performance differs only marginally from that of SMMSE with DET. To that extend, any difference between the reference method without coordination and *ProSched* including coordination can be solely attributed to the interference avoidance and not to the use of different precoders. In addition it can be seen that the reference performance with SMMSE DET differs hardly from that of DPC. This can be attributed first of all to the good performance of SMMSE in general, but also to the user channels being low rank in the WINNER model, causing the DPC to use only one spatial mode most of the time, as with SMMSE DET.

From Figure 16.2 showing the CCDFs of the total system rate it can be seen that *ProSched* performing joint interference avoidance together with the low complex precoder SMMSE increases the probability of achieving high rates in a wide range of rate values. It suffers a slight drawback in peak throughput, likely due to the suboptimality of the precoder. Recall that precoding is done separately for each transmitter but that the presented coordinated scheduler takes into account the predicted interference which depends on the selection of the users. The interference generated is different for each possible user assignment requiring different precoding matrices at each transmitter. However, the *ProSched* interference prediction scheduling requires no additional matrix decompositions during the testing of combinations. Finally, the perfor-

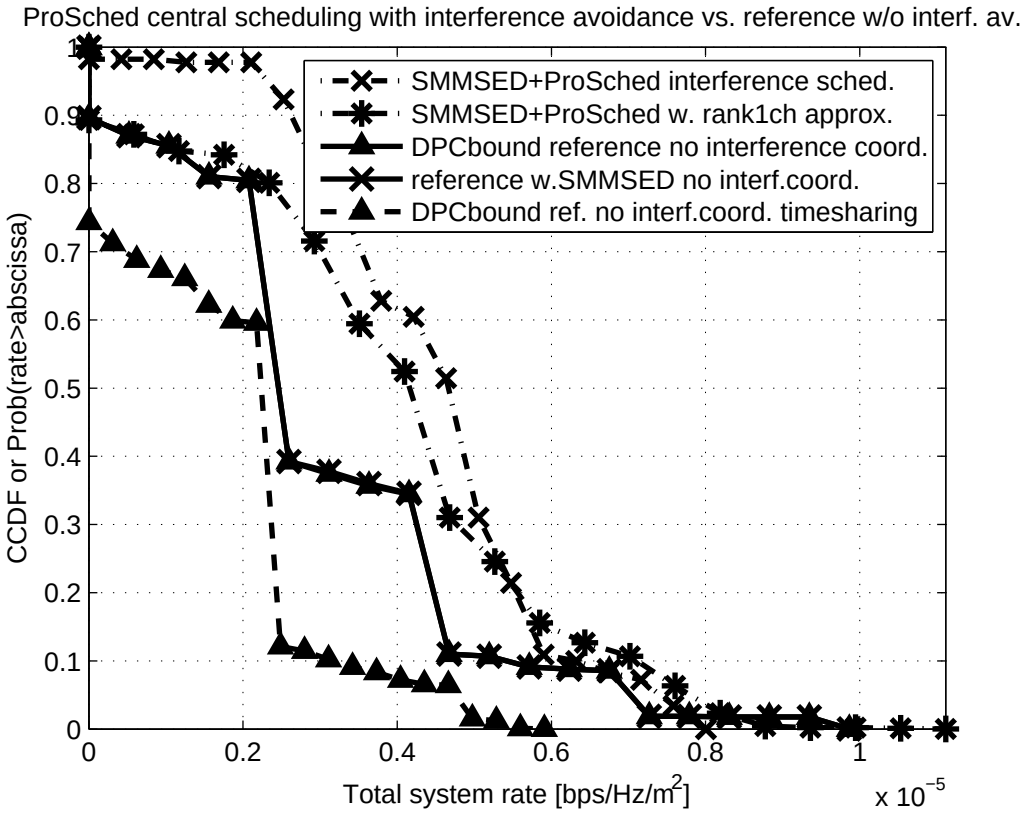


Figure 16.2.: CCDF of total system rate in a REC: *ProSched* interference coordination provides a gain in a wide range of values except for peak rates where the reference without coordination has a slight edge due to the optimal DPC precoder. The gain is entirely due to the coordination, as can be seen from the fact that the DPC reference performs identically to the SMMSE DET based reference. It is still significant when scaled rank one channels are projected in *ProSched* to reduce signaling overhead.

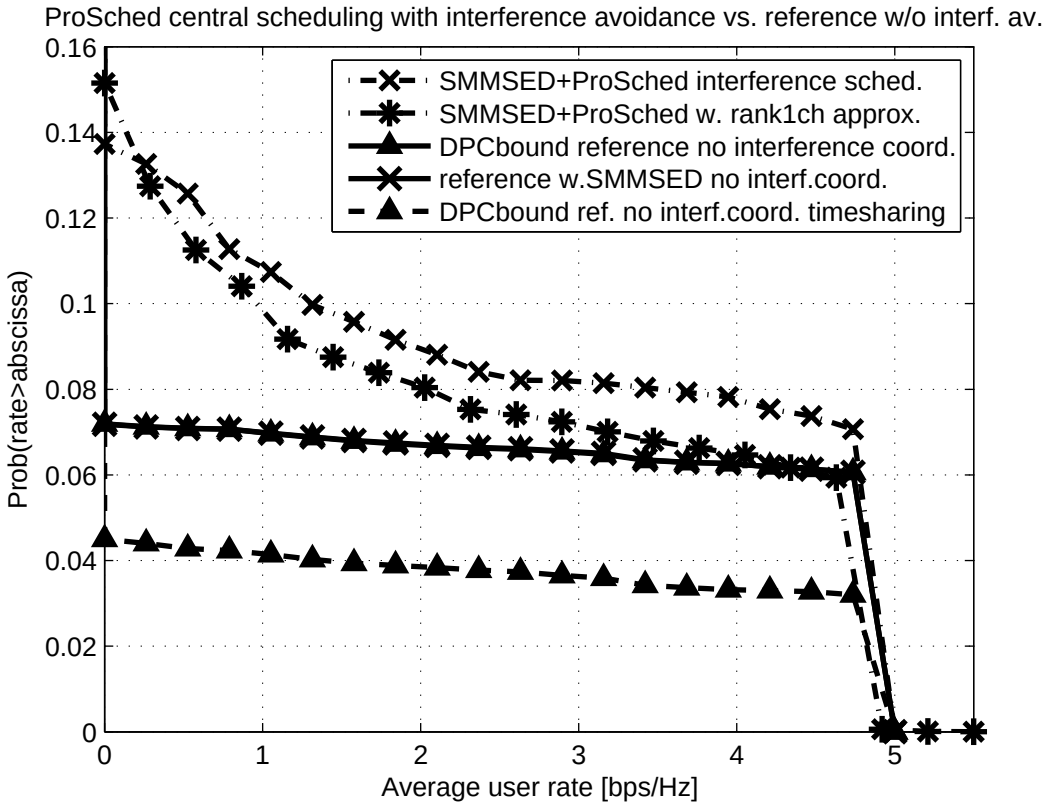


Figure 16.3.: Average user rate in a REC with *ProSched* interference coordination: Similar conclusions as for the total system rate apply. The slope of the reference performance suggests that it might provide higher probabilities for high rates beyond the MCS limit. The overall probability of being served is low due to the lack of system resources.

mance of the proposed reference method is shown taking into account the time sharing between two groups of RNs as described in [95] as a simple interference reduction scheme. In this setting it was observed that this may even reduce the achievable rates compared to the case without any interference avoidance.

Figure 16.3 shows the average user rate. The fact that our reference method limits the number of subcarriers to one causes a lack of system resources in relation to the number of active users as can be seen from the overall low probability of obtaining any throughput at all. Other than this, the same conclusions apply as for the total throughput. From the slope of the reference performance it could be concluded that it should provide higher probabilities for high rates which is, however, not visible due to the MCS limit. The observations described here are heavily scenario dependent. An even higher gain due to interference avoidance should be expected in a more populated scenario offering a higher selection diversity.

17. Summary

In this chapter, a *ProSched* low-complexity scheduling solution was developed for inter-site interference coordination scheduling. An efficient approximation of the interference power was introduced into the scheduling metric with the help of orthogonal projection matrices that can be reused from the single-BS algorithm and, thus, do not represent a significant increase in computational complexity. To find the best user and/or RN combinations to be served simultaneously, an extended tree-based search algorithm was presented in three different variations. One version can be used to treat all transmitting station jointly as a virtual MIMO system. The second can schedule users to multiple independent BSs cooperatively. And finally an approach to schedule links in a REC with SDMA was given. The latter was fully developed only for the two-hop case but directions were given for the case with more hops.

Solutions for soft- and hard handover of users between BSs were described for the coordinated scheduling approach. As of now, which solution to apply is to be decided depending on the scenario, keeping in mind an optimization of antenna positions for the virtual MIMO system by the operator.

Several side results were presented: A method to reduce the algorithm complexity by scheduling only once per subcarrier based on an averaged covariance matrix was investigated. To assess the performance of *ProSched* in a REC with half duplex SDMA RNs, a reference simulation method was conceived to obtain an estimate of the system sum rate when no interference coordination is performed. An approach to reduce the signaling overhead needed for coordinated scheduling was discussed.

All methods were tested with the help of simulations showing that interference avoidance scheduling can provide a gain in system throughput in addition to the gains provided when selecting spatially uncorrelated users to be served simultaneously by the SDMA scheme. *ProSched* can furthermore be used to take into account various QoS constraints as presented in Part II.

Part IV.

