Channel Estimation and Precoding in Closed-Loop Distributed Multi-Antenna Systems

Von der Fakultät Informatik, Elektrotechnik und Informationstechnik
der Universität Stuttgart zur Erlangung der Würde
eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

vorgelegt von

Thorsten Wild

aus Karlsruhe

Hauptberichter:	Prof. Dr.-Ing. Joachim Speidel
Mitberichter:	Prof. Dr. rer.nat. Friedrich K. Jondral
Tag der mündlichen Prüfung:	13.11.2015

Institut für Nachrichtenübertragung
der Universität Stuttgart

2015

Abstract

Today's wireless cellular systems are limited by inter-cell interference. Coordinated multi-point transmission and reception (CoMP) is a promising approach to cope with this problem. Here, multiple base stations act as a distributed antenna system, exchanging coordination information and potentially user data via backhaul in order to reduce this interference, increasing spectral efficiency and generating a more homogeneous user experience throughout the entire cell, especially at the currently weak cell edges.

This dissertation aims at enabling coordinated multi-point systems by dealing with its current realization challenges. Accurate channel knowledge is required both at transmitter and receiver side in order to realize the CoMP gains. Hence, practical channel estimation algorithms are investigated and developed in order to get as accurate channel knowledge as possible with manageable computational complexity under realistic system operation points. Especially the often neglected obtainment of statistical parameter knowledge is included here, which is challenging under dynamic user scheduling conditions. A multi-user multi-cell channel estimator is provided which, even for larger number of coordinated cells, can get receiver performance fairly close to performance with perfect channel knowledge.

Furthermore, different downlink precoding and receive combining strategies are compared against each other under imperfect channel knowledge. As the number of coordinated cells - the cluster size - is limited in practice, mobiles at cluster edges suffer from inter-cluster interference. A novel patented precoding and control signaling scheme is introduced in order to deal with the inter-cluster interference.

Kurzfassung

Heutige zellulare Mobilfunksysteme sind begrenzt durch Interzell-Interferenz. Koordinierte Mehrpunkt-Sende- und Empfangsverfahren (“Coordinated multi-point transmission and reception” - CoMP) sind ein vielversprechender Ansatz, um das Interferenzproblem anzugehen. Hierbei agieren mehrere Basisstationen als ein verteiltes Mehrantennensystem und tauschen Koordinierungsinformationen sowie optional auch Nutzerdaten mit Hilfe des “Backhauls” aus, um den Einfluß der Interferenz zu reduzieren. Dadurch wird die spektrale Effizienz erhöht und eine homogenere erlebte Nutzerqualität über die ganze Zelle geschaffen, was insbesondere die Situation am Zellrand verbessert.

Diese Dissertation zielt darauf ab, Systeme mit koordinierten Mehrpunkt-Verfahren in die Praxis umzusetzen, indem begrenzende Faktoren bei der Umsetzung gelöst werden. Exakte Kenntnis des aktuellen Funkkanals wird sowohl am Sender als auch am Empfänger benötigt, um die CoMP-Gewinne realisieren zu können. Daher werden in dieser Arbeit praktikable Kanalschätzalgorithmen unter machbarem Komplexitätsaufwand untersucht und entwickelt, um möglichst genaues Kanalwissen unter realistischen Systemarbeitspunkten zu erlangen. Besonders die oft vernachlässigte Gewinnung von statistischem Parameterwissen ist in dieser Arbeit inbegriffen, wobei die erschwerende dynamische Nutzerzuweisung berücksichtigt ist. Ein Mehrbenutzer-Mehrzell-Kanalschätzer wird erarbeitet, der selbst für eine größere Anzahl koordinierter Zellen, eine Leistungsfähigkeit des Empfängers erreicht, die nah an perfektes Kanalwissen heranreicht.

Ein weiterer Aspekt dieser Arbeit ist der Vergleich verschiedener Mehrantennen-Sende- und Empfangsstrategien in der Abwärtsstrecke unter unvollkommenem Kanalwissen. Da die Zahl der koordinierbaren Zellen, d.h. die Zellverbunds-Größe, in der Praxis begrenzt ist, erfahren Mobilgeräte am Rand des Zellverbunds Interferenzen von Nachbar-Zellverbänden. In dieser Dissertation wird ein neuartiges Verfahren erarbeitet, das auch patentiert wurde, um diese Interferenz zu reduzieren.

Acknowledgements

The research leading to this dissertation was carried out in Bell Labs Alcatel-Lucent Stuttgart in cooperation with the Institute of Telecommunications (INUE) at the University of Stuttgart. I'd like to express my deepest gratefulness to Prof. Dr.-Ing. Joachim Speidel for giving me the oppurtunity to work under his supervision, for all his support, valuable suggestions and fruitful discussions. Also, I cordially thank Prof. Dr. rer.nat. Friedrich Jondral for the assessment of this thesis.

I'd like to sincerely thank Tod Sizer, who has made this opportunity possible from Bell Labs side and encouraged me to do this dissertation. My special thanks goes to Stephan ten Brink, who inspired me with his true research spirit and the mindset to strive for noble competition. During the time where he was my department head, he provided me the required degrees of freedom and his support. I'd also like to thank my department head Hans-Peter Mayer of Next Generation Wireless Systems for the same kind of support and degrees of freedom.

My sincere gratefulness goes to my colleague Le Hang Nguyen for her support and for all the discussions on channel estimation; I'd like to acknowledge her companionship in simulations. I'd like to thank all my Bell Labs colleagues for providing me a great working environment with a lot of fruitful discussions. A special thanks goes to Cornelis Hoek from whom I learned a lot over the years in physical layer and multi-antenna processing with very pragmatic thinking. My thanks to my colleagues in New Jersey, especially Sivarama Venkatesan and Howard Huang for all the discussions which broadened my mind in channel estimation and various communication topics. My thanks to Frank Schaich for the great teamwork in a manner of friendship in the 5G projects beyond this dissertation. I'd like to thank my colleagues in Stuttgart who made this more than "just a place to work" with their teamspirit and their openness for discussions, especially Mark Doll, Danish Aziz, Paolo Baracca, Uwe Doetsch, and Andreas Weber, just to name a few. Bernd Gloss was a valuable source of help for building the LaTeX-basis of this thesis.

I'd like to deeply thank my longtime friend Jürgen Fischer for his encouragement, emotional support and his valuable lectorate work, as well as my departed friend Gebhard Thierer, who always encouraged me.

This work is only possible due to all the appreciation, emotional support and encouragement by my friends and my family. Thank you heartily Corina. Thank you Oli Hahnenstein, Jörg Paschke and Ralf Bierkandt.

Für meine Familie Corina, Juri und Paul.
Für meine Eltern Ernst und Ruth.

Nomenclature

Abbreviations and Acronyms

3GPP	3rd Generation Partnership Project
AWGN	Additive White Gaussian Noise
BICM	Bit-interleaved coded modulation
BS	Base Station
CAZAC	Constant Amplitude Zero Autocorrelation
CFO	Carrier Frequency Offset
CIR	Channel Impulse Response
CoMP	Coordinated Multi-Point Transmission/Reception
CRC	Cyclic Redundancy Check
CRLB	Cramér-Rao Lower Bound
CS/CB	Coordinated Scheduling / Coordinated Beamforming
CSI	Channel State Information
CSIR	Channel State Information at the Receiver Side
CSIT	Channel State Information at the Transmitter Side
CTF	Channel Transfer Function
CP	Cyclic Prefix
CU	Central Unit for CoMP Processing
DFT	Discrete Fourier Transform
dIRC	simplified IRC algorithm, only taking into account the matrix diagonal
DL	Downlink
DPC	Dirty Paper Coding
Ebf	Eigenbeamforming
EDGE	Enhanced Data Rates for GSM Evolution
FDD	Frequency Division Duplexing
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GEbf	Generalized Eigenbeamforming
GPRS	General Packet Radio Services
GSM	Groupe Spécial Mobile
ICIC	Inter-Cell Interference Coordination
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
IIR	Infinite Impulse Response
IoT	Interference over Thermal Noise Power Ratio
IQ	In-Phase and Quadrature Component

IRC	Interference Rejection Combining
ITU	International Telecommunication Union
JR	Joint Reception
JT	Joint Transmission
LS	Least Squares
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
MAP	Maximum a Posteriori
MAC	Multiple Access Control (also: Multiple Access Channel)
MCS	Modulation and Coding Scheme
MDL	Minimum Descriptive Length
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MMSE	Minimum Mean Squared Error
ML	Maximum Likelihood
MRC	Maximum Ratio Combining
MS	Mobile Station
MSE	Mean Squared Error
MWF	Matrix Wiener Filter
NCRLB	Normalized Cramér-Rao Lower Bound
NMSE	Normalized Mean Squared Error
OFDM	Orthogonal Frequency Division Multiplexing
PAPR	Peak-to-Average Power Ratio
PMI	Precoding Matrix Indicator
PRB	Physical Resource Block
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RF	Radio Frequency
RRH	Remote Radio Head
Rx	Receive
SCME	Spatial Channel Model Extended
SINR	Signal to Interference-plus-Noise Ratio
SIMO	Single Input Multiple Output
SIR	Signal to Interference Ratio
SLNR	Signal to Leakage-plus-Noise Ratio
SNR	Signal to Noise Ratio
SNV	Signal to Noise Variance
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
Tx	Transmit
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telephony System
WCDAM	Wideband Code Division Multiple Access
ZF	Zero Forcing

Mathematical Notation

In this thesis vectors are always column vectors, denoted by lowercase bold symbols, e.g. $\mathbf{a}$. Matrices are represented by uppercase bold symbols, e.g. $\mathbf{A}$. $\mathrm{Tr}(\mathbf{A})$ denotes the trace operator, $\mathbf{A}^T$ the transpose, $\mathbf{A}^H$ the Hermitian (thus conjugate complex) transpose, $\mathbf{A}^{-1}$ the inverse, $||\mathbf{A}||_F$ the Frobenius norm of matrix $\mathbf{A}$ respectively. A diagonal matrix $\mathbf{A}$ consisting of the vector $\mathbf{a}$ on its main diagonal (the other elements are zero) is written as $\mathbf{A} = \mathrm{diag}(\mathbf{a})$. The identity matrix is denoted as $\mathbf{I}$. The matrix filled with zeros is denoted as $\mathbf{0}$. (Any subscript N used with $\mathbf{I}_N$ or $\mathbf{0}_N$ denotes the dimension $N \times N$ of the identity / zero matrix.) $\vec{\mathbf{1}}_{[L\times 1]}$ denotes a vector of length L filled with ones: $\vec{\mathbf{1}}_{[L\times 1]} = [1, 1, \ldots, 1]^T$. Likewise $\vec{\mathbf{0}}$ represents a vector filled with zeros.

Estimates are depicted by a hat, e.g. $\hat{\mathbf{h}}$. $\mathrm{E}\{\ldots\}$ denotes the expected value, $\mathrm{Var}\{\ldots\}$ the variance of a random value. For a discussion of the terminology "covariance" and "correlation" the reader should refer to appendix B.

Sets are written in caligrahpic style, like $\mathcal{S}$. $|\mathcal{S}|$ denotes the cardinality of a set. $\mathbb{C}$ is the set of complex numbers.

$\mathcal{O}$ denotes the Landau-symbol.

Thesis-Specific Notation

a	effective path gain (including transmit power)
$\mathbf{A}$	effective path gain matrix (including transmit power)
α	average path gain
$\mathcal{B}$	set of antennas of coordinated base stations
c	equivalent baseband channel impulse response coefficient
$\breve{c}$	channel impulse response coefficient
γ	SINR
δ	Dirac delta function
Δf	OFDM subcarrier spacing
f	frequency
$\mathbf{F}$	DFT matrix
$\mathbf{g}$	receive combining weight vector
h_{ijkl}	OFDM channel transfer function (CTF) coefficient
$\hat{h}_{ijkl}$	estimate of OFDM channel transfer function (CTF) coefficient
$\tilde{h}_{ijkl}$	OFDM channel transfer function (CTF) coefficient normalized to unit variance
i	receive antenna index
$\mathcal{I}$	set of receive antennas
j	transmit antenna index
$\mathcal{J}$	set of transmit antennas
j	imaginary number
k	OFDM symbol index

κ	LTE-A slot index
$\bar{\kappa}$	LTE-A subframe index
l	OFDM subcarrier index
λ	eigenvalues of a matrix
M	number of downlink receive antennas per mobile station
$\mathbf{n}$	noise vector
N	number of samples in the statistical sense, thus data subsets collected (chapter 3)
N	number of downlink transmit antennas per base station (chapter 4 and 5)
N_{CP}	OFDM cyclic prefix length (in samples)
N_{SC}	number of OFDM subcarriers
p	probability density
$\mathbf{p}$	transmit precoding antenna weight vector
P	number of pilot symbols
$\mathcal{P}$	set of pilot symbol resource elements
$\mathbf{\Phi}$	auto-covariance matrix
$\hat{\mathbf{\Phi}}^{[SC]}$	sample auto-covariance matrix
$\hat{\mathbf{\Phi}}^{[shr]}$	estimated auto-covariance matrix using shrinkage
ρ	transmit power
r	receive signal at antenna combiner output
$\mathbf{r}$	cross covariance vector
$\mathbf{R}$	cross covariance matrix
$\mathbf{R}^{[\mathrm{hh}]}$	channel auto covariance matrix
σ_n^2	noise variance
s	transmit symbol (data or pilot symbol)
$\mathbf{s}$	pilot vector
$\mathbf{S}$	pilot matrix
$\mathbf{t}$	time domain OFDM transmit sample vector
$\mathcal{T}$	downlink set of transmitting base station cells
u	user index
$\mathcal{U}$	set of active allocated users within CoMP cluster
v	base station index
$\mathcal{V}$	set of active allocated users within explicitly considered interfering cluster
$\mathbf{w}$	weight vector (for channel estimator)
$\mathbf{W}$	weight matrix (for channel estimator)
x	precoded transmit symbol (data or pilot symbol)
$\mathcal{X}$	explicitly considered set of interfering base station cells
$\mathbf{y}$	observation vector
$\mathbf{z}$	noise-plus-interference vector

Contents

1. Introduction

Communication is a fundamental necessity of the human being. Digital wireless communication has revolutionized our every day life, allowing us to be closer to our loved ones and supporting us to be embedded in the ubiquitous internet, where we gain access to the knowledge of the world, regardless where we are. This thesis is about enhancing wireless communication systems and improving our understanding in this regard.

With Heinrich Hertz proving the propagation of the electromagnetic wave, based on Maxwell's general theory, and Marconi's pioneering of the first radio transmission, the ground was layed for wireless communication. The fundamental limits of digital communication were derived by Claude Shannon in his famous work in the end of the 1940's, e.g. [Sha48], building up information theory. Digital Communication allows to use source and channel coding and is much more efficient than its analog counterpart. At the same time, Norbert Wiener, working in parallel to Shannon, published his work on the Wiener Filter [Wie49].

It took decades before the time was right for the first digital communication system, supported by the invention of the transistor at AT&T Bell Labs and the integrated circuit technology at Texas Instruments. In 1982, the standardization work began for GSM (Groupe Spécial Mobile). GSM provided an open standard-based cellular network, which spread all over Europe and later the entire world. Whereas it was initially designed for voice traffic, only its later extensions provided some support for data transport, denoted GPRS (General Packet Radio Services) and EDGE (Enhanced Data rates for GSM Evolution or EGPRS). While the single-antenna point-to-point link was in its basic nature already well understood by Shannon, the question of interference and cross-talk between different transmissions is an open challenge until today. The answer of GSM was easy (but clearly suboptimal): Different cells where separated by using different frequencies, requiring frequency planning. Due to propagation attenuation, it was of course possible to reuse frequencies at cell cites with large enough distances, leading to frequency reuse factors of e.g. 7. Mobile users sharing the same cell were separated by using different time slots, the so-called time division multiple access (TDMA).

Counting from the first analog cellular network generation, the next generation to be developed after GSM was the third. 3GPP, the third generation partnership programme, was the standardization organization evolving from GSM. 3GPP began the standardization work for the universal mobile telephony system (UMTS) in the 1990's. The basic signal format was changed from GMSK (Gaussian Minimum Shift Keying) to WCDMA (Wideband Code Division Multiple Access). GMSK has very power-amplifier-friendly properties. These properties were relaxed in 3G for the sake of new benefits: WCDMA, due to its spread spectrum nature, allows to deal with interference in a different way. The robust nature of spreading is able to tolerate higher interference levels, dealing with it by an interference randomization strategy. The extensive frequency planning of GSM could be dropped as far as possible, as UMTS was able to operate at frequency

reuse 1. UMTS was from scratch also designed for carrying data. The packet-switched transmission, which was at that time already established in the internet, found its way into the cellular networks, especially in the HSPA (High Speed Packeted Access) extensions of UMTS.

Additionally in the 1990's, signal processing over multiple antenna elements opened up a new "signal dimension", besides time, frequency and code: the space. Already before the invention of spatial multiplexing, the multi-antenna processing techniques, at that time also known as "smart antennas", were capable of improving the signal-to-noise-plus-interference ratio (SINR) of the link by transmit and/or receive processing. Beamforming, by adjusting phases and amplitudes in the base band signal, makes use of adapting the directivity of the radiation pattern of a multiple-element antenna array. This can be used to improve the receive signal strength and avoid unwanted interference. Spatial diversity techniques add robustness against fast fading by predominantly concentrating on antenna elements which are most reliable in terms of current fast fading conditions of the propagation channel.

A major breakthrough came by the invention of spatial multiplexing, pioneered by people like Paulraj and Kailath in 1993. One of the most famous spatial multiplexing transceiver schemes is the BLAST (Bell Labs Layered Space-Time Architecture) scheme by Foschini. With spatial multiplexing, the transmission of multiple data streams in parallel on the same time-frequency resource was possible by exploiting the dimension space via antenna processing. All the mentioned spatial processing schemes can be treated under the umbrella of MIMO: multiple input multiple output systems.

A natural physical layer signal format complementing MIMO is the orthogonal frequency division multiple access (OFDM) [Cha66, WE71], where the algorithms become especially simple and elegant, because OFDM decomposes the signals into subcarriers, which are narrowband subchannels, experiencing flat fading. They can be processed independently, which reduces complexity. Furthermore, the optimal processing techniques have become well understood for MIMO-OFDM.

The third generation cellular communication standard had its inherent limitations. In order to achieve higher spectral efficiency, required to satisfy growing user demand for data traffic, the WCDMA air interface had to be used with multi-code transmission. Here the orthogonality of the codes was destroyed by multi-path propagation, naturally occuring in radio channels of cellular systems. So self-interference in UMTS limits its performance. The acitivity of 3GPP regarding the preparation of a new air interface was simply called LTE - Long Term Evolution. Here the WCDMA air interface was replaced by OFDM. LTE thus also provided a much better MIMO support, due to OFDM, compared to UMTS. The spatial processing technique linear precoding here was combined with spatial multiplexing, allowing to adaptively trade off spatial multiplexing and beamforming plus diversity advantages. In overall, LTE is much more suited to adapt to the conditions of the radio propagation channel than UMTS, providing higher spectral efficiency.

In order to achieve high spectral efficiency, LTE has to be run in frequency reuse 1 or near 1. This leads to the disadvantage that the user experience greatly varies across the cell. Close to the base station, in good SINR conditions, the data throughput is usually high, while at the cell edge it may be lower by orders of magnitude due to inter-cell interference.

This inter-cell interference limitation is one of the key remaining problems of cellular wireless systems. Simple techniques like fractional frequency reuse and inter-cell interference coordination (ICIC), using cell planning of time-frequency resources proved to be only of very small effectiveness [AL07][1]. Advanced technologies to deal with this problem were already early addressed in academic literature, e.g. by Shamai et al. Network MIMO is the approach to treat a set of multiple cells as a single distributed antenna system, by the help of sharing data over the backhaul network. This allows to generalize MIMO technologies across multiple cells to deal with the intercell-interference problem.

In the extensions of the LTE standard, LTE-Advanced (LTE-A), the techniques dealing with coordination across base stations are called coordinated multi-point transmission and reception (CoMP). CoMP can be interpreted as a superset of Network MIMO. LTE-A is certified by the ITU-R to fulfill the IMT-Advanced requirements (while LTE is not yet capable of achieving this). Thus LTE-A can be seen as a full fourth generation cellular system.

Focus, Contribution and Structure of the Thesis

This thesis deals with important questions in order to enable CoMP and network MIMO as major building blocks, which finally ease problems of inter-cell interference, leading to a better and more homogeneous user experience across the cell area. In Chapter 2 the background of this work is summarized and the problem is formulated.

The success of channel-adaptive spatial processing schemes depends on the accuracy of the available channel knowledge. So one key topic of this thesis is the problem of channel estimation across multiple cells. The contribution of this thesis here is to introduce a practical channel estimator which comes close to perfect channel knowledge under LTE-A-like conditions. This is discussed in chapter 3.

In order to make the channel knowledge available at the transmitter in frequency division duplex systems (FDD), feedback is required, which besides channel estimation errors experiences additional inaccuracies due to quantization and delay. In downlink CoMP joint transmission the design of appropriate precoders is an important aspect. In this thesis, different precoding coordination strategies are uniformly treated and compared under imperfect channel state information (CSI), as well as different reception strategies. Evaluation takes place in chapter 4.

A third practical challenge for CoMP is to deal with the coordination cluster edges. Realizable CoMP clusters are limited in size and thus there are always cluster edges where mobiles within neighboring clusters experience mutual cross-talk. Thus the former cell edge problem becomes a cluster edge problem. The contribution of this thesis is here to come up with a novel precoding solution in order to cope with out-of cluster leakage, which is discussed in chapter 5.

Chapter 6 provides an outlook on the perspectives of CoMP in next generation wireless systems. Finally, the conclusion is given in chapter 7.

[1] For small cell deployments in recent LTE-A, enhanced ICIC (eICIC) plays a more important role to keep the control channels decodeable.

2. Background and Problem Formulation

The background of this work describes the general transceiver chain and an an air interface of a multi-antenna communication system (subsection 2.1.1) within a cellular architecture (subsection 2.1.2). The transmission happens through a wireless propagation channel (subsection 2.1.3), using MIMO transceiver techniques (subsection 2.1.4), potentially distributed across multiple cells. The receiver has to estimate those propagation channels across multiple cells. This challenge is described in section 2.2. For the spatial adaptation of the transmitter to the radio channel, it is required to feedback the estimated channel knowledge. Then appropriate precoding and coordination strategies can be used. The respective challenges are described in section 2.3.

2.1. General Background

2.1.1. Air Interface of Multi-Antenna Communication

In this work we focus in the area of wireless digital communication on algorithms for improving the successful transport of bits and frames. In the language of the Open Systems Interconnection (OSI) model [OSI], this is the physical and link layer.

The processing chain, from transmitter to receiver, starts with the data source, where information bits arrive from higher layers, depicted in Fig. 2.1. A base band signal processing unit deals with the necessary computations for generating the digital transmit signal. This unit typically consists of modules like field programmable gate arrays (FPGA) or digital signal processors (DSP).

The source information bits are segmented/concatenated and encoded by the channel encoder, e.g. using Turbo codes [BG96], convolutional codes or low density parity check codes (LDPC) [Gal63]. Redundancy is added for error protection. To ensure knowledge about correct reception, cyclic redundancy check (CRC) bits are added. Then the bits are mapped to modulation symbols, typically using Gray mapping, where misdetection for neighbor symbols leads to as few bit errors as possible. The choice of the proper modulation and coding scheme (MCS) in this kind of bit-interleaved coded modulation (BICM) approach is done by *link adaptation*: Usually the MCS is adapted to the radio channel conditions using feedback information of a channel quality indicator (CQI).

As we are dealing with MIMO transmission (see subsection 2.1.4), there can be several *data streams* or so-called *spatial layers* in parallel, exploiting spatial multiplexing capabilities offered

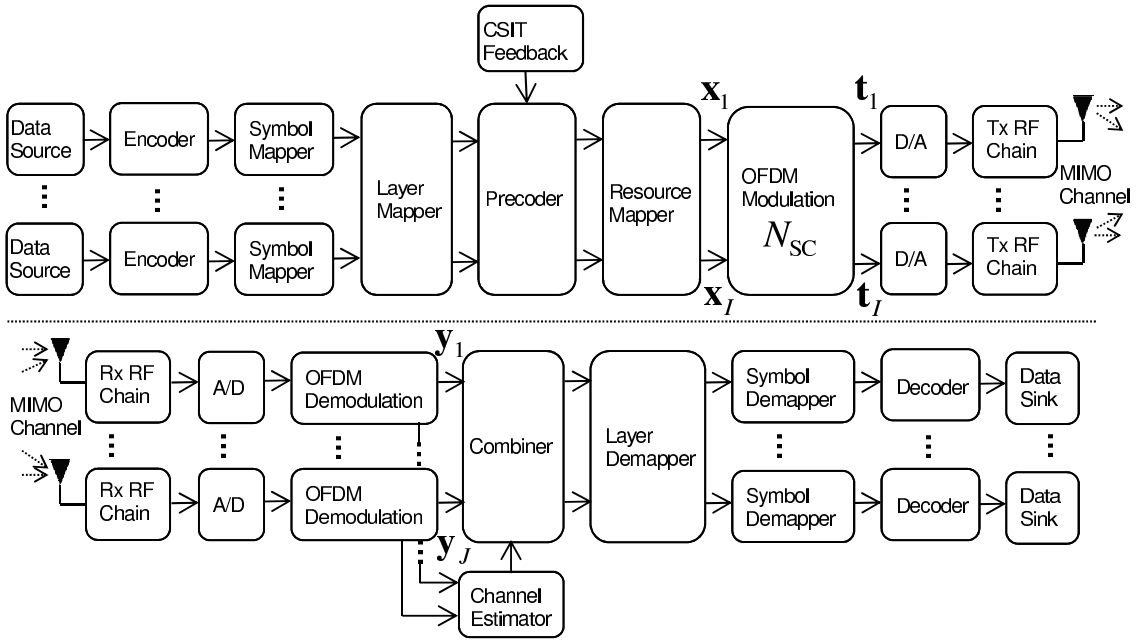

Figure 2.1.: Simplified block diagram of multi-antenna transmit (top) and receive (bottom) chain.

by the MIMO propagation channel (see subsection 2.1.3) intended for one or more users. The spatial transmit processing includes layer mapping, precoding and mapping to radio resources. Usually, spatial processing makes use of the available channel state information at the transmitter (CSIT), e.g. obtained via feedback signaling. A more refined general spatial processing architecture is e.g. described in [OORW08] - the spatial processing concept defined in the European WINNER research project. The standardized parts of the LTE physical layer processing chain are addressed in [3GPc].

In an OFDM system [Cha66, WE71], the basis of LTE, LTE-A and many IEEE 802.11 standards, the modulation symbols $\mathbf{x}$ resulting from the preceding processing steps are mapped to N_{SC} orthogonal subcarriers, flexibly supporting multiple-access and e.g. frequency-selective scheduling. As *resource element*, we denote the pair (k, l) consisting of the single subcarrier l and the single OFDM symbol k. With a fast inverse Fourier transform (IFFT), the time domain signal samples of all superimposed subcarriers are generated. A cyclic prefix (CP) of length N_{CP} is added to prevent inter-symbol and inter-carrier interference, ISI and ICI respectively, in the presence of multipath propagation, provided the maximum channel excess delay does not exceed the CP. The CP ensures that the convolution with a channel impulse response is cyclic, preserving orthogonality. The discrete time-domain OFDM transmit signal of transmit antenna j for symbol k can be expressed as

$$\mathbf{t}_{jk} = \frac{1}{\sqrt{N_{\mathrm{SC}}}} \mathbf{\Xi}_{\mathrm{T}} \mathbf{F}^H \mathbf{x}_k \tag{2.1}$$

with $\mathbf{x}_k \in \mathbb{C}^{N_{\mathrm{SC}} \times 1}$, $\mathbf{t}_{jk} \in \mathbb{C}^{(N_{\mathrm{SC}}+N_{\mathrm{CP}}) \times 1}$ and the matrix $\mathbf{\Xi}_{\mathrm{T}} = \left[[\mathbf{0}_{N_{\mathrm{CP}} \times (N_{\mathrm{SC}}-N_{\mathrm{CP}})}, \mathbf{I}_{N_{\mathrm{CP}}}]^T \mathbf{I}_{N_{\mathrm{SC}}} \right]^T$ of dimension $(N_{\mathrm{SC}}+N_{\mathrm{CP}}) \times N_{\mathrm{SC}}$ taking care for CP addition (the subscript of $\mathbf{\Xi}_{\mathrm{T}}$ is a mnemonic

for transmitter). The DFT matrix $\mathbf{F}$ is given by[1]

$$\mathbf{F} = \frac{1}{\sqrt{N_{\mathrm{SC}}}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \xi & \xi^2 & \cdots & \xi^{N_{\mathrm{SC}}-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi^{N_{\mathrm{SC}}-1} & \xi^{2(N_{\mathrm{SC}}-1)} & \cdots & \xi^{(N_{\mathrm{SC}}-1)^2} \end{bmatrix} \tag{2.2}$$

for $\xi = \exp(-j2\pi/N_{\mathrm{SC}})$. More details on LTE / LTE-A OFDM signal formats are given in appendix A.

For transportation to a remote radio head (RRH), e.g. via optical fiber, the digital Common Public Radio Interface (CPRI) [CPR11] can be used. For antenna elements which are co-located to the base band processing this is not required.

The next processing steps are digital-to-analog (D/A) conversion and up-conversion to the radio carrier frequency (RF) using local oscillators and mixers (some-times using intermediate frequencies - IF). Filtering and matching is done. In FDD systems this involves a duplex filter, allowing to use the same antenna for transmission and reception. Then the analog RF signal is fed into an antenna network. When later on in this thesis a "single antenna element" is referred, it may consist of several dipoles with RF phase shifters to obtain the desired directivity. So the term "single antenna" is meant from the perspective of the base band signal processing, having in this case a single output.

The signal is now transmitted by electromagnetic waves via the radio propagation channel, which is described in the subsection 2.1.3.

This transmit chain is able to transmit to multiple users. The receiver chain of one user is illustrated in Fig. 2.1. After RF processing and A/D conversion, OFDM demodulation is done by discarding the cyclic prefix and doing an FFT. The receiver estimates the channel based on known pilot symbols, also known as training or reference symbols. Now the MIMO spatial receive processing can occur. Fig. 2.1 shows the simplest option of a linear receive combiner and equalizer which enhances the useful signal and suppresses unwanted contributions, leading to an increased signal-over-interference-plus-noise-ratio (SINR) compared to the antenna input. More options and details are provided in subsection 2.1.4. Finally the different spatial layers are decoded.

2.1.2. Cellular Network Architecture

In [Mac79], the cellular concept was identified as an efficient way for wireless communication. Fig. 2.2 shows an example cellular network architecture. While 2G and 3G systems still had base station controllers / radio network controllers inbetween backhaul and core network, the modern architecture starting with HSPA and continued in LTE, is flat for lean all-IP (Internet Protocol) support. Between BS and "enhanced packet core" (EPC) - in 3GPP terminology - there are no intermediate network elements.

[1] The product of normalization factors of DFT and IDFT has to be $\frac{1}{N_{\mathrm{SC}}}$, regardless of all possible conventions. Furthermore we follow the convention that Fourier matrices are unitary and the DFT normalization factor is 1.

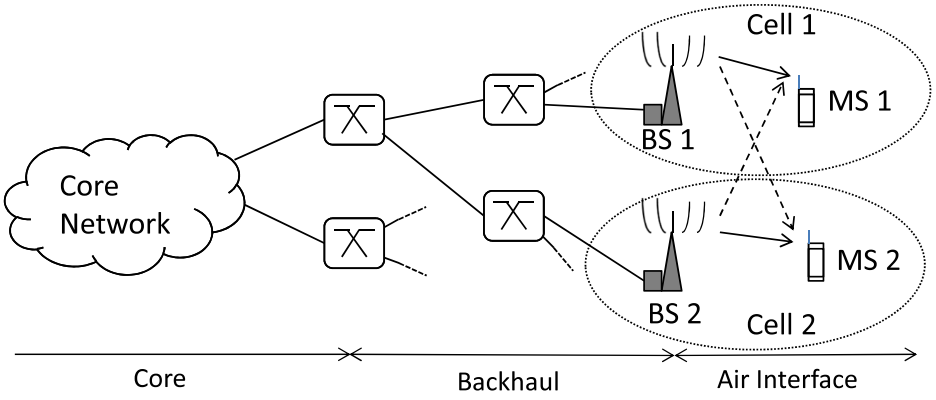

Figure 2.2.: Example cellular network architecture.

The logical interface between base stations is called X2 in LTE terminology. The logical interface between base stations and core network is called S2 interface. Physically, both partly share the same backhaul connections. The backhaul can be realized e.g. via Ethernet, passive optical networks (PON), Digital Subscriber Lines (DSL) or microwave links.

Classical cellular networks experience a limitation in their air interface, caused by inter-cell interference, depicted in Fig. 2.2 by the dashed arrows. CoMP copes with this limitation by exchanging user data and or coordination control information via the backhaul. CoMP-specific aspects of backhaul are handled in subsection 2.1.4.3.

2.1.3. Wireless Channels and Channel Models

The channel of interest in this thesis h_{ijkl} is the complex-valued equivalent baseband representation in the frequency domain, based on a well-designed[2] OFDM system. It describes the input-output relation of transmit antenna j, receive antenna i, at time instant k for subcarrier l. In order to understand this relation, we need to consider radio propagation (characterized by the channel impulse response $\breve{c}_{ij}(\tau, t)$) the equivalent base-band representation (characterized by $\mathbf{c}_{ijk} \in \mathbb{C}^{(N_{\mathrm{SC}}+N_{\mathrm{CP}})\times 1}$) and the elegant OFDM processing steps, leading to the simple resource-element-wise description by h_{ijkl}. Throughout this section, the statistical properties of the channel will be discussed as well.

Radio propagation via electromagnetic waves can be understood via the wave equation by Hertz. In the far field of a radiating antenna, a transversal electromagnetic wave is observed. Line-of-sight conditions have a direct propagation path to the mobile receive antenna. However, the majority of cellular links has non-line-of-sight conditions. Receive signals are obtained by reflections at scattering objects (like buildings, cars, trees etc.), by diffraction (at e.g. roof-edges) etc. Movements of mobile stations cause Doppler shifts.