Adaptive Information Reuse and a Model for Iterative Channel Estimation in Turbo Equalization for MIMO OFDM

18. Introduction

When coded information is transmitted over a channel with spatial, temporal or inter-carrier interference, the otherwise undesired interfering signal components can be exploited as a useful source of redundancy by applying the idea of Turbo Decoding to equalization. The channel is regarded as an additional (implicit) inner encoder, which is serially concatenated with the (explicit) channel coder. At the receiver, two corresponding decoding blocks iteratively exchange their information.

Research such as [101, 102] and earlier articles in the context of Code Division Multiple Access (CDMA) have alleviated the originally intractable computational complexity of the Turbo equalization approach by replacing the inner maximum a-posteriori (MAP) channel equalizer by channel adaptive filters or a Soft Interference Cancelation (SIC) step.

This part of the work develops several new methods for Turbo Equalization in the context of Multiple Input Multiple Output (MIMO) Orthogonal Frequency Division Multiplexing (OFDM) systems with focus on the spatial interference components. A general introduction to MIMO OFDM can be found in Part I along with a summary of the contributions of the other parts of the thesis. The example system and notation in this part broadly follow [103], where a Minimum Mean Squared Error (MMSE) filter combined with SIC acts as the inner equalizer at the receiver.

The contributions are threefold:

Firstly, an adaptive algorithm for reusing a-priori information across Turbo iterations is proposed (Chapter 20), and its performance gain over the conventional scheme is demonstrated. As an enabler for the algorithm, Extrinsic Information Transfer Charts [104] are briefly revisited.

Secondly, we model the effect of channel estimation errors at the receiver on the performance of the Turbo equalizer (Chapter 21). Specifically, a practical semi-blind channel estimation scheme is considered, where after each Turbo iteration, new reliably decoded data symbols are used to improve the channel estimate. We show how the behavior of this scheme can be captured in a standard Extrinsic Information Transfer (EXIT) chart.

Finally, we investigate the combination of a Turbo equalizer with Singular Value Decomposition (SVD) based MIMO precoding. It is shown that, although precoding and scheduling as in Parts II and III follow a complementary approach of balancing and suppressing spatial interference, the combination with a a Turbo equalizer is worth while. This is the case because in realistic systems, Channel State Information (CSI) at the transmitter (CSIT) is imperfect and can cause residual inter-stream interference that is sufficiently pronounced to justify a Turbo equalizer.

All three aspects are illustrated through simulation results.

19. System model

In our examples we use a MIMO link with OFDM modulation. It is assumed that the cyclic prefix is longer than the channel delay and that synchronization is perfect. It is known that this results in a frequency flat MIMO channel $\boldsymbol{H}_n \in \mathbb{C}^{M_\mathrm{R} \times M_\mathrm{T}}$ on every subcarrier with integer number $n \in [1, \ldots, N_\mathrm{c}]$ having a receive vector

$$\boldsymbol{\gamma}_n = \boldsymbol{H}_n \boldsymbol{M}_n \boldsymbol{\beta}_n + \boldsymbol{n}_n \in \mathbb{C}^{M_\mathrm{R} \times 1} \,. \tag{19.1}$$

A number of parallel information bit streams $\boldsymbol{u}_s$, $s \in [1, \ldots, r]$, are independently convolutionally encoded, interleaved, and then mapped to complex baseband constellations to form the data streams $\boldsymbol{b}_s$. A block of r parallel symbols at position n in the coded streams is spatially multiplexed and forms a transmit vector $\boldsymbol{\beta}_n = \begin{bmatrix} b_{1,n} & \cdots & b_{r,n} \end{bmatrix}^\mathrm{T}$. The absolute time index in the stream increases with the number of the current OFDM symbol, but for simplicity we only use one index n. This notation may both represent a multi-user downlink with non-cooperating receivers having M_R antennas each, or an uplink with a total of M_T transmit antennas at the terminals. Also for simplicity we restrict ourselves to Binary Phase Shift Keying (BPSK) modulation, i.e., the elements of $\boldsymbol{\beta}_n$ are $\in \{-1, 1\}$. The power of a transmit vector is, thus, $\mathrm{trace}\left\{\boldsymbol{\beta}_n \boldsymbol{\beta}_n^\mathrm{H}\right\} = r$.

We allow for a precoding matrix $\boldsymbol{M}_n \in \mathbb{C}^{M_\mathrm{T} \times r}$ to adjust phase and amplitude of the transmitted data symbols. It includes the allocation of the total transmit power P_T on the symbols such that $\sum_{n=1}^{N_\mathrm{c}} \mathrm{trace}\left\{\boldsymbol{M}_n \boldsymbol{M}_n^\mathrm{H}\right\} = P_\mathrm{T}$ (as was before). The total noise power per receive antenna is σ_N^2, and the noise vectors $\boldsymbol{n}_n$ have independently complex Gaussian distributed elements following $\mathcal{CN}\left(0, \frac{\sigma_\mathrm{N}^2}{N_\mathrm{c}}\right)$.

Before the receive vectors are concatenated to form received streams $\boldsymbol{y}_s$ and the streams are decoded, an MMSE-SIC step according to [103] is performed (see next chapter) on every subcarrier after OFDM demodulation.

This OFDM based system model was chosen to study the behavior of a Turbo equalizer when the only interference present is between the spatially multiplexed receive symbols. It resembles the notation used before except for the fact that the transmitted vector is now called $bm\beta$ instead of $\boldsymbol{d}$ to be more in-line with existing literature in the field and to emphasize that, here, the transmitted symbols b stems from coded information bits, which was not necessarily the case before.

20. A modification of the Turbo principle

20.1. Introducing adaptive a-priori information reuse

In a Turbo structure, the building blocks have to be able to provide soft estimates $\hat{b}_{s,n}$ of the transmitted bits, commonly in the form of log-likelihood ratios (L-values) defined such as $L(\hat{b}_{s,n}) = \ln\left\{\frac{p(b_{s,n}=1|\boldsymbol{y}_s)}{p(b_{s,n}=-1|\boldsymbol{y}_s)}\right\}$ for each received stream $\boldsymbol{y}_s$. In the case of systematic codes this L-value can be decomposed into three distinct terms as stated, e.g., in [105]:

$$L(\hat{b}_{s,n}) = L_\mathrm{c}(y_{s,n}|b_{s,n}) + L_\mathrm{e}(\hat{b}_{s,n}) + L_\mathrm{a}(b_{s,n})\,. \tag{20.1}$$

The channel L-value $L_\mathrm{c}(y_{s,n}|b_{s,n}) = \ln\left\{\frac{p(y_{s,n}=1|b_{s,n}=1)}{p(y_{s,n}=-1|b_{s,n}=-1)}\right\}$ reflects the channel quality. The second L-value is called extrinsic because it contains the information that other symbols in the received codeword give on the value of the symbol n. The a-priori information $L_\mathrm{a}(b_{s,n}) = \ln\left\{\frac{p(b_{s,n}=1)}{p(b_{s,n}=-1)}\right\}$ about the occurrence of a symbol is delivered as an estimate by the other decoder in the system.

In traditional Turbo decoding, not the full $L(\hat{b}_{s,n})$ that was produced in the current building block is used as a-priori information in the next block. Instead, the previous a-priori value that was used for generating $L(\hat{b}_{s,n})$ is subtracted to prevent passing (possibly erroneous) information back to the decoder from which it was produced [105]. Also, L_c is subtracted to remove "common" information. Only the newly produced extrinsic information is passed on to the next block.

In our approach, however, we abstain from subtracting L_a as soon as the a-priori L-values contain more mutual information than the current iteration was able to produce, i.e., when $I_\mathrm{a} \geq I_\mathrm{e}$. This situation may occur during the iterations and may obviously be exploited to increase convergence speed and reduce the risk of early iterations getting stuck. However, previous proposals in [106, 107] are unable to detect this point. Instead, they suggest to re-use L_a in all iterations with an empirically determined constant weighting factor < 1, which implies a re-use of a-priori information also in early iteration even when it still violates the condition $I_\mathrm{a} \geq I_\mathrm{e}$, and does not permit to fully re-use the a-priori information in higher iterations. We propose to track the iterations with the help of EXIT charts to enable a full re-use as soon as it is beneficial for both the MMSE-SIC filter and the decoder. The latter is also in contrast to [106, 107] where the effect of a constant re-use is shown only for the filter stage.

20.2. EXIT charts revisited

An EXIT chart consists of Mutual Information Transfer Functions (MITFs) $I_{\mathrm{a,in}} \rightarrow I_{\mathrm{e,out}}$ for each Turbo building block which are combined to one chart such that the output of one block becomes the input of the next. One may then predict the iteration behavior by connecting the curves with straight lines, starting from $I_{\mathrm{a,in}} = 0$. The required MITFs can be estimated because of the observation that the distribution of L_{e} at the output of a MAP decoder and a MMSE-SIC filter can be approximated with a single parameter normal distribution $f_{L|B=\pm 1} = \mathcal{N}\left(\pm\sigma_{\mathrm{i}}^2/2, \sigma_{\mathrm{i}}^2\right)$ [104, 108]. To obtain such estimated transfer curves, a Turbo building block is fed with artificial sequences of L-values having mutual information

$$I(L,b) = \frac{1}{2} \sum_{b\in\{-1,1\}} \int_{\mathbb{R}} f_{L|B=b}(l) \, \log \frac{2 \, f_{L|B=b}(l)}{f_{L|B=+1}(l) + f_{L|B=-1}(l)} \, \mathrm{d}l \tag{20.2}$$

in the entire range $[0, 1]$, generated by varying σ_{i} in the above distribution. For each value of σ_{i}, a histogram of the output L-value distribution $f_L(l)$ is recorded and integrated to obtain the corresponding mutual information that the L-value as a continuous random variable has about the discrete transmitted symbols. To invert $I\left(L, b\right)$ (in order to obtain the matching σ_{i} for a given I) it can be solved numerically and a table of values created [109]. Approximations are often used in the literature to allow for an easy inversion.