[2]Well-designed here means non-frequency selective channels per subcarrier, the cyclic prefix length L_{CP} is long enough to avoid inter-symbol interference and time-frequency alignment of signals is perfect, so that orthogonality among subcarriers does hold.

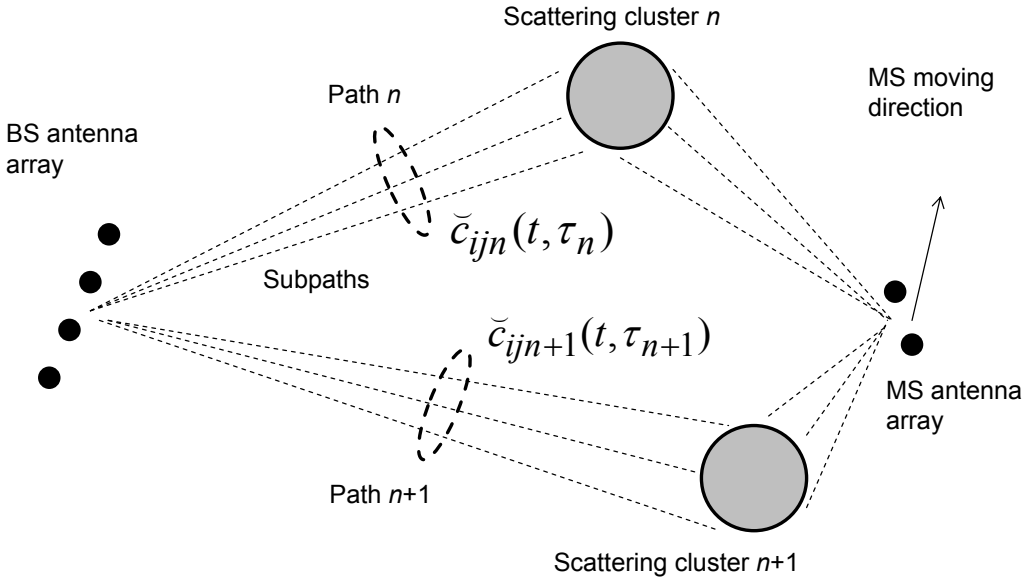

Figure 2.3.: Illustration of a spatial channel model.

Due to multi-path propagation, those signal components arrive with different delays and cause a *frequency-selective* channel. The n-th component $\breve{c}_{ijn}$, corresponding to the delay $\tau_n(t)$, is already a superposition of non-time-resolvable[3] sub-path contributions. The mobility of transmitters and/or receivers causes *time-variant* channels, thus fast fading. Fig. 2.3 illustrates the spatial channel with scattering clusters, which cause the superpositions of sub-paths.

The frequency-selective, time-varying radio channel is characterized by its real-valued time-domain channel impulse response (CIR) $\breve{c}(\tau, t)$ with

$$\breve{c}_{ij}(\tau, t) = \sum_{n=1}^{N_{\text{path}}} \breve{c}_{ijn}(t)\delta(\tau - \tau_n(t)). \tag{2.3}$$

The corresponding frequency-domain representation of $\breve{c}_{ij}(\tau, t)$ is

$$\breve{h}_{ij}(f, t) = \sum_{n=1}^{N_{\text{path}}} \breve{c}_{ijn}(t)e^{-j2\pi f\tau_n(t)}. \tag{2.4}$$

A continuous time input signal $\breve{x}_j(t)$ from the j-th transmit antenna generates the time-domain output at receive antenna i

$$\breve{y}_i(t) = \breve{x}_j(t) * \breve{c}_{ij}(\tau, t) + \breve{z}_i(t) = \int_{-\infty}^{+\infty} \breve{c}_{ij}(\tau, t)\breve{x}_j(t - \tau)\mathrm{d}\tau + \breve{z}_i(t), \tag{2.5}$$

where $*$ denotes the linear convolution and $\breve{z}_i(t)$ contains all further contributions from other transmit antennas, interference and noise.

[3] Whether components are not time-resolvable depends on the signal bandwidth.

We are interested in the complex equivalent baseband representation of the signal. RF modulation and demodulation happens with two components: in-phase (with sine) and quadrature (with cosine). Thus we obtain so-called I- and Q-branch outputs at the receiver which can be represented in the complex domain. The signal undergoes sampling, analog-to-digital conversion (ADC) and receive filtering. Thus we denote this complex baseband representation, incorporating all these processing steps, as $c_{ij}(\tau, t)$ in order to differentiate from $\breve{c}_{ij}(\tau, t)$ introduced above. The sampling creates a time discretization, which can be viewed as a multiplication with $\sum_{\bar{m}=-\infty}^{+\infty} \delta\left(t - \bar{m}T_s\right)$. The discrete time instants $\bar{m}T_s$ have the sampling interval $T_s = \frac{1}{\Delta f N_{\mathrm{SC}}}$. Applying the temporal structure of OFDM symbols, we use OFDM symbol indices $k = \lfloor \frac{\bar{m}}{N_{\mathrm{SC}}+N_{\mathrm{CP}}} \rfloor$. The samples in each symbol k are indexed $m = 0, 1, \ldots, N_{\mathrm{SC}} + N_{\mathrm{CP}} - 1$ with $m = \bar{m} \mod (N_{\mathrm{SC}} + N_{\mathrm{CP}})$. Assuming that the radio channel is approximately constant over one OFDM symbol we attribute the complex-valued equivalent-baseband time-discretized CIR of symbol k to the vector elements c_{ijkm} with

$$\mathbf{c}_{ijk} = \left[c_{ijk0}, c_{ijk1}, \ldots, c_{ijkL-1}, \vec{\mathbf{0}}_{[1\times(N_{\mathrm{SC}}+N_{\mathrm{CP}}-L)]}\right]^T. \tag{2.6}$$

Here, $\mathbf{c}_{ijk}$ acts as a finite impulse response (FIR) filter with L non-zero elements. Its l-th subcarrier frequency domain representation at frequency position $l\Delta f$, likewise to 2.4 with discretized delays $\tau_m = m/(\Delta f N_{\mathrm{SC}})$, can be expressed as

$$h_{ijkl} = \sum_{m=0}^{L-1} c_{ijkm} e^{-j2\pi lm/N_{\mathrm{SC}}}. \tag{2.7}$$

The convolution of the channel, in linear algebra notation, can also be represented by a multiplication with a Toeplitz matrix $\mathbf{C}_{ijk} \in \mathbb{C}^{(N_{\mathrm{SC}}+N_{\mathrm{CP}})\times(N_{\mathrm{SC}}+N_{\mathrm{CP}})}$, whose first column is $\mathbf{c}_{ijk}$. In an OFDM system the frequency domain receive subcarrier signal $\mathbf{y}_{ik} = [y_{ik1}, \ldots, y_{ikl}, \ldots, y_{ikN_{\mathrm{SC}}}]$, based on the time-domain transmit signal $\mathbf{t}_{jk}$ from (2.1), passing the channel from antenna j to i, is described as

$$\mathbf{y}_{ik} = \sqrt{N_{\mathrm{SC}}}\mathbf{F}\boldsymbol{\Xi}_{\mathrm{R}}\mathbf{C}_{ijk}\mathbf{t}_{jk} + \mathbf{z}_{ik} = \underbrace{\mathbf{F}\boldsymbol{\Xi}_{\mathrm{R}}\mathbf{C}_{ijk}\boldsymbol{\Xi}_{\mathrm{T}}\mathbf{F}^H}_{\mathbf{D}_{ijk}}\mathbf{x}_k + \mathbf{z}_{ik}. \tag{2.8}$$

The DFT matrix $\mathbf{F}$ is defined in (2.2). $\mathbf{z}_{ik}$ contains contributions from all other transmit antennas, interference and noise. The matrix $\boldsymbol{\Xi}_{\mathrm{R}} = [\mathbf{0}_{N_{\mathrm{SC}}\times N_{\mathrm{CP}}}, \mathbf{I}_{N_{\mathrm{SC}}}]$ of dimension $N_{\mathrm{SC}} \times (N_{\mathrm{SC}} + N_{\mathrm{CP}})$ carries out the CP removal (the subscript is a mnemonic for receiver) . The usage of cyclic prefix has turned the linear convolution of the channel into a cyclic convolution [Gol05] and thus $\mathbf{D}_{ijk}$ becomes diagonal and contains the channel transfer function (CTF) of the respective OFDM symbol where the transmission is inter-symbol interference free[4]. For OFDM symbol time index k, subcarrier index l, transmit antenna j, receive antenna i, the CTF thus provides a frequency-domain representation of the channel:

$$\mathbf{D}_{ijk} = \mathrm{diag}\underbrace{\left[(h_{ijk0}, h_{ijk1}, \ldots, h_{ijkl}, \ldots, h_{ijkN_{\mathrm{SC}}-1})^T\right]}_{:=\mathbf{h}_{ijk}} \tag{2.9}$$

[4]Assuming $L \leq N_{\mathrm{CP}}$; the channel delay spread is within CP duration

Thus in OFDM, because the cyclic prefix has turned the linear convolution of the propagation channel into a circular convolution, the impact of the channel is simply a scalar complex multiplication per subcarrier. Hence, in the absence of inter-symbol and inter-carrier interference[5], each resource element at time k, subcarrier l can be treated independently from the others, due to the elegance of OFDM. The CTF is simply the Fourier transform[6] of the CIR:

$$\mathbf{h}_{ijk} = \sqrt{N_{\mathrm{SC}}}\mathbf{F}\mathbf{c}_{ijk}, \tag{2.10}$$

confirming (2.7) for OFDM processing. The wideband channel thus has been decomposed by OFDM into orthogonal narrowband[7] subchannels, which are correlated in time and frequency.

Let us now have a look on statistical properties of the channel. In wireless communications, due to the changing nature of the propagation channel we obtain temporal variations. The channel can be considered as a random variable. A single instant Rayleigh fading coefficient for a narrowband channel can simply be modeled by the realization of a complex Gaussian distribution.

The temporal variation of the channel is caused by mobility of transmitter, receiver and/or scatterer. Depending on the direction of components, different Doppler shifts occur and can cause a Doppler spectrum. In the classical Jakes [JC94], [Cla68] assumption, the mobile is surrounded by a circle of scatterers, leading to the famous Jakes Doppler spectrum and a Rayleigh distributed amplitude of the channel. This is known as Rayleigh fading. According to [Cla68], in case of isotropic scattering, the temporal auto-correlation of a channel with maximum Doppler shift f_d can be expressed using the zero-order Bessel function J_0 of the first kind as

$$R(\tau) = J_0\left(2\pi f_d \tau\right) \tag{2.11}$$

with τ denoting the time-lag, depicted in Fig. 2.4. This auto-correlation is linked to the Jakes Doppler spectrum by a Fourier transform.

Let us now focus on the statistical properties of the channel in the frequency domain. The channel covariance[8] between channel coefficients at frequencies $f_1 = l_1\Delta f$ and $f_2 = l_2\Delta f$, having OFDM subcarrier indices l_1 and l_2 with subcarrier spacing Δf, can be computed from (2.7) as

$$\begin{aligned}
\mathrm{E}\left\{h_{ijkl_1}h^*_{ijkl_2}\right\} &= \mathrm{E}\left\{\left(\sum_{m=0}^{L-1} c_{ijkm}e^{-j2\pi l_1 m/N_{\mathrm{SC}}}\right)\left(\sum_{m=0}^{L-1} c^*_{ijkm}e^{j2\pi l_2 m/N_{\mathrm{SC}}}\right)\right\} \\
&= \mathrm{E}\left\{\sum_{m=0}^{L-1} |c_{ijkm}|^2 e^{-j2\pi(l_1-l_2)m/N_{\mathrm{SC}}}\right\} \\
&= \sum_{m=0}^{L-1} \mathrm{E}\left\{|c_{ijkm}|^2\right\} e^{-j2\pi(l_1-l_2)m/N_{\mathrm{SC}}},
\end{aligned} \tag{2.12}$$

[5] Inter-carrier interference in LTE is avoided by strict 3GPP oscillator requirements of 0.05 ppm for BS and 0.1 ppm for MS [3GPe]. ISI can be avoided by carefully tailoring the multi-carrier numerology to the envisaged propagation scenario with a long-enough cyclic prefix.

[6] Note that the DFT matrix already contains the pre-factor $1/\sqrt{N_{\mathrm{SC}}}$ and the DFT definition with normalization factor 1 thus requires $\sqrt{N_{\mathrm{SC}}}\mathbf{F}$ for the DFT operation.

[7] In LTE the comparatively narrow subcarrier spacing of $\Delta f = 15\,\mathrm{kHz}$ makes sure that the narrowband assumption also holds for high delay spreads of macro cells.

[8] The terminology of "covariance", especially in the non-zero mean case, is discussed in appendix B

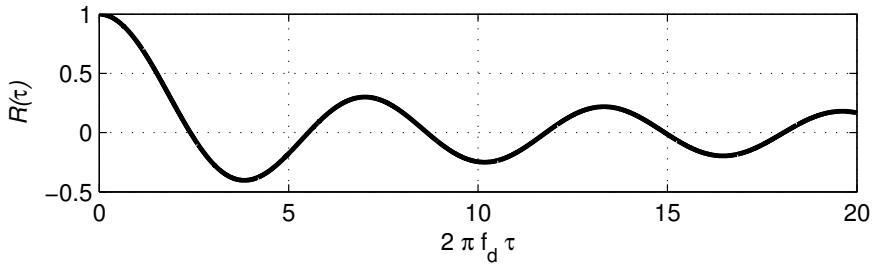

Figure 2.4.: Temporal auto-correlation of Clarke/Jakes isotropic scattering as a function of time lag τ and maximum Doppler frequency f_d.

where the first line is based on inserting (2.7) and the second line follows from statistical independence of fading on different channel delay taps m and $\mu \neq m$. It is important to notice that the absolute frequency does not matter for the frequency covariance, only the relative frequency difference plays a role. Furthermore by looking at (2.12), one can observe that the power delay profile, with path powers $\mathrm{E}\{|c_{ijkm}|^2\}$ sufficiently describes the frequency covariance.

Classical models for time-variant, frequency selective channels are the ITU-R models for Vehicular A/B, Pedestian A/B etc [ITU]. Here, the power delay profile, consisting of delays τ_m and powers $\mathrm{E}\{|c_{ijkm}|^2\}$, is given for a certain number of fixed channel taps. The fast fading per tap is modeled as Rayleigh fading.

The frequency selectivity and the time variance of radio channels are characterized by *coherence bandwidth* and *-time* of the channel, respectively. For those terms, a variety of definitions exists [Skl97]. A typical definition for coherence bandwidth is e.g. the frequency interval, where the correlation has dropped to 0.9, resulting in $f_{coh} \approx \frac{0.02}{\sigma_D}$, where σ_D is the root mean square (RMS) delay spread. E.g. for the ITU Vehicular A model, $\sigma_D \approx 0.37$ μs, thus $f_{coh} \approx 54$ kHz, corresponding to roughly 4 LTE subcarrier spacings[9]. A similar definition in time direction with 0.9 correlation results in $T_{coh} \approx \frac{0.1}{f_d}$. The Doppler shift for a relative velocity of $v = 108$ km/h and a carrier frequency of $f_c = 2$ GHz is $f_d \approx \frac{v \cdot f_c}{c} = 200$ Hz, leading to a coherence time of $T_{coh} = 0.5$ ms, thus one LTE slot with 7 OFDM symbols, see appendix A.

As ITU models are designed for SISO channels, an extension to MIMO is required. Regarding spatial correlation, antennas which are far apart from each other (more than several wavelengths) can be assumed to fade independently; otherwise their fast fading may be correlated. The angular spread, describing the spatial dispersion of angle of arrivals or departures at the antenna, is an important factor for the correlation of fading between different antenna elements. Closely-spaced (e.g. half a wavelength) elements with low angular spread have a highly correlated fading. In contrast to that, even with a very close element spacing, high angular spreads decorrelate the fading of multiple antennas.

For modeling of spatial correlation one approach is the Kronecker model [KSP+02], where the MIMO channel correlation is constructed by a Kronecker product of a given transmit and receive correlation matrix. E.g. this is used in 3GPP for test purposes [3GPb]. A drawback is that explicit

[9]For the perspective of a center subcarrier going into positive and negative directions until the correlation has reduced to 0.9 this is a "window" of roughly 7 subcarriers.

correlation matrices are required. Simulations with multiple users require different matrices (unless all users shall share the same position). Furthermore, it can be shown that Kronecker models fail for particular kinds of propagation scenarios like "pin-hole channels" [CFV00].

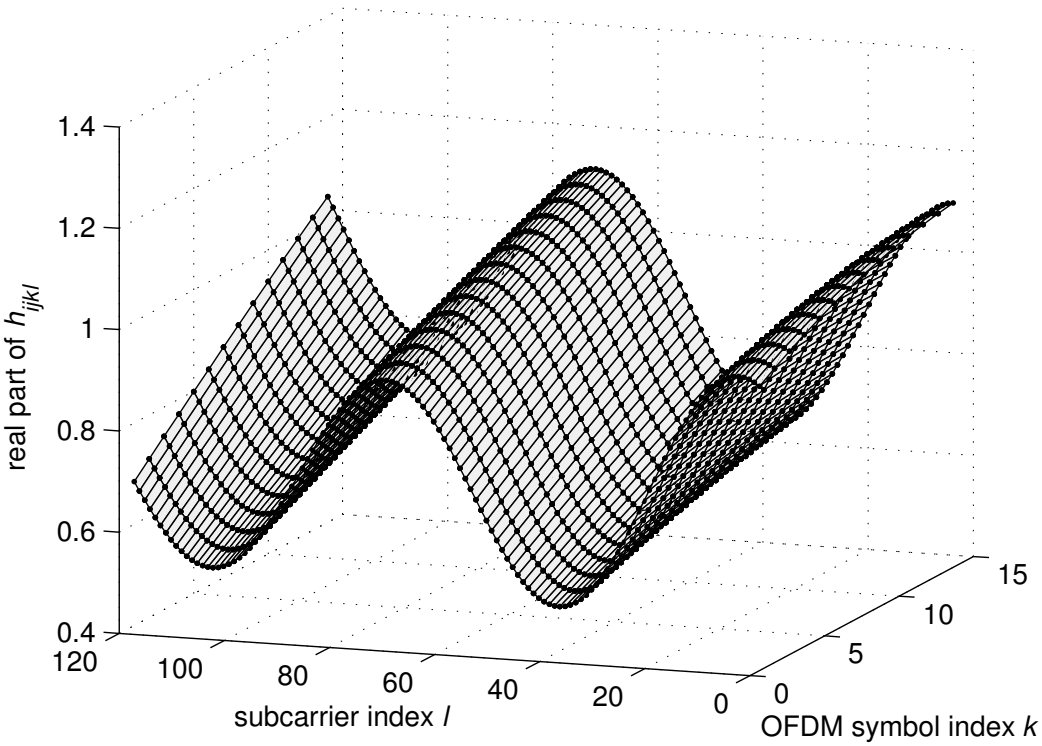

Figure 2.5.: Example SCME urban macro channel realization for one antenna-to-antenna link (for a transmit antenna j to a receive antenna i): Real part of the complex-valued OFDM channel transfer function h_{ijkl} for 10 LTE physical resource blocks (120 subcarriers, spaced 15kHz each), in one 1ms subframe (14 OFDM symbols), 50km/h MS velocity at 2GHz carrier frequency

Recent realistic modeling for spatial channels [3GPa, BHS05, WIM06] is thus doing a different approach: The channel implicitly results from a statistically modeled geometrical scenario as shown in Fig. 2.3[10]. Each path, corresponding to a particular propagation delay of the CIR, is modeled by a set of non-time-resolvable sub-paths. The current models use a simplifying single bounce assumption, thus per path, a single scattering cluster is acting as a reflector for incoming and outgoing sub-paths. The angular spread at the base station (BS) and mobile station (MS) depends on the distance and size of the scattering clusters. Hereby, implicit BS and MS antenna correlations are generated by the models, depending on geometrical properties. The models were parametrized by measurement campaigns.

In order to generate channel realizations with the model, at first, large scale parameters (channel tap delays, scattering cluster angular positions etc.) have to be initialized, being static during

[10]Note that the models focus on the generation of coefficients for the complex equivalent baseband CIR $c_{ij}(\tau, t)$

a particular Monte Carlo "drop" of users. Using these large scale parameters, the small scale fading coefficients are generated, providing a time-variant realization of the channel impulse response.

The "Spatial Channel Model" (SCM) [3GPa] by 3GPP was created for 5MHz channel bandwidth at carrier frequencies around 2 GHz. The "Spatial Channel Model Extended" (SCME) [BHS05] is an extension for 10 MHz bandwidth; an example channel realization is depicted by Fig. 2.5. Further refinements were made within the European WINNER research project, leading to the "WINNER model 2" (WiM2) [WIM06], valid for up to 20 MHz with a variety of outdoor and indoor propagation channel scenarios. While for macro cellular channels, the SCM uses 6 paths (with 20 subpaths each), the larger supported system bandwidth of the WiM2 uses around 20 paths for proper modeling. This number is well chosen and can be confirmed by own Alcatel-Lucent MIMO measurements [WHHK13][11].

2.1.4. MIMO, Coordinated Multi-Point and Network MIMO

2.1.4.1. SU-MIMO and MU-MIMO Overview

While first cellular systems like GSM were single input, single output (SIS0) systems with just single transmit and receive antennas, the usage of multiple receive antennas in a single input, multiple output system (SIMO) clearly provided benefits. Maximum ratio combining (MRC) [Bre03] maximizes the SINR at the receive combiner output in the presence of uncorrelated noise. More advanced combining techniques like optimum ratio combining [BZ71] [Win84] are able to suppress also spatially colored interference.

Multiple transmit antennas, e.g. in multiple input single output systems (MISO), can be used for beamforming. This means modifying the directivity pattern of an antenna array, e.g. by complex weights in the base band signal processing, in order to transmit or receive in wanted signal directions and potentially do "null-steering", e.g. [Ana81], towards unwanted signal sources. While those first smart antenna techniques were using direction estimation techniques, like MUSIC [Sch86] or ESPRIT [PRK85], later solutions made use of the statistical properties of the spatial channel [VT02].

Other typical MISO techniques are transmit diversity techniques using space-time or space-frequency block codes, like the well known Alamouti scheme [Ala98].

While all those techniques aim at improved SINR at the receiver side and/or additional robustness against fast fading dips, the introduction of spatial multiplexing in MIMO was required to make full use of the channel capacity [FG98] [Tel99] of single-user MIMO (SU-MIMO). The channel capacity here can be achieved by obtaining the transmit covariance matrix by singular value decomposition of the MIMO channel matrix, followed by power allocation based on waterfilling over the singular values [PNG03][CT06].

A now well-known approach to MIMO was realized and demonstrated with the Bell Laboratories Layered Space-Time (BLAST) scheme [Fos96]. Designed for open-loop transmission in the

[11] In [WHHK13] information-theoretic model order selection criteria [WK85] were used, determining around 20 paths as having the shortest descriptive length [Mac03] of the radio channel, thus characterize it well enough.

absence of channel knowledge at the transmitter, in Horizontal (H-)BLAST, each spatial stream is transmitted by one antenna and the receiver has a linear front-end, followed by successive interference cancellation (SIC).[12]

Fig. 2.6 (a) shows the setup for SU-MIMO. On an individual time-frequency resource just a single user is served per cell. A good overview on those classical MIMO techniques can be found in [Spe05] and [PNG03].

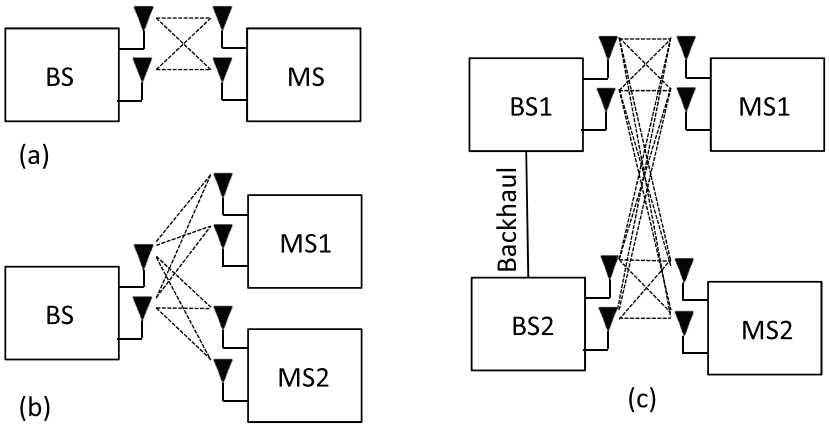

Figure 2.6.: Schematic comparison of (a) SU-MIMO, (b) MU-MIMO and (c) CoMP.

SU-MIMO is today also well established within the LTE standard [3GPc], starting from Release 8 on. With the offered downlink closed-loop linear precoding mode, a rank indicator is fed back by the mobile, indicating the number of spatial streams which should be used for transmission in the current radio propagation environment for best estimated throughput. The most important part of the MIMO feedback is the so-called Precoding Matrix Indicator (PMI) which provides the index of a preferred weight matrix. The transmit powers per spatial stream are kept equal, but the choice of the MCS offers adaptation possibilities. A 4 bit PMI offers the choice of 16 different weight matrices per rank, thus allowing a coarse adaptation to the channel conditions.

When there is more than one active user per cell, several users can be served on the same time-frequency resource, which is known as multi-user MIMO (MU-MIMO), illustrated in Fig. 2.6 (b). In the terminology of information theory, uplink MU-MIMO is the MIMO multiple access channel (MAC)[13]. The single antenna two-user case has a pentagonal two-user rate region [CT06]. Vertices of the capacity region can be achieved with the users transmitting at full power and the receiver using an MMSE detector with SIC [Ver89][Tel99].

The MU-MIMO downlink corresponds in information theory terminology to the MIMO broadcast (BC) channel. With full channel knowledge at the transmitter, the dirty paper coding (DPC)

[12]Remark on SIC complexity : The workload of current LTE mobiles, which do not use SIC for reception, shows that more than 60% of the receiver processing load is consumed in the decoder, around 30% in the RF front-end, and only 10% for demodulation, detection and channel estimation [vB09]. As SIC would multiply the decoder complexity for multiple iterations, it is still not widespread.

[13]This abbreviation should not be confused with medium access control, from the OSI layer model.

technique [Cos83] uses sequential user encoding and, similar to an interference cancellation approach at the transmit side, ensures that a given user experiences interference only from users encoded after it. The MIMO extension of DPC is able to achieve the MIMO BC rate region [WSS06]. [14]

Unfortunately, DPC is not a practical technique e.g. due to high complexity and issues with peak power constraints. A feasible non-linear precoding approach is Tomlinson Harashima precoding [GFH95] , but for the sake of complexity, usually only linear precoding is considered by the MU-MIMO research community. Asymptotically, with large number of users, a linear zero forcing (ZF) precoder has the same sum-rate scaling as DPC [YG06]. When the angular spread at the base station is low, propagation channels from transmit antennas in a linear array configuration are correlated. In this case also a combination of beamforming and user selection can enable spatial division multiple access (SDMA) as a simple form of MU-MIMO [WIN05].

The performance potential of MU-MIMO is regarded to be larger than SU-MIMO. Especially when there is a certain set of users available for transmission, *user selection* aims at selecting spatially orthogonal users [DS05][Wil06]. While SU-MIMO has to operate with channel conditions of a single user and the respective channel rank with corresponding multiplexing capabilities, MU-MIMO in conjunction with user selection allows to actively influence the composite MIMO channel and thus increase the multiplexing gain, in case channel knowledge is available at the transmitter.

In the literature, e.g. [BKBO13], there is strong evidence that for MU-MIMO with multiple receive antennas, a single spatial stream per MS is enough to yield the full performance potential. This means the degrees of freedom at the transmitter are used for transmitting to multiple MS with one spatial stream each, instead of using the degrees of freedom for providing some MS with more than one spatial stream.

LTE system simulations, e.g. [AL07], have shown that current MIMO-OFDM systems need to operate at frequency reuse 1 (or close to 1) in order to be spectrally efficient. This means that neighbor cells are operating at the same time-frequency resources. Thus at the cell edges, strong inter-cell interference can occur. As a result, for cell edge users, the SINR at the receive antenna is low, and hence, MIMO cannot efficiently provide multiplexing gains, as the multi-antenna resources are required in order to improve the poor SINR.

For currently deployed LTE systems, as soon as the cell load will increase, the network performance suffers from inter-cell interference limitations.

2.1.4.2. Network-MIMO and CoMP Overview

In CoMP, a set of base station sectors/cells, the *coordination cluster* works together. In this set, the *serving cell* from the perspective of a particular MS, is the one which the MS would be attached in the absence of CoMP mode operation. The *supporting cells* are the additional cells of the CoMP cluster which are included to increase the performance of this MS.

[14] There is a remarkable duality between MIMO BC and MIMO MAC [VJG03],[WSS06].

When base stations of multiple cells are able to exchange information, the inter-cell interference limitation problems can directly be handled by spatial processing algorithms. Wyner came up in 1994 with the fundamental limits in the uplink when jointly receiving across multiple cells [Wyn94] for particular scenarios [15]. This type of joint reception is in the literature also known as *network MIMO*. Wyner also showed that an MMSE receiver is in many cases (nearly) as good as an optimal receiver. Uplink network MIMO using MMSE receivers was e.g. considered in [Ven07]. An alternative to this approach is to establish a distributed interference cancellation approach across the entire network. In the so-called network interference cancellation (NICE) approach [BKKR11], each cell first decodes, if possible, its strongest users and forwards dominant interferer user data to neighbor cells. Exchange of user data is clearly less backhaul-consuming than complex-valued receive signal samples.

With an unconstrained backhaul (no latency issues and unlimited backhaul capacity), uplink network MIMO from an information theoretic perspective is identical to the MIMO MAC with antennas distributed across the different cell sites.

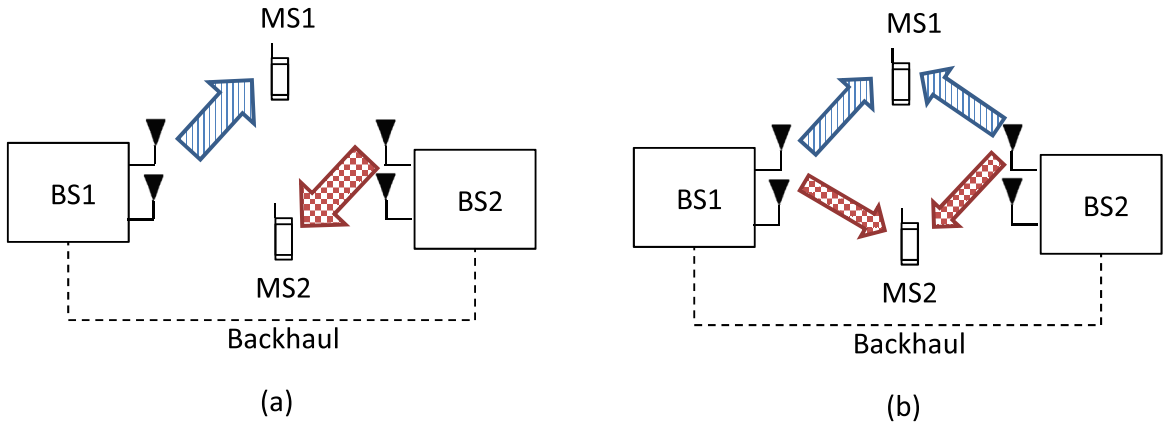

Figure 2.7.: Schematic picture of (a) Downlink coordinated scheduling/beamforming (CS/CB), (b) Downlink joint transmission (JT) a.k.a. Network MIMO.

In the downlink, [SZ01] suggested to use dirty paper coding across multiple cells, in order to do joint transmission from those cells to the mobiles in the coordination area. Again, when the backhaul is unconstrained, the downlink network MIMO setting, shown in Fig. 2.7 (b) is in fact the information-theoretic MIMO BC with particular transmit power constraints. Similar as in MU-MIMO, DPC is not considered as practical. The non-linear precoding alternative could be Tomlinson-Harashima precoding, as discussed e.g. in [SSF+10]. Linear precoding here also is the more simple alternative, e.g. in [HTH+09] downlink network MIMO is investigated with linear precoding and limited sets of coordinating cells.

A different notion of base station coordination is when the cell sites only share CSI and control information, but no user data. Base stations can now use their transmit precoding in order to spatially avoid interference. This is known as coordinated beamforming or multi-cell beamforming.

[15] Base stations and users in a line, or in hexagonal layout.

The principle is depicted in Fig. 2.7 (a). In information theory terminology, this is the Interference Channel [CT06]. Its capacity is still unknown, even in the Gaussian case. Recently, a lot of research activities in this Interference Channel setting has come up on so-called *Interference Alignment*, e.g. [CJ08]. The basic idea of interference alignment is to ensure per receiver that all interference is received within the same subspace[16], so that the useful signal part has the remaining subspace available as degrees of freedom, e.g. for spatial multiplexing capabilities. While the idea is promising, one can criticize the usual performance metric used by the Interference Alignment community, focusing on maximizing the degrees of freedom. This gives the slope of the capacity curve, asymptotically for very large SNR. But: This might not be the relevant operation point for cellular systems. So there are also skeptical opinions about potential gains with larger number of cells [MGL12]. Interference alignment will not be in the scope of this thesis, as it is a large topic on its own, considered not yet close to applicability to LTE-A.

With the 3GPP standardization activities for LTE-Advanced (3GPP release 10+) being prepared beginning early 2008, multi-cell coordination schemes were considered as very promising candidate technologies in order to improve the LTE system performance, especially at the cell edge. Coordinated multi-point transmission and reception (CoMP) was the term used within 3GPP [3GPf]. 3GPP distinguished between two cases, illustrated in Fig. 2.7. First, Joint transmission (JT) and reception (JR) schemes, requiring exchange of data, CSI and control information, being essentially network MIMO. Second, coordinated scheduling/beamforming (CS/CB) techniques, only requiring exchange of data and control information.