In MMSE-SIC Turbo equalization, a method to estimate L-values at the output of the MMSE-SIC block is needed [110] (see also next section) which can further be exploited to directly obtain σ_{i} for the EXIT chart generation, replacing the need for L-value measurements (helpful for real time applications but not used here). Note that the MITFs of the MMSE-SIC filter are strongly dependent on the channel and may, therefore, be averaged along the subcarriers. They also need to be generated for each Signal to Noise Ratio (SNR) operating point.

20.3. Iterative receiver summary

The receiver consists of an MMSE-SIC block (acting as a replacement for the inner decoder) followed by decoders for each stream where we use the BCJR algorithm for convolutional codes [111] in the numerically more stable Log-MAP form [112]. Flowcharts for the two blocks are given in Figures 20.1. Recall that one stream's decoder operates on a block of symbols, collected from multiple subcarriers or even multiple OFDM symbols depending on the channel and code parameters, whereas a filter block exists for every subcarrier.

In the following description we omit the iteration index for simplicity:

As a first step, expected values of the transmitted symbols are generated from the L-values

(a) The MMSE-SIC block for one subcarrier n.

(b) The decoder block.

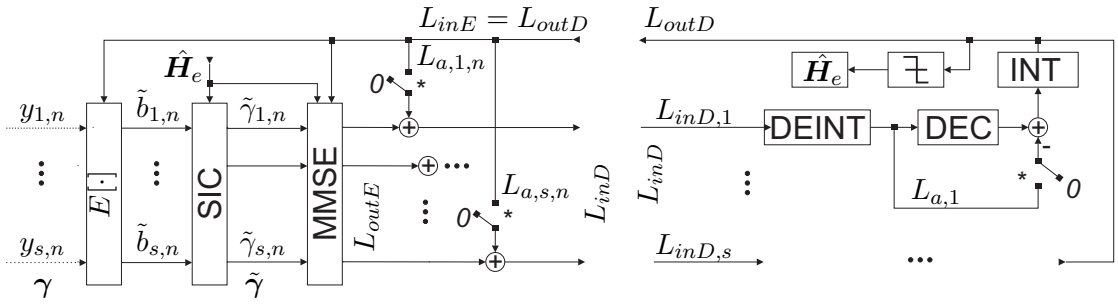


Figure 20.1.: Block diagrams of the MMSE-SIC and decoder parts of the iterative receiver. The adaptive a-priori information re-use mechanism is shown as switches marked with an asterisk.

input to the MMSE-SIC block L_{inE}:

$$\begin{aligned}\tilde{b}_{s,n} &= (-1)\cdot p(b_{s,n}=-1|y_{s,n}) + (+1)\cdot p(b_{s,n}=+1|y_{s,n}) \\ &= \tanh\left(L_{\mathrm{inE},s,n}/2\right) , \forall s \in [1,\ldots,r] , n \in [1,\ldots,Nc] \; . \end{aligned} \tag{20.3}$$

The input information comes from the decoder block except in the first iteration where the symbols $+1, -1$ are assumed to be equally likely and the $L_{\mathrm{inE},s,n}$ are set to zero. (The second line of equation (20.3) holds with the definition of L-values, but only when the symbols are uncorrelated due to ideal interleaving.) These estimates are then used to create replicas of the spatial interference components present in each element of a receive vector by passing $\tilde{\boldsymbol{\beta}}_{s,n} = \left[\tilde{b}_{1,n},\ldots,\tilde{b}_{s-1,n},0,\tilde{b}_{s+1,n},\ldots,\tilde{b}_{r,n}\right]^{\mathrm{T}}$ through an estimate $\hat{\boldsymbol{H}}_{\mathrm{e},n}$ of the effective channel $\boldsymbol{H}_{\mathrm{e},n} = \boldsymbol{H}_n\boldsymbol{M}_n$ that acted on the transmitted symbols, leaving out $\tilde{b}_{s,n}$ corresponding to the current transmit symbol. (We refer to this estimate also as Channel State Information at the Receiver (CSIR)). The replicas are then subtracted from the received symbols:

$$\tilde{y}_{s,n} = y_{s,n} - \hat{\boldsymbol{H}}_{\mathrm{e},n}\tilde{\boldsymbol{\beta}}_{s,n} \; . \tag{20.4}$$

To come to an estimate $\hat{b}_{s,n} = \boldsymbol{w}_{s,n}^{\mathrm{H}}\tilde{\boldsymbol{\gamma}}_n$ where more of the remaining interference has been suppressed, a Wiener MMSE filter vector $\boldsymbol{w}_{s,n}$ is employed such that $\mathrm{E}\left\{\left\|\boldsymbol{w}_{s,n}^{\mathrm{H}}\tilde{\boldsymbol{\gamma}}_n - b_{s,n}\right\|^2\right\} \rightarrow \min$. Note that there is no gain from defining the filter over multiple adjacent carriers due to the assumption of ideal interleaving leading to uncorrelated symbols and due to orthogonal carriers. Following, e.g., [110], the optimal weight vectors for the current iteration, assuming that the

channel estimate is perfect, are

$$\boldsymbol{w}_{s,n}^{\mathrm{opt}} = \left(\hat{\boldsymbol{H}}_{\mathrm{e},n}\boldsymbol{\Lambda}_{s,n}\hat{\boldsymbol{H}}_{\mathrm{e},n}^{\mathrm{H}} + \hat{\sigma}_{\mathrm{N}}^2/N_{\mathrm{c}}\boldsymbol{I}\right)^{-1}\hat{\boldsymbol{H}}_{\mathrm{e},n}\boldsymbol{e}_s \tag{20.5}$$

where $\boldsymbol{e}_s$ is a canonical basis vector of $\mathbb{R}^r$ used to select the respective column from $\hat{\boldsymbol{H}}_{\mathrm{e},n}$ and

$$\boldsymbol{\Lambda}_{s,n} = \mathrm{E}\left\{(\boldsymbol{\beta}_n - \tilde{\boldsymbol{\beta}}_n)(\boldsymbol{\beta}_n - \tilde{\boldsymbol{\beta}}_n)^{\mathrm{H}}\right\} \tag{20.6}$$

is a covariance matrix that is diagonal due to the symbols being independent. The elements of $\boldsymbol{\beta}_n$ have $\mathrm{E}\{b_{s,n}\} = \tilde{b}_{s,n}$ as above and variance $1 - \tilde{b}_{s,n}^2$ since they follow a discrete distribution. This holds also for $\boldsymbol{\beta}_n - \tilde{\boldsymbol{\beta}}_n$ because of $\tilde{\boldsymbol{\beta}}_n$ being a constant shift. Commonly, the covariance of element number s is disregarded by setting it to one because of the notion that the information used was extrinsic only (provided by other symbols), resulting in $\boldsymbol{\Lambda}_{s,n} = \mathrm{diag}\left(1 - \tilde{b}_{1,n}^2, \ldots, 1 - \tilde{b}_{s-1,n}^2, 1, 1 - \tilde{b}_{s+1,n}^2, \ldots, 1 - \tilde{b}_{r,n}^2\right)$. We stick to this principle even when our proposed modification is in effect. There is room to tune or regularize the MMSE filter at this point, which is not further investigated here. A pragmatic reason to keep this inconsistency is that, to compute EXIT charts the building blocks need to be independent.

As in [110], to generate L-values from the filter output it is modeled as an Additive White Gaussian Noise (AWGN) channel $\hat{b}_{s,n} = \mu_{s,n}b_{s,n} + \eta_{s,n}$ with a real valued transmission factor $\mu_{s,n}$ and $\eta_{s,n} \sim \mathcal{CN}(0, \nu_{s,n}^2)$ of which the extrinsic L-values are readily obtainable and yield

$$\begin{aligned} L_{\mathrm{outE},s,n} &= \frac{4 \cdot \mathrm{Re}\left\{\hat{b}_{s,n}\right\}}{1 - \mu_{s,n}}\,; \text{with}\, \mu_{s,n} = \mathrm{E}\left\{\hat{b}_{s,n}b_{s,n}\right\} \\ \mu_{s,n} &= \mathbf{e}_s^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{e}}{}^{\mathrm{H}}\left(\hat{\boldsymbol{H}}_{\mathrm{e}}{}^{\mathrm{H}}\boldsymbol{\Lambda}_{s,n}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{e}} + \hat{\sigma}_{\mathrm{N}}^2/N_{\mathrm{c}}\boldsymbol{I}\right)^{-1}\hat{\boldsymbol{H}}_{\mathrm{e}}\mathbf{e}_s\,. \end{aligned} \tag{20.7}$$

Since this value is already the extrinsic one, there is no need to subtract any a-priori L-values. Instead, if a-priori knowledge from the decoder is to be re-used as we propose in Section 20.1, it has to be added. This may be different for the decoder depending on the algorithm chosen. With the BCJR, a-priori knowledge is contained in the output $L_{\mathrm{outD},s,n}$ and needs to be subtracted. Note that, to generate an MITF the variances of the $L_{\mathrm{outE},s,n}$ do not need to be measured from a histogram. They may be computed due to the above AWGN model as $\sigma_{i,outE}^2 = \sqrt{8\mu_{s,n}/\left(1 - \mu_{s,n}\right)}$.