The discussion on CoMP began in Release 10 and led to inclusion of new downlink reference symbols, denoted "channel state information reference symbols" (CSI-RS). They support the CSI measurement of additional cells beyond the classical single serving cell. Additionally, "demodulation reference symbols" (DMRS) were introduced, which are precoded in the same way as the data for joint transmission. This allows to measure the effective precoded channel, regardless of needing to know which transmission point was actually involved in the transmission. LTE Release 11 continued the investigations on CoMP [3GPg], including also different evaluation scenarios. The outcome of Release 11 standardization at the end of 2012 was the inclusion of additional feedback reporting options, mainly tailored for CS/CB, allowing to report also CSI for neighbored cells. As this CSI still uses the classical PMI codebooks defined in [3GPc], this feedback is still rather coarse, supporting interference avoidance by beamforming, like in 2.7 (a), but not yet supporting coherent downlink JT, as in 2.7 (b). This release 11 multiple PMI reporting allows to get an impression which precoding index ("beam") from a neighbor cell generates the largest interference. This kind of reporting mechanism was defined by the author of this thesis as "worst companion" [WH10], suggested in the standard in [AL08] with discussion of performance results in [AL10] and [GAB$^+$11]. At the cell edge, throughput gains larger than 30% were obtained in system level simulations, as the worst companion feedback can be used to spatially avoid worst case interference situations for a mobile in the downlink using scheduling restrictions.

In Release 12 (started in 2013) and beyond, CoMP investigations are ongoing, potentially adding also full support for downlink JT. Explicit feedback as well as implicit feedback will be considered as options. Explicit here means a description of the channel which does not make any

[16] The subspace dimensions might be multiple antennas, but also multiple access resource sets or a combintation of both.

assumptions on used receiver types. In contrast to that, implicit feedback, as used e.g. in release 8 PMI reporting, selects the preferred precoding based on a particular receiver type assumption at the MS side, like Maximum Ratio Combining.

Accompanying the discussion within 3GPP, also in the research literature, there are system level performance studies on uplink CoMP LTE available, e.g. [MFS10].

A good general overview on research outcomes and existing challenges in coordinated multi-cell multi-antenna networks is given in [MF11] and [GHH+10].

2.1.4.3. Backhaul Aspects of CoMP

The existing celluar architecture was touched in subsection 2.1.2, illustrated in Fig. 2.2. Depending on the actual coordination scheme, the resulting backhaul rates can become rather high, see e.g. [HFM11][17]. Furthermore, with the refraction index of fiber around 1.5, the propagation delay is 5μs per km. Switching latencies add up to the overall delay budget. But: From a technical point of view, according to [FD11], "today's backhaul technologies are powerful enough to support CoMP", as long as the coordination cluster is limited in size, e.g. below several tens of kilometers. For existing architectures, as depicted in Fig. 2.2, it is proposed to add additional interconnection links on different aggregation levels, e.g. connect the rightmost switches in the figure directly in order to reduce latency.

While the illustrative figures, shown in this thesis up to now, imply that fully functional base stations exchange information, there are a manifold of possible architecture realizations for CoMP. A few examples:

- In a network of fully functional base stations, base stations are adopting master or slave roles, exchanging information via X2 interface. (The per-user master in uplink joint reception might be the BS of the serving cell.)
- A very straightforward approach is to deploy remote radio heads (RRH) in a distributed manner, realizing several radio cells and connect them via fiber with the CPRI interface [CPR11] to a central processing unit (CU). This CU, equipped with powerful processing capabilities, acts now as one "super base station", handling all the CoMP processing. A drawback might be that this deployed fiber might be inefficiently used and the coordination areas are completly fixed, defined by the RRH wiring. (The links from the baseband to the RRHs are denoted "front-haul".)
- A sophisticated approach for saving backhaul resources is an architecture with *split processing* [DDM+13]. Instead of fully centralizing the processing at a certain point, [DDM+13] analyses different potential processing split points in the LTE baseband processing chain. E.g. with the so-called "PHY2"-split there will be distributed network elements handling the physical layer processing, starting from resource-mapping/demapping in Fig. 2.1, while central processing units handle the per-user processing in the outer parts of the processing chain.

[17] For 3 cooperating BS with 20MHz bandwidth: While for CS/CB the resulting rates are in the order of 1MBit/s, for JT and JR (network MIMO), backhaul rates of more than 1GBit/s may be required.

An information theoretical analysis of the rate-constrained backhaul is discussed e.g. in [Mar10]. Here the trade-off between capacity gains and required backhaul achievable with various CoMP concepts is discussed.

Based on existing literature cited here, the conclusion for this thesis regarding the entire CoMP backhaul aspects are:

- The required rates will be supported already by today's technology. So it is a matter of each cellular network operator's existing and planned infrastructure, thus a matter of cost, where CoMP-based systems can be deployed.
- The latency poses a more fundamental restriction due to the limits of light speed propagation. This means that only a limited set of cells can reasonably cooperate. This leads us to the concept of *coordination clusters*, discussed in section 2.3.

Throughout this thesis, besides reasonably limiting coordination cluster sizes, we do not further consider additional limitations of the backhaul.

2.1.4.4. Channel Estimation for MIMO and CoMP

Certain simple communication strategies, like differential binary phase shift keying (DBPSK) in conjunction with non-coherent detection, do not rely on accurate channel knowledge at the receiver. In contrast, for spectrally efficient communication with higher order modulation, like coherent detection of 64 quadrature amplitude modulation (64-QAM), channel knowledge is needed, so that the receiver can take account for frequency-selective amplitude and phase variations of the propagation channel for proper symbol detection. Furthermore, for MIMO, in order to achieve the full ("closed-loop") MIMO capacity [PNG03], also channel state information at the transmitter is required. In frequency division duplexing (FDD) systems, the downlink channel is not reciprocal to the uplink, due to the duplexing frequency distance generating different fading realizations. Hence, for the adaptation of the transmitter, channel estimation at the receive side is required used for feedback generation. In time division duplexing systems (TDD), for exploiting channel reciprocity, channel estimation in the reverse link direction is required.

In the MIMO research community, a large part of the existing literature only concentrates on narrowband channels, or block-flat channels, where a single complex scalar coefficient is sufficient to describe one antenna-to-antenna link. But existing cellular communication systems, like LTE, are wideband and OFDM based and it is very likely that also future systems will be multi-carrier systems. This thesis will thus take into account OFDM, using LTE-A system parameters, where needed. Considered channels are assumed realistically fully frequency selective and time-variant. These properties are essential for the CoMP setting, as described below.

While there is a certain history in the literature on blind channel estimation [BHP02], not requiring training sequences, in this thesis, we restrict ourselves to pure *pilot-based channel estimation*. All noteworthy existing cellular communication systems have pilot / training / reference / synchronization sequences, not only for channel estimation, but also for measurement purposes (e.g. for cell selection and handover), for synchronization etc. The LTE-A uplink pilot pattern and structure is described in appendix A.

The design of the pilot grid follows the aimed mobility and cell size support, defining Doppler and delay spread and thus coherence time and coherence bandwidth of the channel (see subsection 2.1.3). This means that usually the 2-D sampling theorem has to hold for proper pilot placements. Exceptions are possible in the presence of sparsity, where a reduced pilot density may be possible by using compressed sensing approaches.

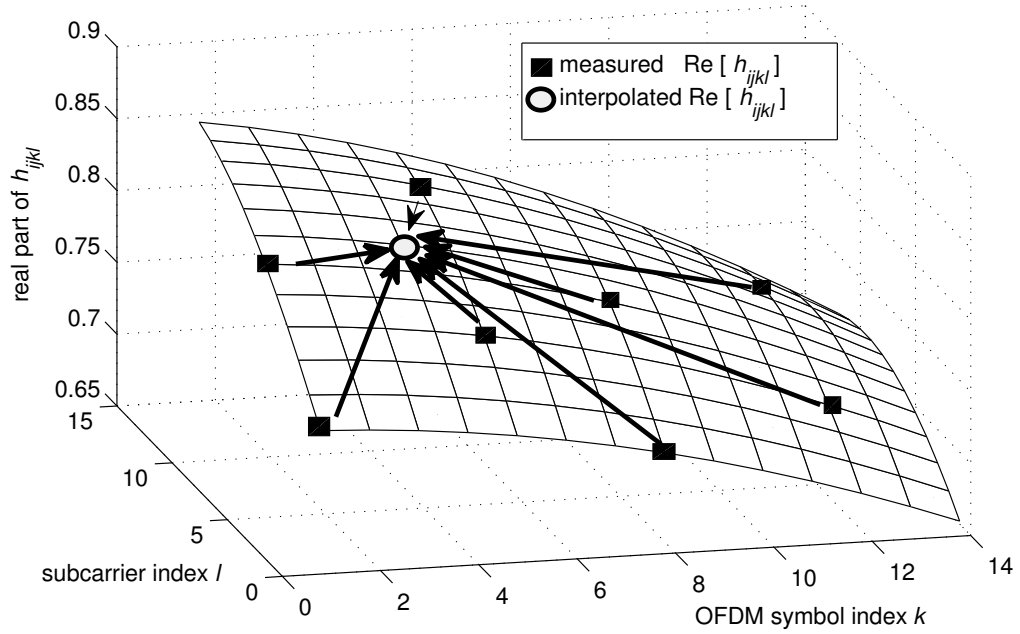

Figure 2.8.: Real part of time-variant, frequency-selective channel coefficient h_{ijkl} for one transmit antenna j to one receive antenna i, illustrating OFDM pilot-aided channel estimation in the LTE DL pilot grid: The squares represent the channel at the pilot symbol resource element positions, the circle represents a particular data symbol resource element position, for which the channel shall be estimated by interpolation.

A good first choice for OFDM channel estimation is to use Least Squares [Kay93] [BLM03]. This has proven to perform usually well in OFDM systems. In time-variant frequency-selective channels, the linear estimator, minimizing the mean squared error, based on statistical knowledge, is the two-dimensional (w.r.t. time and frequency) Wiener filter, described in [HKR97] for SISO OFDM.

Fig. 2.8 depicts the estimation task: Estimation of h_{ijkl} at the squared pilot symbol resource element positions (the pairs of l and k) is a *smoothing* task, while at the example circle data symbol position we face an *interpolation* task. For data symbol positions at the edge resource elements lying "outside" of the pilot grid, the estimator has to do *prediction* (or extrapolation). All this is treated in a unified way by the Wiener filter.

With MIMO OFDM, when different antenna links are uncorrelated[18], one 2-D Wiener filter per

[18] Which is the case for the commonly deployed two-element cross-polarized BS antennas, in conjunction with high angular spreads at the MS antennas

antenna-to-antenna link ("MIMO subchannel") is still optimal in the presence of orthogonal per-transmit-antenna pilot symbols. Otherwise, for antenna correlations, the channel estimator benefits when performing a 3-D estimation with the additional spatial dimension [WA06]. In this thesis, the considered estimators neglect antenna correlations. This is because in CoMP, antennas at multiple sites have uncorrelated fading. Co-located BS antennas are typically uncorrelated when they are deployed in the usual cross-polarized configuration and MS antennas are uncorrelated because of high angular spreads, as the MS in an urban macro cell environment is usually much stronger surrounded by scatterers than the BS.

The Wiener filter requires statistical parameter knowledge on the radio channel. This is not always easy to obtain, especially in the CoMP case. A practical realization of the Wiener filter in a simplified form with stripped down required parameter knowledge is done in [SS00].

MU-MIMO channel estimation is discussed e.g. in [WA06]; here the proposed 3-D MAP-based solution coincides with the Wiener filter solution. The actual multi-user channel estimation is here simplified by assigning different OFDM resource element positions (different subcarriers and/or OFDM symbols) to the different users, where all other users transmit zeros, meaning that the multi-user situation has been transformed to a single user situation. All the statistical parameters are assumed to be known by full knowledge of the underlying channel model and its parameters like angular spread, delay spread and Doppler spectrum.

In [BLM03] optimal training design for MIMO OFDM in conjunction with LS channel estimators is studied: In order to minimize the MSE, the sequences in time-variant, frequency-selective channels have to be equipowered, equispaced, and phase-shift-orthogonal in frequency domain, corresponding to circular shift orthogonality in the time domain.

In coordinated multi-point, the multi-cell channel estimation research challenge for proper performance is addressed in [ZTW+11]. In [TSS+08] for the downlink it is proposed to use Walsh-Hadamard sequences stretched over time in order to orthogonalize the incoming pilot sequences from different cells, presuming that the channel is approximately not changing over time. This allows a very simple correlation-based channel estimator.

Beyond pilot-aided channel estimation, powerful iterative techniques exist, which use the help of data symbols for improving the channel estimation quality. The prize has to be paid on the complexity side. Regarding turbo equalization complexity, according to [TS11], trellis-based schemes have a complexity order of the modulation symbol alphabet size to the power of delay spread length, while MMSE based algorithms have an order of squared equalizer support size. According to [vB09], the decoder is by far the most computation time intensive part of the base band processing. In case of iterative channel estimation, exploiting the decoder, this part becomes a multiple. For the sake of complexity, in this thesis, iterative and data-aided techniques are excluded.

2.2. Problem Statement: Channel Estimation for Coordinated Multi-Point

The former sections showed the theoretical benefits of CoMP. One objective of this thesis is a practical implementation of multi-cell channel estimation. The scenario is depicted in a simplified manner in Fig. 2.9. The focus is on uplink joint reception in the LTE-A context. In addition to conventional single-cell channel estimation, the potentially weak cross-links from neighbor cells, depicted by dashed lines in Fig. 2.9, also have to be estimated in order to coherently collect useful signal contributions and suppress unwanted interference across the distributed antennas of an entire set of cells. For the sake of complexity, this thesis restricts the set of channel estimators to linear pilot-aided estimation. [19]

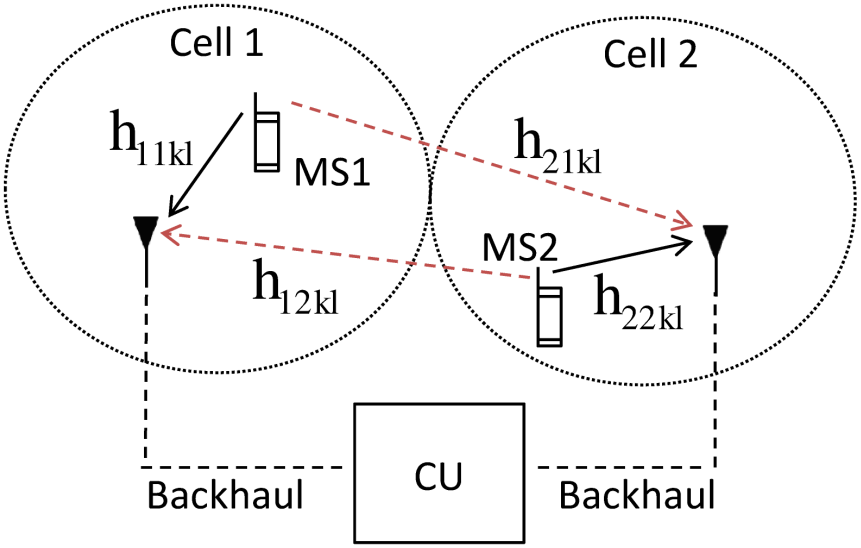

Figure 2.9.: Schematic picture multi-cell channel estimation: 2-cell, 2-user uplink joint reception example, single Tx/Rx with a central processing unit.

While multi-user channel estimation for CoMP is rather similar to the MU-MIMO case, there are particularities: The SINR operation points for mobiles outside of the own serving cell are much lower than in the typical MIMO OFDM case. For satisfying spectral efficiency, the system shall operate in full frequency reuse 1, meaning for the estimation process that a large number of mobiles is generating co-channel interference. Here the thesis deals with a significant challenge, which is typically neglected by academic research and becomes an important issue in the particular CoMP operation points: Obtaining the required *parameter knowledge* for channel estimation in realistic time-variant frequency-selective propagation channels.

In block-flat radio channels, which are often assumed in the research community, the magnitude response of the radio channel is constant over a whole time-frequency block and the parame-

[19] The achieved performance results in chapter 3 will also justify that this kind of restriction does not cause significant performance losses. Note that the linear MMSE (LMMSE) estimator is the Wiener filter. LMMSE estimation does not necessarily coincide to an MMSE estimator, which is not restricted to linear processing. This will be discussed in chapter 3.

ter knowledge collapses mainly to the knowledge of noise variance. When taking into account the time-variant frequency-selective nature of channels, the channel changes at least slightly for each OFDM resource element. Parameters like the time-frequency autocovariance of each user's channel are required for reasonable performance. Furthermore, due to those varying channels, the orthogonality of pilot sequences is impacted: At the receive side the sequence orthogonality is not preserved when the sequences of multiple users superimpose, each with a varying channel over the entire sequence length. This will be investigated in this thesis. So in this multi-user multi-cell context, we cannot neglect the choice of the pilot sequences, as they impact the estimation quality. This thesis thus will also look at different assignment strategies of pilot sequences and their interplay with the channel estimation algorithms.

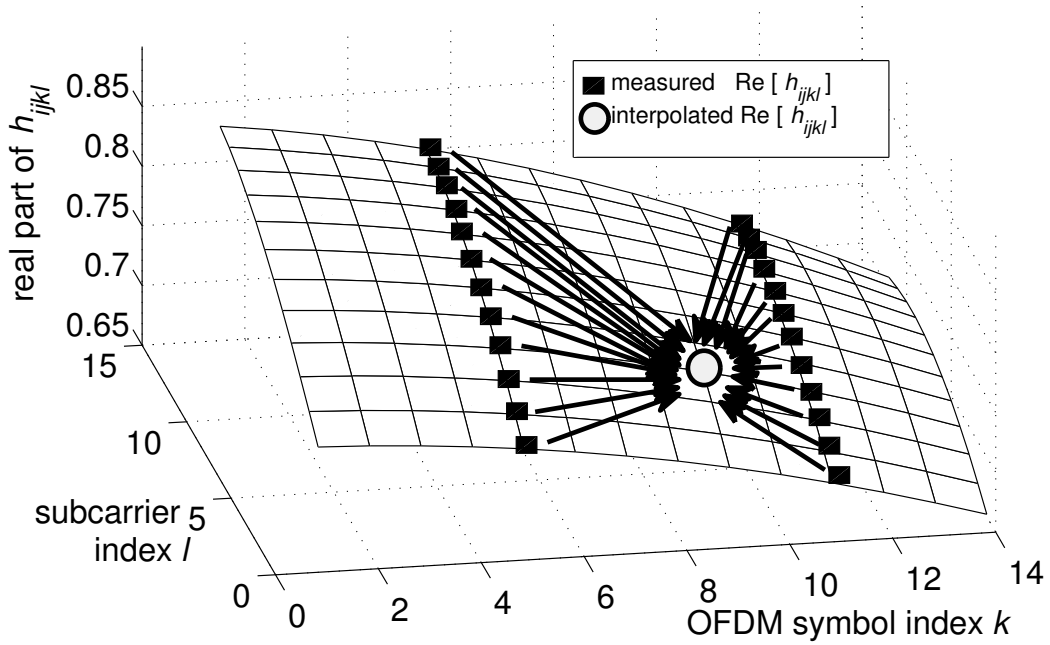

Figure 2.10.: Real part of time-variant, frequency-selective channel coefficient h_{ijkl} for one transmit antenna j to one receive antenna i, illustrating OFDM pilot-aided channel estimation in the LTE UL pilot grid: The squares represent the channel at the pilot symbol resource element positions, the circle represents a particular data symbol resource element position, for which the channel shall be estimated by interpolation.

Fig. 2.10 shows the estimation scenario: The time-variant, frequency-selective channel transfer function h_{ijkl} between user antenna j and base station antenna i of a particular data resource element (shown as circle), located at OFDM symbol k and subcarrier l, has to be estimated based on the received signal at the pilot symbol resource elements (shown as squares). The known pilot sequences are fully interfered by neighbor cell pilot signals, transmitted by other MS antennas j'. In CoMP, when channels of supporting cells shall be estimated, the received pilots of interest may be much weaker than the interfering signals (e.g. 10 dB below). In addition, noise is present.

In this work, for uplink we focus on MS with a single transmit antenna. Hence, in the system model of chapter 3 there is no precoding and superposition of spatial layers at the transmitter. The modulation symbols x_{jkl} at the input of the OFDM modulator are simply the data symbols s_{jkl}.

While there are no particular assumptions on the statistical properties of the propagation channels, the main focus in this thesis lies on zero-mean channels. The derived linear estimators are also applicable to non-zero mean channels. In this case the terminology for "covariance" has to be taken carefully, as discussed in appendix B.

As the cellular system has to support dynamic scheduling of users, and as users generate limited amounts of traffic, there is only limited time to collect statistical channel knowledge until large-scale radio propagation conditions have changed. The set of scheduled users on a particular time-frequency resource even changes much faster, usually each subframe. Furthermore we are facing a "chicken-and-egg problem" here: In order to properly separate the users, we need statistical knowledge and in order to obtain this knowledge, we have to separate the users.

As a performance metric for channel estimation, in literature often the *mean squared error* (MSE) of the channel estimates $\mathrm{E}\left\{|\hat{h}_{ijkl} - h_{ijkl}|^2\right\}$ is used. This metric, being purely focused on the estimates, is not considered to be sufficient for this thesis. Here we want to evaluate the impact of a certain channel estimation error on CoMP performance after multi-antenna combining. Thus, the main performance metric used throughout this thesis is the SINR at the output of the joint reception receive combiner with estimated channels.

A further subproblem handled within this thesis is the need for a simplified model of channel estimation errors. While sections 3.2 to 3.4 of this thesis take into account all details of physical layer processing including all OFDM pilot and data resource elements, for system level simulators this way is not feasible, as the simulation times would be intractably long when there are e.g. 57 cells simulated with about 1000 users with every processing detail taken into account. The modeling thus requires simplifications and abstractions. This simplification challenge is discussed and solved in section 3.5. A further objective for the simplified model in this thesis is the support of performance prediction for different particular types of transmit or receive algorithms for both uplink and downlink.

2.3. Problem Statement: Downlink Closed-Loop Coordinated Multi-Point Transmission

In this (chapter 4 and 5) part of the thesis, we will investigate downlink linear precoding and receive combining strategies for CoMP in the presence of imperfect channel knowledge.

In order to obtain the channel knowledge at the transmitter side for the channel-adaptive precoder weight computation in a frequency division duplexing (FDD) system we require a closed-loop, as depicted in Fig. 2.11. This means, we have to estimate the channel at the receive side and feed channel state information back to the transmitter. The estimator has impairments due to estimation errors, the feedback brings in additional quality reductions due to quantization and

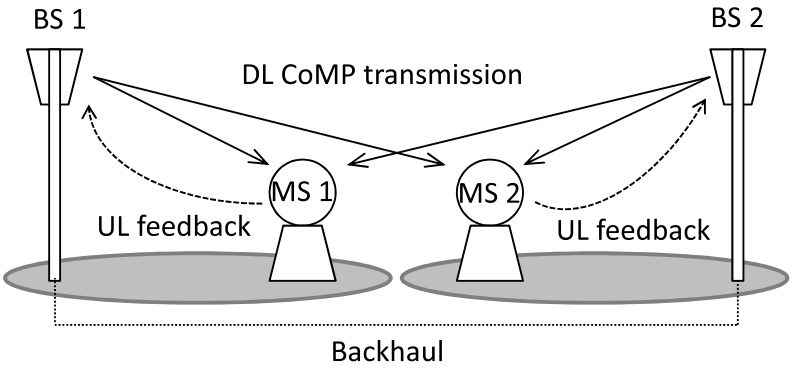

Figure 2.11.: Schematic picture of downlink closed-loop coordinated multi-point transmission.

delays. For the investigations, we use the error models generated in section 3.5, based on the problem formulations of the previous section 2.2.

While base stations will be fully able to exchange information via backhaul, mobiles will not be able to directly share information. Thus we also do not consider a joint design of precoders and receive combiners, as one mobile has no access to receive combining weights or channel knowledge of other mobiles, because this would result in prohibitive signaling overhead.

For precoding, we do not consider any numerical optimization schemes, like convex optimization [BV04], for the sake of manageable complexity in real-time mobile communication systems. We pick up a promising precoder strategy known from multi-user MIMO [STS07] and show its universal applicability both to joint transmission as well as interference avoidance by coordinated beamforming.

Furthermore, this thesis deals with the problems occuring at the CoMP cluster edges. As discussed in subsection 2.1.4, at least fundamental limits from physics will restrict the sizes of coordination clusters. In the uplink, user-individual cell clusters are possible, as the received signals can be processed user-individually. For each user all relevant signal contributions can be collected for a cell area centered around that user. However, in the downlink we have a different, more difficult situation. Transmitting to a certain user means investing in transmit power for it, which more or less prevents a direct solution for user-individual clusters, as a set of users shares available power resources.

This means that the former inter-cell interference problem for systems without CoMP, is now replaced by an inter-cluster interference problem. It is obvious that clusters should not exchange user data, as then they would become a new “super-cluster”, leading the cluster size restriction ad absurdum. Moreover, the scheduling entities of the clusters should be allowed to operate completely decoupled, otherwise again the separation into clusters would be harmed. This thesis shows, even under those strict requirements, how the inter-cluster interference can be mitigated by optimizing the precoder of each cluster after the scheduling decisions have been made. Inbetween the clusters, a limited exchange of channel knowledge for cluster edge mobiles is required. The thesis presents a solution introducing minimum additional latency into the system.

3. Channel Estimation for Coordinated Multi-Point

This chapter will address the challenge of channel estimation for CoMP, in particular in the context of a LTE-A or near LTE-A setting (see appendix A). In 3.1, the general *signal model* is introduced and the objective function of the estimator is defined.

Section 3.2 will focus on an uplink multi-cell *single user* scenario. This scenario will be of interest for operation in a noise-limited regime in the absence of interference (or similarly an indication of the impact if the receiver treats all the other signals as noise[1]). This includes parameter estimation, like noise variance and channel covariance estimation. Furthermore the issue of required "warm-up time"[2] for parameter estimation is investigated, looking at advanced methods to reduce the required observation time.

In the *multi-user* scenario, the choice of pilot sequences becomes important. Section 3.3 will compare different sequence assignments and answer the question "how orthogonal" the sequences have to be if different types of estimators are used. Those estimators are from two different classes: Either the channels are estimated *jointly* for all users within the coordination cell set, or, *separately* per user.

In a further step, section 3.4 deals with the parameter estimation challenge for the multi-user case, which is even more demanding. Noise estimation is revisited, as simpler methods may fail due to high interference levels in multi-user CoMP. With different combinations of users and potentially fragmented activity, the need for short "warm-up times" becomes even more important as in the single user case. All parameters are to be estimated from the receive signal, no prior knowledge is assumed.

Computationally intensive simulations, like system level simulations demand abstraction and simplification of actual channel estimation in order to get performance figures of complete MIMO and CoMP systems. A channel estimation error model is introduced in section 3.5 which is simple and easy to use.

Finally, section 3.6 provides a short wrap-up on lessons learned in channel estimation for CoMP throughout this thesis.

[1] Treating interference as noise might be of interest with impartial information and distributed receiver processing, when estimating channels *separately* per user.

[2] We define warm-up time as the number of subframes where statistical knowledge was already collected *before* the actual channel estimation starts.

3.1. Uplink Signal Model and Linear Estimator

3.1.1. Input-Output Relation and Normalization

With the channel transfer function h_{ijkl}, introduced in (2.9), the receive OFDM symbol k at subcarrier index l and receive antenna i is:

$$y_{ikl} = \sum_{j \in \mathcal{J}} h_{ijkl} x_{jkl} + n_{ikl} \tag{3.1}$$

The transmit antenna index j stems from the set $\mathcal{J}$ of $|\mathcal{J}| = J$ transmit antennas of different mobile stations which actively transmit, thus are scheduled by the base stations. Their signals superimpose at each element of the BS antenna set $\mathcal{I}$, indexed i, containing $|\mathcal{I}| = I$ total number of antennas. $\mathbf{z}_{ik}$ from the previous chapter is now replaced by explicit modeling of interference and an additional Gaussian distributed noise term n_{ikl} with variance σ_n^2, representing AWGN (from e.g. thermal noise) and all unmodeled sources of interference.

The mobiles transmit symbols s_{jkl} with power $\mathrm{E}\{|x_{jkl}|^2\} = \rho_j$ over the channel with average path gain $\mathrm{E}\{|h_{ijkl}|^2\} = \alpha_{ij}$ from MS j to BS i. The joint scaling effect of transmit power and path gain is denoted by $a_{ij} = \rho_j \alpha_{ij}$ and in this thesis denoted as *effective path gain*. Thus we have:

$$y_{ikl} = \sum_{\forall j \in \mathcal{J}} h_{ijkl} s_{jkl} + n_{ikl} \tag{3.2}$$

In the LTE-A uplink, as detailed in appendix A, the P pilot positions of all users fully overlap. The positions are described here as a set of ordered pairs of time/frequency indices $(k_i, l_i) \in \mathcal{P}$ with $\mathcal{P} = \{(k_1, l_1), (k_2, l_2), \ldots, (k_P, l_P)\}$. In case of single transmit antenna mobiles, the set of users $\mathcal{U}$ equals the set of transmit antennas $\mathcal{U} = \mathcal{J}$.

In a vectorized form, one could think of expressing the channel as a huge vector comprised of all the four index dimensions. Due to tractability and accessibility, in this thesis, channel representations in different dimensions are used. In section 2.1.4.4, it was discussed that channel estimators in case of uncorrelated receive antennas, fitting our scenario, can be run separately per receive antenna i. The used representations for a channel are:

- A vector $\mathbf{h}_{ikl}^{[\mathrm{U}]}$ in *user dimension* is appealing to collect the output of multi-user channel estimators at particular resource elements k,l.
- A vector $\mathbf{h}_{ij}^{[\mathrm{P}]}$ in *pilot dimension* is helpful to handle exploiting channel correlations in time-variant, frequency-selective channels for each transmit-to-receive antenna link i,j.
- A vector $\mathbf{h}_{jkl}^{[\mathrm{B}]}$ in base station *receive antenna dimension* is useful to deal with the spatial processing of receive antenna combining for each user j at particular resource elements k,l.

$$\begin{aligned}\mathbf{h}_{ikl}^{[\mathrm{U}]} &= [h_{i1kl}, h_{i2kl}, \ldots, h_{iJkl}]^T \in \mathbb{C}^{J\times 1} \\ \mathbf{h}_{ij}^{[\mathrm{P}]} &= [h_{ijk_1l_1}, h_{ijk_2l_2}, \ldots, h_{ijk_Pl_P}]^T \in \mathbb{C}^{P\times 1} \\ \mathbf{h}_{jkl}^{[\mathrm{B}]} &= [h_{1jkl}, h_{2jkl}, \ldots, h_{Ijkl}]^T \in \mathbb{C}^{I\times 1}\end{aligned} \tag{3.3}$$

Here the diagonal matrices $\mathbf{A}_{\mathbf{i}}^{[\mathrm{U}]} = \mathrm{diag}\left([a_{i1}, a_{i2}, \ldots, a_{iJ}]\right)$ and $\mathbf{A}_{\mathbf{j}}^{[\mathrm{B}]} = \mathrm{diag}\left([a_{1j}, a_{2j}, \ldots, a_{Ij}]\right)$ carry the impact of path gains and transmit powers.

The receive vector at antenna $i \in \mathcal{I}$ in pilot dimension can be written as

$$\mathbf{y}_i = \sum_{\forall j\in\mathcal{J}} \mathbf{S}_j \mathbf{h}_{ij}^{[\mathrm{P}]} + \mathbf{n}_i. \tag{3.4}$$

The per-user pilot symbols, with time-frequency positions defined by $\mathcal{P}$ are arranged in the diagonal matrix $\mathbf{S}_j \in \mathbb{C}^{P\times P}$, $\mathbf{S}_j = \mathrm{diag}\left([s_{jk_1l_1}, s_{jk_2l_2}, \ldots, s_{jk_Pl_P}]\right) = \mathrm{diag}\left(\mathbf{s}_j\right)$[3] with $\mathbf{s}_j$ as the pilot sequence vector. Both the observation vector $\mathbf{y}_i = [y_{ik_1l_1}, y_{ik_2l_2}, \ldots, y_{ik_Pl_P}]^T$ and the noise vector $\mathbf{n}_i = [n_{ik_1l_1}, n_{ik_2l_2}, \ldots, n_{ik_Pl_P}]^T$ are defined based on the pilot positions $\mathcal{P}$.

This receive vector (3.4) can be further rearranged into a *generalized linear model* (in the sense of [Kay93]) by stacking all pilot matrices into $\mathbf{S}^{[\mathrm{all}]} = [\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_J] \in \mathbb{C}^{P\times(UP)}$. Next, the channel vectors in pilot dimension $\mathbf{h}_{ij}^{[\mathrm{P}]}$ of all users are stacked into one column vector $\mathbf{h}_i^{[\mathrm{all}]} = \left[\mathbf{h}_{i1}^{[\mathrm{P}]\,T}, \mathbf{h}_{i2}^{[\mathrm{P}]\,T}, \ldots, \mathbf{h}_{iJ}^{[\mathrm{P}]\,T}\right]^T \in \mathbb{C}^{(UP)\times 1}$. This leads to the generalized linear model

$$\mathbf{y}_i = \mathbf{S}^{[\mathrm{all}]} \mathbf{h}_i^{[\mathrm{all}]} + \mathbf{n}_i. \tag{3.5}$$

Note that while (3.5) is helpful to demonstrate the properties of the underlying model, when directly used for the estimators, it leads to very "clumsy" dimensions, unfavorable for matrix inversions and additionally requiring a lot of unnecessary parameters in later statistical estimation. This will be discussed further in section 3.3.

In case of single antenna BS, the BS set $\mathcal{B}$ is equal to the set of receive antennas $\mathcal{I}$, thus $\mathcal{B} = \mathcal{I}$. In the thesis-related publications [WNtB12] and [WNtB13], for simplicity it was considered that each BS serves one MS, supported by capturing the signals of N_{supp} supporting base stations via a perfect backhaul. This results in $\mathcal{U} = \mathcal{J} = \mathcal{B} = \mathcal{I} = \{1, \ldots, N_{\mathrm{supp}} + 1\}$.

Despite this simple indexing used here, one has to keep in mind that the scheduling process of the users and their different data traffic patterns result into different combinations of each user per LTE-A subframe, indexed $\bar{\kappa}$. The scheduler takes into account buffer status, latency requirements, channel conditions etc for all users and decides on resource allocation of subcarriers l and OFDM symbols k. Even if a user has enough data in the buffer at time $\bar{\kappa}$ and could continuously transmit in $\bar{\kappa} + 1$ etc , it might very well be allocated on a different frequency position in $\bar{\kappa} + 1$ than in $\bar{\kappa}$. Thus, it is reasonable to consider a *per-subframe* channel estimation only taking over statistical knowledge from previous subframes.

[3] The notation with a diagonal matrix enables the required element-wise multiplication of pilot vector and channel vector, also known as Hadamard product.