When a desired convergence criterion is reached, e.g., in terms of EXIT chart predicted bit error rate [109], hard decisions of either the MMSE-SIC or the decoder output L-values may be generated (in our case this is only a thresholding operation) and these values then convolutionally decoded.

21. Influence of channel estimation errors

In a Frequency Division Duplexing (FDD) system, an estimate $\hat{\boldsymbol{H}}_n$ of the channel for the transmit precoder may be obtained via feedback of a pilot assisted measurement, causing a delay before it can be used. In a Time Division Duplexing (TDD) system, measurements on the other link direction are often assumed to be reciprocal and used for precoding on the current link. According to [113], the residual error of a Least Squares (LS) estimator is additive when the noise in the system model is complex Gaussian and when the elements of $\boldsymbol{H}_n$ are i.i.d., i.e., $\hat{\boldsymbol{H}}_n = \boldsymbol{H}_n + \boldsymbol{E}_n$ where the entries of $\boldsymbol{E}_n$ are i.i.d. $\sim \mathcal{CN}\left(0, \frac{\sigma_\mathrm{N}^2/N_\mathrm{c}}{N_{p,n} \cdot P_p}\right)$. The pilot power P_p on the reverse link may differ from the transmit power on the precoded link. For simplicity we assume it is the same, arguing that the noise figure at the reverse link receiver may also be lower if it is, e.g., a base station. The parameter $N_{p,n}$ is the number of consecutively available pilot symbols in time direction while the channel can be considered unchanged (assuming it is the same on each antenna or spatial stream). This model is widely used also for correlated channels and may be extended with an empirical error floor to model interpolation between a sparse pilot grid [30].

We propose to model also the error in $\hat{\boldsymbol{H}}_{\mathrm{e},n}$ used at the MMSE-SIC receiver with this approach. In our notation, the transmit power allocation is contained in $\boldsymbol{H}_{\mathrm{e},n}$ through $\boldsymbol{M}_n$ and, therefore, $P_p = 1$. In the following it is described how this receiver estimation error approach can be extended to also include the effect of the well known concept of semi-blind channel estimation in an EXIT chart. This could serve as a tool to optimize parameters of semi-blind estimation, such as symbol reliability thresholds based on EXIT charts (not investigated here). The extension consists in using the error model for the channel estimate during the MITF generation for the MMSE-SIC block in the following way: When a symbol is reliable enough (e.g., when its expected value exceeds a certain threshold $|\tilde{b}_{s,n}| > c$), it may be used as an additional pilot for estimation to enhance $\hat{\boldsymbol{H}}_{\mathrm{e},n}$ for the next iteration. To reflect this in the output information measurements for an input σ_i, the number of available pilots $N_{p,n}$ in the error model is set based on the number of symbols that have been predicted to be, on average, above the threshold based on the previously simulated value of σ_i. Since the reliability of the symbols from different antennas or spatial layers may differ, an $N_{p,n}$ parameter should be used per layer and the residual estimation error generated separately for each row in the error matrix.

22. Simulation examples

Due to the assumption of ideal interleaving, any frequency selectivity that could affect the code through burst errors would be whitened out. Thus, we only look at the frequency flat case $N_\mathrm{c} = 1$. We use a Kronecker channel model where the receive and transmit correlation matrices have only one parameter [114], set to 0.5. The dimensions are $M_\mathrm{R} = M_\mathrm{T} = 4$. A channel realization is used for 30 consecutive data symbols, inspired by the duration of the rectangular resource chunks in time and frequency used in the Wireless World Initiative New Radio (WINNER) II fourth generation (4G) project [30] which have been defined based on measured coherence times. The SNR is set to $P_\mathrm{T}/\sigma_\mathrm{N}^2 = 5\,\mathrm{dB}$ and the code used is $[7, 5]_8$.

The first example in Figure 22.1 illustrates the effect of the proposed adaptive information reuse with EXIT charts. The system is with perfect CSIR and without any transmit precoding nor power loading, i.e., $\boldsymbol{M}_n = \boldsymbol{I}\sqrt{\frac{P_\mathrm{T}}{N_\mathrm{c} \cdot r}}$ with $r = M_\mathrm{T}$. The average MITF of all spatial layers is displayed for the decoder and the MMSE-SIC as well as the predicted iterations. With information reuse, a greater slope is visible in the MITFs leading to faster convergence. A bigger tunnel between the MITFs is also provided which is important in low SNR situations. Additionally, a comparison was made with the method of fixed a-priori re-use for the MMSE-SIC block only, as in [106, 107], where $0.5L_\mathrm{a}$ of the previous iteration is subtracted in all iterations (see also Chapter 20). In order not to overload the figure, these results are given in the form of Bit Error Rates (BERs) as a function of the number of iterations in Table 22.1, for 5 dB as well as 0 dB. With the newly proposed adaptive reuse the iterations approach the $I = 1$ point much faster and converge at a lower BER.

In the next example in Figure 22.2 we investigate whether it is worthwhile combining precoding based on erroneous Channel State Information at the Transmitter (CSIT) with a Turbo receiver, using the estimation error model discussed in Chapter 21. The information reuse scheme discussed above is applicable in this case also, however, to visualize the effects separately, it is not used in this example. The precoding matrix contains a basis of the eigenmodes of the channel (obtainable, e.g., via a singular value decomposition), weighed with equal power. It is assumed that only one pilot symbol was used to estimate $\boldsymbol{H}_n$ which is a worst case assumption. With precoding, the MMSE-SIC MITF has a slightly lower slope and end point than in the case without precoding, because the precoder already suppresses part of the otherwise useful spatial interference. Under the unrealistic assumption of perfect CSIT, the slope would be even less (not shown) because of the zero interference condition due to the orthogonality of the eigenmodes.

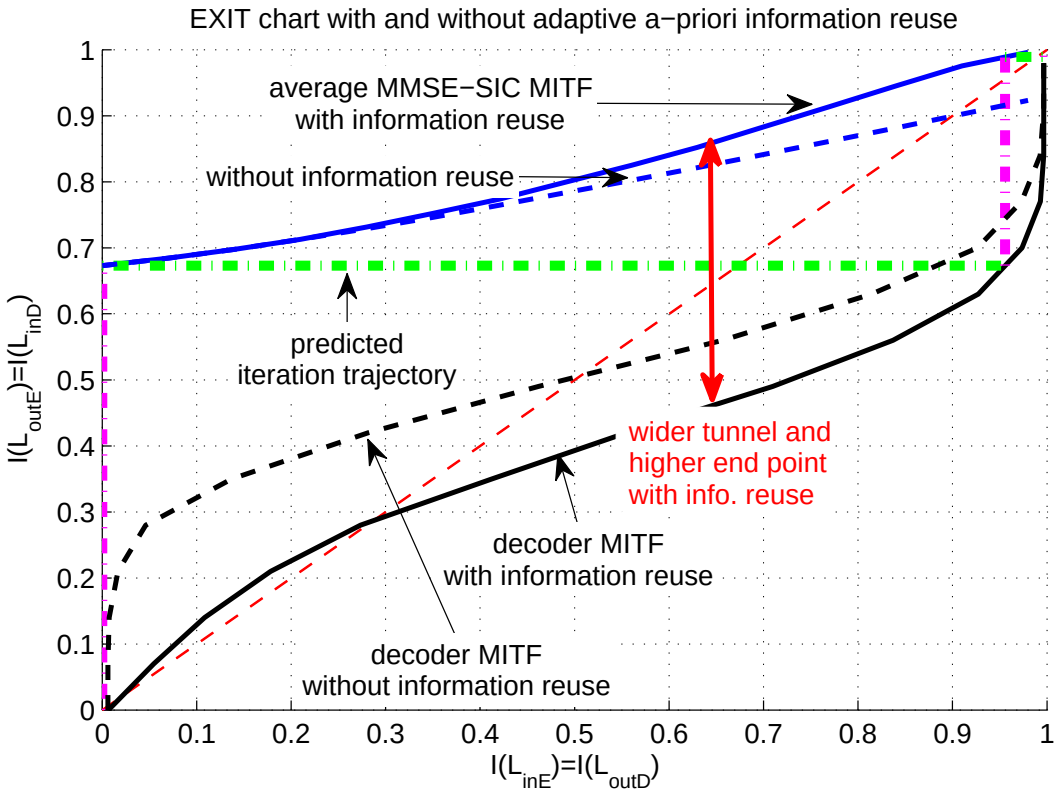


Figure 22.1.: This example illustrates the effect of the proposed adaptive information reuse scheme. A greater MITF slope and a wider tunnel is visible, leading to faster convergence. The $I = 1$ point corresponding to approx. zero BER is only reached in the case with reuse.

BER at 5 dB after iteration	a-priori information re-use		
	none	fixed, filter only	adaptive, filter&dec.
1	$9.398e-2$	$9.396e-2$	$9.396e-2$
2	$2.783e-2$	$7.743e-3$	$2.714e-3$
3	$1.862e-2$	$8.24e-4$	≈ 0
4	$1.856e-2$	$8.21e-4$	≈ 0
5	$1.856e-2$	$8.21e-4$	≈ 0
at 0 dB			
1	0.1686	0.1686	0.1686
2	0.1228	$5.997e-2$	$5.814e-2$
3	0.1041	$4.364e-3$	$3.702e-4$
4	$9.879e-2$	$3.837e-3$	$1.128e-4$
5	$9.789e-2$	$3.836e-3$	$1.128e-4$

Table 22.1.: BERs after each iteration at 5 dB and 0 dB, showing also a result with fixed information re-use which is not in the figures.

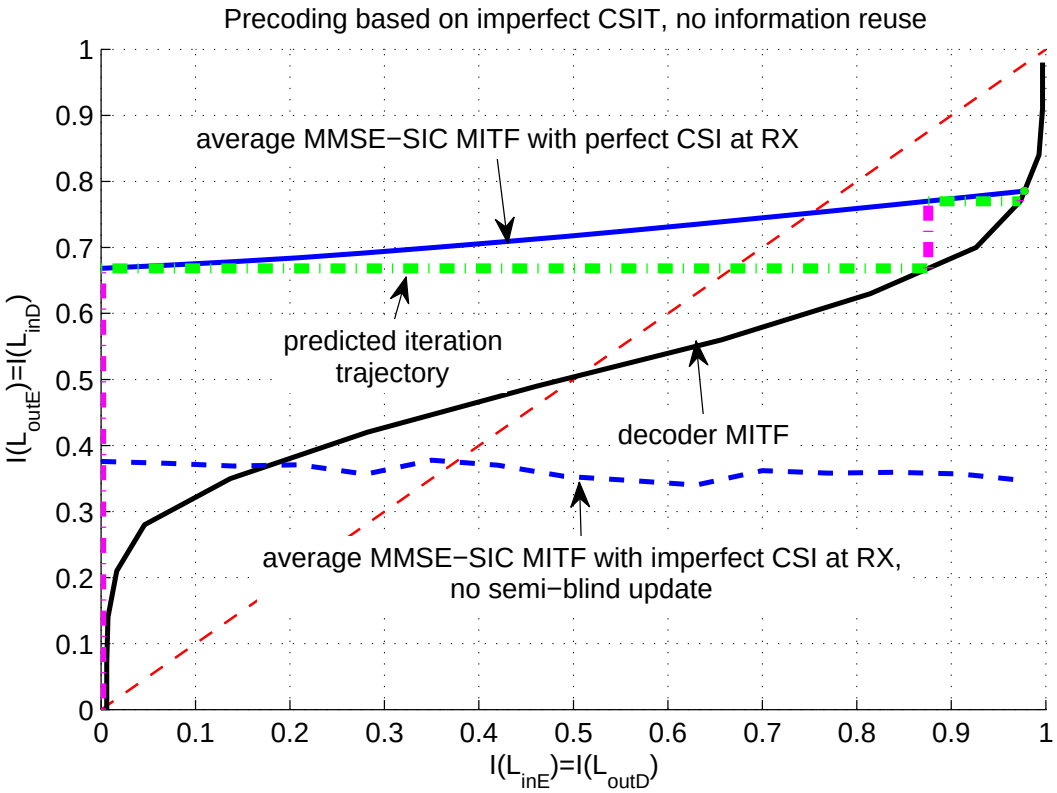


Figure 22.2.: When precoding based on imperfect CSIT is used, the MMSE-SIC MITF has a lower slope because already after the first MMSE filter the residual spatial interference is less than in the case without precoding. In addition, the dashed line shows that CSI errors at the receiver can easily move the filter curve below that of the decoder, prohibiting any iteration gain.

When the CSIR is also imperfect, the MITF starts lower and is almost flat compared to perfect CSIR, because the MMSE filter is not able to diagonalize the channel further due to the mismatched CSIR. Again one transmitted pilot was assumed. (An improvement of the situation is possible with semi-blind estimation, see the next example). We conclude that there is a gain from Turbo iterations even when precoding is used because of the imperfect suppression of interference at the transmitter in the case with imperfect CSIT, offering an opportunity to reduce the number of pilots, but that this gain is easily lost when the CSIR is mismatched. The latter observation is in line with other literature such as [109] where it was already shown that the filter stage is rather sensitive against CSI errors. The third example in Figure 22.3 shows how our proposed model for soft channel estimation during the Turbo iterations acts within the EXIT chart, with $c = 0.5$. The system is the same as in Figure 22.1 with an initial CSI error at the receiver corresponding to one pilot transmitted. Due to this error, the MMSE-SIC MITF

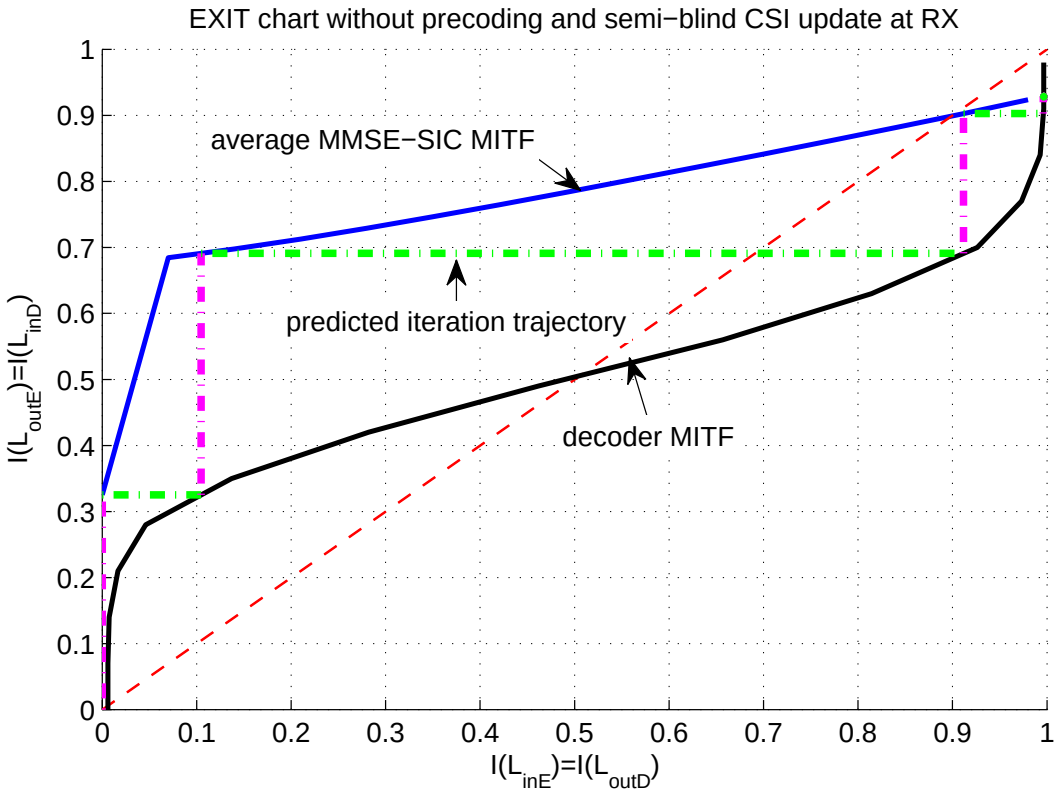


Figure 22.3.: This example without precoding shows the effect of the model reflecting the recomputation of the CSI after every iteration based on reliably detected symbols. Due to the initial CSI error, the MMSE-SIC MITF starts at a lower point than in the case without CSI error but thanks to the constantly improving CSI quality a higher slope of the MMSE-SIC MITF is predicted until all available symbols in the block exceed the reliability threshold.

starts at a lower point but thanks to the constantly improving CSI quality a higher slope of the MMSE-SIC MITF is predicted in the early iterations until all 30 symbols in the block exceed the reliability threshold (i.e., no more new pilots are available). Almost the same end point is reached as in Figure 22.1.

23. Conclusion

The proposed adaptive a-priori information reuse scheme can increase convergence speed and iteration stability of many Turbo systems compared to not reusing a-priori knowledge and compared to a reuse with a fixed weighting factor. The method is based on tracking the convergence of the equalizer with an EXIT chart.

With the help of a channel estimation error model we also investigated if a Turbo equalizer is beneficial when MIMO precoding is performed at the transmitter based on imperfect channel state information leading to residual spatial interference. We can conclude that this interference may be sufficiently pronounced so that Turbo iterations are helpful also in this case, offering a potential to reduce the number of transmitted pilot symbols. In addition it was confirmed again that the filter stages of a Turbo receiver are especially sensitive to CSI errors at the receiver.

Furthermore, a method applicable to OFDM systems to model the effect of semi-blind receiver channel estimation in an EXIT chart was described. These developments were submitted for publication in [14].

Part V.

Conclusions and possible future work

The first two main parts of this work deal with low-complexity approaches for multi-user scheduling in the downlink of Multiple Input Multiple Output (MIMO) systems with Space-Division Multiple Access (SDMA).

When SDMA is employed, each resource element in space, frequency, and time can be used to serve a different subset of terminals. In order not to spoil the complexity advantages that linear channel adaptive precoders have over non-linear ones, the scheduling algorithm should avoid the computation of the precoding weights for all users in all combinations to be tested. It was discussed how selecting the right subset of users can boost the system performance close to the theoretical limit of Dirty Paper Coding (DPC).

The presented algorithm *ProSched* achieves this by using an estimate of a user's Zero Forcing (ZF) precoding rate and a dendrogram based algorithm to traverse the search space. The ZF precoding rate can be considered related to other precoders' rates in the sense that they converge to the ZF solution for increasing Signal to Noise Ratios (SNRs). Using rate estimates, *ProSched* can identify the optimum user subset size while considering the division of the total available transmit power. The detrimental effect of correlated user channel subspaces on the efficiency of the precoding weights is also inherently reflected by the rate estimates.

The name *ProSched* stems from the fact that these rates can be computed with the help of a novel interpretation of ZF precoding which is based on orthogonal projection matrices into the null spaces of other users. The key to the presented complexity reduction is the approximation of the involved projection matrices with the help of so-called repeated projections into the single users' null spaces. The latter are independent of the user combination and can thus, be computed once and re-used throughout one run of the search.