3.1.2. Linear Estimation Approach

The objective of the estimator is now to estimate all channel coefficients $h_{ijkl} = \sqrt{a_{ij}}\tilde{h}_{ijkl}$ of interest between user antenna i, base station antenna j, for OFDM symbol k and subcarrier l. For this estimation the received signal (3.4) at pilot positions $\mathcal{P}$ is used, as well as the knowledge of the user-individual pilot sequences $\mathbf{S}_j$.

As this thesis focuses on linear channel estimation, due to this linearity the estimate $\hat{h}_{ijkl}$ is a weighted sum of $\mathbf{y}_i$ components, hence a complex-valued inner product of receive vector $\mathbf{y}_i$ and a weight vector
$\mathbf{w}_{ijkl} = \left[w_{ijkl}^{(k_1,l_1)}, w_{ijkl}^{(k_2,l_2)}, \ldots, w_{ijkl}^{(k_P,l_P)}\right]^T \in \mathbb{C}^{P\times 1}$:

$$\hat{h}_{ijkl} = \mathbf{w}_{ijkl}^H \mathbf{y}_i. \tag{3.6}$$

The weight vector depends on i, j, k, l. Each element of this vector is attributed to a particular pilot position (k_1, l_1). The proper choice of the weight vector is crucial for the performance of the estimator and is provided by different designed and investigated algorithms throughout chapter 3.

Sections 3.3 and 3.4 are dealing with multi-user channel estimation algorithms, where the aim is to obtain $\mathbf{h}_{ikl}^{[\mathrm{U}]}$ in "one shot", using the weight matrix $\mathbf{W}_{ikl}$ which can be stacked columnwise from per-user weight vectors $\mathbf{W}_{ikl} = [\mathbf{w}_{i1kl}, \mathbf{w}_{i2kl}, \ldots, \mathbf{w}_{iJkl}]$. The linear multi-user estimate is thus

$$\hat{\mathbf{h}}_{ikl}^{[\mathrm{U}]} = \mathbf{W}_{ikl}^H \mathbf{y}_i. \tag{3.7}$$

Section 2.1.4.4 already has introduced the terms *smoothing*, *interpolation*, and *prediction*. If our resource element position (k, l) of interest for estimation is part of the pilot set $\mathcal{P}$, then we are facing a smoothing task. Smoothing is in our case only relevant for intermediate supporting steps of algorithms as we are interested in the channel characteristic at data resource element positions, which are different from pilot resource element positions. For the 2-D time-frequency estimation, the discrimination between interpolation and prediction is not always fully clear. An interpolation task is given when our resource element position (k, l) of interest for estimation is inside the outer border of the 2-D area of the pilot set $\mathcal{P}$. The remaining region outside the outer border of the pilot set is a prediction or extrapolation task, where we can usually expect the worst performance as the channel coefficients of those resource elements are typically less correlated to the channel coefficients at the pilot positions. For the special case of a LTE-A subframe, smoothing happens at OFDM symbols 3 and 10 (counting from 0 on), interpolation at OFDM symbols 4 to 9 and prediction/extrapolation at symbols 0 to 2 (when the previous subframe remains unused by the user(s) of interest or is not available) and 11 to 13.

Whether linear estimation imposes a performance drawback (compared to any general non-linear approach) or is still optimal will be discussed in section 3.2.1.2 on Wiener filtering.

3.2. Single-User Multi-Cell Channel- and Parameter Estimation

While the system model in the previous section 3.1 describes the general multi-user case, we first start with a single user setting: The signal of one transmitting MS is received by several cooperating BS, as depicted by Fig. 3.1. Thus we are looking at (1) a noise limited case, where CoMP is merely used for enhancing coverage of the system. Alternatively this scenario can also be interpreted as a (2) multi-user system, where we treat interference as noise, thus do not use any particular multi-user algorithms (and the superposition of many interferers provide an interference characteristic similar to white Gaussian noise).

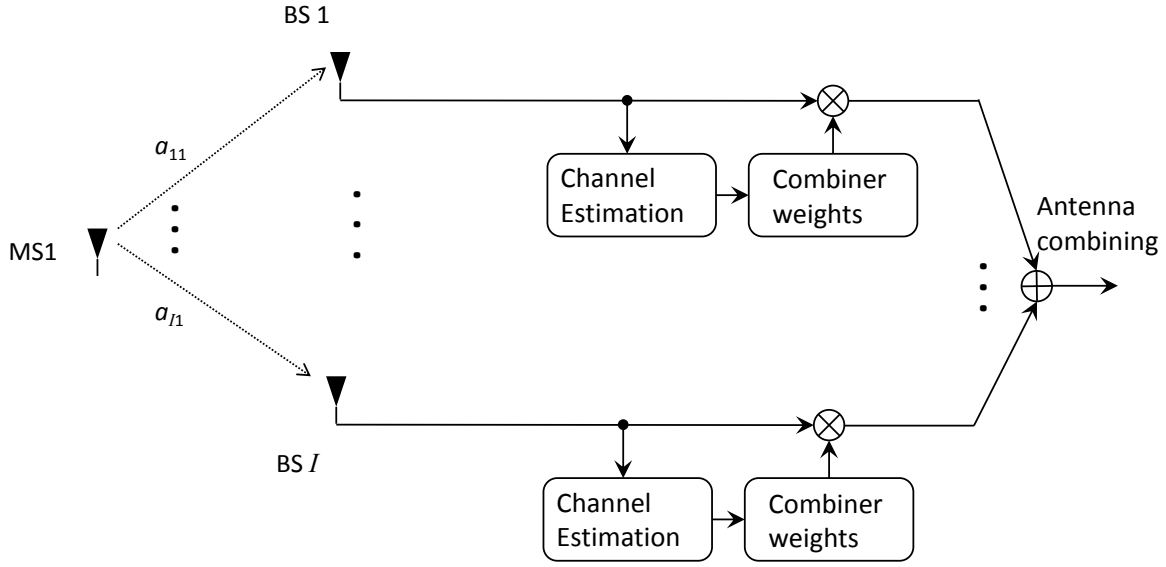

Figure 3.1.: Schematic picture of uplink single-user coordinated multi-point reception, including channel estimation and receive antenna combining.

An important CoMP-specific property is the potentially *strong imbalance of the different effective path gains* a_{ij} in (3.2) and (3.4), when the coordination set becomes larger. This is due to the fact that we are facing a distributed antenna system instead of a co-located one. Especially the effective path gains a_{ii} from a MS to the serving cell might be much stronger than the $a_{ij}, \forall i \neq j$ from a MS to the supporting cells[4]. For both scenario interpretations (1) and (2), the *low signal-to-noise ratio* (SNR) operation point is the most relevant one. Coping with (1) coverage limitations is obviously more urgent in low SNR, as well as (2) the multi-user system with Gaussian-like interference behaviour, emulated by AWGN.

The results developed throughout this section 3.2 will illustrate the strong sensitivity of uplink CoMP performance on the quality of the underlying channel estimation. Additionally, the access

[4] Another difference in CoMP is that propagation of signals to different non-co-located distributed antennas happens with completely different power-delay profiles (which is not the case for single-cell reception with co-located antennas). This prevents reusing statistical knowledge across non-co-located antennas.

to knowledge of required supporting parameters, like (3.19), described in 3.2.1, poses strong challenges to achieve satisfying accuracy, which will be addressed in subsection 3.2.2. Some results of this section are published in [NRWtB11].

Key Questions of Interest
After introducing the algorithms for least-squares OFDM channel estimation approaches and Wiener filtering in the context of CoMP, the often-neglected question of parameter estimation is discussed. Parameter estimation refers to getting access to required statistical knowledge for channel estimation. First the particularities in CoMP cellular systems are shown, which demand an assessment of parameter estimation. Then channel covariance and noise variance estimation are discussed, bringing up advanced methods to improve performance, especially when the available statistical observation time is short.

3.2.1. Channel Estimation Algorithms for Single-User

With just one explicitly modeled single transmit antenna user $\mathcal{U} = \mathcal{J} = \{1\}$, the pilot signal vector (3.4) for our single user setting of Fig. 3.1 simplifies to

$$\mathbf{y}_i = \mathbf{S}_1 \mathbf{h}_{i1}^{[\mathrm{P}]} + \mathbf{n}_i = \sqrt{a_{i1}} \mathbf{S}_1 \tilde{\mathbf{h}}_{i1}^{[\mathrm{P}]} + \mathbf{n}_i. \tag{3.8}$$

First, least squares based approaches will be discussed, followed by the optimal Wiener filtering based approach.

3.2.1.1. Least Squares Based Approaches

The Least Squares (LS) algorithm is a comfortable approach which does not require any prior knowledge and in literature has proven to perform reasonably for OFDM in many cases, e.g. [MS02].

Let us first consider the smoothing problem, thus estimating the channel vector $\hat{\mathbf{h}}_{i1}^{[\mathrm{P}]}$ at the respective pilot positions of a time-variant frequency-selective channel. Here, the least squares minimization problem can be formulated as

$$\underset{\hat{\mathbf{h}}_{i1}^{[\mathrm{P}]}}{\arg\min} ||\mathbf{y}_i - \mathbf{S}_1 \hat{\mathbf{h}}_{i1}^{[P]}||^2 \tag{3.9}$$

having the solution $\hat{\mathbf{h}}_{i1}^{[\mathrm{P}]} = (\mathbf{S}_1^H \mathbf{S}_1)^{-1} \mathbf{S}_1^H \mathbf{y}_i$, using e.g. [Mey00]. This means, for a single pilot resource element k, l, the channel estimate is based on the observation y_{i1kl} at this particular resource element. And, as $\mathbf{S}_1$ is diagonal, it is just a plain scalar pilot inversion $\hat{h}_{i1kl} = \frac{s_{1kl}^*}{|s_{1kl}|^2} y_{i1kl}$. Thus the weight vector $\mathbf{w}_{ijkl}$ from (3.6) consists of all zeros, except at the pilot position of interest. It is obvious that this provides no noise suppression and thus is just recommended for very high SNR operation points, not suitable for CoMP.

A simple improvement can be achieved by sliding-window filtering techniques. In order to illustrate and justify such an approach, let's first rearrange the system model for a special case

where the channel is constant over the entire pilot set $\mathcal{P}$. Now looking at such a special case the channel can simply be described as a complex scalar for the entire pilot set $\mathcal{P}$. Here, the receive vector model (3.4) for a single user $j = 1$ can be rewritten as

$$\mathbf{y}_i = h_{i1}^{[\mathrm{P}]}\mathbf{s}_1 + \mathbf{n}_i. \tag{3.10}$$

Now the least squares channel estimation problem changes to

$$\arg\min_{\hat{h}_{i1}^{[\mathrm{P}]}} ||\mathbf{y}_i - \hat{h}_{i1}^{[\mathrm{P}]}\mathbf{s}_1||^2 \tag{3.11}$$

The solution is $\hat{h}_{i1}^{[\mathrm{P}]} = \frac{\mathbf{s}_1^H}{||\mathbf{s}_1||_2^2}\mathbf{y}_i$. Thus the pseudo-inverse has transformed into averaging of the channel after compensating phases and amplitudes of the pilot symbols. This averaging provides a gain over the per-resource-element noise contributions and shows similarities to a matched filter in the sense that a linear filter $\mathbf{w} = \frac{\mathbf{s}_1^H}{||\mathbf{s}_1||_2^2}$ is "matched" to the pilot sequence by combining observation vector contributions weighted with the complex conjugate pilot symbols.

In section 2.1.3 the principles of coherence bandwidth and coherence time were discussed, refering to frequency and time intervals where the CTF is approximately constant. We can call such a time-frequency area as *coherence block*.

Now focusing on pilot structures of the LTE-A uplink of appendix A, the pilot resource elements are directly neighboring in frequency at two different distant OFDM symbols within the subframe. A self-evident approach in this case is to smooth the channel estimates in frequency direction of the respective pilot OFDM positions and perform a linear interpolation for the channel estimates at the data resource element positions. Smoothing for a frequency-selective channel, using the discussions of a constant channel (3.10), can be handled in a simple manner by a *sliding window* filter. Let's assume the channel is constant for a block of L neighboring subcarriers. For the frequency position l of interest, the least squares approach compensates the phase and amplitude variations of the received pilot symbols and averages the result over the block from subcarrier $l - \frac{L-1}{2}$ to $l + \frac{L-1}{2}$, with L being an odd number. For the smoothing operation at OFDM symbol k, with a certain window size L a cut-out $\breve{\mathbf{y}}_{ikl}$ of the observation vector $\mathbf{y}_i$ centered around subcarrier l, as well as a cut-out $\breve{\mathbf{s}}_{1l}$ of the pilot sequence vector $\mathbf{s}_1$ (of user 1) is considered as

$$\breve{\mathbf{y}}_{ikl} = \left[y_{ik(l-(L-1)/2)}, y_{ik(l-(L-1)/2)+1)}, \ldots, y_{ik(l+(L-1)/2)}\right]^T \tag{3.12}$$

$$\breve{\mathbf{s}}_{1l} = \left[s_{1k(l-(L-1)/2))}, s_{1k(l-(L-1)/2+1)}, \ldots, s_{1k(l+(L-1)/2)}\right]^T \tag{3.13}$$

Thus the smoothed channel estimate can now be obtained with least squares, assuming the channel to be constant over the window, as

$$\hat{h}_{i1kl} = \frac{1}{L}\breve{\mathbf{s}}_{1l}^H\breve{\mathbf{y}}_{ikl} \tag{3.14}$$

with unit amplitude pilot symbols.

At the frequency edges of a resource allocation block, the window lacks subcarriers for averaging, as they lie outside of the allocation block and thus no pilot symbols for the MS of interest are available. One can expect performance losses from edge effects here.

The question is how to design the window size L. The expected radio propagation environment for our system deployment scenarios of interest is providing some guidelines here. The delay spread defines the coherence bandwidth and thus the choice of L. A window size of $L = 7$, tailored to typical delay spreads, as discussed in section 2.1.3 serves as a working assumption for this section.

Note that in principle also different windows than rectangular ones can be used. In order to avoid a lot of discussion of heuristical window shapes, the optimal weights (and thus "windows") are discussed in the next section.

3.2.1.2. The Wiener Filter

While least squares is considered to be a reasonable heuristical approach in many cases, it provides no optimality in general [Kay93]. Optimality in terms of minimizing the cost function of a mean squared error (MSE) $\mathrm{E}\left\{|\hat{h}_{ijkl} - h_{ijkl}|^2\right\}$ is provided by the MMSE estimator. Given our observation pilot vector $\mathbf{y}_i$, the estimate is the conditional mean of the channel:

$$\hat{h}_{ijkl} = \mathrm{E}\left\{h_{ijkl}|\mathbf{y}_i\right\} \tag{3.15}$$

The expectation is carried out with respect to the posterior PDF, using [Kay93] chapter 14.2, as

$$p(h_{ijkl}|\mathbf{y}_i) = \frac{p(\mathbf{y}_i|h_{ijkl})p(h_{ijkl})}{\int p(\mathbf{y}_i|h_{ijkl})\mathrm{d}h_{ijkl}} \tag{3.16}$$

While the MMSE estimator uses the mean of (3.16), an alternative optimization criterion is provided by the Maximum A Posteriori (MAP) estimator. With this criterion, $\hat{h}_{ijkl}$ is chosen as the value that maximizes $p(h_{ijkl}|\mathbf{y}_i)$. The underlying cost function of this approach is sometimes referred to as "hit-or-miss" cost function [Kay93]. This means if the absolute error $|\hat{h}_{ijkl} - h_{ijkl}|$ is below a small threshold ϵ_{Err}, the cost is zero, otherwise it is 1. All errors above the threshold ϵ_{Err} are treated equal just as a "miss", while also all "hits" are treated equal.

In general, an MMSE estimator is not linear. In case of complex Gaussian distributed h_{ijkl} and $\mathbf{y}_i$, which is not generally true in the scope of this thesis, the linear MMSE estimator (LMMSE) coincides with the MMSE estimator and the MAP estimator[5] [Kay93]. The LMMSE solution is also known as the Wiener filter, named after Norbert Wiener [Wie49].

As (3.5) has shown that we can attribute the problem of CoMP channel estimation within this thesis to a general linear model, we again have the nice property that, using [Kay93] chapter 14.3, MMSE and MAP coincide into the LMMSE solution. This means, for the CoMP scenario in this thesis, we do not lose optimality by restricting the estimator to a linear one[6].

The actual weight vectors can be obtained in the following way: According to the *orthogonality principle* [Kay93],[Say08], the mean squared error (MSE) $\mathrm{E}\left\{|\hat{h}_{ijkl} - h_{ijkl}|^2\right\}$ gets minimized when the estimation error $\hat{h}_{ijkl} - h_{ijkl}$ does not have any residual correlation with the observation vector $\mathbf{y}_i$ [7]. This *statistical orthogonality* between the error and the observation can be expressed

[5] In case of a Gaussian distribution, the maximum of the pdf and the expected value coincide.

[6] This statement of optimality requires zero mean channels. The non-zero mean case is discussed in appendix B

[7] Otherwise, the MSE would not be minimal.

as

$$\mathrm{E}\left\{(\hat{h}_{ijkl} - h_{ijkl})\mathbf{y}_i^H\right\} = \vec{0}. \tag{3.17}$$

Inserting our linear approach (3.6) in (3.17) we can derive the weights as

$$\begin{aligned}
\mathrm{E}\left\{(\mathbf{w}_{ijkl}^H\mathbf{y}_i - h_{ijkl})\mathbf{y}_i^H\right\} &= \vec{\mathbf{0}} \;\Rightarrow\; \mathbf{w}_{ijkl}^H\,\mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\} - \mathrm{E}\left\{h_{ijkl}\mathbf{y}_i^H\right\} = \vec{\mathbf{0}} \\
\Rightarrow\; \mathbf{w}_{ijkl}^H &= \mathrm{E}\left\{h_{ijkl}\mathbf{y}_i^H\right\}\mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\}^{-1} \\
\Rightarrow\; \mathbf{w}_{ijkl} &= \mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\}^{-1}\mathrm{E}\left\{h_{ijkl}^*\mathbf{y}_i\right\}
\end{aligned} \tag{3.18}$$

The complex $P \times P$ auto-covariance matrix is denoted $\mathbf{\Phi}_i = \mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\}$. The complex $P \times 1$ cross-covariance vector is $\mathbf{r}_{ijkl} = \mathrm{E}\left\{h_{ijkl}^*\mathbf{y}_i\right\}$.

In [HKR97], regarding SISO OFDM, this kind of estimation for time-variant, frequency-selective channels, is called 2-D Wiener filtering, as the CTF $\mathbf{h}_{ij}^{[P]}$ for the pilot resource elements from transmit antenna j to receive antenna i depends on the dimensions time and frequency. Thus its channel estimate for a certain resource element position (k, l) depends on the two dimensions k and l.

Looking at (3.18), in order to obtain the weight vector which minimizes the MSE, the statistical parameters $\mathbf{\Phi}_i$ and $\mathbf{r}_{ijkl}$ are required. The auto-covariance $\mathbf{\Phi}_i$ of the receive signal at the pilot positions $\mathcal{P}$, the receive pilot vector auto-covariance, for one transmitting user $\mathcal{J} = \{j\}$ can be written as

$$\begin{aligned}
\mathbf{\Phi}_i &= \mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\} \\
&= \mathbf{S}_j \underbrace{\mathrm{E}\left\{\mathbf{h}_{ij}^{[P]}\mathbf{h}_{ij}^{[\mathrm{P}]\,H}\right\}}_{\mathbf{R}_{ij}^{[\mathrm{hh}]}} \mathbf{S}_j^H + \sigma_n^2\mathbf{I}.
\end{aligned} \tag{3.19}$$

The second line follows from inserting (3.4), using independence of noise vector $\mathbf{n}_i$ and channel vector $\mathbf{h}_{ij}$. With $\mathbf{R}_{ij}^{[\mathrm{hh}]} = \mathrm{E}\left\{\mathbf{h}_{ij}^{[\mathrm{P}]}\mathbf{h}_{ij}^{[\mathrm{P}]\,H}\right\}$ we denote the channel auto covariance. This channel auto-covariance for pilot symbols on the same OFDM symbols k and different subcarriers l_1 and l_2 can be computed using (2.12) with the knowledge of the power-delay profile, thus the average per-tap power $\mathrm{E}\left\{|c_{ijkm}|^2\right\}$ of the equivalent baseband channel impulse response.

Also, when treating the fading of a single subcarrier as a narrowband channel with isotropic scattering, using the Jakes model, we can compute for a fixed l the correlation of two different OFDM symbols k_1 and k_2 with the Bessel function of (2.11). The channel covariance between an OFDM resource element at (k_1, l_1) and another element at (k_2, l_2), based on (2.12) and (2.11), thus becomes

$$\begin{aligned}
R((k_1, l_1), (k_2, l_2)) &= \mathrm{E}\left\{h_{ijk_1l_1}h_{ijk_2l_2}^*\right\} \\
&= \underbrace{\left(\sum_{m=1}^{L}\mathrm{E}\left\{|c_{ijkm}|^2\right\}e^{-j2\pi(l_1-l_2)m/N_{\mathrm{SC}}}\right)}_{frequency-domain\ caused} \underbrace{J_0\left(2\pi f_d(k_2-k_1)\Delta T\right)}_{time-domain\ caused}
\end{aligned} \tag{3.20}$$

with f_d and ΔT denoting maximum Doppler shift and OFDM symbol duration respectively. An important property of (3.20) is that this covariance just depends on the *relative* differences in time $k_2 - k_1$ and frequency $l_2 - l_1$, each of them contributing a factor in (3.20).

The pilot resource element channel auto covariance matrix $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ is here written as

$$\mathbf{R}_{ij}^{[\mathrm{hh}]} = \begin{pmatrix} R((l_1,k_1),(l_1,k_1)) & \dots & R((l_1,k_1),(l_P,k_P)) \\ \vdots & \ddots & \vdots \\ R((l_P,k_P),(l_1,k_1)) & \dots & R((l_P,k_P),(l_P,k_P)) \end{pmatrix}. \tag{3.21}$$

The cross covariance vector can be computed as

$$\begin{aligned} \mathbf{r}_{ijkl} &= \mathrm{E}\left\{h_{ijkl}^{*}\mathbf{y}_i\right\} \\ &= \mathrm{E}\left\{h_{ijkl}^{*}\mathbf{S}_j\mathbf{h}_{ij}^{[\mathrm{P}]} + h_{ijkl}^{*}\mathbf{n}_i\right\} \\ &= \mathbf{S}_j\,\mathrm{E}\left\{h_{ijkl}^{*}\mathbf{h}_{ij}^{[\mathrm{P}]}\right\} \\ &= \mathbf{S}_j\left[R((l,k),(l_1,k_1)),\dots,R((l,k),(l_P,k_P))\right]^T \end{aligned} \tag{3.22}$$

by inserting (3.8) and using the fact that noise has zero mean. The computation of (3.22) can happen similarly to (3.20) with access to the channel model parameters.

This section has analytically obtained the second order statistics $\mathbf{\Phi}_i$ and $\mathbf{r}_{ijkl}$ of the receive signal by exploiting genie-aided knowledge of the model parameters $\mathrm{E}\left\{|c_{ijkm}|^2\right\}$ (associated with τ_n), f_d and σ_n^2. However, those parameters can not be easily obtained in practice. As they also require some kind of "channel knowledge" to obtain them we are facing a "chicken and egg" problem, because - simply speaking - the parameters are required to obtain channel knowledge and the channel knowledge is required to obtain the parameters. This problem is addressed in the next section 3.2.2.

Furthermore, in [HKR97], there are already different realization options for the Wiener filter discussed, which for the LTE-A uplink case turn into:

- Full 2-D Wiener filtering: 2-D refers to time- and frequency dimension
- 1-D Wiener smoothing at pilot positions in frequency direction followed by linear inter- / extrapolation in time direction
- 2x1-D Filtering: 1-D Wiener smoothing at pilot positions in frequency direction followed by 1-D Wiener inter-/extrapolation in Time Direction

The complexity of the computation of the filter weights (3.18) is dominated by the inversion of the auto-covariance matrix. With a filter based on P pilots, the complexity order of the inversion is upper bounded by $\mathcal{O}(P^3)$. In the LTE-A uplink, per subframe we have two pilot OFDM symbols. One 1-D smoothing operation in frequency domain is thus executed on $\frac{P}{2}$ pilot symbols, having approximately by factor $2^3 = 8$ less complexity than 2-D Wiener filtering.

3.2.2. Practical Parameter Estimation for the Single-User Case

We have discussed two basic families of channel estimation algorithms: Least squares based approaches and the Wiener filter, where MMSE and MAP solutions coincide. The latter requires access to the following parameters:

- Receive signal auto covariance (3.19)
- Cross covariance (3.22) between receive signal vector and channel coefficient at resource element of interest
- Noise variance, appearing in (3.19)

The following subsections deal with the question how to efficiently estimate these parameters.

3.2.2.1. Basic Auto-Covariance Estimation

A natural approach to obtain the auto-covariance is to use the available past observation vectors over N different subframes, which are indexed by κ. This circumvents the need for the unknown model parameters in (3.20).

Due to scheduling in the MAC-layer of the communication system, the MS may be allocated to different frequency sub-band positions over time. Fortunately, due to the dependency on just *relative* differences in frequency, as illustrated in (3.20), all the observations, regardless of their sub-band position, can be used to aggregate knowledge on the auto-covariance.

The maximum likelihood (ML) estimate of covariance matrices using Gaussian distributed zero-mean observation vectors is known to be

$$\hat{\mathbf{\Phi}}_i^{[SC]} = \frac{1}{N} \sum_{\kappa=1}^{N} \mathbf{y}_i(\kappa) \mathbf{y}_i^H(\kappa). \tag{3.23}$$

In the case of zero-mean observations, this ML estimate is an unbiased estimate and identical to the *sample covariance matrix* (SC).[8] In the following, we will just use the term sample covariance matrix. When there are enough observations (so that they are sufficient in a statistical sense), then it is known that the sample covariance is a good choice. However, in mobile communications, we are facing several difficulties, which question whether we are able to achieve enough observations:

- User activity, based on amount and type of data traffic, how filled data buffers are etc. From inactive users, no samples can be gained.

[8] Otherwise, for non-zero-mean channels, the unbiased sample covariance is $\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]} = \frac{1}{N-1} \sum_{\kappa=1}^{N} (\mathbf{y_i}(\kappa) - \bar{\mathbf{y}}_\mathbf{i})(\mathbf{y_i}(\kappa) - \bar{\mathbf{y}}_\mathbf{i})^H$, as the *sample* (instead of true) mean $\bar{\mathbf{y}}_\mathbf{i}$ is involved. Due to *Bessel's correction*, $N-1$ is in the denominator to compensate for the bias in the sample variance stemming from the sample mean. Then, the maximum likelihood estimate and the sample covariance estimate results coincide only for large number of samples $N \to \infty$.

- The large scale change in radio environments, when due to mobility, dominant scatters significantly change. E.g. a user moves around a buidling corner etc. This results in a change in the actual covariance matrix we want to observe.
- Handover and thus also changes in the set of cells used in the CoMP cluster. For new radio links, the observation has to begin from scratch.

Furthermore, those effects combine detrimentally. This means, a moving user with rather sporadic activity, e.g. due to web browsing, has certain periods of inactivity. When the location of the user has considerably changed, from (3.20) we see that a certain change in channel tap powers and corresponding delays $\mathrm{E}\{|c_{ijkm}|^2\}$ (associated with τ_n) directly changes the covariance. Alike, acceleration or a change in movement direction (e.g. affecting the velocity vector component radial to the BS) changes f_d. From those aspects, it is clear that we are usually far from exploiting an infinite history of past channel observations.

Intuition suggests that the relation between the size P of the observation vector and the number of available observations N will play an important role for the quality of the covariance estimation. Our required computations for the Wiener filter weights (3.18) require an inversion of the auto covariance matrix. The sample covariance (3.23) is formed by a sum of N rank-1 matrices. Due to sub-additivity of matrix rank[9], a full rank sample auto-covariance matrix is only possible if there are at least as many observations as the size of the observation vector: $N \geq P$. With less samples, $\mathbf{\Phi}_i$ will not be invertible. E.g. in LTE-A, with a typical allocation size of 5 PRBs, we have $P = 120$ pilot symbols and thus a 120×120 auto-covariance matrix $\mathbf{\Phi}_i$. In this case, at least 120 subframes are required, which takes - even if the user is continuously transmitting in every subframe - a time of 120 ms as "warm-up" in order to be able to execute the Wiener filter (3.18). We will see in the following sections that with this minimum amount of time for collection of statistics, there is still a lot of room for improvement in terms of the accuracy of the covariance estimate and the resulting quality of the channel estimation.

3.2.2.2. Improved Auto-Covariance Estimation

Any alternative covariance estimation to the sample covariance method has to fulfill several criteria for our purposes:

- Invertibility of the matrix has to be ensured.
- The focus lies on operation with a comparatively small amount of samples (past subframes with observation vectors)

With fewer observations than number of pilots considered in the auto-covariance matrix, $N < P$, $\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]}$ from (3.23) will not be invertible. A typical heuristical approach here would be *diagonal loading*, thus adding a real-valued constant term onto the main diagonal of $\hat{\mathbf{\Phi}}_i^{[SC]}$, similar in effect as some virtual additional noise. This can restore invertibility. But it leaves open questions: What would be the best choice of the size of this additional term?

In financial mathematics, similar issues of covariance estimation occur as in mobile communications. The observation vectors $\mathbf{y_i}(\kappa)$ are often larger than the amount of available samples in

[9] $\mathrm{rank}(A + B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$

a time interval with approximately constant "large scale environment". So it is no surprise that in this field, a solution arose, called *shrinkage* [LW04b][LW04a].[10] A good overview on this technique from a signal processing perspective is given in [CWEH10].

This approach is now applied to the CoMP auto covariance estimation problem here in this thesis, extending the state-of-the-art in communications. Shrinkage applied to CoMP provides optimal answers to the diagonal loading problem. Assume, we know that all the received observation vectors are just uncorrelated noise vectors. Then our estimate of the auto-covariance would simply be

$$\hat{\mathbf{D}}_i = \frac{\mathrm{Tr}(\hat{\mathbf{\Phi}}_i^{[SC]})}{P}\mathbf{I}, \tag{3.24}$$

as uncorrelated noise has a diagonal auto-covariance matrix and it is scaled here with the power contained in the observed sample auto-covariance, captured by the trace-operator. $\hat{\mathbf{D}}$ is called the *shrinkage target* in shrinkage terminology. Clearly it is most well conditioned.

To form a trade-off between the sample auto-covariance $\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]}$, optimal for large number of samples, and the structured regularizing $\hat{\mathbf{D}}$, which can be interpreted as diagonal loading term, shrinkage considers the following types of estimators:

$$\hat{\mathbf{\Phi}}_i^{[\mathrm{shr}]} = (1-\eta)\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]} + \eta\hat{\mathbf{D}}_i. \tag{3.25}$$

The so-called shrinkage coefficient η is the tunable parameter for finding the best combination. Typically, it aims at minimizing the MSE of the covariance estimate

$$\mathrm{E}\left\{||\hat{\mathbf{\Phi}}_i^{[\mathrm{shr}]} - \mathbf{\Phi}_i||_F^2\right\}. \tag{3.26}$$

When η is chosen between 0 and 1, (3.25) is a convex combination of the two components.

In [CWEH10], several algorithms are discussed in order to pick the right shrinkage parameter η in an MMSE sense. The "Oracle estimator" requires Genie-aided knowledge, thus cannot be implemented in practice. The Ledoit-Wolf estimator [LW04b] has the drawback that it requires to store *all* past observation vectors, which becomes clumsy for our particular problem, as we would need to store e.g. hundreds of LTE-A subframes. When the observation vectors $\mathbf{y_i}(\kappa)$ are approximately Gaussian distributed, the Ledoit-Wolf Rao-Blackwell (LWRB) estimator, derived in [CWEH10], can be used - which is done in the further studies throughout this thesis.[11] The shrinkage coefficient, aiming at minimizing the squared error (3.26) according to LWRB, can be computed as

$$\eta_{RBLW} = \frac{\frac{(N-2)}{N}\,\mathrm{Tr}(\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]}\hat{\mathbf{\Phi}}_i^{[SC]\,H}) + \mathrm{Tr}^2(\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]})}{(N+2)\left[\mathrm{Tr}(\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]}\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]\,H}) - \frac{1}{P}\,\mathrm{Tr}^2(\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]})\right]}. \tag{3.27}$$

This approach has the advantage that it is purely based on the current sample covariance matrix.

[10] "Honey, I shrunk the sample covariance matrix" is the astonishing title of one of the publications.

[11] In [CWEH10] there is also a proposed iterative technique: Oracle approximation shrinkage (OAS), which performs best but is rather complex. It was not considered in this thesis as the example results of Fig. 8 in [CWEH10] are similar to LWRB estimator.

Now the application of shrinkage onto the CoMP auto-covariance estimation is assessed based on simulations using the single user signal model (3.8) and picking out a particular link between MS and one BS. Fig. 3.2 and Fig. 3.3 show the performance gains obtained from shrinkage. We consider a 1-D Wiener filter example case for LTE-A with 2 PRBs, thus having $P = 24$ pilots in frequency domain. An ITU Vehicular A channel with block fading is used. Block fading can be considered as a best case situation for the basic sample covariance estimation, because this leads to completely independent samples between different blocks. 9600 overall samples are considered.

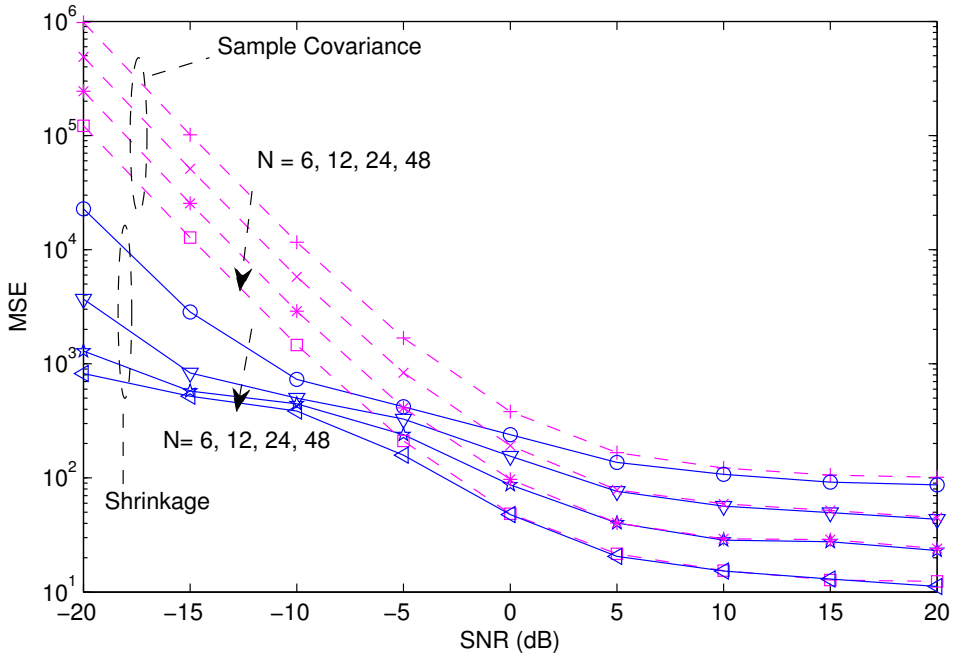

Figure 3.2.: Covariance estimation accuracy: Comparison of sample covariance estimate (dashed lines) vs shrinkage (solid lines): MSE $\mathrm{E}\left\{||\hat{\mathbf{\Phi}}^{[\mathrm{shr}]} - \mathbf{\Phi}||_F^2\right\}$ over SNR for $N = 6$ to $N = 48$ samples and $P = 24$ pilots.

Clearly, strong gains of shrinkage (in terms of MSE compared to the classical sample covariance estimation) can be observed, especially in lower SNR and with smaller number of samples, but also for $N > P$ gains are visible. Note that the negative SNR range is not uncommon[12] in the CoMP setting for links between MS and supporting BS (thus links which cannot be exploited in a non-CoMP system, which only relies on received signals at the serving cell).

In Fig. 3.3 the SNR is set to -3dB, corresponding to a cell edge case, which is an intended operation point for CoMP. Again, some gains can be observed even when $N > P$. When $N > 5P$ the gains are approaching 0, as in this case the shrinkage coefficient in (3.25) has converged to 1, meaning that the best choice for estimating the covariance is to purely rely on the sample covariance (as it has become reliable due to enough collected samples) and setting

[12] At least when modeling interference as AWGN.

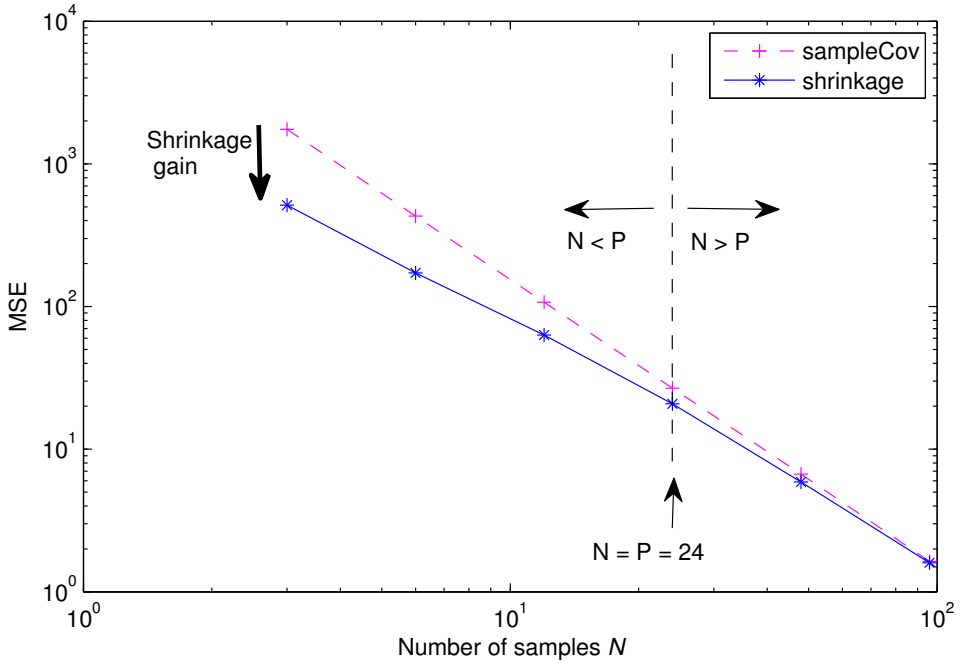

Figure 3.3.: Covariance estimation accuracy: Comparison of sample covariance estimate (dashed line) vs shrinkage (solid line): MSE over number of samples for an SNR of -3dB.

the diagonal loading term to zero. Formally speaking: In (3.27), with $N >> P$, it results $\eta \to 1$, hence in (3.25) with $\eta \to 1$, it results $\hat{\mathbf{\Phi}}_i^{[\mathrm{shr}]} \to \hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]}$.

Conversely, as long as we do not have the case that the amount of collected samples N becomes very large against the dimension P of the covariance matrix, shrinkage can make a difference in estimation accuracy. In LTE-A with allocation sizes of 5 PRBs, we have $P = 60$ for 1-D Wiener filtering and $P = 120$ for 2-D Wiener filtering. This means that shrinkage is still attractive when we have collected pilot sequence observation data of even hundreds of subframes, thus even for fractions of seconds of "warm-up time" with continuous data transmission.

3.2.2.3. Cross-Covariance Estimation

The second componenent required to compute the Wiener filter weights in (3.18) is the cross-covariance vector $\mathbf{r}_{ijkl}$ which is described in (3.22).

In case we are facing a smoothing task, thus estimating the channel at a pilot position, hence $(k, l) \in \mathcal{P}$, the knowledge of the pilot sequence and access to the channel covariance $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ directly solves (3.22). In case of inter- or extrapolation, we need to extract additional model parameters to obtain $\mathbf{r}_{ijkl}$ via (3.20) or use just linear interpolation.

In both ways it is required to get access to the channel covariance $\mathbf{R}_{ij}^{[\mathrm{hh}]}$. In the previous sections we have discussed that the auto-covariance can be directly learned from observing a certain

number of samples of the pilot receive vector. From this auto-covariance, we can compute the channel covariance by rearranging (3.19) into

$$\hat{\mathbf{R}}_{ij}^{[\mathrm{hh}]} = \mathbf{S}_j^H \left(\hat{\mathbf{\Phi}}_i - \hat{\sigma}_n^2 \mathbf{I} \right) \mathbf{S}_j. \tag{3.28}$$

This step is achieved by $\mathbf{S}_j \mathbf{S}_j^H = \mathbf{S}_j^H \mathbf{S}_j = \mathbf{I}$, exploiting the constant-amplitude property of LTE-A pilot sequences and treating the estimates as approximation of the true variables. Here it is recommended that $\hat{\mathbf{\Phi}}_i$ is obtained by shrinkage, thus $\hat{\mathbf{\Phi}}_i = \hat{\mathbf{\Phi}}_i^{[\mathrm{shr}]}$. Note, that (3.28) additionally needs an estimate of the noise variance σ_n^2, discussed in the subsequent section.

For inter- or extrapolation in an LTE-A uplink, the task is simplified to one dimension as the pilot symbols cover the entire frequency range of a user allocation. This means that inter- / extrapolation extends only in the time direction. 1-D temporal Wiener interpolation / -prediction can be carried out here in the following way: As the MS is usually surrounded by a rich scattering environment, the assumption of a Jakes Doppler spectrum, discussed in section 2.1.3, holds. In this case the channel auto-correlation between different OFDM symbols follows the Bessel function depicted in Fig. 2.4. Hence we compute an empirical correlation between different pilot OFDM symbols (3.29), which allows to estimate the maximum Doppler shift f_d[13]. Plugging f_d into (2.11) provides the overall time-correlation behaviour of the channel between different pilot symbols and thus also allows the temporal Wiener inter-/extrapolation task.

Using simply the empirical correlation of N_s time-neighboring pilot symbols, being 7 OFDM symbols apart from each other, as done in [NRWtB11], leads to

$$\hat{R}\left(7\Delta T\right) = \frac{\frac{1}{P(N_s-1)} \sum_{i=1}^{N_s-1} \tilde{\mathbf{y}}_i^H \tilde{\mathbf{y}}_{i+1}}{\left(\frac{1}{2P(N_s-1)} \sum_{i=1}^{N_s-1} \left(||\tilde{\mathbf{y}}_i||^2 + ||\tilde{\mathbf{y}}_{i+1}||^2 \right) \right) - \hat{\sigma}_n^2}. \tag{3.29}$$

3.2.2.4. Noise Estimation

In (3.28) we have seen that the noise variance σ_n^2 is an auxillary parameter, required to obtain the cross-covariance vector $\mathbf{r}_{ijkl}$. A directly related problem is to determine the signal-to-noise ratio (SNR) as it is the ratio of total average receive power minus noise variance over noise variance:

$$\gamma = \frac{\frac{1}{P} \mathrm{E}\left\{ \mathbf{y}_i^H \mathbf{y}_i \right\} - \hat{\sigma}_n^2}{\hat{\sigma}_n^2} \tag{3.30}$$

In this section, several algorithms for noise variance / SNR estimation are compared against each other and against the theoretical bound and their suitability for single user channel estimation in CoMP is discussed.

[13] This assumes that the pilot spacing properly samples the radio channel, which is true for LTE up to speeds of around 250 km/h in the 2 GHz carrier frequency range. The inverse of the Bessel function is only carried out for its beginning nonambiguous part, see Fig. 2.4.

The Cramér-Rao lower bound (CRLB) [Kay93] is the theoretical limit of the mean squared error of an unbiased estimator. For estimating the SNR in an AWGN channel, in the literature the normalized CRLB is known [Gap08, Ala01] to be

$$NCRLB(\gamma) := \frac{CRLB(\gamma)}{\gamma^2} = \frac{1}{\bar{P}}(1 + \frac{2}{\gamma}) \tag{3.31}$$

in case it is "transmit data-aided", thus using in total $\bar{P}$ pilot symbols with an SNR of γ on the observation vector. In case of using N subframes with P pilot symbols each, we have a total of $\bar{P} = NP$ symbols for the noise estimation.

SNV Algorithm

When it comes to actual algorithms for SNR estimation, it is claimed in [PB00] that the signal-to-noise variance (SNV) algorithm comes close to this bound. In the terminology of [PB00], the noise estimation problem of this thesis is "transmit-data-aided", as we rely on pilot symbols. Of course, complex valued input is used. In [PB00] (32) and (34) the algorithm is given for real valued transmit-data-aided and complex valued receive-data-aided input. From those two equations it is straightforward to deduce the solution for complex valued transmit-data-aided input:

$$\hat{\gamma}_{SNV} = \frac{\left[\frac{1}{\bar{P}} \sum_{\kappa=1}^{N} \sum_{p=1}^{P} \mathrm{Re}\left\{y_p^*(\kappa) s_p(\kappa)\right\}\right]^2}{\frac{1}{\bar{P}-3/2} \sum_{\kappa=1}^{N} \sum_{p=1}^{P} |y_p(\kappa)|^2 - \frac{1}{\bar{P}(\bar{P}-3/2)} \left[\frac{1}{\bar{P}(\bar{P}-3/2)} \sum_{\kappa=1}^{N} \sum_{p=1}^{P} \mathrm{Re}\left\{y_p^*(\kappa) s_p(\kappa)\right\}\right]^2} \tag{3.32}$$

For ease of notation, the antenna index i has been dropped and the indices k, l have been replaced by the pilot index p which represents the appropriate resource element (k, l). The enumerator represents the coherent average over the pilot signal contributions[14]. The denominator is the total receive signal power minus the estimated useful signal contribution and thus the noise.

Subspace Noise Estimation

In case of OFDM signal formats and multi-path propagation, [XJY05] has proposed a subspace noise estimation method. It makes use of the specific OFDM signal properties indicated in (2.8) and even works blindly. The general idea is to decompose the total receive signal covariance matrix[15] in a useful signal subspace and a noise subspace by means of eigenvalue decomposition and information-theorectic model order selection. Then, the contained power in each subspace is measured.

Let us have a look at the particular structure of the pilot auto-covariance matrix (3.19) in frequency direction for one OFDM symbol k for one user $\mathcal{J} = \{j\}$. With a discrete multi-path channel (2.7), when all considered pilot symbols lie in the same OFDM symbol k, we obtain

$$\begin{aligned} \mathbf{\Phi}_i^{Freq} &= \mathbf{S}_j \, \mathrm{E}\left\{\mathbf{h}_{ijk} \mathbf{h}_{ijk}^H\right\} \mathbf{S}_j^H + \sigma_n^2 \mathbf{I} \\ &= \mathbf{S}_j \bar{\mathbf{F}} \, \mathrm{E}\left\{\mathbf{c}_{ijk} \mathbf{c}_{ijk}^H\right\} \bar{\mathbf{F}}^H \mathbf{S}_j^H + \sigma_n^2 \mathbf{I}. \end{aligned} \tag{3.33}$$

[14] The notation with the real part operator already indicates that this is only useful in phase-compensated flat fading or AWGN channels.

[15] This does not require estimating new helper parameters, as we have dealt with those already in sections 3.2.2.1 and 3.2.2.2 which we already require for the Wiener filter.

The second line uses the Fourier transform relation (2.10) between CIR and CTF, where $\bar{\mathbf{F}}$ represents the $P \times N_{\mathrm{SC}}$ cut-out matrix[16] of $\mathbf{F}$ (2.2) for the respective subcarrier indices of the P pilots. With $\mathbf{A} = \mathbf{S}_j\bar{\mathbf{F}}$ and $\mathbf{B} = \mathrm{E}\left\{\mathbf{c}_{ijk}\mathbf{c}_{ijk}^H\right\}$ we have the covariance structure

$$\mathbf{\Phi}_i^{Freq} = \underbrace{\mathbf{ABA}^H}_{\mathbf{\Psi}} + \sigma_n^2\mathbf{I}. \tag{3.34}$$

This structure of the input covariance matrix is assumed in [WK85] and exploited to determine the most likely "model order".

Note that $\mathbf{\Phi}_i^{\mathrm{Freq}}$ in (3.34) has always full rank P due to the noise term $\sigma_n^2\mathbf{I}$. On the other hand $\mathbf{B}$, being responsible for the subspace of the signal, is only of rank r (at maximum L, the number of multi-path components of the CIR). Thus $\mathbf{\Psi}$ has rank r as well. r can be interpreted as the rank of the *signal subspace*. Spectral[17] decomposition in linear algebra [Mey00] allows to rewrite (3.34), following [WK85], as

$$\mathbf{\Phi}_i^{\mathrm{Freq}} = \mathbf{\Psi} + \sigma_n^2\mathbf{I} = \underbrace{\sum_{\mu=1}^{r}(\lambda_\mu - \sigma_n^2)\mathbf{v}_\mu\mathbf{v}_\mu^H}_{\text{rank } r} + \sigma_n^2\mathbf{I}. \tag{3.35}$$

The eigenvalues of $\mathbf{\Phi}_i^{\mathrm{Freq}}$are $\lambda_1, \lambda_2, \ldots, \lambda_P$, the eigenvectors are $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_P$. The sum power contained in the covariance matrix $\mathrm{tr}(\mathbf{\Phi}_i^{\mathrm{Freq}}) = \sum_{\mu=1}^{P}\lambda_\mu$ is preserved in its eigenvalues. The specific structure of (3.35) causes that the smallest $P - r$ eigenvalues are all identical and equal to the noise variance σ_n^2 - they form the *noise subspace*. The strongest r eigenvalues are different in power, depending on the multi-path channel characteristics. Thus the strategy of a subspace noise estimation algorithm is to extract those $P - r$ identical noise-space eigenvalues, which provide the noise variance and thus the SNR.

The obstacle here is that we do not have access to the true covariance, but only access to an empirical covariance, which does not have the property of having a set of fully identical eigenvalues. The second problem is that the actual rank r of the useful signal subspace is unknown. In [WK85], those issues are addressed by information-theoretic criteria, AIC (Akaike information criterion) and MDL (minimum descriptive length). As the latter one in the literature is considered as the better approach[18], we focus on MDL.

Let us view this from a hypothesis testing perspective: *Occam's razor* is known in science as a principle of parsimony in selecting hypothesises. The most likely hypothesis is the one with the smallest number of assumptions. MDL can be seen as the formal application of this principle on the compression of observed data. Any structure or regularity of the data, as expressed in (3.35), can be used to compress these data and thus reduce the number of bits required to describe the data. Among different models, the model should be selected which requires the smallest number of bits to describe it (in an ideal virtual message). The amount of bits here is coming from two different parts: (1) Indicating which model is used and (2) describing the data within the model. For the total message length, this means a trade-off between a parameter message part and a data

[16] $\bar{\mathbf{F}}$ for a user allocated to subcarrier indices $\{\bar{l}, \bar{l}+1, \ldots, \bar{l}+P-1\}$ is defined as the rows $\bar{l}$ to $(\bar{l}+P-1)$ of $\mathbf{F}$.

[17] This refers to spectrum as a set of eigenvalues of a matrix in contrast to the spectrum in the sense of frequency domain representation of a signal.

[18] AIC tends to overestimate the dimension of signal space $\hat{r}$, according to [WK85].

message part. The longer the parameter message part, which provides the "compression" model, the shorter is the required data message part which handles the residuals. MDL is handling this trade-off by searching for the minimum of the total description length.

In case of our model selection problem of noise estimation using the covariance structure (3.35), the different hypothesises of Occam's razor are the different possible ranks r of $\mathbf{\Psi}$. Once we have the most probable $\hat{r}$, we can directly estimate the noise variance from the $P - \hat{r}$ remaining Eigenvalues by simple averaging. Indicating the model and providing its parameters is in [WK85] handled by counting the degrees of freedom of the eigenvalues and eigenvectors of a particular $\hat{r}$. In [WK85] the log-likelihood of the respective model order is computed, based on the assumption of i.i.d. complex Gaussian samples:

$$L(r) = \log \left(\frac{\prod_{\mu=r+1}^{P} \hat{\lambda}_{\mu}^{1/(P-r)}}{\frac{1}{P-r} \sum_{\mu=r+1}^{P} \hat{\lambda}_{\mu}} \right)^{(P-r)N} \tag{3.36}$$

It is interesting to note that the fraction in brackets represents the ratio of geometric mean of noise space eigenvalues over arithmetic mean. The MDL criterion defines the total descriptive length as the negative $L(r)$ plus a correcting term for the degrees of freedom, which is [WK85]

$$\mathrm{MDL}(r) = -L(r) + \frac{1}{2} r(2P - r) \log N. \tag{3.37}$$

This gives us the following steps for estimating the SNR:

- Obtain the sample auto-covariance $\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]}$ according to section 3.2.2.1, (3.23).
- Perform an eigenvalue decomposition of $\hat{\mathbf{\Phi}}_i^{[\mathrm{SC}]}$, obtaining the eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_P$
- Compute $\mathrm{MDL}(r)$ (3.37) for $r = 0, 1, \ldots, P$.
- $\hat{r}$ is the r which minimizes the MDL and thus also the estimated dimension of our signal space.
- The noise variance hence consists of all the power contained in the noise space $\frac{1}{P-\hat{r}} \sum_{\mu=\hat{r}+1}^{P} \hat{\lambda}_{\mu}$.
- The remaining power is attributed to the useful signal and the SNR follows (3.30).

An additional extension of this scheme to make it more robust is investigated in the subsequent noise estimation simulations, denoted as "subspace- based, setting $\hat{r} \geq 1$". This ensures that at least one Eigenvalue is attributed to the useful signal space, thus $\hat{r} \geq 1$. This is a safety mechanism in case the MDL fails to pick the right model order, which can happen in extremely low SNR and/or with few number of samples.

Filtering-Based SNR Estimation

The basic idea of this simple and easy to implement approach is to take the receive signal vector $\mathbf{y}_i$, compensate for pilot rotations by a multiplication with $\mathbf{S}_j^H$ and filter this "pilot-compensated" receive signal vector in the same way as a basic channel estimator does, e.g. from section 3.2.1.1 according to (3.14). This provides us a noise-reduced channel estimate. When comparing this result with a noisy unfiltered channel estimate, the difference approximately contains the noise. The result should be averaged over the N available subframes of interest. With a length l filter,

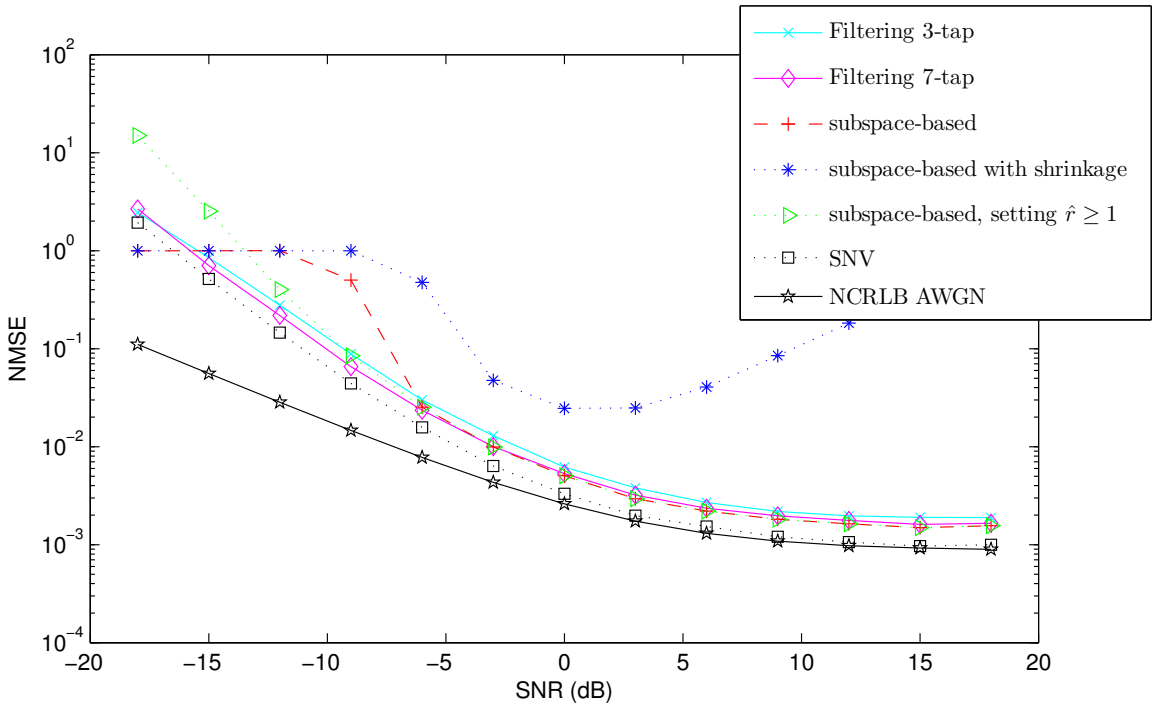

Figure 3.4.: Comparison of different SNR estimation techniques for AWGN channel: NMSE (3.39) over SNR γ. Filtering-based techniques with 3 and 7 taps (3.38) are compared with subspace-based techniques, the SNV algorithm (3.32) and the normalized Cramér-Rao bound (3.31).

represented by the Toeplitz matrix $\mathbf{T}_l$, implementing the sliding window in (3.14), the noise variance estimate becomes [NRWtB11]

$$\hat{\sigma}_n^2 = \sum_{\kappa=1}^{N} \frac{l}{l-1} ||\mathbf{T}_l \mathbf{S}_j^H \mathbf{y}_i - \mathbf{S}_j^H \mathbf{y}_i||_F^2 \tag{3.38}$$

Residual errors of this scheme occur in case of high delay spreads with a large channel variation over frequency. Additionally, the filtered output still has residual noise on top of it.

Numerical Results

For this thesis, in order to pick the appropriate SNR estimation technique for CoMP, simulations were performed in order to compare the techniques against each other and against the CRLB. The normalized mean squared error (NMSE) of the estimated SNR $\hat{\gamma}$ serves as performance metric:

$$\text{NMSE} = \frac{\text{E}\left\{(\hat{\gamma} - \gamma)^2\right\}}{\gamma^2} \tag{3.39}$$

The following parameters have been used: $P = 24$ pilots, $N_{sam} = 48$ samples, 2000 realizations, thus 9600 overall samples.

The results are shown in Fig. 3.4 for the AWGN channel. For positive SNR, the SNV algorithm indeed comes close to the theoretical bound. But, in the specific distributed antenna scope of

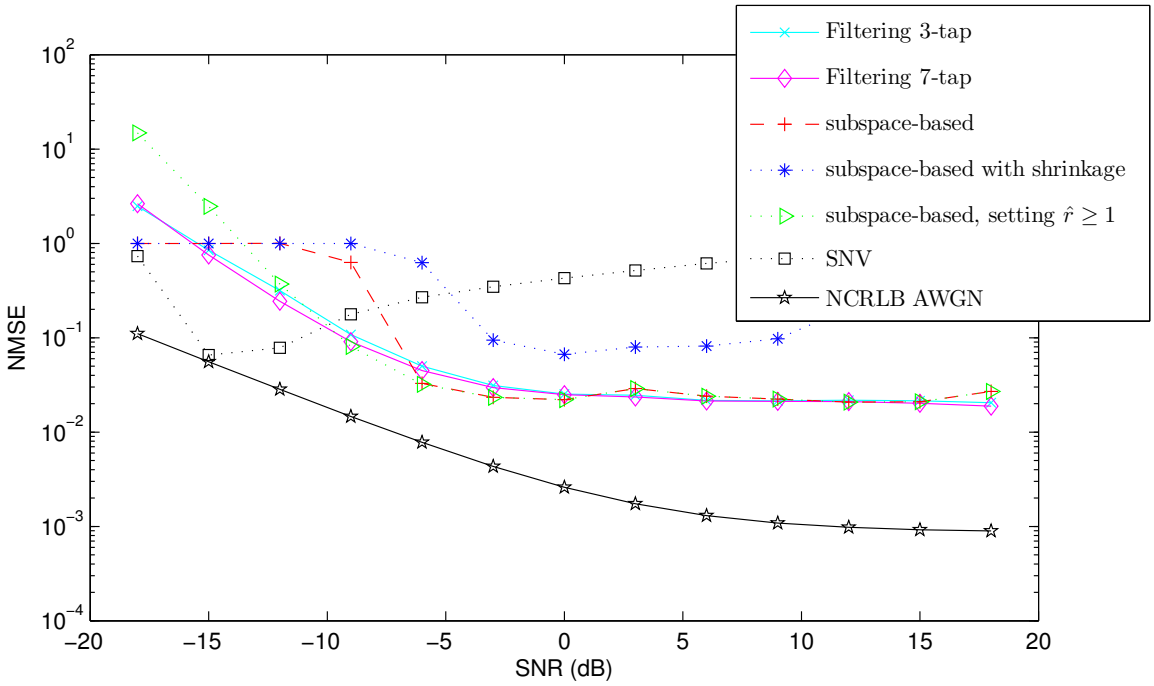

Figure 3.5.: Comparison of different SNR estimation techniques for Vehicular A channel: NMSE (3.39) over SNR γ. Filtering-based techniques with 3 and 7 taps (3.38) are compared with subspace-based techniques, the SNV algorithm (3.32) and the normalized Cramér-Rao bound (3.31).

this thesis (CoMP), we have to consider also negative SNRs for the supporting cell links. Here the gap to the bound gets larger for SNV and becomes comparable to other algorithms.

Two filtering-based approaches are considered: “Filtering 3-tap” and “Filtering 7-tap”, denoting the respective sliding window lengths. For low SNR they perform similar as SNV, for high SNR they show a larger gap to the bound as SNV. The 7-tap filter performs better due to better averaging and, as AWGN has no delay (and no Doppler) spread and thus no channel variation in frequency (and time) direction, which could introduce additional approximation errors for longer filter lengths.

Subspace noise estimation, for an SNR of -6dB and above performs very similar to the filtering-based approaches. For the extremly low SNR cases, problems of the MDL algorithm are occuring: The straight line outcome of a normalized MSE equal to one is remarkable and occurs at SNR of -12 dB and below. Here MDL fails in properly discriminating signal and noise subspace and falsely attributes the entire receive signal as noise: The subspace estimator (incorrectly) outputs $\hat{r} = 0$ in most cases. Knowing from the scheduler that there is a user present, we can set a minimum of $\hat{r} = 1$. The optional improvement “subspace- based, setting $\hat{r} \geq 1$” does this, but still provides inaccurate SNR estimates at very low SNR, not being as good as the simple filtering based approaches. In section 3.2.2.2 shrinkage was discussed as an alternative to improve the covariance estimation accuracy. Results of subspace noise estimation in conjunction with shrinkage are shown in the curve “subspace-based with shrinkage”. The motivation of in-

vestigating this combination was to obtain a more robust covariance estimate as a better basis for subspace noise estimation. Unfortunately, here shrinkage completely worsens the outcome, making it useless. This can be explained by the fact that the diagonal loading in (3.25) artificially changes the eigenspace structure of the covariance, which causes MDL to fail in discriminating noise and signal subspace. Thus, in conjunction with MDL, only the sample covariance matrix should be used as input.

Fig. 3.5 gives the results for an ITU Vehicular A channel. Here the picture changes. The SNV algorithm fails in coping with the frequency variations of the channel, because it is not designed for it, and thus should not be used. For SNRs of -6dB and above, the filtering based methods again perform very similar to the more complex subspace based methods. Below -6dB, subspace-based SNR estimation has the same problems as in the AWGN channel and performs poor. Again, shrinkage completely causes subspace methods to fail.

In summary, the most recommendable approaches for this single user setting are - suprisingly - the simple filtering based schemes. SNV is just a special case solution for AWGN, which is not something we can rely on in real world propagation channels. The more complex subspace noise estimation performs similar to the filtering based schemes, but cannot cope with SNRs below -6dB. As lookahead: In the multi-user parameter estimation section 3.4 we will see that there, subspace noise estimation performs well and the filtering approaches perform poor.

3.2.3. Single-User Simulation Results

Building upon the achieved parameter estimation work of this thesis, this section 3.2.3 summarizes simulation results, published in [NRWtB11]. It is centered around the challenge of collecting small "power crumbs".

Channel Estimation Simulation Scenario
In order to stress CoMP-specific properties in single user $\mathcal{J} = \{j\}$ noise-limited channel estimation the following scenario has been chosen in [NRWtB11]: Half the useful signal power is collected by the serving cell base station. The other half is equally split onto a set of $N_{\text{supp}} = I-1$ supporting base stations. So this chosen scenario stresses the particular CoMP setting with unbalanced link attenuations. When increasing this number N_{supp}, one can see the point up to where CoMP in conjunction with particular channel estimation algorithms still provides benefits. With a large number of N_{supp} and in conjunction with imperfect channel estimation, CoMP does not provide benefits over single cell reception any more.

It is well known that the optimal receive antenna combining algorithm in the presence of white noise is maximum ratio combining (MRC) [Bre03]. In MRC, the combining weights, illustrated in Fig. 3.1, are proportional to the conjugate channel coefficient in 3.3. Thus MRC emphasizes the output of antennas with stronger propagation channels and de-rotates the per-antenna phase shifts for coherent combining:

$$\mathbf{g}_{jkl}^{H} \quad \propto \quad \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \tag{3.40}$$

$$\hat{\mathbf{g}}_{jkl}^{H} \quad \propto \quad \hat{\mathbf{h}}_{jkl}^{[\mathrm{B}]\,H} \tag{3.41}$$

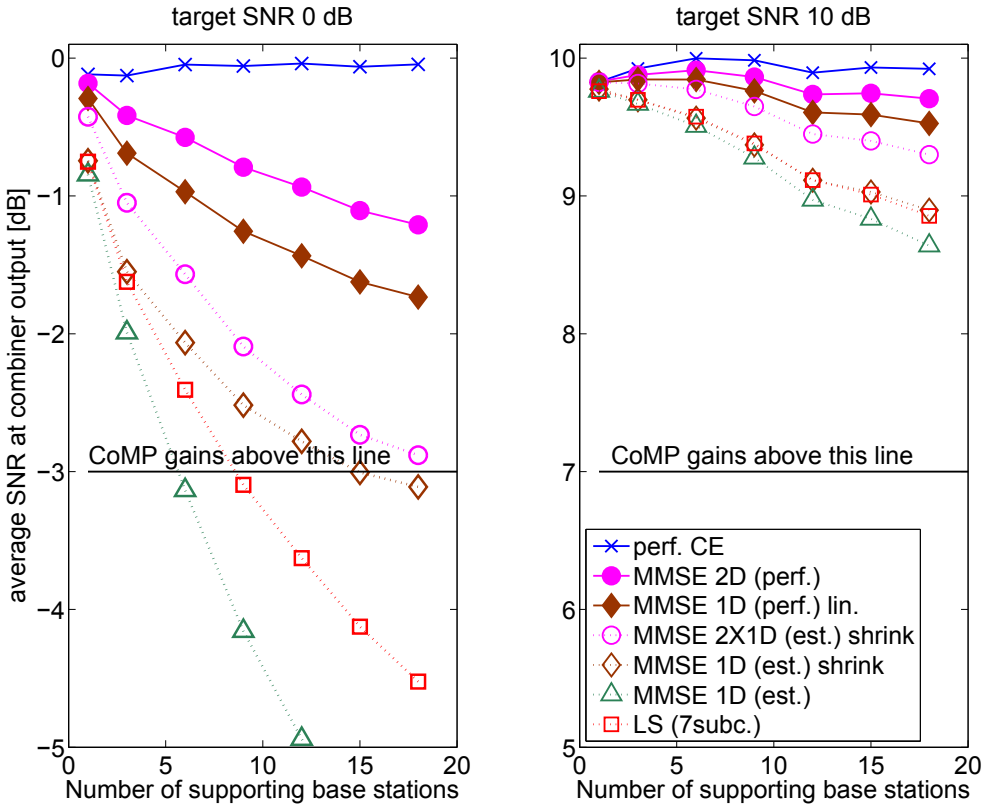

Figure 3.6.: SNR at receive combiner output over number of supporting cells for single-user multi-cell channel estimation. Source: [NRWtB11], Fig. 4.

The estimated optimal combiner weights $\hat{\mathbf{g}}_{jkl}^H$ are based on the outcome $\hat{\mathbf{h}}_{jkl}^{[\mathrm{B}]}$ of the channel estimation.