Furthermore, the *ProSched* rate estimates can be conveniently combined with existing Quality of Service (QoS) and fairness mechanisms such as proportional fairness. It was shown how the estimates can be obtained from second order channel statistics both in frequency and time direction when no meaningful instantaneous Channel State Information (CSI) is available, or to reduce the number of scheduling runs by exploiting correlation of neighboring subcarriers.

The presented search algorithm can be used to solve both the user grouping as well as the user selection problem corresponding to different system models. It can be modified to reduce the complexity even further by scheduling all subcarriers jointly with the help of a virtual user concept while tracking the solution in time to exploit temporal correlation.

The uniqueness of the approach and the *ProSched* solution has been patented in Germany and has an international patent pending.

The used approximation techniques are certainly limited by the fact that the exact strength of each user's spatial modes is not known during the scheduling process as presented here. The identified user selection could, thus, be inferior in scenarios where the users' channels differ

greatly in rank and could allow more low-rank users to be scheduled than could be estimated. On the one hand, this could be augmented with the help of modifications such as the following: Virtual users could be generated representing copies of the existing users where the total channel norm is estimated to be condensed into one single mode. In this way, the optimum number of modes or each user could be searched for at the expense of more possible combinations. On the other hand, the significance of such a search can be doubted when the results of the Wireless World Initiative New Radio (WINNER) project are kept in mind. There, one preferred precoding scheme is Successive MMSE (SMMSE) [35] with Dominant Eigenmode Transmission (DET), which is certainly a conclusion of the channel measurements carried out in the project showing rank deficient channels in the scenarios of interest.

Interesting future work could also be to develop a *ProSched* like solution for the uplink, or to verify to which extend the downlink solution can adopted regardless of the fact that in practical systems the sum power constraint present in the downlink does not exist for the uplink transmissions.

In Part III, *ProSched* scheduling solutions were developed for systems with multiple serving stations that can be coordinated by a central intelligence. The core of the chapter is an augmentation of the *ProSched* rate metric by a low complexity estimate of the total interference power that one station generates for users assigned to other stations (inter-cell interference). Although all stations may use adaptive SDMA precoding, the interference estimate does not require any further matrix decompositions. Instead, it re-uses the single user orthogonal projection matrices computed for the single Base Station (BS) solution.

Search algorithms were developed to cover three different deployment strategies: a virtual MIMO system where all antennas in the system are used to precode cooperatively, a system with multiple BS coordinating only their scheduling, and finally Relay Enhanced Cells (RECs) with Time Division Duplexing (TDD) and full duplex SDMA relays. The latter was fully developed for the case of two hops only, with directions given for a more general case. The required overhead to signal CSI from the Relay Nodes (RNs) to the central intelligence was briefly addressed and a method to reduce it with the help of a scaled rank one channel approximation was given. A possibility to implement both hard or soft handover of users between BSs using the very same search tree was shown.

To be able to fully exploit the benefits offered by interference coordination, such systems raise the interest in low latency measurement compression and quantization methods.

As a benchmark, a simulation chain was discussed which is able to give an estimate of the achievable system sum rate of a REC without coordination with the help of optimal DPC algorithms, making the estimate independent of the choice of precoding algorithm.

Issues in Turbo Equalization for MIMO Orthogonal Frequency Division Multiplexing (OFDM)

were investigated in Part IV. Iterative Turbo receivers aim at exploiting interfering signal components as a source of redundancy by passing so-called soft symbol estimates between the building blocks of the receivers. Commonly, a-priori knowledge about symbols is not passed on to the next building block in order to generate new and uncorrelated information in each iteration. Although it is known that this principle can be violated, previously no adaptive control mechanism was available. Our work proposes an adaptation of a feedback coefficient with the help of online Extrinsic Information Transfer (EXIT) charts. In the future it might be of interest to translate this approach to scenarios with only partially or unknown interference components.

Along the lines it was shown that iterative receivers offer a gain in expected Bit Error Rate (BER) even when precoding at the transmitter was used which aims at reducing signal correlation. This is due to the fact that erroneous CSI at the transmitter leads to suboptimal precoding weights resulting in residual interference that is sufficiently pronounced to drive an iterative receiver. Iterative receivers also allow to greatly enhance channel estimation at the receiver (which in turn is beneficial for the precoder). A model for OFDM systems was shown that allows to include the iterative enhancement of CSI in the most commonly used analysis tools for iterative systems, the EXIT chart. The model is constructed around a channel estimation error model for channel matrices with uncorrelated entries. Although the latter has been widely used in the literature also for correlated channels, the derivation of a proper error model for the correlated case is desirable in the future. In the case of single carrier systems, modeling of channel estimation in EXIT charts will require more complex statistical analyses. This is because reliably detected symbols that can serve as additional pilots may occur irregularly in the entire data frame which may span multiple channel coherence times as opposed to the OFDM symbol based processing where this is, by design, not the case, which allows to produce block wise expected values of the number of reliable symbols in each iteration.

Appendix A.

Glossary of Acronyms, Symbols and Notation

Acronyms

3D	three-dimensional
3G	third generation
4G	fourth generation
AWGN	Additive White Gaussian Noise
BD	Block Diagonalization
BD THP	Block Diagonalization Tomlinson-Harashima precoding
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
CCDF	Complementary Cumulative Distribution Function
CDMA	Code Division Multiple Access
CSI	Channel State Information
CSIR	Channel State Information at the Receiver
CSIT	Channel State Information at the Transmitter
CQI	Channel Quality Indicator
DET	Dominant Eigenmode Transmission
DFT	Discrete Fourier Transform
DM	Diversity Mode
DOA	Direction Of Arrival
DPC	Dirty Paper Coding
EVD	EigenValue Decomposition
EXIT	Extrinsic Information Transfer
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
GoB	Grid of Beams
ICE	Iterative Channel Estimation
LOS	Line Of Sight
LS	Least Squares
LTE	Long Term Evolution
MAP	maximum a-posteriori
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
MITF	Mutual Information Transfer Function
MMSE	Minimum Mean Squared Error
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access

OSTBC	Orthogonal Space-Time Block Code
PACE	Pilot Assisted Channel Estimation
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RBD	Regularized Block Diagonalization
REC	Relay Enhanced Cell
RN	Relay Node
RR	Round Robin
SDMA	Space-Division Multiple Access
SIC	Soft Interference Cancelation
SISO	Single Input Single Output
SM	Spatial Multiplexing
SMMSE	Successive MMSE
SMMSE THP	Successive Minimum Mean Squared Error Tomlinson-Harashima precoding
SNR	Signal to Noise Ratio
SPQR	Sparse Pivoted QR decomposition approximation
SVD	Singular Value Decomposition
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
THP	Tomlinson-Harashima precoding
TxWF	Transmit Wiener Filtering
UCA	Uniform Circular Array
ULA	Uniform Linear Array
WINNER	Wireless World Initiative New Radio
ZF	Zero Forcing
ZMCSCG	Zero Mean Circular Symmetric Complex Gaussian

Symbols and Notation

$a,\ b,\ c$	scalars
$\boldsymbol{a},\ \boldsymbol{b},\ \boldsymbol{c}$	column vectors
$\boldsymbol{A},\ \boldsymbol{B},\ \boldsymbol{C}$	matrices
δ_{ab}	Kronecker δ-symbol, $\delta_{ab} = \begin{cases} 1 & \text{if } b = a \\ 0 & \text{if } b \neq a \end{cases}$
f	frequency
$h,\ \boldsymbol{h},\ \boldsymbol{H}$	channel coefficient, vector, matrix
$\boldsymbol{I}$	Identity matrix
M_{R}, M_{T}	number of antennas at the receiver and transmitter, respectively
n	discrete time index
$\boldsymbol{n}$	complex Gaussian receiver noise vector
N_{c}	total number of OFDM subcarriers
N_{u}	total number of users in the system
$\boldsymbol{M}$	precoding matrix
P_{T}	maximum total transmit power

$\boldsymbol{R}$	correlation matrix
σ_n^2	total noise power per receive antenna in entire band
$\boldsymbol{t}_g$	a vector of complex symbols transmitted to a user number g
T_s	duration of one OFDM symbol
$\tilde{.}$	designates a (channel- or projection) matrix of other users than the current
$.^{(0)}$	designates a basis for or projection into a nullspace
$.^{(1)}$	designates a basis for or projection into a signal space
$.^\mathrm{T}$	matrix transpose
$.^\mathrm{H}$	Hermitian transpose sign
$\|\cdot\|_\mathrm{F}^2$	Frobenius norm
$\mathcal{N}(a, b)$	Normal distribution of a random variable with mean a, variance b
$\mathcal{CN}(a, b)$	complex valued Normal distribution

Symbols specific to Part IV

$\boldsymbol{b}_s$	symbol stream number s before serial-parallel conversion
$\boldsymbol{\beta}_n$	vector of transmit symbols on one subcarrier
c	received pilot symbol reliability threshold parameter
$f_X(x)$	probability density function of a random variable X
$\boldsymbol{\gamma}_n$	receive vector on one subcarrier
$\hat{.}$	denotes an estimated value
$\boldsymbol{H}_n$	frequency flat MIMO channel matrix on subcarrier n
I	mutual information
r	total number of parallel MIMO streams
$L(\cdot)$	log-likelihood value of a received symbol
$\boldsymbol{M}_n$	MIMO precoding matrix for one subcarrier n
n	index of current subcarrier
$\boldsymbol{n}_n$	complex Gaussian receiver noise vector on current subcarrier
$N_{p,n}$	number of consecutive pilots per coherence time on current carrier
$p(\cdot)$	probability of an event
σ_i^2	variance used in approximating an information distribution
$\boldsymbol{u}_s$	uncoded data stream number s before serial-parallel conversion
$\boldsymbol{y}_s$	received streams after serial-parallel conversion and MMSE-SIC filtering

Bibliography

Publications and Technical Documents as Co- or First Author

[1] A. Osseiran, V. Stankovic, E. Jorswieck, T. Wild, M. Fuchs, and M. Olsson, "A MIMO framework for 4G systems: WINNER concept and results," in *Invited paper at the VIII IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Helsinki, Finland, June 2007.