The ratio of combined useful signal power over combined noise power provides the SNR at the output of the receive combining over the entire distributed antenna system:

$$\gamma = \frac{\mathrm{E}\left\{|\mathbf{g}_{jkl}^H \mathbf{h}_{jkl}^{[\mathrm{B}]} s_{jkl}|^2\right\}}{\mathrm{E}\left\{|\mathbf{g}_{jkl}^H \mathbf{n}_{kl}^{[\mathrm{B}]}|^2\right\}} \tag{3.42}$$

The noise vector $\mathbf{n}_{kl}^{[\mathrm{B}]} = [n_{1kl}, n_{2kl}, \ldots, n_{Ikl}]^T$ is formed along the dimension of the I base station receive antennas. Replacing the ideal antenna weights $\mathbf{g}_{jkl}^H$ by the estimated ones $\hat{\mathbf{g}}_{jkl}^H$ gives an upper bound for the combiner performance in the presence of channel estimation error. This output SNR (3.42) will be the key performance indicator for the investigations in this subsection.

Results

In [NRWtB11] multi-cell simulations have been carried out. The WINNER II channel model [WIM06], as time-variant and frequency-selective spatial channel model, has been used with a user velocity of 50 km/h. Two system operation points have been chosen: A total target output

SNR of 0dB (corresponding to -3 dB input SNR at the serving cell) represents the noise limited case. A target SNR of 10 dB (thus 7 dB input SNR at the serving cell) represents an interference limited system[19]. As we have not included interference throughout the entire section 3.2, this provides some insights of the situation in conjunction with a perfect multiuser channel estimator and combiner. The user allocation size is 2 physical resource blocks, which consists of 48 pilot symbols per subframe in total.

Fig. 3.6 shows the performance when the channel estimator has already collected statistical knowledge for 200 subframes, thus a "warm-up time" of 200 ms. Perfect channel knowledge, thus $\hat{h}_{ijkl} = h_{ijkl}$, provides a constant performance of 0 dB output SNR, independent of the number of supporting cells N_{supp}. "MMSE 2-D (perf.)" means that the channel estimator has perfect statistical knowledge of the channel, thus $\hat{\mathbf{\Phi}}_i = \mathbf{\Phi}_i$ and $\hat{\mathbf{r}}_{ijkl} = \mathbf{r}_{ijkl}$, providing the perfect estimator weights, according to (3.18). As a growing number of supporting links N_{supp} decreases the SNR per link in the normalized scenario, the output SNR performance decays, because the receive combining is increasingly mismatched. "MMSE 1-D (perf.) lin." uses the same knowledge; the difference is that a 1-D MMSE smoothing is carried out in frequency domain at the pilot positions and then the result is linear inter- and extrapolated in order to obtain the channel estimate at the data resource element position. There is some loss compared to 2-D MMSE, as with the velocity of 50 km/h, which can be seen as rather high for a CoMP scenario, the linear interpolation is not as good as a fully fledged 2-D (time+frequency based) estimation. For 9 supporting cells and 0 dB target SNR, a perfect MMSE estimator loses about 0.8 dB against perfect CSI; with 1-D instead of 2-D estimation there are additional 0.4 dB of losses. Least squares based channel estimation in the 0 dB target SNR scenario performs very poor; for a scenario with 9 and more (weak) supporting links, the CoMP joint reception is inferior to single cell reception. The warm-up time of 200 subframes (see also Fig. 3.7) is still not sufficient that MMSE with conventionally estimated auto covariance, according to section 3.2.2.1, can achieve at least the LS channel estimator performance. With auto covariance estimation based on shrinkage, according to section 3.2.2.2 this changes, and MMSE clearly becomes superior to LS. The two cascaded 1-D Wiener filters ("MMSE 2x1D") add an additional small performance gain over frequency direction 1-D Wiener smoothing ("MMSE 1D"), followed by linear interpolation

In the 10 dB target SNR scenario, the losses from channel estimation are much smaller. Here, the choice of the algorithm is also not that critical and also least squares could be used as channel estimator.

Fig. 3.7 shows to impact of the number of available past samples (here: subframes) for estimating the auto- and cross covariance. This is called the "warm-up time". For the low target SNR operation point of 0 dB, without shrinkage, the losses of short warm-up times to maximum performance are even larger than 2 dB in terms of post-combining SNR. In order to reach its full performance potential, clearly more than 1000 subframes have to be collected for auto- and cross-covariance estimation for 1-D MMSE. Even with full buffer traffic, this means more than a second of contiguous transmission. When the mobile station has non-full-buffer traffic and a certain mobility (with changing scattering environment and thus channel statistics) this becomes a problem. The parameters $\hat{\mathbf{\Phi}}_i$ and $\hat{\mathbf{r}}_{ijkl}$ might already have changed significantly in the new

[19] While interference is not modeled throughout this section, still insights can be gained at this high SNR operation point. Namely, how the performance looks like with channel estimators which are not impacted by interference at all, thus representing a "perfect multiuser receiver".

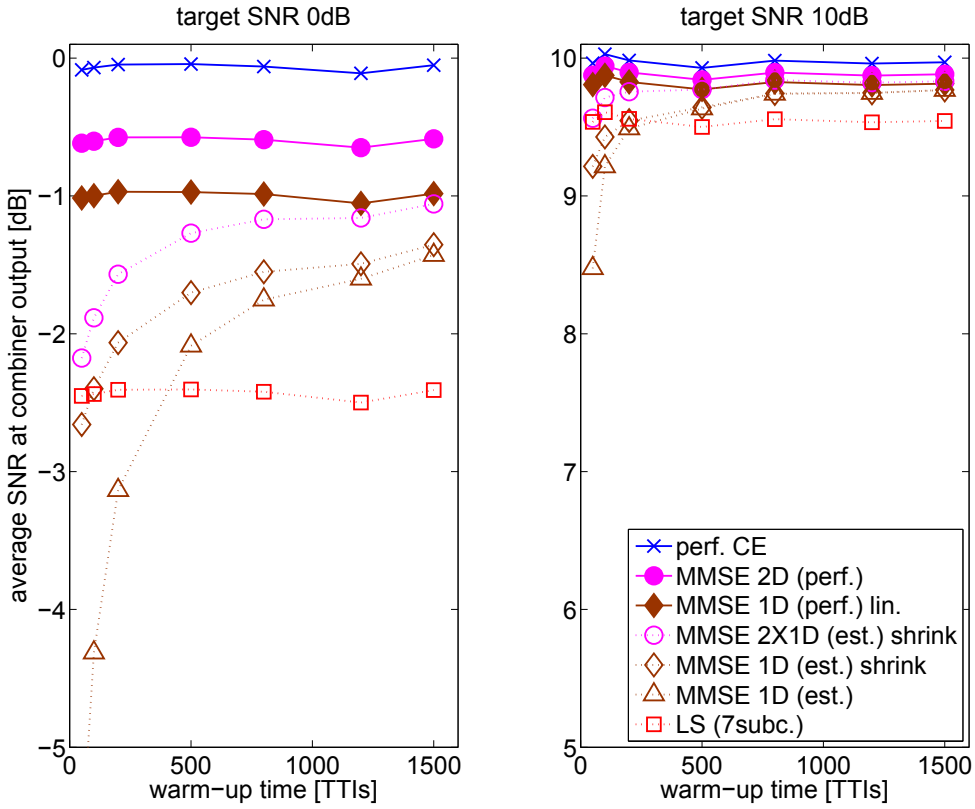

Figure 3.7.: SNR at receive combiner output over "warm-up time" for single-user multi-cell channel estimation. Source: [NRWtB11], Fig. 5.

radio environment. In this case the loss due to inaccurate statistical knowledge of the channel even cannot be overcome with infinite time for collection of channel statistics[20]. Shrinkage reduces this "warm-up gap" to about 1 dB. With about 500 subframes, it has collected enough channel statistics. This is still a large number, but the main reason for this is the low SNR operation point.

The results for 10 dB show a much faster warm-up behaviour. Without shrinkage, we have a warm-up gap of 1 dB and require about 250 subframes of statistical knowledge for the full performance. With shrinkage, the gap becomes less than half a dB.

3.2.4. Summary of Single-User CoMP Channel and Parameter Estimation

Interplay of Parameter Estimation and Channel Estimation Algorithms in CoMP

[20]Example: A user with 10% activity of transmission, moving with 36 km/h = 10 m/s has moved 100 meters on average for 1000 subframes of data transmission. Thus depending on the environment, the parameters P_n and τ_n in (3.20) which define $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ might already have changed noticeably.

The introduced least squares-based OFDM channel estimation approaches are known to work fine for single-cell and do not need additional parameters to be estimated. However, when looking at the simulation results in section 3.2.3, it is clear that these approaches fail in CoMP. This is because of low SNR operation points and imbalanced effective path gains. Satisfying performance requires the usage of Wiener filters. Here, statistical knowledge is required to de-weight unreliable sources and profit from any possible correlation between the data symbol resource element of interest and any available pilot resource element. Those pilot resource elements may potentially be far away in terms of time-frequency resources. This statistical knowledge, the channel cross- and auto-covariances as well as the noise variance, is obtained by collecting observations of a MS over different subframes. Collecting these parameters in a noise-limited environment requires a long "warm-up time" for the channel estimation algorithm, until satisfying performance is achieved, as shown in section 3.2.3. This is a strong drawback.

Summary for Parameter Estimation

The results in sections 3.2.2.2 and 3.2.3 show that the shrinkage technique strongly helps to shorten the "warm-up time" and improve the quality of the parameter estimation. Regarding noise variance estimation, according to section 3.2.2.4 , the simple filtering based schemes perform best in this noise-limited setting. The more complex Subspace noise estimation performs similar, but with weaknesses in very-low SNR operation points.

3.3. Multi-User Multi-Cell Channel Estimation and Pilot Sequence Assignment

While the previous section 3.2 has dealt with the problem of estimation and combining for a single user in multiple cells, now this section addresses multiple users. Our system of interest is operated at full frequency reuse in order to maximise spectral efficiency. Thus all signals from all cells fully interfere with each other. The receive combiner has to take into account interference and suppress it. For a two user, two cell scenario, in appendix C, the post-combining SINR for perfect channel knowledge is analytically computed for narrowband channels as a function of a path loss parameter. This parameter represents the "isolation" between different cells. Appendix C shows that a CoMP system with fully interfering cell edge mobiles can achieve even better performance than completely interference-free single cell reception, due to interference-aware combiners and the multi-cell antenna gain.

Using the LTE-A signal structure, described in appendix A, the pilot sequences of different MS (located in different cells) fully superimpose. Thus the sequence properties become important, in terms of their orthogonality and cross-correlation. This section will discuss both standard-compliant pseudo-random sequence assignments, as well as a potential extension with orthogonal sequences.

The considered estimator strategies discussed here are as follows: The channel estimation is performed (1) separately per user as well as (2) jointly for all users. The former is more suitable for distributed processing architectures. The latter requires a central processing point, but offers a higher performance potential. An important aspect of this section is the investigation of the

interrelations between the pilot sequence assignment and the channel estimator strategies, thus, the suitability of certain channel estimation approaches to particular pilot sequence assignments across the different coordinated cells.

We have seen from the previous section 3.2 that considering the parameter knowledge in conjunction with the channel knowledge is important. This aspect will be handled in the subsequent section 3.4 for the multi-user case. In order to keep the discussion manageable, we assume perfect parameter knowledge throughout this section 3.3. A simpler approximated MMSE receiver with less parameters is introduced and discussed as well.

Some of the work contained in this section was published in [WNtB12].

3.3.1. Scenario and Notation

A set of I base station antennas is coordinated in a cluster in order to jointly process the receive signals coming from J transmitting antennas. The set $\mathcal{J}$ denotes the selected users of a time-frequency radio resource of interest. The users may be selected by the scheduler out of a larger set of active mobile stations.

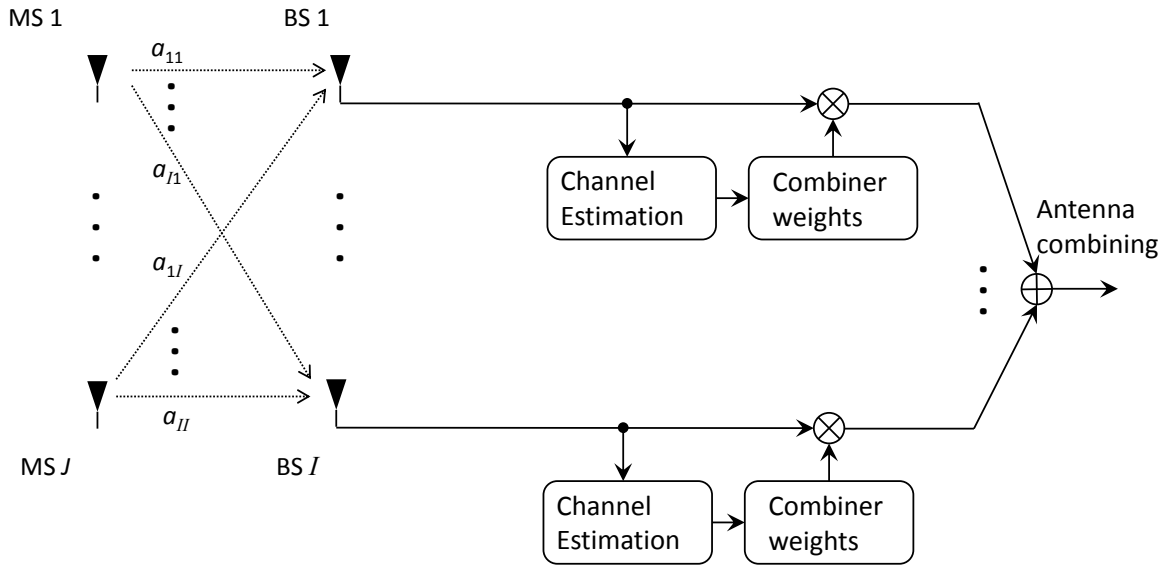

Figure 3.8.: Schematic picture of uplink multi-user coordinated multi-point reception scenario.

According to (3.4) the receive vector in pilot dimension at BS antenna i can be written as

$$\mathbf{y}_i = \sum_{\forall j \in \mathcal{J}} \mathbf{S}_j \mathbf{h}_{ij}^{[\mathrm{P}]} + \mathbf{n}_i = \sum_{\forall j \in \mathcal{J}} \sqrt{a_{ij}} \mathbf{S}_j \tilde{\mathbf{h}}_{ij}^{[\mathrm{P}]} + \mathbf{n}_i. \tag{3.43}$$

Note that in this notation, the general linear model (3.5) is not directly used, because of the resulting dimensions in the channel estimator derivation, leading to unnecessary statistical knowledge

to be collected, further increasing warm-up time. In the estimator subsection 3.3.2 the decoupling of the users by this treatment is discussed. The outcome based on the notation (3.43) is more suited to varying user combinations than the generalized linear model.

Fig. 3.8 illustrates the setting for the user set $\mathcal{J} = \{0, 1, \ldots, N_{\text{supp}}\}$. Each user is associated to one serving base station, while the other N_{supp} base stations act as supporting cells.

3.3.1.1. Path Gain and Power Control Modeling

Appendix C indicates that the coupling parameters $\sqrt{a_{ij}}$ from MS to BS will play a certain role in the combiner performance, but the actual post-combining SINR based on perfect channel knowledge remains rather insensitive to the realization of $\sqrt{a_{ij}}$, according to Fig. C.1.

Clearly, the differences in receive power levels of different users, based on the combined effect of path gains and power control $\sqrt{a_{ij}}$, will impact the channel estimator performance. Hence, for the evaluations based on simulations, $\sqrt{a_{ij}}$ must be chosen carefully.

For the path gain properties, downlink geometry measurements [SHC+10] of a dense urban deployment (which is an attractive setting for the introduction of CoMP) give indications. It was observed that the median path gain drops by roughly 4dB from the serving cell to the strongest interfering cell. A further drop of roughly 4dB is observed from strongest to second strongest interfering cell etc.

In LTE / LTE-A systems, an open loop power control is applied with fractional path loss compensation [3GPd] which compensates from a fraction of the path loss variation up to the full path loss. In non-CoMP systems, a non-full path loss compensation is beneficial, because it limits the interference caused to neighbor cells. This inter-cell interference plays an important role for the system operation point in terms of the *interference-over-thermal-noise-power-ratio* (IoT) . IoT at BS antenna i is defined as the ratio of the sum of interference and noise power over the noise power

$$\text{IoT}_i = \frac{\sum_{\forall \nu \in \mathcal{J} \setminus \{i\}} a_{i\nu} + \sigma_n^2}{\sigma_n^2} \tag{3.44}$$

For the ease of notation, as we have one user per cell and we do not consider scheduling here, we write the user indices identical to their serving cell indices. Unit pilot sequence power and unit channel attenuation is assumed.

A further important parameter to characterize the level of inter-cell interference is the signal-to-interference-ratio (SIR) at the i-th BS antenna input for a link from MS antenna j to BS antenna i:

$$\text{SIR}_{ij} = \frac{a_{ij}}{\sum_{\forall \nu \in \mathcal{J} \setminus \{j\}} a_{i\nu}} \tag{3.45}$$

As we will see from the simulation results throughout section 3.3.3, CoMP deals well with increased IoT, due to the interference-suppression capabilities of the coordinated distributed antenna system. This motivates the target of full path loss compensation, as the cell edge mobiles will benefit from it in terms of increased throughput. Thus, it will play out one of the dominant advantages of CoMP: Balancing out fairness in cells by providing increased cell edge rates. So

for this CoMP scenario, a full path loss compensation is assumed. Furthermore it is assumed that the peak power limitation will not play a dominant role.

In order to model the receive power level distributions, based on the path gain observations from measurements and the power control strategy discussed above, a discrete probability density distribution is used, which can be parametrized, based on the number of supporting cells N_{supp} and the tuning parameters β and α_0:

$$p\left(a_{ij}|N_{\text{supp}},\beta,\alpha_0\right) = \frac{1}{N_{\text{supp}}} \sum_{k=1}^{N_{\text{supp}}} \delta\left(a_{ij} - \alpha_0\beta^{-k}\right). \tag{3.46}$$

In an actual LTE system, an IoT control mechanism is active which impacts the transmit powers of all mobiles in order to roughly achieve a desired IoT operation point per cell. This is taken into account here by the parameter α_0, which models a general impact on receive power levels from non-serving cells. In the simulation results subsection, the impact of varying IoT, thus varying α_0, is shown.

The distribution implies that the coordination cluster is properly chosen to be the set of cells N_{supp} which are received best in downlink. The larger the coordination set and thus N_{supp}, the more weak links are likely to occur in the system, making channel estimation more challenging. The parameter β determines how strong the magnitude of the additional path gain decays when the coordination cluster becomes larger. The measurements [SHC+10] motivate $4\,\text{dB} \approx 10\log_{10} e$, so we set $\beta = e$.

The direct links to the serving cells are set to unit receive power levels with $a_{ii} = 1$ and for the supporting cell links the $a_{ij}, \forall i \neq j$ are drawn out of the parametrized distribution (3.46).

3.3.1.2. Receive Combining and Resulting SINR

Signals received at different OFDM resource elements are mutually orthogonal. Hence, receive combining can be executed for each resource element separately in the antenna dimension (see definitions of (3.3)). The receive vector for a resource element at subcarrier l and OFDM symbol k can be written as

$$\mathbf{y}_{kl}^{[\text{B}]} = \sum_{\forall j \in \mathcal{J}} \mathbf{h}_{jkl}^{[\text{B}]} s_{jkl} + \mathbf{n}_{kl}^{[\text{B}]} \tag{3.47}$$

As we consider linear combining, for user j the received symbol after combining is $r_{jkl} = \mathbf{g}_{jkl}^{H}\mathbf{y}_{kl}^{[\text{B}]}$. For the single-user scenario, maximum ratio combining (3.41) was used. Clearly this is suboptimal in the presence of interference. So let's now discuss the optimal linear combiner in MMSE sense when the instantaneous channel is known. The general solution to this problem is known and used e.g. in [SB05]. It is derived here for the scenario of this section 3.3 to provide the comprehension for the reader.

The condition for achieving the minimum mean squared error, the orthogonality principle [Kay93] [Say08], states that error between actual transmitted symbol and the estimated symbol at com-

biner output has to be statistically orthogonal to the receive vector:

$$\begin{aligned}
\mathrm{E}\left\{(r_{jkl} - s_{jkl})\mathbf{y}_{kl}^{[\mathrm{B}]\,\mathrm{H}}\right\} &= \vec{\mathbf{0}} \\
\mathrm{E}\left\{(\mathbf{g}_{jkl}^{H}\mathbf{y}_{kl}^{[\mathrm{B}]} - s_{jkl})\mathbf{y}_{kl}^{[\mathrm{B}]\,\mathrm{H}}\right\} &= \vec{\mathbf{0}} \;\Rightarrow\; \mathbf{g}_{jkl}^{H}\,\mathrm{E}\left\{\mathbf{y}_{kl}^{[\mathrm{B}]}\mathbf{y}_{kl}^{[\mathrm{B}]\,\mathrm{H}}\right\} - \mathrm{E}\left\{s_{jkl}\mathbf{y}_{kl}^{[\mathrm{B}]\,\mathrm{H}}\right\} = \vec{\mathbf{0}} \\
\Rightarrow\; \mathbf{g}_{jkl}^{H} &= \mathrm{E}\left\{s_{jkl}\mathbf{y}_{kl}^{[\mathrm{B}]\,\mathrm{H}}\right\}\mathrm{E}\left\{\mathbf{y}_{kl}^{[\mathrm{B}]}\mathbf{y}_{kl}^{[\mathrm{B}]\,\mathrm{H}}\right\}^{-1} \\
\Rightarrow\; \mathbf{g}_{jkl} &= \mathrm{E}\left\{\mathbf{y}_{kl}^{[\mathrm{B}]}\mathbf{y}_{kl}^{[\mathrm{B}]\,\mathrm{H}}\right\}^{-1}\mathrm{E}\left\{s_{jkl}^{*}\mathbf{y}_{kl}^{[\mathrm{B}]}\right\} \\
\Rightarrow\; \mathbf{g}_{jkl} &= \left(\sum_{\forall\mu\in\mathcal{J}}\mathbf{h}_{\mu kl}^{[\mathrm{B}]}\mathbf{h}_{\mu kl}^{[\mathrm{B}]\,H} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{h}_{jkl}^{[\mathrm{B}]} \\
\Rightarrow\; \mathbf{g}_{jkl} &= \underbrace{\xi\left(\sum_{\forall\mu\in\mathcal{J},\mu\neq j}\mathbf{h}_{\mu kl}^{[\mathrm{B}]}\mathbf{h}_{\mu kl}^{[\mathrm{B}]\,H} + \sigma_n^2\mathbf{I}\right)^{-1}}_{\mathbf{Z}_{jkl}}\mathbf{h}_{jkl}^{[\mathrm{B}]}.
\end{aligned} \tag{3.48}$$

The assumptions are uncorrelated data symbols between users with unit norm, so $\mathrm{E}\left\{s_{jkl}^{*}\mathbf{y}_{kl}^{[\mathrm{B}]}\right\} = \mathbf{h}_{jkl}^{[\mathrm{B}]}$ holds. The last step is proven in the appendix D. The spatial structure of interference and noise is collected in matrix $\mathbf{Z}_{jkl}$ and ξ is a normalization factor in order to achieve $||\mathbf{g}_{jkl}||_2 = 1$.

This approach is often denoted as *interference rejection combining* (IRC).

For the SINR computation, we first decompose the signal at the receive combiner output:

$$r_{jkl} = \mathbf{g}_{jkl}^{H}\mathbf{y}_{kl}^{[\mathrm{B}]} = \underbrace{\mathbf{g}_{jkl}^{H}\mathbf{h}_{jkl}^{[\mathrm{B}]}s_{jkl}}_{r_{jkl}^{[U]}} + \underbrace{\mathbf{g}_{jkl}^{H}\left(\sum_{\forall\mu\in\mathcal{J},\mu\neq j}\mathbf{h}_{\mu kl}^{[\mathrm{B}]}s_{\mu kl} + \mathbf{n}_{kl}^{[\mathrm{B}]}\right)}_{r_{jkl}^{[I+N]}} \tag{3.49}$$

The SINR at the receive combiner output, can be computed as the useful signal power $\mathrm{E}\left\{|r_{jkl}^{[U]}|^2\right\}$ over the noise and interference contributions $\mathrm{E}\left\{|r_{jkl}^{[I+N]}|^2\right\}$ at the combiner output[21]:

$$\gamma_{jkl} = \frac{\mathbf{g}_{jkl}^{H}\mathbf{h}_{jkl}^{[\mathrm{B}]}\mathbf{h}_{jkl}^{[\mathrm{B}]\,H}\mathbf{g}_{jkl}}{\mathbf{g}_{jkl}^{H}\mathbf{Z}_{jkl}\mathbf{g}_{jkl}}. \tag{3.50}$$

For the performance evaluations, this post-combining SINR is averaged over the entire user allocation, i.e. $j = \{1, 2, \ldots, J\}$,and used as a performance metric.

Within this section a simplified, less complex IRC algorithm is investigated as well which we denote “dIRC”. dIRC just uses the diagonal of $\mathbf{Z}_{jkl}$ for taking account of the interference. This means the weighting is just a real-valued scaling factor based on the inverse interference-plus-noise power level per antenna. The advantage is that the matrix inversion becomes trivial, as it is just a division per diagonal element. As we will see later on in the simulation section, this dIRC

[21] Note that the expectation is done w.r.t. the transmitted symbols and the noise realizations.

algorithm can also provide useful results, near the full-blown IRC, in combination with some channel estimation algorithms. The dIRC weights are:

$$\mathbf{g}_{jkl}^{\mathrm{dIRC}} = \xi_{\mathrm{dIRC}} \cdot \mathrm{diag}\left[Z_{jkl}(1,1), Z_{jkl}(2,2), \ldots, Z_{jkl}(J,J)\right]^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} \tag{3.51}$$

In case of non-perfect CSI, we again consider the receive combiner as a spatial filter and compute useful signal power over noise-plus-interference power. This leads to (3.50) with receive weights $\hat{\mathbf{g}}_{jkl}$ based on estimated channels $\hat{\mathbf{h}}_{jkl}^{[\mathrm{B}]}$, thus being mismatched, compared to weights $\mathbf{g}_{jkl}$ based on perfect CSI.

3.3.1.3. Pilot Sequence Assignments

In the multiuser scenario, the assignment of pilot sequences to the users becomes an issue, potentially impacting channel estimation performance. Thus we first discuss LTE-A standard compliant pilot sequences assignments, as well as possible extensions close to the LTE-A standard.

Standard-Compliant Assignments
The LTE-A standard uses *demodulation reference symbols* (DMRS) as pilots, presented in detail in appendix A. As LTE was not designed for CoMP from scratch, classical single cell reception was anticipated with reduced complexity receivers, targeting an interference randomization strategy on the pilot sequences. This can be seen from the correlation properties

$$|\rho_{\mu\nu}| = |\frac{1}{L}\mathbf{s}_{\mu}^{H}\mathbf{s}_{\nu}| \tag{3.52}$$

between pairs of length L sequences $\mathbf{s}$, indexed μ and ν. The Zadoff-Chu sequences, according to (A.1), provide a very well defined cross-correlation property when comparing sequences with different roots, $\mu \neq \nu$, as shown by the blue curve marked with "x" in Fig. 3.9. The correlation of sequences is an important indicator on residual interference contained in the channel estimation result.

For a correlation length of $L = 60$, corresponding to an allocation size of 5 PRBs, we observe that the correlation outcome is almost the same, regardless of the sequence root[22] chosen. If we compare this result to a correlation with a random sequence (constant amplitude, random phase), the cdf shows that around 30% of the random sequences have worse outcomes, which may even have a correlation up to 0.3 and beyond. The clear-defined cross-correlation property of Zadoff-Chu prevents such worst case spikes. It can be straightforwardly deduced that the average correlation absolute of pure random unit amplitude sequences is $1/\sqrt{L} \approx 0.13$ for length $L = 60$.

However, when the actual channel estimator operates on a smaller cutout of the sequence, e.g. a part where the channel remains approximately constant due to the coherence bandwidth, then the cross-correlations change drastically. Fig. 3.9 shows what happens, when considering cutouts of a length $L = 12$ (1 PRB): The Zadoff-Chu sequences now behave similar to random sequences.

[22] According to appendix A (A.1)

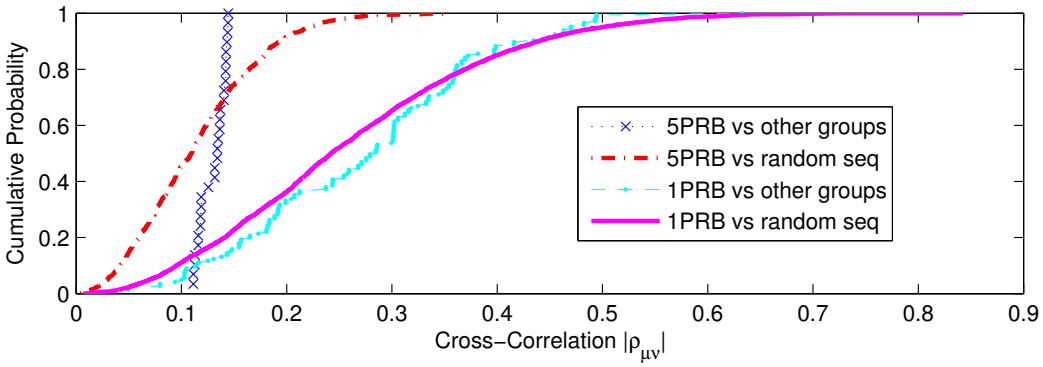

Figure 3.9.: Absolute cross-correlation $|\rho_{\mu\nu}|$ (3.52) of pilot sequence groups: Group $\mu = 0$ correlated with groups $\nu = \{1, 2, \ldots, 29\}$ over L=60 ("5PRB vs other groups") and L=12 ("1PRB vs other groups") or correlated with random phase constant amplitude sequences over L=60 ("5PRB vs random seq') and L=12 ("1PRB vs random seq").

When looking at the correlation properties at the receive side, one has to take into account also the phase rotations and amplitude variations of the propagation channel. As the sequences in LTE-A uplink are stretching over the frequency dimension, the well-defined cross-correlation property over long sequence lengths (e.g. $L = 60$) can only be uphold in the case of frequency-flat channels, e.g. caused by line-of-sight propagation. Let us consider the case of received signals from two users, if one of them has a frequency-selective fading: The resulting cross-correlation of the pilot sequences which have passed the channel $|\rho_{\mu\nu}^{\mathrm{Rx}}| = |\frac{1}{L}\mathbf{h}_{i\mu}^H \mathbf{S}_\mu^H \mathbf{S}_\nu \mathbf{h}_{i\nu}|$ to receive antenna i is completely different from the pure sequence correlation $|\rho_{ij}|$. When one of the two example users e.g. has subcarrier-wise-independent i.i.d. Rayleigh fading, the cross-correlation observed from a receiver perspective is very similar to the one generated from a pure random pilot sequence.

To sum up the observed sequence properties at the receiver: In case of processing only a cutout of the entire sequence, or in the likely case of frequency-selective fading, the sequence design and assignment provided by the LTE/LTE-A standard results in a sequence cross-correlation behaviour similar to pure random sequences.

Assignments with possible Standard Extensions

For uplink multi-user MIMO transmission within one cell, the LTE-A standard allows to use Zadoff-Chu sequences with the same roots with different cyclic sequence shifts, see (A.3). One nice property of Zadoff-Chu sequences is that they have *constant amplitude, zero autocorrelation* (CAZAC) and thus a sequence is orthogonal to any cyclic shift of itself. This means, several users within a cell transmit on the same time frequency resources with the same root sequence, but with different cyclic shifts. In case of true orthogonality, the channel estimation, even with single-user strategies would be possible interference free[23]. This approach now can be easily generalized to CoMP: Within a coordinated set of cells, the base station central control unit can assign the same

[23]Note that in LTE, by design, imperfections are already included, as the subcarrier allocation size is a multiple of 12. As Zadoff-Chu sequences require prime number lengths, certain parts of the sequences are repeated at the end.

root indices with different cyclic shifts to users spread across the cells of the CoMP cluster, which transmit on the same time-frequency resources. This is not standard-compliant. However, in the investigations of this thesis, this provides a simple potential extension to the LTE-A standard which can be investigated as an orthogonal sequence assignment alternative in CoMP.

While the sequences inhabit zero cross-correlation due to orthogonality, again, at the receive side the picture becomes different. With the propagation channel varying differently over the frequency for different users, orthogonality gets lost and the cross-correlation goes up. From receiver perspective, the above extreme-case[24] of subcarrier-wise-independent i.i.d. Rayleigh fading, results in the same the same statistical behavior of the cross-correlation as in the standard-compliant case, similar to random reference sequences.

Furthermore, the realization of this orthogonal sequence assignment for CoMP has its limitations and restrictions. The number of possible shifts is limited. First there is the above mentioned frequency selectivity effect, thus the shift length should be smaller than the coherence bandwidth. As orthogonal sequences are limited in number this limits the number of supported users. This provides a bad scaling over the area of many cells. At the edges of CoMP clusters there is the possibility to run out of orthogonal sequences. Dealling with this, where possible, will result in more control overhead due to signaling of assigned sequence indices, as the limited amount has to be shared by several users not being simultaneously active. Additionally, to upkeep orthogonality, all the resource allocation block sizes have to be aligned over the users in multiple cells, as partly overlapping resource allocations will consume further limited orthogonal sequences. Nevertheless, being aware of all practical limitations, we will simulate and consider this case as a benchmark case.

A third alternative appears as a compromise between the standard-compliant solution and the orthogonal assignment based on shifts. The CoMP control unit tries to keep as much users as possible within the same root, orthogonalize them by different shifts. A different sequence root is used preferably when users are isolated as much as possible by path loss differences. Note that this requires extensive sequence planning, which increases control signaling overhead.

Further (not Investigated) Alternatives for Pilot Sequences
When dropping the restriction that only Zadoff-Chu sequences will be used, one can discuss further alternatives. There is a whole bunch of orthogonal sequence sets like Walsh-Hadamard sequences or DFT sequences. However, a sequence length L still provides only L different orthogonal sequences which result in no advantages over Zadoff-Chu sequences with different shifts in terms of available number of sequences. The CAZAC property in general is important for a reference sequence, due to their auto-correlation properties for timing-offset estimation, frequency-offset estimation etc. From the family of CAZAC sequences, the best known sequences are Zadoff-Chu, but there are possible alternatives, like e.g. Björck sequences [BS11]. There is no indication that the results will significantly differ when other CAZAC sequences than Zadoff-Chu are used. As discussed above, the effect of the propagation channel "randomizes out" the intricacies of potential design choices.

Cases Considered within this Thesis
This thesis purely focuses on Zadoff-Chu sequences, being closest to the standard and comprising the powerful CAZAC property. Three different cases are considered:

[24] With LTE OFDM parameters requiring ultra-high delay spreads, only potentially possible in high mountain terrain

- Separation of users by different roots: Standard-compliant non-orthogonal sequences
- Separation of users by different shifts: Possible extension of the standard with orthogonal sequences.
- Separation of users by using 2 different roots and different shifts within each root: Possible extension of the standard, more easy to realize than different shifts only, with orthogonality within each of the two sequence groups.

3.3.2. Multi-User Channel Estimation Algorithms

In our multi-user channel estimation scenario we have two possible classes of estimators:

- *Separate*: Channel estimation is done separately per user, based on algorithms obtained from single user channel estimation. Its basic approaches are described in 3.2.1.
- *Joint*: Channel estimation is done jointly for all users, thus multi-user channel estimation algorithms are required.

The separate channel estimation approach allows to better distribute the processing load. In the example of a master-slave CoMP concept, discussed in 2.1.4.3, where each serving cell only processes its served user, the processing for channel estimation at a particular base station can be restricted purely to users of interests, while other base stations might handle the channel estimation of other users based on e.g. forwarded IQ data. Separate channel estimation is also less complex. A further advantage is the ability to handle a user which has been transmitting data for already a long time and now gets interfered by just recently active users, not having transmitted beforehand. For those interferers, there might be no or poor statistical knowledge available. Thus it might be better to initially *treat interference as noise* and process the long-time-active user on its own, where the statistical knowledge is much more accurate.

The joint channel estimation approach offers a higher performance potential, as it will be able to deal also with interference[25]. This is beneficial in the absence of full sequence orthogonality at the receive side. The price to be paid is higher complexity. Whether joint channel estimation will be worth the effort is a question which will be answered throughout the remaining section 3.3.

3.3.2.1. Separate Channel Estimation

The least squares based approach of section 3.2.1.1 can be directly applied also to the multi-user scenario with the risk of potentially low performance. This means, our separate LS estimate is obtained according to (3.14).

The Wiener Filter, derived in section 3.2.1.2, now has to account at least for the receive power levels of the other users. By treating other users as spectrally white noise, which might be reasonable with a large number of non-dominant interfering users, we can replace the noise variance σ_n^2 in the receiver formulation of 3.2.1.2 by the overall variance of interference σ_i^2 plus

[25] In contrast, for seperate channel estimation we treat interference as AWGN.

noise, denoted $\bar{\sigma}^2$, with $\bar{\sigma}^2 = \sigma_n^2 + \sigma_i^2$. Thus the Wiener filter weights of the single-user case (3.18) are now modified for the multi-user case with separate (per-user) channel estimation to:

$$\mathbf{w}_{ijkl} = \bar{\mathbf{\Phi}}_i^{-1} \bar{\mathbf{r}}_{ijkl} \tag{3.53}$$

When treating the interfering users as noise the auto-covariance at receive antenna $i \in \mathcal{I}$ can be synthesized here as

$$\bar{\mathbf{\Phi}}_i = \mathbf{S}_j \underbrace{\mathrm{E}\left\{ \mathbf{h}_{ij}^{[P]} \mathbf{h}_{ij}^{[\mathrm{P}]\,H} \right\}}_{\mathbf{R}_{ij}^{[\mathrm{hh}]}} \mathbf{S}_j^H + \bar{\sigma}^2 \mathbf{I}. \tag{3.54}$$

where $j \in \mathcal{J}$ is the user of interest and $\bar{\sigma}^2 = \sigma_n^2 + \sum_{\nu=1,\nu\neq j}^{J} a_{i\nu}$ can be composed by obtaining knowledge on the sum interference receive power levels, where all interfering users contribute with their path gain and transmit power a_{ij}.

3.3.2.2. Joint Channel Estimation

A multi-user channel estimator aims at obtaining all per-resource-element channel coefficients in parallel for all users, thus $\mathbf{h}_{ikl}^{[\mathrm{U}]}$ from (3.3). According to the discussion in the signal model section 3.1, the general linear model (3.5) holds. Hence, the linear MMSE estimator coincides with the MMSE estimator and a Wiener filter is the optimal solution in a mean squared error sense. In contrast to the single user solution (section 3.2.1.2) or the separate per-user solution (section 3.3.2.1) the estimator output is not scalar, but vector-valued. This means that the Wiener filter has to be a $J \times P$ weight matrix $\mathbf{W}_{ikl}$, which we call Matrix Wiener Filter (MWF). So the joint estimates using a linear estimator can be written as

$$\hat{\mathbf{h}}_{ikl}^{[\mathrm{U}]} = \mathbf{W}_{ikl}^H \mathbf{y}_i. \tag{3.55}$$

Full Matrix Wiener Filter

The derivation of the weight matrix follows by the orthogonality principle, with $\mathbf{0}_{[J\times P]}$ being a $J \times P$ matrix filled with zeros, as

$$\begin{aligned} \mathrm{E}\left\{ (\mathbf{W}_{ikl}^H \mathbf{y}_i - \mathbf{h}_{ikl}^{[\mathrm{U}]}) \mathbf{y}_i^H \right\} &= \mathbf{0}_{[J\times P]} \;\Rightarrow\; \mathbf{W}_{ikl}^H \mathrm{E}\left\{ \mathbf{y}_i \mathbf{y}_i^H \right\} - \mathrm{E}\left\{ \mathbf{h}_{ikl}^{[\mathrm{U}]} \mathbf{y}_i^H \right\} = \mathbf{0}_{[J\times P]} \\ \Rightarrow\; \mathbf{W}_{ikl}^H &= \mathrm{E}\left\{ \mathbf{h}_{ikl}^{[\mathrm{U}]} \mathbf{y}_i^H \right\} \mathrm{E}\left\{ \mathbf{y}_i \mathbf{y}_i^H \right\}^{-1} \\ \Rightarrow\; \mathbf{W}_{ikl} &= \mathrm{E}\left\{ \mathbf{y}_i \mathbf{y}_i^H \right\}^{-1} \mathrm{E}\left\{ \mathbf{y}_i \mathbf{h}_{ikl}^{[\mathrm{U}]\,\mathrm{H}} \right\}. \end{aligned} \tag{3.56}$$

The complex $P \times P$ auto-covariance $\mathbf{\Phi}_i = \mathrm{E}\left\{ \mathbf{y}_i \mathbf{y}_i^H \right\}$ of the receive signal at pilot positions can

be computed using (3.43):

$$\begin{aligned}
\mathbf{\Phi}_i &= \mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\} \\
&= \mathrm{E}\left\{\left(\sum_{\forall j\in\mathcal{J}}\sqrt{a_{ij}}\mathbf{S}_j\tilde{\mathbf{h}}_{ij}^{[\mathrm{P}]}\right)\left(\sum_{\forall j\in\mathcal{J}}\sqrt{a_{ij}}\mathbf{S}_j\tilde{\mathbf{h}}_{ij}^{[\mathrm{P}]}\right)^H\right\}+\sigma_n^2\mathbf{I} \\
&= \sum_{\forall j\in\mathcal{J}} a_{ij}\mathbf{S}_j\,\mathrm{E}\left\{\tilde{\mathbf{h}}_{ij}^{[\mathrm{P}]}\tilde{\mathbf{h}}_{ij}^{[\mathrm{P}]\,H}\right\}\mathbf{S}_j^H+\sigma_n^2\mathbf{I} \\
&= \sum_{\forall j\in\mathcal{J}} \mathbf{S}_j\mathbf{R}_{ij}^{[\mathrm{hh}]}\mathbf{S}_j^H+\sigma_n^2\mathbf{I}. \qquad (3.57)
\end{aligned}$$

The second line bases on the independence of the noise vector $\mathbf{n}_i$ to all channel vectors. In the third line, all the mixed terms $\mathrm{E}\left\{\tilde{\mathbf{h}}_{i\mu}\tilde{\mathbf{h}}_{i\nu}^H\right\}, \forall\mu\neq\nu$ are assumed to disappear, which assumes uncorrelated fast fading realizations of transmit antennas belonging to different users μ and ν. According to section 2.1.3, a distance of several wavelengths between different antennas is sufficient to fulfill this assumption[26]. Furthermore the channel typically has high angular spread at the mobile stations, which even more reduces the distance beyond which the channels of different transmit antennas become uncorrelated. This in conjunction with polarization due to different possible orientations of different user's antennas makes sure that this fast fading decorrelation of different users can be assumed to be always fulfilled in our chosen scenario. This means we can use the final line of (3.57) as a near-perfect approximation of the receive signal auto-covariance.

This decoupling on users also allows us to use the following strategy suited for a practical system using dynamic scheduling and non full-buffer traffic. As user combinations will change, we can track the channel covariances $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ of each user separately and handle their composite auto-covariance by the summation term in (3.57). Here users can add up independently and there are no cross-terms left to handle. This greatly simplifies the handling of the multi-user statistics and the complexity of the channel estimator.

Just for comparison, the autocovariance based on the generalized linear model results in:

$$\begin{aligned}
\mathbf{\Phi}_i &= \mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\} \\
&= \mathrm{E}\left\{\left(\mathbf{S}^{[all]}\mathbf{h}_i^{[\mathrm{all}]}+\mathbf{n}_i\right)\left(\mathbf{S}^{[\mathrm{all}]}\mathbf{h}_i^{[\mathrm{all}]}+\mathbf{n}_i\right)^H\right\} \\
&= \mathbf{S}^{[\mathrm{all}]}\,\mathrm{E}\left\{\mathbf{h}_i^{[\mathrm{all}]}\mathbf{h}_i^{[\mathrm{all}]\,H}\right\}\mathbf{S}^{[\mathrm{all}]\,H}+\sigma_n^2\mathbf{I}. \qquad (3.58)
\end{aligned}$$

Note that the composite channel covariance $\mathrm{E}\left\{\mathbf{h}_i^{[\mathrm{all}]}\mathbf{h}_i^{[\mathrm{all}]\,H}\right\}$ has now to be rearranged for every possible user combination. Furthermore, it is of dimension $PJ\times PJ$ instead of $P\times P$ of the user-individual $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ in (3.57). Moreover, the approach (3.58) lacks the modularity of (3.57), suitable for dynamically changing user combinations by scheduling. The proposed approach (3.57) just can add an additional contribution for each user, while in (3.58) dimensions have to

[26] At a carrier frequency of 2GHz, the wavelength is $\lambda = 15$ cm.

be changed on the fly, which is clumsy or even prohibitive e.g. when collecting the empirical covariance. In summary, (3.57) is the preferable approach over (3.58)[27].

The complex $P \times J$ cross-covariance matrix is denoted $\mathbf{R}_{ikl} = \mathrm{E}\left\{\mathbf{y}_i \mathbf{h}_{ikl}^{[\mathrm{U}]\,\mathrm{H}}\right\}$, where each column contains the cross-covariance vector to the channel of user j, as in (3.22) referred to as $\mathbf{r}_{ijkl}$. This is true when the mixed terms $\mathrm{E}\left\{\tilde{\mathbf{h}}_{i\mu}\tilde{\mathbf{h}}_{i\nu}^{H}\right\}, \forall \mu \neq \nu$ disappear, as discussed for the auto-covariance.

Simplified Small-Window Approximation

For successful operation, the full MWF requires knowledge on the parameters $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ and σ_n^2. The details of the multi-user parameter estimation challenge are discussed in section 3.4; the single user case was handled in section 3.2.2. Already here, we can discuss the available options, when access to $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ is not available. As $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ for an example allocation size of 5PRBs is of dimension 120×120, a certain warm-up time in terms of available subframes for observation is required, as discussed in 3.2.2. As this warm-up time is in some cases not available, an option is presented here which allows to skip the necessity for acquiring the full $\mathbf{R}_{ij}^{[\mathrm{hh}]}$.

So, in case the receiver has no access of the full large-dimensional $\mathbf{R}_{ij}^{[\mathrm{hh}]}$, an alternative is to rely on the effective path loss a_{ij}, being defined as a combination of path loss and transmit power. This is much easier to obtain as it is a scalar for a particular MS-to-BS link.

For a small frequency distance, the radio channel CTFs at different subcarriers are very highly correlated, approximately with correlation of 1. The actual frequency distance where this approximation holds is a function of coherence bandwidth, discussed in section 2.1.3. The chosen LTE subcarrier spacing of 15 kHz, even for macro channels, lets this assumption hold for several subcarriers, e.g. about one PRB, depending on the delay spread. For a cut-out window of L pilots (of the same OFDM symbol), which are arranged in frequency direction, we can approximate the channel covariance matrix as

$$\breve{\mathbf{R}}_{ij}^{[\mathrm{hh}]} \approx a_{ij} \vec{\mathbf{1}}_{[L\times 1]} \vec{\mathbf{1}}_{L\times 1}^{T}, \tag{3.59}$$

being filled with identical values due to the correlation of (almost) 1. With this approximation, the following receive strategy can be pursued:

1. For each pilot OFDM symbol (thus per LTE slot) 1-D small window (in frequency-direction) MMSE channel estimates for smoothing at the pilot positions (weights are given in (3.63))
2. "Sliding" this window for each slot in frequency direction over all pilot subcarriers.
3. Linear interpolation / extrapolation in time direction for obtaining the channel estimates at the data symbol positions.

The per-slot sliding-window estimates rely on a synthesized auto-covariance. Inserting (3.59) in (3.43) yields

$$\breve{\mathbf{\Phi}}_{il,[L]} = \sum_{\forall j \in \mathcal{J}} a_{ij} \breve{\mathbf{S}}_{jl}^{[L]} \vec{\mathbf{1}}_{[L\times 1]} \vec{\mathbf{1}}_{[L\times 1]}^{T} \breve{\mathbf{S}}_{jl}^{[L]\,H} + \sigma_n^2 \mathbf{I}. \tag{3.60}$$

[27] When strictly forcing (3.58) to a block diagonal structure by setting the small off-block-diagonal values to zero, it becomes an equivalent solution to (3.57).

Note that $\breve{\mathbf{S}}_{jl}^{[L]\,H}$ represents the appropriate "cut-out" of the pilot sequence for the particular window of interest and thus $\breve{\mathbf{\Phi}}_{il,[L]}$ has a frequency-dependency based on the known pilot sequences.

The cross-covariance for smoothing, using (3.59) in (3.22), can be expressed as

$$\breve{\mathbf{R}}_{ikl} = \left[a_{i1} \breve{\mathbf{S}}_{1l}^{[L]} \vec{\mathbf{1}}_{[L\times 1]}, ..., a_{iJ} \breve{\mathbf{S}}_{Jl}^{[L]} \vec{\mathbf{1}}_{[L\times 1]} \right] \tag{3.61}$$

which are the small window cross-covariances stacked columnwise per user.

The MWF for the 1-D sliding window, estimating the channel $\mathbf{h}_{ikl}^{[\mathrm{U}]}$ jointly for all users, uses the weight matrix $\breve{\mathbf{W}}_{ikl}^{[L]}$, where $[L]$ denotes the length L window, for the observation window pilot vector $\breve{\mathbf{y}}^{[L]}$:

$$\hat{\mathbf{h}}_{ikl}^{[U]} = \breve{\mathbf{W}}_{ikl}^{[L]\,H} \cdot \breve{\mathbf{y}}_{i}^{[L]} \tag{3.62}$$

In order to obtain the weights, the derivation in (3.56) with (3.60) and (3.61) is used:

$$\breve{\mathbf{W}}_{ikl}^{[L]} = \breve{\mathbf{\Phi}}_{il,[L]}^{-1} \breve{\mathbf{R}}_{ikl} \tag{3.63}$$

At the edges of a user allocation, the smoothing window cannot be centered anymore on the pilot resource elements of interest, as neighboring pilot resource elements beyond the allocation edges belong to different users or remain unused. For the allocation edges, the sliding window is reduced in size respectively. E.g. a sliding window which is usually of size $L = 7$ changes into $L = 4$ at the allocation edge as the 3 other subcarriers lie outside the allocation. This edge effect reduces the performance a little bit, but stays small for typical LTE allocation sizes of 5PRBs, thus 60 subcarriers.

3.3.3. Simulation Results

In order to assess the impact of the different pilot sequence assignments in conjunction with different receiver strategies, simulations have been made. The considered sequence assignments use the different cases introduced at the end of section 3.3.1.3

The following parameters are used: The WINNER II spatial channel model [WIM06] is used for urban macro channels (C2) with a low velocity of 3km/h, as typically assumed for CoMP. 1000 large-scale-parameter drops have been used with a duration of 20 ms each. The allocation size is 5 PRBs and all user allocations across all considered cells are assumed to be identical, thus fully overlapping (without partial overlap between users). As in section 3.2.1.1, the window size for the Least Squares estimator is $L = 7$, which is used for the small-window MWF approximation as well. *All statistical parameters are perfectly known to the estimators* within this section 3.3. The SNR, excluding interference, for a typical LTE-A macro cell system can be assumed to be 10 dB, as the system is still interference limited. This SNR is chosen for the simulation setup.

In contrast to the single user scenario, discussed in section 3.2, we now introduce inter-cell interference into the system, by having one user per cell active on the same time-frequency resource, as described in section 3.3.1 in detail. Two parameters characterize this inter-cell interference level: SIR and IoT. Both parameters will be varied in this investigation (in Fig. 3.12 and

Fig. 3.13). Besides that, if it is not otherwise stated, typical LTE-A macro cell values are picked, by setting $\beta = e, \alpha_0 = 1$ in (3.46). When increasing N_{supp} for a cluster with a few cells, the resulting SIR approaches 2.3 dB and the IoT approaches 9.4 dB, which is a clearly interference limited system as occuring in practical urban / sub-urban LTE deployments [SHC+10].

In Figure 3.10, the distribution of the normalized mean squared error of the estimated channel of the data symbol resource elements[28] at the channel estimator output

$$\text{NMSE}_{ij} = \frac{\text{E}\left\{|h_{ijkl} - \hat{h}_{ijkl}|^2\right\}}{\text{E}\left\{|h_{ijkl}|^2\right\}} \tag{3.64}$$

is shown for the 3 different investigated sequence assignments. (The depicted MSE refers to the links of the mobiles to their serving cell.) $N_{\text{supp}} = 7$ supporting cells with one active user per cell are forming the considered CoMP cluster.

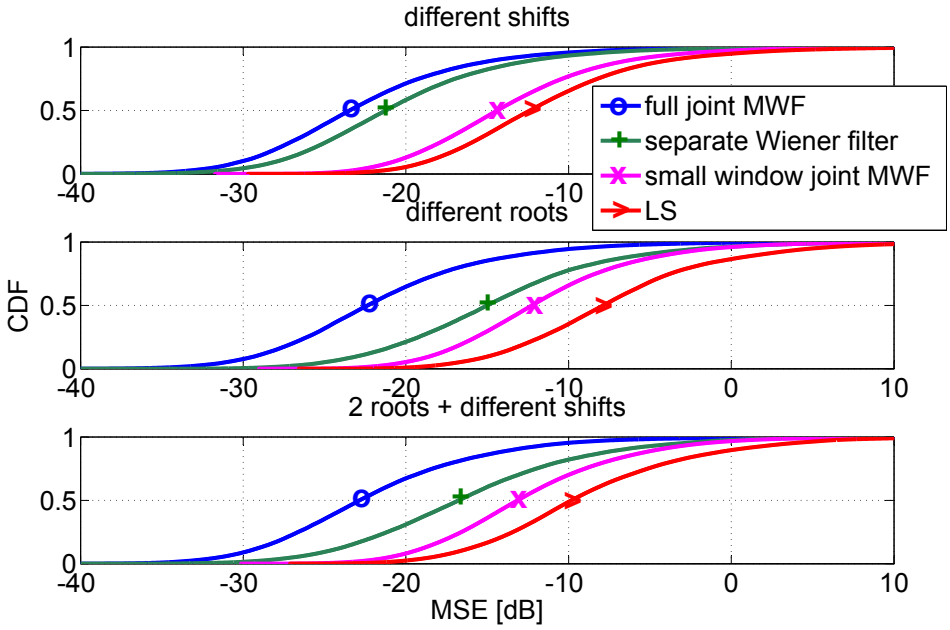

Figure 3.10.: Distribution of channel estimation normalized mean squared error (3.64), resulting from different estimators. Different pilot sequence assignments separate users by different shifts (top), different roots (mid) and two different roots, each with different shifts (bottom). $N_{\text{supp}} = 7$, SNR 10 dB, one user per cell, MSE refers to serving cell links.

We see for the joint full MWF (3.56), that the MSE distribution is almost the same, regardless of the chosen sequence assignment. This means that the full MWF is able to deal very efficiently with the cross-talk from multiple interfering users from the different cells of the CoMP cluster.

[28] The expectation for the NMSE is done for all resource element index pairs (k, l) of the allocated data symbols for a link from transmit antenna j to receive antenna i.

The allocation size considered here is typical 5 PRBs. This means the full MWF can exploit all 120 pilot symbols within the allocation to estimate the channel for each data symbol. This requires the receive pilot vector and the statistical knowledge gained e.g. from past observations. This statistical knowledge is, as mentioned, assumed here to be perfect.

In case of orthogonal sequences ("different shifts"), the separate Wiener filtering channel estimation, based on (3.53) and (3.54), is almost as good as the joint full MWF. It uses almost the same approach as a single-user channel estimator and is thus simpler to realize than the joint full MWF. In the non-orthogonal, standard compliant case ("different roots"), the separate MMSE already suffers from performance degradation. The median MSE goes up from $-22\,\mathrm{dB}$ to $-15\,\mathrm{dB}$. Even when there are two orthogonal subgroups ("2 roots + different shifts"), the median MSE is already at $-17\,\mathrm{dB}$.

These results show that least squares based channel estimation performs poor. For "different shifts", the median MSE is at $-12\,\mathrm{dB}$, while for "different roots" it is around $-8\,\mathrm{dB}$, with "two roots plus shifts" it is inbetween at $-10\,\mathrm{dB}$. Around 20% of all users have a MSE even larger than $0\,\mathrm{dB}$, thus the estimation error magnitudes in these cases already exceed the channel coefficient magnitudes. Clearly, the inter-cell interference drastically impacts the LS approach.

The small-window approximation of the joint MWF ("small window joint MWF") deals much better with inter-cell interference than LS. With "different shifts" it is roughly speaking as good as the full separate Wiener filter with "different roots", with a median MSE of $-15\,\mathrm{dB}$. With "different roots" the MSE just goes up slightly to $-12\,\mathrm{dB}$ (in case of "two roots plus shifts" it is $-13\,\mathrm{dB}$. Of course, due to the limited window size, the estimation cannot be built upon that many pilots as for the full window sizes, but it can deal comparatively well with inter-cell interference, and - discussed in further detail in the next section - it requires much less statistical parameter knowledge as the full MWF.

Figure 3.11 now compares additionally different sizes of CoMP clusters with different N_{supp}. The performance metric is the average output SINR (3.50) at the IRC receiver (3.48). This means, in addition to Fig. 3.10, the impact of the channel estimation on the receive combiner, using IRC interference suppression, is taken into account. As the total level of interference increases with number of cells, the combiner output SINR decreases likewise, as can be observed from the Fig. 3.11. The performance with perfect channel knowledge serves as a benchmark and upper bound. Joint full MWF performs in all 3 different sequence assignment cases similarly well and has only small SINR gap versus perfect CSI. This gap is in the order of 0.3 dB for larger cluster sizes.

Performance of separate Wiener filter, in case of "different shifts", thus orthogonal pilot sequences, is not far from joint full MWF in terms of performance. It loses around half a dB to joint full MWF for $N_{\mathrm{supp}} = 7$. In case of "different roots" (or even "2 roots with different shifts", the losses are much larger, as this channel estimator is not tailored to handling the cross-talk from different users' sequences.

The small window joint MWF is more stable in performance. In case of "different roots" it outperforms the separate Wiener filter solution due to its better interference-handling capabilities, while for "different shifts" it is weaker because the estimated channel of one data resource element is relying on less pilot symbols (as the auto-covariance is based on a flat channel approximation (3.59)).

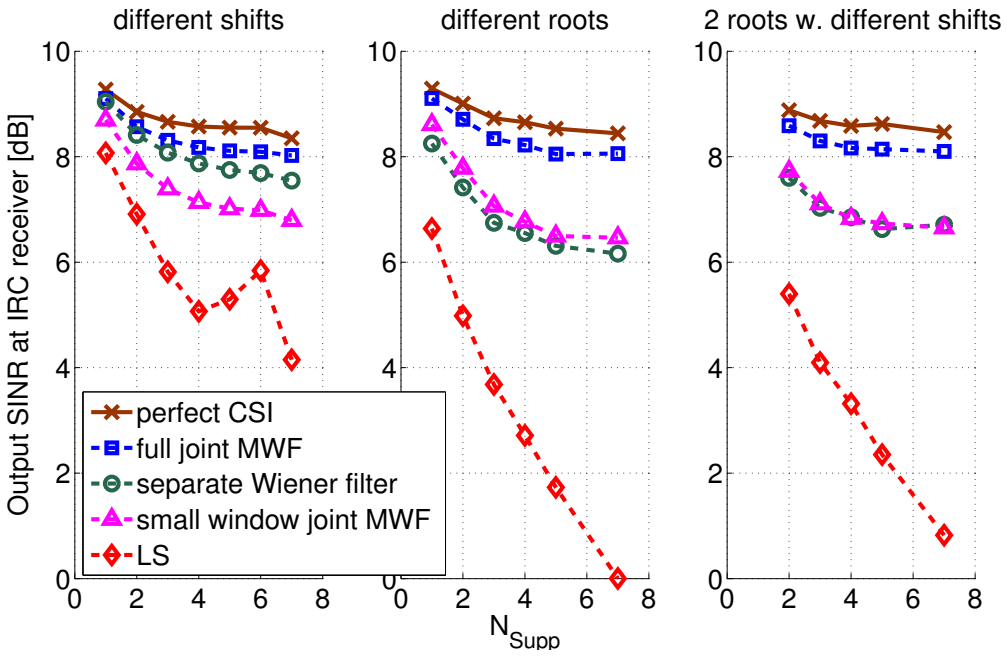

Figure 3.11.: SINR at the output of an IRC combiner as a function of N_{supp} with input SNR 10 dB. Pilot sequence assignments for different users use different sequence shifts (left), different sequence roots (mid), two different roots, each with different shifts (right).

Least squares in general performs poor, except for small N_{supp} and orthogonal sequences ("different shifts"). The performance behavior for "different shifts" shows a remarkable "zig-zag shape". This is no simulation artifact, but comes from the residual cross-correlation of the sequences when observed with a particular receiver window. While the sequences themselves are orthogonal, their truncated form is likely to be not, when processed with a sliding window of a certain size. For $N_{\text{supp}} = 4$ with 5 different shifts the sequence cross-correlations for the LS window size of $L = 7$ are particularly high. The same holds for $N_{\text{supp}} = 7$ with 8 shifts. Both numbers of shifts thus create those downward spikes in the output SINR of Fig. 3.11 (left subplot). On the contrary, for $N_{\text{supp}} = 6$, for 7 shifts the cross correlation is particularly low (providing even perfect orthogonality, except allocation edge effects). Here the observation window matches best to the sequence properties, leading to this upward spike.

Figure 3.12 shows for a given large cluster size of $N_{\text{supp}} = 7$ the post-combining SINR as a function of IoT. The serving cell path gain remains constant, equal to 1, and thus the SIR, when adding more and more cells, as mentioned above, saturates around $2.3\,\text{dB}$. The variation in IoT is here achieved by simulating different operation points of σ_n^2. In case of perfect channel knowledge, we can observe that the output SINR, asymptotically shows a linear increase with increasing IoT. The Gaussian noise power, limiting the performance of an antenna system which is capable of interference suppression by the IRC algorithm, linearly decreases with increasing IoT. We can further observe that a joint full MWF almost achieves the same performance as perfect CSI with small losses in high IoT operation points.

When a small window joint MWF is used, a notable gap can be observed and the output SINR as

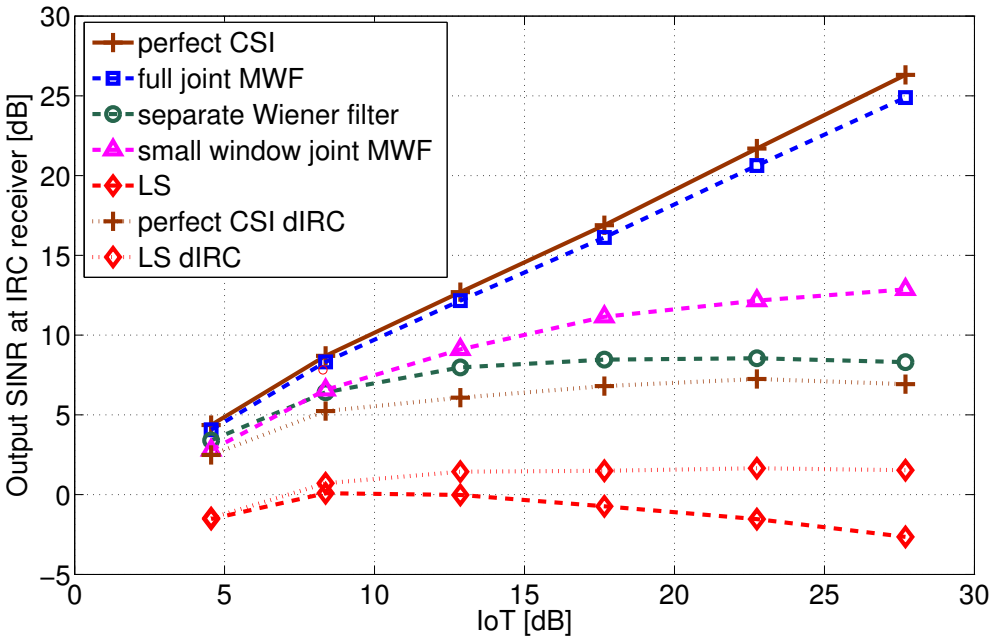

Figure 3.12.: SINR at the output of an IRC combiner over IoT (3.44), by varying σ_n^2, for $N_{\text{supp}} = 7$. Pilot sequence assignments use different sequence roots per user. The simplified dIRC combiner, according to (3.51), is shown in dotted lines.

a function of IoT begins to flatten for larger IoT values.The number of exploited pilots is smaller, due to the necessary small window. This leads to the effect that the inter-cell interference cannot be perfectly suppressed in the channel estimator and thus the resulting channel estimating errors lead to mismatched receive combining weights, limiting the output SINR.

The separate Wiener filter algorithm here already saturates earlier, as it fully lacks built-in interference-suppression capabilities.

Least squares algorithm in this figure performs drastically poor, even losing performance when the IoT increases. This is because of missing interference suppression capabilities. The channel estimation errors here lead to a poorly conditioned interference matrix $\mathbf{Z}_{jkl}$ in (3.48), which would benefit with added robustness by the "diagonal loading" effect of higher values of σ_n^2. The simplified IRC combiner (3.51) is also depicted. By ignoring the off-diagonal elements, channel estimation algorithms which are vulnerable to interference (mainly LS) benefit from this more robust combining approach. As dIRC also lacks full interference suppression capabilities, the output SINR even flattens out with perfect CSI, as we can see from the brown dotted curve.

Figure 3.13 shows the output SINR over varying antenna-input SIR. This variation is achieved by modifying the supporting cell cross-talk parameter α_0. The SNR is kept constant at $10\,\text{dB}$. With $N_{\text{supp}} = 7$ again a large CoMP cluster size is considered. Looking at the performance curve for perfect CSI, a characteristic with remarkable inflection points can be observed. This is also confirmed in the theoretical computations in appendix C for two users. The curvatures reflect the trade-offs happening in the interference suppression of the IRC receiver. At low input SIR, the

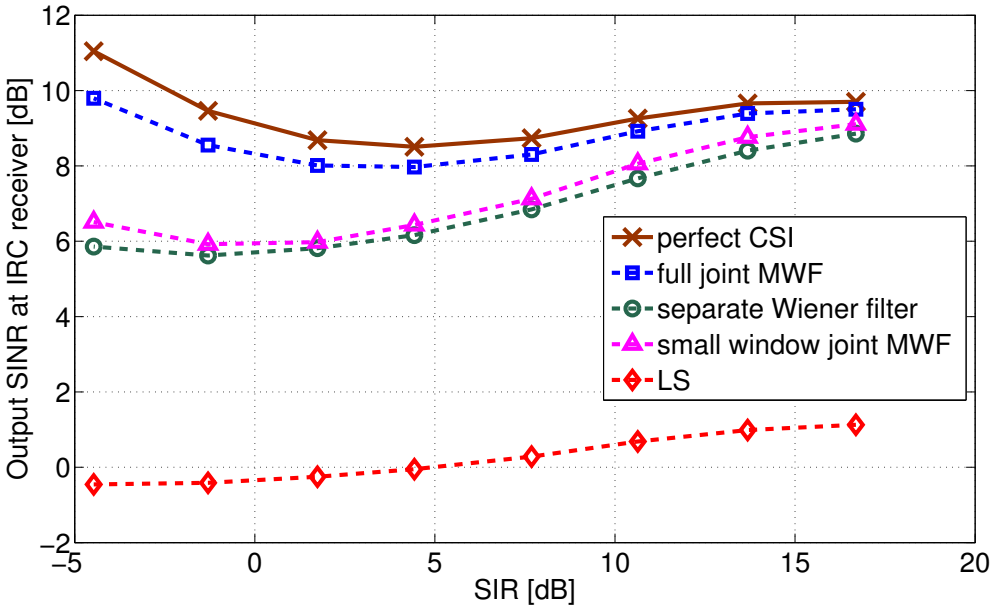

Figure 3.13.: SINR at the output of an IRC combiner over SIR (3.45), by varying α_0 in (3.46), for $N_{\text{supp}} = 7$ and input SNR fixed to 10 dB. Pilot sequence assignments use different sequence roots per user.

level of interference is high. The receive combiner has enough degrees of freedom to deal with interference, because the number of transmit antennas equals the number of receive antennas in the considered scenario. Hence, this inter-cell interference can be efficiently suppressed at this operation point. When the input SIR is increased in conjunction with constant input SNR, the relative share of noise, which cannot be suppressed, gets larger. The performance results of the different channel estimators also contain the characteristic shape of the curve for perfect CSI, Fig. C.1 in appendix C. The algorithms show similar relative performance ranking as in the previously shown figures. For large SIR, the MMSE-type estimators get close to the performance for perfect CSI.