[2] 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, Oct 2004, pp. 145–148.

[3] M. Haardt, M. Fuchs, and G. Del Galdo, "Verfahren zur Auswahl einer optimierten Teilnehmerzahl in Mobilfunksystemen," *Deutsches Patent- und Markenamt, Aktenzeichen 10 2005 046 911.6*, 22. September 2005, internationally pending under number WO 2007/033997 using the title "Method for selection of an optimized number of subscribers in mobile radio systems".

[4] 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, Signal Processing (ICASSP)*, vol. 3, Philadelphia, PA, March 2005, pp. 1121–1124.

[5] ——, "Low complexity space-time-frequency scheduling for MIMO systems with SDMA," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 2775 – 2784, September 2007.

[6] M. Haardt, A. Alexiou, M. Fuchs *et al.*, "WWRF white paper on smart antennas, MIMO systems, and related technologies: Spatial scheduling for MIMO systems with SDMA," in *Proc. of the 14-th Meeting of the Wireless World Research Forum (WWRF)*, San Diego, CA, July 2005.

[7] M. Haardt, V. Stankovic, G. Del Galdo, and M. Fuchs, "Efficient multi-user MIMO downlink precoding and scheduling," IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2005), Dec 2005, puerto Vallarta, Mexico. distinguished keynote presentation, co-author of presentation slides only.

[8] IST-4-027756, WINNER II, "D3.4.1, The WINNER II Air Interface: Refined Spatial-Temporal Processing Solutions," Framework Programme 6, Tech. Rep., November 2006. [Online]. Available: http://www.ist-winner.org/deliverables.html

[9] F. Roemer, M. Fuchs, and M. Haardt, "Distributed MIMO systems with spatial reuse for high-speed indoor mobile radio access," in *Proc. 20-th Meeting of the Wireless World Research Forum WWRF*, Ottawa, Canada, April 2008.

[10] M. Fuchs, G. Del Galdo, and M. Haardt, "Low complexity spatial scheduling ProSched for MIMO systems with multiple base stations and a central controller," in *Proc. International ITG/IEEE Workshop on Smart Antennas (WSA)*, Guenzburg, Germany, March 2006.

[11] IST-4-027756, WINNER II, "D3.5.2, Assessment of relay based deployment concepts and detailed description of multi-hop capable RAN protocols as input for the concept group work," Framework Programme 6, Tech. Rep., June 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[12] ——, "D3.5.3, final assessment of relaying concepts for all CGs scenarios under consideration of related WINNER L1 and L2 protocol functions," Framework Programme 6, Tech. Rep., September 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[13] M. Fuchs, C. Kaes, U. Korger, and M. Haardt, "Comparison of MIMO relaying deployment strategies in a Manhattan grid scenario," in *Proc. International ITG/IEEE Workshop on Smart Antennas (WSA)*, Vienna, Austria, February 2007.

[14] M. Fuchs, B. Bandemer, and M. Haardt, "Adaptive a-priori information reuse and a modeling approach for iterative channel estimation in turbo equalization for MIMO OFDM," December 2007, submitted for review to European Transactions on Telecommunications.

References by Other Authors

[15] "WINNER: Wireless World Initiative New Radio," European 6th Framework programme project. [Online]. Available: http://www.ist-winner.org

[16] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge: Cambridge University Press, 2003.

[17] A. Goldsmith, *Wireless Communications*. Cambridge: Cambridge University Press, 2005.

[18] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 684–702, June 2003.

[19] S. Alamouti, "A simple transmit diversity scheme for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, October 1998.

[20] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 46, no. 7, p. 1804–1824, July 2002.

[21] M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, vol. 29, pp. 439–441, May 1983.

[22] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. IEEE International Conference on Communications ICC 2002*, vol. 3, April 2002, p. 1444–1450.

[23] W. Yu and J. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," in *Proc. 2002 IEEE International Symposium on Information Theory*, June 2002, p. 498.

[24] G. Caire and S. Shamai(Shitz), "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Information Theory*, vol. 49, pp. 1691–1706, July 2003.

[25] 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.

[26] V. Stankovic, M. Haardt, S. Gale, and A. Jeffries, "Linear and non-linear multi-user MIMO downlink precoding," in *Proc. of the 17-th Meeting of the Wireless World Research Forum (WWRF)*, Heidelberg, Germany, Nov. 2006.

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

[28] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, October 2003.

[29] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1570–1580, April 2005.

[30] IST-4-027756, WINNER II, "D6.13.7, WINNER II Test scenarios and calibration cases issue 2," Framework Programme 6, Tech. Rep., December 2006. [Online]. Available: http://www.ist-winner.org/deliverables.html

[31] ——, "D6.13.12, Final CG "wide area" description for integration into overall System Concept and assessment of key technologies," Framework Programme 6, Tech. Rep., November 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[32] ——, "D6.13.11, Final CG "metropolitain area" description for integration into overall System Concept and assessment of key technologies," Framework Programme 6, Tech. Rep., October 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[33] ——, "D6.13.12, Final CG "local area" description for integration into overall System Concept and assessment of key technologies," Framework Programme 6, Tech. Rep., October 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[34] 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)*, January 2004, pp. 403 – 408.

[35] 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, Nov. 2004.

[36] ——, "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, Nov. 2006.

[37] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 3, pp. 528–541, March 2006.

[38] V. Stankovic, "Multi-user MIMO wireless communications," Ph.D. dissertation, Ilmenau University of Technology, March 2007. [Online]. Available: http://www.db-thueringen.de/servlets/DocumentServlet?id=8620

[39] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 953–961, March 2008.

[40] C. Ahara and J. Rossbach, "The scheduling problem in satellite communications systems," *IEEE Transactions on Communications*, vol. 15, no. 3, June 1967.

[41] D. Bartolomé, "Fairness analysis of wireless beamforming schedulers," Ph.D. dissertation, Department of Signal Theory and Communications Universitat Politècnica de Catalunya (UDPC), Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), January 2005, advisor: Dr. A. I. Pérez, http://www.cttc.es/docs/dbartolome/PhDThesis.pdf.

[42] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine 39*, vol. 39, pp. 150–154, February 2001.

[43] L. Becchetti, S. Diggavi, S. Leonardi, A. Marchetti-Spaccamela, S. Muthukrishnan, T. Nandagopal, and A. Vitaletti, "Parallel scheduling problems in next generation wireless networks," in *Proc. ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, 2002, pp. 238–247.

[44] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. of the IEEE Int. Comp. Conf. (ICC95)*, Seattle, WA, June 1995, pp. 331 –335.

[45] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Information Theory*, vol. 48, pp. 1277–1294, June 2002.

[46] P. Svedman, S. K. Wilson, L. J. Cimini, Jr., and B. Ottersten, "A simplified opportunistic feedback and scheduling scheme for OFDM," in *Proc. IEEE 59th Vehicular Technology Conference, VTC 2004-Spring* , vol. 4, Milan, Italy, May 2004, pp. 1878 – 1882.

[47] D. Avidor, J. Ling, and C. Papadias, "Jointly opportunistic beamforming and scheduling (JOBS) for downlink packet access," in *IEEE International Conference on Communications (ICC 2004)*, June 2004.

[48] P. W. C. Chan and R. S. K. Cheng, "Optimal power allocation in zero-forcing MIMO-OFDM downlink with multiuser diversity," in *Proc. 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.

[49] 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.

[50] D. Bartolome, A. Pascual-Iserte, and A. I. Perez-Neira, "Spatial scheduling algorithms for wireless systems," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, China*, May 2003.

[51] T. Maciel, "Suboptimal resource allocation for multi-user MIMO-OFDMA systems," Ph.D. dissertation, Technical University of Darmstadt, Germany, April 2008, submitted for approval.

[52] T. Maciel and A. Klein, "A low-complexity SDMA grouping strategy for the downlink of multi-user MIMO systems," in *Proc. International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Helsinki, Finland, September 2006.

[53] ——, "A convex quadratic SDMA grouping algorithm based on spatial correlation," in *Proc. IEEE International Conference on Communications (ICC)*, June 2007.

[54] F. M. Wilson and A. W. Jeffries, "Adaptive SDMA downlink beamforming for broadband wireless networks," in *Proc. Wireless World Research Forum Meeting 14, San Diego, CA*, July 2005.

[55] Z. Shen, R. Chen, J. G. Andrews, J. R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3658 – 3663, September 2006.

[56] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Communications Magazine, special issue on MIMO Systems*, vol. 42, pp. 60–67, October 2004.

[57] A. Tarighat, M. Sadek, and A. H. Sayed, "A multi user beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios," in *IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP), 2005, Philadelphia, PA*, March 2005.

[58] Z. Shen, R. Chen, J. G. Andrews, J. R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct. 30 - Nov. 2, 2005.

[59] ——, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *submitted for publication to IEEE Transactions on Signal Processing*, September 2005.

[60] G. Dimic and N. C. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Trans. on Signal Processing*, vol. 53, no. 10, pp. 3857–3868, Oct. 2005.

[61] T. Wild, "Successive user insertion for long-term adaptive beams using short-term CQI," in *InOWo*, Hamburg, Germany, August 2006.

[62] 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.

[63] I. Halperin, "The product of projection operators," *Acta. Sci. Math. (Szeged)*, vol. 23, pp. 96–99, 1962.

[64] M. Berry, S. Pulatova, and G. Stewart, "Computing sparse reduced-rank approximations to sparse matrices," *Univ. of Tennessee Dep. of Computer Science Technical Reports*, vol. 525, 2004.

[65] IST-2003-507581 WINNER, "D5.4, Final Report on Link Level and System Level Channel Models," Framework Programme 6, Tech. Rep., November 2005. [Online]. Available: http://www.ist-winner.org/deliverables_older.html

[66] IST-4-027756, WINNER II, "D1.1.2, WINNER II channel models," Framework Programme 6, Tech. Rep., September 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[67] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," in *Proc. of the Fifth European Wireless Conference EW*, Barcelona, Spain, February 2004.

[68] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operat. Res. Society*, vol. 49, pp. 237–252, 22. September 1998.

[69] J. Holtzman, "Asymptotic analysis of the proportional fair algorithm," in *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 2, 2001, pp. 33–37.

[70] M. Schacht, "System performance gains from smart antenna concepts in CDMA," Ph.D. dissertation, University Duisburg-Essen, Germany, February 2005, appearing in: Selected Topics on Communications, Shaker Verlag.

[71] P. Svedman, S. K. Wilson, and B. Ottersten, "A QoS-aware proportional fair scheduler for opportunistic OFDM," in *Proc. IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall*, vol. 1, 2004, pp. 558 – 562.

[72] A. Tulino, A. Lozano, and T. Guess, "Capacity-achieving input covariance for single-user multiantenna channels," *IEEE Trans. on Wireless Communications*, 2005.

[73] V. Stankovic and M. Haardt, "Multi-user MIMO downlink beamforming over correlated MIMO channels," in *Proc. International ITG/IEEE Workshop on Smart Antennas (WSA ´05)*, Duisburg, Germany, April 2005.

[74] Q. H. Spencer, "Transmission strategies for wireless multi-user, multiple-input, multiple-output communication channels," Ph.D. dissertation, Brigham Young University, March 2004, http://contentdm.lib.byu.edu/ETD/image/etd378.pdf.

[75] "The *IlmProp* - a flexible geometry-based Multi-User MIMO Channel Modelling framework." More information on the model, as well as the source code and some exemplary scenarios can be found at http://tu-ilmenau.de/ilmprop.

[76] G. Del Galdo, "Geometry-based channel modeling for multi-user MIMO systems and applications," Ph.D. dissertation, Ilmenau University of Technology, Ilmenau, Germany, 2007, ISBN: 978-3-938843-27-7, available to download at http://www.delgaldo.com.

[77] IST-4-027756, WINNER II, "D1.1.1, WINNER II interim channel models," Framework Programme 6, Tech. Rep., November 2006. [Online]. Available: http://www.ist-winner.org/deliverables.html

[78] ——, "D2.2.3, Modulation and Coding schemes for the WINNER II System," Framework Programme 6, Tech. Rep., November 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[79] P. Mogensen, I. W. N. Kovacs, F. Frederiksen, A. Pokhariyal, K. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the Shannon bound," in *Proc. of the IEEE 65th Vehicular Technology Conference VTC2007-Spring.*, Dublin, April 2007, pp. 1234–1238.

[80] IST-4-027756, WINNER II, "D6.13.1, WINNER II Test scenarios and calibration cases issue 1," Framework Programme 6, Tech. Rep., June 2006. [Online]. Available: http://www.ist-winner.org/deliverables.html

[81] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation," in *Proc. RAWCON*, August 2003.

[82] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Information Theory*, vol. 49, pp. 1912–192, August 2003.

[83] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," *IEEE Transactions on Information Theory*, vol. 50, p. 145–151, January 2004.

[84] S. Jagannathan, C. Hwang, and J. Cioffi, "Resource allocation in uplink multi-carrier MIMO systems for low-complexity transceivers," in *Proc. 41st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, November 2007.

[85] Y. J. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Trans. on Communications*, vol. 53, pp. 107–116, January 2005.

[86] B. Bandemer and S. Visuri, "Resource allocation in uplink multi-carrier MIMO systems for low-complexity transceivers," in *Proc. IEEE International Conference on Communications ICC '07*, Glasgow, UK, June 2007, pp. 785–790.

[87] M. Dohler, A. Gkelias, and H. Aghvami, "A resource allocation strategy for distributed MIMO multi-hop communication systems," *IEEE Communications Letters*, vol. 8, no. 2, pp. 99 – 101, February 2004.

[88] J. Mietzner and P. Hoeher, "On the duality of wireless systems with multiple cooperating transmitters and wireless systems with correlated antennas," in *Proc. 14th IST Mobile & Wireless Communications Summit*, June 2005.

[89] ——, "Equivalence of spatially correlated and distributed MIMO systems," in *Proc. ITG/IEEE Workshop on Smart Antennas (WSA)*, Guenzburg, Germany, March 2006.

[90] "IST-2003-507581 WINNER: D2.10 final report on identified RI key technologies, system concept, and their assessment; Appendix G.2.5 Performance of SMMSE precoding," December 2005.

[91] D. Bartolome and A. I. Perez-Neira, "Multiuser spatial scheduling in the downlink of wireless systems," in *Proc. IEEE SAM04 Conference*, Barcelona, Spain, July 2004.

[92] M. Wódczak, "On routing information enhanced algorithm for space-time coded cooperative transmission in wireless mobile networks," Ph.D. dissertation, Faculty of Electrical Engineering, Poznan University of Technology, March 2006.

[93] ——, "Extended REACT - Routing information Enhanced Algorithm for Cooperative Transmission," in *Proc. 16th IST Mobile and Wireless Communications Summit*, Budapest, Hungary, July 2007.

[94] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 1774–1786, May 2007.

[95] IST-4-027756, WINNER II, "D3.5.1, Relaying concepts and supporting actions in the context of CGss," Framework Programme 6, Tech. Rep., October 2006. [Online]. Available: http://www.ist-winner.org/deliverables.html

[96] M. Herdin and T. Unger, "Performance of single and multi-antenna amplify-and-forward relays in a Manhattan street grid scenario," in *IST Mobile and Wireless Communications Summit*, Mykonos, Greece, June 2006.

[97] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 29–43, January 2005.

[98] C. Lo, S. Vishwanath, and J. R.W. Heath, "Rate bounds for MIMO relay channels using precoding," in *Global Telecommunications Conference, 2005. IEEE GLOBECOM '05*, Dec. 2005.

[99] H. Bolcskei, R. Nabar, O. Oyman, and A. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 1433 – 1444, June 2006.

[100] M. Kobayashi and G. Caire, "Resource allocation in uplink multi-carrier MIMO systems for low-complexity transceivers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP '07*, vol. 3, Honolulu, HI, April 2007, pp. 5–8.

[101] M. Tüchler and J. Hagenauer, "Turbo equalization using frequency domain equalizers," in *Proc. Allerton Conf.*, Monticello, IL, U.S.A., Oct. 2000, pp. 1234 – 1243.

[102] G. Bauch and N. Al-Dhahir, "Reduced-complexity space-time turbo-equalization for frequency-selective MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 819 – 828, Oct. 2002.

[103] T. Abe and T. Matsumoto, "Space-time turbo equalization in frequency-selective MIMO channels," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 3, pp. 469–475, May 2003.

[104] S. ten Brink, "Convergence behaviour of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, no. 10, p. 1727–1737, Oct. 2001.

[105] J. Hagenauer, P. Robertson, and L. Papke, "Iterative ("Turbo") decoding of systematic convolutional codes with the MAP and SOVA algorithms," in *Proc. of the ITG Conf. on Source and Channel Coding (SCC'94)*, Munich, Germany, Oct. 1994, pp. 21–29.

[106] C. Douillard, C. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalisation," *European Transactions on Telecommunications*, vol. 6, no. 5, p. 507 – 511, 1995.

[107] B. Lu and X. Wang, "Iterative receivers for multiuser space-time coding systems," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, p. 2322 – 2335, 2000.

[108] R. Otnes and M. Tüchler, "EXIT chart analysis applied to adaptive turbo equalization," in *Proc. of the 5th Nordic Signal Processing Symposium, NORSIG-2002*, on board Hurtigruten, Norway, October 2002.

[109] M. Tüchler, R. Koetter, and A. Singer, "Turbo equalization: Principles and new results," *IEEE Transactions on Signal Processing*, vol. 21, no. 1, pp. 67–80, January 2004.

[110] D. Reynolds and X. Wang, "Low complexity turbo-equalization for diversity channels," *Signal Processing, Elsevier Science Publishers*, vol. 81, no. 5, pp. 989–995, May 2002.

[111] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, March 1974.

[112] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proc. IEEE International Conf. on Communications (ICC)*, vol. 2, Seattle, WA, June 1995, pp. 1009 – 1013.

[113] D. Samardzija and N. Mandayam, "Pilot assisted estimation of MIMO fading channel response and achievable data rates," in *Proc. DIMACS Workshop on Signal Processing for Wireless Transmission*, Rutgers University, Piscataway, NJ, October 7-9 2002.

[114] A. van Zelst and J. S. Hammerschmidt, "A single coefficient spatial correlation model for multiple-input multiple-output (MIMO) radio channels," in *Proc. URSI XXVIIth General Assembly*, Aug. 2002.

Martin Fuchs-Lautensack

ISBN 978-3-938843-43-7