3.3.4. Conclusion on Sequence Assignment and Estimation Strategies

When the per-user pilot sequences across the multiple cells of a CoMP cluster are separated by different roots, this is the LTE-A release 10 standard compliant case. In macro cell propagation environments, the typical delay spreads cause coherence bandwidths clearly smaller than the allocation size of e.g. 5PRBs, see section 2.1.3. Hence, correlation properties over sequence subsections, located in frequency subportions of a few subcarriers are more important than the typically considered ones over entire sequences[29]. Fig. 3.9 shows that, in this sequence subsec-

[29]This fact might change with a different multi-carrier numerology, compared to LTE, e.g. with denser subcarrier placements.

tion case, sequence correlations are similar to random phase sequences. This means that in our considered multi-user scenario there will be cross-talk between the sequences at the receiver-side. The effect of this cross-talk on the SINR at the output of an IRC receive combiner can be observed in the figures 3.11, 3.12 and 3.13. The matrix Wiener filter (3.56), which jointly estimates the channel for all users, as it exploits the full correlation properties across all pilot symbols of an entire allocation can handle this cross-talk very well and performs close to perfect channel knowledge, even for large CoMP cluster sizes (of e.g. 7 cells). As it requires substantial parameter knowledge, to be discussed in the subsequent section 3.4, a simplified joint MWF solution, based on a small sliding window processing for estimating within channel coherence bandwidth, is introduced which requires much simpler parameter knowledge. It can deal rather well with inter-cell interference, but loses significantly in performance compared to perfect channel knowledge and full joint MWF, when CoMP cluster sizes become larger. The reason is that the channel estimate for one data symbol is based just on a few pilot symbols, so that the narrowband approximation holds.

In case of different roots per user, the separate per-user estimation approaches perform poor, as they do not explicitly treat the multi-user (and thus multi-cell) interference. The separate Wiener filter (3.53) might still be considered for smaller cluster sizes, but separate least squares (using sliding windows) is not recommended. It fails due to missing interference suppression and small window sizes (for the narrowband approximation (3.11)).

The different alternative options for sequence assignment are discussed in section 3.3.1.3. A clear strategy to improve performance is to aim for orthogonal sequences over the cells while still preserving the CAZAC property of the sequences. This is achieved by using Zadoff-Chu sequences with different shifts, an approach which could considered in possible LTE-A standard extensions. The full joint MWF works already very well with different roots and thus benefits only slightly from this change. All other estimators clearly gain from using different shifts, especially the separate ones. Separate Wiener filter becomes almost as good as joint MWF. LS is still not recommended, besides in very small cluster sizes with modest performance.

As using different shifts is difficult to realize in practice over the area of cells and CoMP clusters, combinations of shifts and roots are investigated as well. Already when two different roots are combined with different shifts, we observe clear performance degradations for the separate estimators.

In total, the conclusion is that a full joint MWF is required to have high performance under all conditions. In this case, sequence assignment eases, as a separation by different roots performs nearly as good as any other (e.g. orthogonal) approach. As we have seen in the single user case in section 3.2.4, the extraction of required parameter knowledge is a difficult issue which cannot be neglected. So this question in the multi-user scenario is discussed in the subsequent section 3.4.

3.4. Parameter Estimation for Multi-User Multi-Cell CoMP

Section 3.2 discussed the CoMP channel estimation problem in the single user setting. It showed that parameter knowledge (like time-frequency-covariance matrices of the channel) plays a very important role for the performance of the estimator; e.g. when the time for collecting statistical data on propagation channels of particular users is short. The previous section 3.3 introduced the multi-user setting. As the focus was on pilot sequence assignments (more or less orthogonal ones vs. non-orthogonal and different channel estimation approaches (joint estimators vs. separate ones), the open question on parameter knowledge in the multi-user case was set aside. This will now be discussed within this section.

Let us first recapture the motivation and challenge from a cellular system level perspective. In case the transmitting set of users in a CoMP cluster remains static for a long time, then the parameter estimation becomes easy, e.g. simply using (3.23), and asymptotically achieves the situation of perfect knowledge. This might be the reason that the question of how to get access to the required statistical parameters is ignored in the vast majority of publications on channel estimation. With CoMP this issue cannot be ignored: Actual data traffic conditions and the usage of frequency selective scheduling will lead to situation that - for a given subband of interest - the set of transmitting users in our CoMP cluster will quickly change, e.g. may be different each 1 ms subframe. Furthermore, for a user which previously has been inactive for a certain time, no prior knowledge on the channel parameters can be assumed, and so they have to be extracted from the received samples. Practical channel estimators for CoMP will face the challenge that there often may be much less number of subframe observations available than the length of the pilot sequence is, spanning the pilot auto-covariance $\mathbf{\Phi}_i$.

The objective of this section is to provide a practical estimator for CoMP channel estimation, including also the estimation of the related parameters under the above described difficult conditions. The results of this section have been published in [WNtB13]. While most academic work treats the statistical parameters as perfectly known, related work [CMSH05] of the past already dealt with the issue of estimating the parameters from the receive signal for MMSE channel estimation. [CMSH05] focuses on a single-user MIMO scenario and models the channels as block flat fading, thus constant over a certain time-frequency chunk. In the work of this section, a multi-user CoMP scenario with time-variant frequency-selective channel is adressed. Thus, certain simplifications and assumptions done in [CMSH05] do not hold here.

3.4.1. Scenario and Parameters of Interest

The multi-user scenario is identical to the one in 3.3.1, illustrated by Fig. 3.8. The pilot sequence assignment in this section 3.4 uses the LTE / LTE-A standard-compliant variant with Zadoff-Chu sequences with different roots per user, explained in 3.3.1.3. We have learned from the results of the previous subsection that a MWF delivers the best performance. Furthermore this performance is almost the same for the different investigated sequence assignment variants. So this choice of different roots per user, being the standardized one and easiest to execute, is of general value here. It also accounts for the worst case of the considered sequence assignment

strategies, as the simpler channel estimation algorithms, like least squares, are weaker in this non-orthogonal pilot sequence setting.

The difference of this section 3.4 to the previous 3.3 is that the statistical parameters required for the channel estimation process now are not perfectly known, but are estimated from a limited amount of receive signal samples. The parameters of interest are:

- σ_n^2, the noise variance
- The auto-covariance matrix $\mathbf{\Phi}_i$ and the per-user cross-covariance vectors $\mathbf{r}_{ijkl}$. From (3.57) and (3.22), by the fact that the pilot sequences $\mathbf{S}_j$ are known and the noise variance σ_n^2 is estimated in the above step, this results in getting access to the user-individual channel auto covariance matrices $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ (as $\mathbf{\Phi}_i$ and $\mathbf{r}_{ijkl}$ can be synthesized when the $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ of the involved users are known).
- Effective path loss a_{ij}: From the simplified small-window MWF in (3.61), we have seen that a_{ij} is an additional important helper parameter.

With those parameters, the weights of the linear joint (thus multi-user) channel estimator of section 3.3.2.2 can be computed. With the obtained channel knowledge, the distributed multi-antenna receive combining with interference suppression (IRC) can be carried out, as discussed in section 3.3.1.2.

3.4.2. General Approach for Parameter and Channel Estimation

In statistics, when states and parameters jointly have to be estimated, a method to find maximum likelihood or maximum a posterioiri estimates of parameters is the expectation-maximization (EM) algorithm [DLR77]. EM is an iterative approach. An example where EM was used for jointly estimating channel and noise variance is [ZMH10]. The Bayesian extension of the EM approach are so-called Variational Bayesian methods [Mac03]. Those advanced methods are not of practical relevance in our specific problem of obtaining a practical channel estimator. Note that the dimensionality of our parameters is very large, e.g. for each user involved we face a 120×120 channel covariance matrix $\mathbf{R}_{ij}^{[\mathrm{hh}]}$ (compared e.g. to the one-dimensional parameter required in [ZMH10]) and so for complexity reasons a simpler approach is required.

The strategy pursued in this thesis for practical channel and parameter estimation is a multi-stage approach, as depicted in Fig. 3.14:

- In the previous section 3.3, we have seen that the full joint MWF delivers the best performance. This shall be the *final stage* of our multi-stage estimator.
- The output of the *second last stage* has to be the the channel covariances $\mathbf{R}_{ij}^{[\mathrm{hh}]} = \mathrm{E}\left\{\mathbf{h}_{ij}^{[P]}\mathbf{h}_{ij}^{[\mathrm{P}]\,H}\right\}$, required in the final stage. We use actual sample channel estimates $\hat{\mathbf{h}}_{ij}^{[P]}$ based on a simpler estimation approach (with less required parameters) together with past observations for computing the covariances as $\hat{\mathbf{R}}_{ij}^{[\mathrm{hh}]} = \mathrm{E}\left\{\hat{\mathbf{h}}_{ij}^{[P]}\hat{\mathbf{h}}_{ij}^{[\mathrm{P}]\,H}\right\}$. This basic strategy is similar to one of the approaches considered in [CMSH05], but more challenging. As this task happens, in contrast to [CMSH05], under a strong multi-user

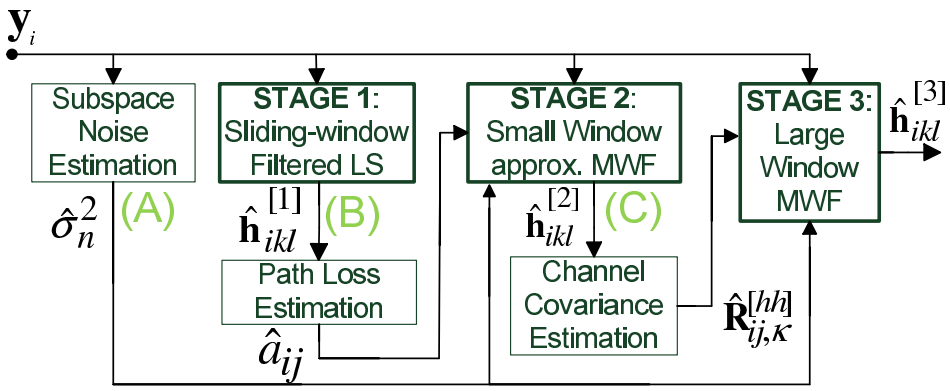

Figure 3.14.: Block diagram of multi-stage channel- and parameter estimation approach.

multi-cell interference situation with time-variant frequency-selective channels, the simpler estimator already requires good interference suppression capabilities. From the results in 3.3, we have seen that the small window approximation of the MWF (3.62) has this interference suppression feature and requires just one scalar parameter per user: a_{ij}. Additionally σ_n^2 is needed, which is anyway required in our estimation process for IRC and full MWF.

- Thus, in our third last or *first stage*, we need an estimate of the receive power level of the signal. An additional prerequisite is a noise variance estimate.

This means, three stages are used here to get the final channel estimates. Fig. 3.14 illustrates this approach. As the effect of *error propagation* has to be carefully assessed due to imperfections within the multiple stages, Fig. 3.14 contains three reference points (A), (B) and (C). For those points, either the actual estimator output from its preceding block is used, or a reference output, based on perfect knowledge. This reference output is used in the simulations as a benchmark for determining error propagation effects. It is denoted "perf. A/B", when the outputs at points (A) and (B) are replaced by outputs generated with perfect knowledge, and likewise "perf. A/C" with perfect knowledge for points (A) and (C).

3.4.3. Noise Variance Estimation in a Multi-User Environment

In section 3.2.2.4, the noise estimation problem for CoMP in a single-user scenario was discussed. The conclusion was that simple filtering-based approaches are the best choice for extremely low SNR. For other SNR operation points they perform similar to subspace-based noise estimation methods. Unfortunately, these filtering-based approaches cannot be used reasonably in the multi-user setting. They compare the difference of unfiltered raw channel estimates with filtered ones and infer on the noise variance from that. This works well in the presence of pure noise. In the multi-user setting, with interference from other cells, this approach cannot distinguish noise from interference. But this distinction is very important, as we want to synthesize the

auto-covariance from individual interference contributions and the noise. This synthesizing is required for coping with the dynamically changing user allocations, thus a non-static combination of users. Here, the subspace-based noise estimation is clearly powerful, as it attributes particular eigenvalues of the receive auto-covariance to noise, due to its particular structure (3.34).

The entire discussion of the single user parameter estimation in section 3.2.2 is centered around a low SNR operation point (either from a coverage-limited perspective and/or the simplification of treating a mix of many weak interfering users as noise). Here, the required number of subframes for observing channel statistics, the"warm-up" times, have shown to be very long, as e.g. in Figure 3.7.

3.4.3.1. Issues with Basic Sub-Space Methods

In a multi-user multi-cell scenario, the focus shifts towards a higher SNR operation point, as such cellular systems are naturally operated interference-limited. The SNR operation point is e.g. 10 dB, which section 3.3 has used as a realistic typical LTE-A value for urban macro cells. This allows for shorter "warm-up" times.

These desired shorter "warm-up" times are causing problems to the subspace-based methods, discussed in section 3.2.2.4. The problem is that one has to rely on the observation of the sample auto-covariance (3.23) in order to assess its structure (3.34), using information-theoretic methods like MDL (3.37). Using our typical allocation size example of 5 PRBs, each slot contains 60 allocated subcarriers and thus $P = 120$ pilot symbols. Thus a 120×120 sample auto-covariance is used. In order to have a matrix with full rank[30], at least 120 subframes are required as observation samples. The restriction to practical schedulers would be too severe to constrain them allocating the same set of users in the same frequency band across all involved cells in the CoMP cluster for such a long time. Problems occur when the users have not enough data in their buffer to occupy the radio resources for such a long time. Furthermore, fast fading might result in the fact that the currently used frequency band becomes more and more disfavorable and thus allocating the user to another subband might be advantageous and demanding. So this thesis has to address a solution which is working with much shorter warm-up times.

3.4.3.2. Improved Subspace-based Methods with Warm-Up Time Reduction

In a first simple step, one can exploit the structure of the LTE-A pilot pattern and reduce the dimension of the considered covariance, as the two slots within a subframe have identical relative pilot positions, as all subcarriers located on the center symbol of the slot. We use the $\frac{P}{2}$-sized pilot observation vector of each slot, thus the respective $\frac{P}{2}$ elements of $\mathbf{y}_i$, as a sample (instead of processing entire subframes). We use the two subsequent slots, contained in one subframe, just as two different samples of this pure frequency-direction auto-covariance. This reduces the matrix dimension by a factor of 2 and increases the number of samples per subframe by a factor of 2 (with one sample per slot). For the above example of 5 PRB allocation size, we now consider

[30] Without full rank, the eigenvalue decomposition cannot be done.

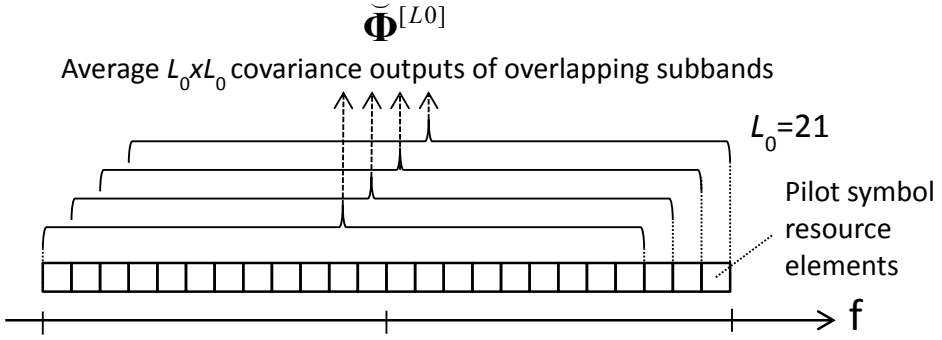

Figure 3.15.: Exploiting frequency shift-invariance of channel covariance by sliding reduced observation window in overlapping subbands for increased sample output. Example for $P/2 = 24$, $L_0 = 21$, and thus 4 shifts.

a 60×60 matrix (instead of 120×120), requiring at least 60 slots, thus 30 subframes[31]. Even though the savings can be seen as significant, this number is still too high.

In array-processing, a technique called *spatial smoothing* [KV96] is known. Here, an antenna array is split into overlapping subarrays, and the covariances of those different subarrays are averaged, assuming the steering vectors remain unchanged. This can be seen like a "sliding window over the antennas". For our problem of CoMP channel estimation we are using a similar trick: The pilot sequences in LTE-A uplink are contiguously stretched across the different subcarriers. We now split this allocation into overlapping subsets and average the covariance over those subsets. A nice property of the radio channel is its frequency-shift invariant covariance, as shown in (2.12). Thus we use a "sliding window over the subcarriers", as depicted in Fig. 3.15.

Instead of executing the subspace estimation, based on a observation window of $\frac{P}{2}$ symbols, reduced observation windows $\breve{\mathbf{y}}_{\kappa l}^{[L0]}$ of size L_0 are used, providing an increase of factor $\frac{P}{2} - L_0 + 1$ in terms of sample output. The windows are indexed l for the frequency position at slot number κ. For the examples of $\frac{P}{2} = 60$ and $L_0 = 55$, the sample output is 6-fold, reducing the minimum warm-up for a full rank matrix to 5 subframes. For $L_0 = 40$, with 21-fold sample output, a single subframe is sufficient to obtain all the samples for a full rank result[32].

The computation for this reduced-dimension sample covariance using N slots, carrying out the sliding window shifts, formed out of overlapping subbands, can be expressed as

$$\breve{\breve{\mathbf{\Phi}}}^{[L0]} = \frac{1}{N(\frac{P}{2} - L_0 + 1)} \sum_{\kappa=1}^{N} \sum_{l=1}^{\frac{P}{2}-L_0+1} \breve{\mathbf{y}}_{\kappa l}^{[L0]} \breve{\mathbf{y}}_{\kappa l}^{[L0]\,H}. \tag{3.65}$$

[31] Note that those steps imply two concatenated 1-D estimation steps, instead of a full 2-D estimation in time and space.

[32] While the minimum requirement for subspace noise estimation is a full rank matrix, as the true $\mathbf{\Phi}_i^{\mathrm{Freq}}$ in (3.35) is of full rank due to the noise, the accuracy of the estimation increases with increased number of samples, which is shown in 3.4.5.

For estimating the noise variance, the subsequent processing steps are identical to the single-user case in section 3.2.2.4: Eigenvalue decomposition of (3.65) (instead of (3.23)), computation of the MDL criterion (3.37), where the minimum value of the model order hypothesis provides the discrimination between signal/interference subspace and noise subspace.

3.4.4. Multi-Stage Channel Estimation

Fig. 3.14 illustrates the selected multi-stage approach which will be described in detail below. The output variables of stage 1, 2, 3 are denoted $\hat{\mathbf{h}}_{ikl}^{[1]}$, $\hat{\mathbf{h}}_{ikl}^{[2]}$, $\hat{\mathbf{h}}_{ikl}^{[3]}$ with dimension $J \times 1$, J being the number of transmitting antennas. As discussed above, the points (A), (B), (C) serve for intermediate check results when using (A) perfect knowledge of noise variance $\hat{\sigma_n}^2 = \sigma_n^2$, (B) perfect channel knowledge as input for path loss estimation $\hat{\mathbf{h}}_{ikl}^{[1]} = \mathbf{h}_{ikl}^{[1]}$ or (C) perfect channel knowledge as input for covariance estimation $\hat{\mathbf{h}}_{ikl}^{[2]} = \mathbf{h}_{ikl}^{[2]}$.

3.4.4.1. Remark on Parameter Tracking

As in previous sections, all actual channel estimation happens within a single subframe (1 ms). All statistical parameters taken from previous observations can simply be averaged or tracked with a first order infinite impulse response (IIR) low pass filter with a certain forgetting factor ρ_{IIR} tuned on the fastest variation of the large scale parameters (shadowing and scattering environment) which shall be supported.

Parameter tracking, e.g. for effective path loss a_{ij} can be done in the following way: Any current parameter measurement result $C(\kappa)$, together with the previous low-pass filtered result $\bar{C}(\kappa-1)$, provides the current low pass filter output

$$\bar{C}(\kappa) = (1 - \rho_{IIR})C(\kappa) + \rho_{IIR}\bar{C}(\kappa - 1). \tag{3.66}$$

Channel measurements [WHHK13] for low velocities indicate a high stability of large scale parameters in rather large time scales, such as 500 ms, indicating the large timescales which can be addressed by large (near one) values of the parameter ρ_{IIR}. For this reason, actual large scale parameter tracking, once initial channel statistics is obtained, is not considered to be an issue and thus is not further investigated in detail in this thesis. Beyond this first order IIR approach (3.66), Kalman filters could be considered here as well. The main focus of the thesis here lies on initial parameter estimation, when no prior knowledge is available, e.g. when the user has not been active for a longer time.

3.4.4.2. Stage 1 for Path Loss Estimation

The first stage does not use prior knowledge. The idea is to get a coarse channel estimation per user from which the effective path loss a_{ij} can be extracted. The self-evident basic approach here is the single user least squares method described in 3.2.1.1, using (3.14). While we have learned that the accuracy of the channel estimates in these case is not satisfying for CoMP, we will see

in section 3.4.5 that it is accurate enough to provide the required coarse knowledge on path loss for the second stage, without inducing notable error propagation through the different stages.

For a particular user j, its stage one output by (3.14), is

$$\hat{h}_{ijkl}^{[1]} = \frac{1}{L} \breve{\mathbf{s}}_{jl}^{H} \breve{\mathbf{y}}_{ikl} \tag{3.67}$$

at frequency position l in the center of the observation window $\breve{\mathbf{y}}_{ikl}$, with the appropriate cut-out of the pilot sequence $\breve{\mathbf{s}}_{jl}$.

The path loss estimate then is simply

$$\hat{a}_{ij} = \mathrm{E}\left\{ \hat{h}_{ijkl}^{[1]} \right\} \tag{3.68}$$

where the expectation is carried out by averaging over time and frequency indices k, l of the current and past observations of $\hat{h}_{ijkl}^{[1]}$. (For tracking of changing path loss conditions, (3.66) can be used.)

3.4.4.3. Stage 2 for Initial Channel Estimation

Our full MWF of stage 3 requires the channel covariance matrix for each user. One could already use the channel estimates of the first stage to compute this channel covariance, but the strong multi-cell interference, which is ignored in the LS approach of stage 1 would provide a flawed result, see simulation results in section 3.3.3. Stage 1 is just good enough for path loss estimation. Thus an estimator which can deal with this multi-user multi-cell situation is required. We have seen from section 3.3 that the small-window approximation of the MWF (3.63) works already fine. Its required scalar parameters come from stage 1, the effective path loss a_{ij} and from the subspace noise estimation.

In contrast to (3.60), the weight computation in this section now does not use perfectly known a_{ij} and σ_n^2, but their estimates in order to compute the small-window auto-covariance:

$$\hat{\breve{\mathbf{\Phi}}}_{il,[L2]} = \sum_{j=0}^{N_{\mathrm{supp}}} \hat{a}_{ij} \breve{\mathbf{S}}_{jl}^{[L2]} \vec{\mathbf{1}} \vec{\mathbf{1}}^{T} \breve{\mathbf{S}}_{jl}^{[L2]\,H} + \hat{\sigma}_n^2 \mathbf{I}. \tag{3.69}$$

In the same way, the cross-covariance estimate, for smoothing at the pilot positions, becomes:

$$\hat{\breve{\mathbf{R}}}_{ikl}^{[yh]} = \left[\hat{a}_{i0} \breve{\mathbf{S}}_{0l}^{[L]} \vec{\mathbf{1}}, ..., \hat{a}_{iN_{\mathrm{supp}}} \breve{\mathbf{S}}_{N_{\mathrm{supp}}l}^{[L]} \vec{\mathbf{1}} \right] \tag{3.70}$$

With those parameter estimates, the small-window MWF estimator weights follow from (3.63) as $\hat{\breve{\mathbf{W}}}_{ikl}^{[L]} = \hat{\breve{\mathbf{\Phi}}}_{il,[L2]}^{-1} \hat{\breve{\mathbf{R}}}^{[yh]}$ and the channel estimates are

$$\hat{\mathbf{h}}_{ikl}^{[2]} = \hat{\breve{\mathbf{W}}}_{ikl}^{H} \breve{\mathbf{y}}_{i} \tag{3.71}$$

where $\hat{\mathbf{h}}_{ikl}^{[2]}$ denotes the stage 2 output vector in user dimension $J \times 1$.

3.4.4.4. Stage 3 for Final Channel Estimation

The channel estimation for the final stage should be a matrix Wiener filter (3.56) which builds upon the entire user allocation, instead of using a small window as in stage 2. This is because we have seen from the simulation results in section 3.3.3 that the small window MWF almost loses 2 dB in post-combining SINR, compared to the full MWF, when the CoMP cluster sizes become large, e.g. with $N_{\text{supp}} = 7$ supporting cells.

The receive weights (3.56), which in section 3.3 were known, based on perfect statistical knowledge, now have to be computed based on estimated parameters:

$$\hat{\mathbf{W}}_{ikl} = \hat{\mathbf{\Phi}}_i^{-1} \hat{\mathbf{R}}_{ikl} \tag{3.72}$$

Due to our design target w.r.t. suitability for dynamically changing user constellations, it makes no sense to get the auto-covariance $\hat{\mathbf{\Phi}}_i$ in a conventional way. This conventional approach observes a particular static set of transmitting users, based on the pilot observation vector $\mathbf{y}_i(\kappa)$ over many slots, indexed κ and computes the sample covariance for it. Optionally it improves the covariance estimate with shrinkage, as introduced in section 3.2.2.2). As this static user set is not given in practice, $\hat{\mathbf{\Phi}}_i$ has to be obtained differently, namely *per user*.

From (3.57) we see that by knowing the pilot sequences and using a noise variance estimate, as discussed in section 3.4.3, we can form a composite auto-covariance estimate

$$\hat{\mathbf{\Phi}}_i = \sum_{\forall j \in \mathcal{J}} \mathbf{S}_j \hat{\mathbf{R}}_{ij}^{[\text{hh}]} \mathbf{S}_j^H + \hat{\sigma}_n^2 \mathbf{I} \tag{3.73}$$

with the help of user-individual channel covariance estimates $\hat{\mathbf{R}}_{ij}^{[\text{hh}]}$. Those $\hat{\mathbf{R}}_{ij}^{[\text{hh}]}$ can now be explicitly obtained by using the stage 2 output $\hat{\mathbf{h}}_{ikl}^{[2]}$ for constructing $\hat{\mathbf{h}}_{ij}^{[P]}$, dimensioned $P \times 1$ over the pilot resource elements, as:

$$\hat{\mathbf{R}}_{ij}^{[\text{hh}]} = \mathrm{E}\left\{ \hat{\mathbf{h}}_{ij}^{[P]} \hat{\mathbf{h}}_{ij}^{[P]\,H} \right\} \tag{3.74}$$

Note that this way of obtaining the covariances has some similarities to the approach in [CMSH05], called "channel-correlation based MMSE". The mentioned differences in this thesis are the time-variant, frequency-selective channels and the high-interference multi-user CoMP scenario, requiring different algorithmic approaches.

Stage 3 also has particular "warm-up" issues due to the construction of $\hat{\mathbf{R}}_{ij}^{[\text{hh}]}$. As in the discussion for noise estimation in section 3.4.3, we like to have the receiver operational with reasonable performance in a very short time (e.g. a few subframes). So in stage 3 processing, we also require novel algorithmic enhancements in order to speed-up the extraction of $\hat{\mathbf{R}}_{ij}^{[\text{hh}]}$.

As first simple step, comparable to the noise estimation section 3.4.3, the structure of the LTE-A pilot pattern is exploited. As it is identical in terms of relative position within the two slots of a subframe, one can reduce the Wiener filtering dimension to the frequency direction by a factor of 2 and collect in the same way 2 samples per subframe, one per slot. As discussed for the single-user Wiener filter in section 3.2.1.2, this means that our 2-D (time-frequency) MWF now becomes a 1-D MWF smoothing in frequency direction. The time-direction can be processed

with either linear interpolation or 1-D temporal Wiener interpolation. As CoMP targets the low mobility case, the linear interpolation is fair enough, as discussed in 3.2.1.2.

Despite this reduction in required samples, further steps are necessary. The idea here is to again use a similar technique than spatial smoothing [KV96]. Instead of the full MWF over the entire band, just a large-window MWF is used, where the covariance is obtained from several overlapping subbands. Here the property of frequency-shift-invariance of the channel frequency covariance (2.12) is exploited, which allows for the averaging of overlapping subband-portions of the same size.

With a stage 3 (large window) size of L_3 instead of the full band size $\frac{P}{2}$, the number of generated samples becomes larger by a factor of $\frac{P}{2} - L_3 + 1$. We define the reduced dimension channel covariance for the transmission of antenna j to base station antenna i for the large window by

$$\hat{\breve{\mathbf{R}}}_{ij}^{[3]} = \begin{bmatrix} \hat{\breve{R}}_{ij}^{[3]}(1,1) & \dots & \hat{\breve{R}}_{ij}^{[3]}(1,L_3) \\ \vdots & \ddots & \vdots \\ \hat{\breve{R}}_{ij}^{[3]}(L_3,1) & \cdots & \hat{\breve{R}}_{ij}^{[3]}(L_3,L_3) \end{bmatrix}. \tag{3.75}$$

The elements m, n of this channel covariance are computed by "sliding the observation window in frequency direction" using $N_s = \frac{P}{2} - L_3 + 1$ number of shifts and averaging over all available N_t time samples[33]:

$$\hat{\breve{R}}_{ij}^{[3]}(m,n) = \frac{1}{(N_s+1)N_t} \sum_{k=1}^{N_t} \sum_{\mu=m}^{m+N_s} \hat{h}_{ijk\mu}^{[2]} \hat{h}_{ijk(\mu+m-n)}^{[2]\,*} \tag{3.76}$$

Running stage 3 with a large window MWF instead of the full window now can use similar processing steps: Synthesize the auto-covariances (3.77) with the large window matrices instead of the full matrices, as well as the weight computation (3.72) and actual channel estimation with reduced dimension.

The approach to avoid edge effects of the sliding window is as follows: The large window always uses its full size, and the window position is selected such that it is as centered as possible around the data resource element position of interest. With a data symbol subcarrier of interest at the allocation edge, still the entire large window can be used, as the available channel covariance is of dimension $L_3 \times L_3$, allowing to make use of $L_3 - 1$ neighbor subcarriers.

Finally, the stage 3 auto-covariance becomes

$$\hat{\mathbf{\Phi}}_{il,[L3]} = \sum_{\forall j \in \mathcal{J}} \breve{\mathbf{S}}_{jl}^{[L3]} \hat{\breve{\mathbf{R}}}_{ij}^{[3]} \breve{\mathbf{S}}_{jl}^{[L3]\,H} + \hat{\sigma}_n^2 \mathbf{I} \tag{3.77}$$

Cross-covariance estimation now follows section 3.2.2.3, based on the Jakes' model. The empirical correlation between different pilot-position channel estimates of different slots allow to estimate the maximum Doppler shift f_d. Using f_d in (2.11) provides the overall time-correlation behavior of the channel inbetweeen different pilot symbols for temporal Wiener inter-/extrapolation. Linear temporal interpolation is a valid alternative for low (or even medium) velocities.

[33] Note that this averaging "over all available time samples" should be limited to the duration where the large scale environment approximately remains constant, as done within the simulations in section 3.4.5, or replaced by the tracking approach of (3.66).

3.4.5. Simulation Results

Computer simulations have been carried out to compare the multi-stage approach with perfect channel knowledge, with a full MWF with perfect parameter knowledge (see section 3.3.2.2) and the stage 2 baseline receiver with only a small amount of parameter knowledge.

The simulation parameters follow the ones in 3.3.3: WINNER II spatial channel model [WIM06], with velocities of 3km/h, 1000 large-scale-parameter drops with a duration of 20 ms each, 5 PRB allocation size. For observing the impact of delay spreads, besides the urban macro channels (C2), additionally bad urban (C3) is used.

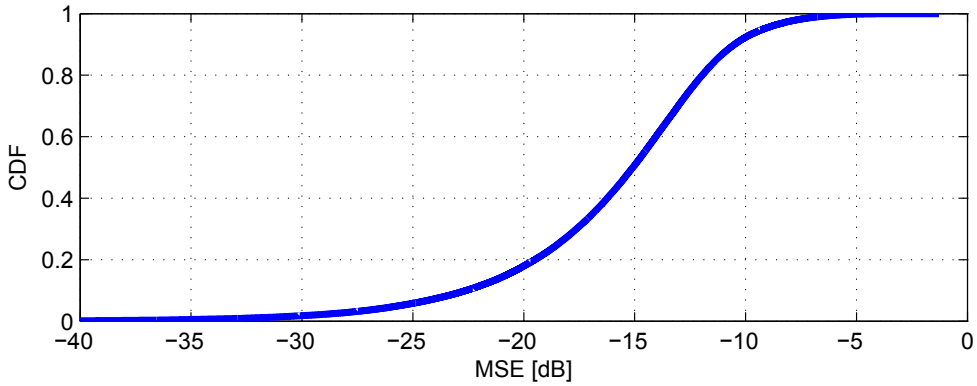

Figure 3.16.: Subspace noise variance estimation: Distribution of normalized mean squared error (3.78) for $L_0 = 55$ and WINNER II C2 Channel model.

First, the performance of subspace noise estimation is assessed. The normalized mean squared error of the subspace noise estimation is defined as

$$\mathrm{NMSE} = \frac{\mathrm{E}\left\{(\sigma_n^2 - \hat{\sigma}_n^2)^2\right\}}{\sigma_n^2}. \tag{3.78}$$

As Fig. 3.16 shows, the NMSE is low, its median is below -15dB. This demonstrates that the selected noise estimation solution is very powerful and can cope with the present mix of noise and interference.

The NMSE of the path loss estimation is defined as

$$\mathrm{NMSE}_{ij} = \frac{\mathrm{E}\left\{(\hat{a}_{ij} - a_{ij})^2\right\}}{a_{ij}}. \tag{3.79}$$

The path loss estimates, depicted in Fig. 3.17, show that the links of MS to their serving cell, $i = j$, can be estimated with much lower NMSE than the links to the supporting cells, $i \neq j$. The reason is the smaller link attenuation to the serving cells and thus the higher receive power levels, leading to more accurate results. In total this path loss estimation is not very accurate, but, as we will see from the subsequent stage 3 results, it is accurate enough to minimize error propagation to a negligible amount.

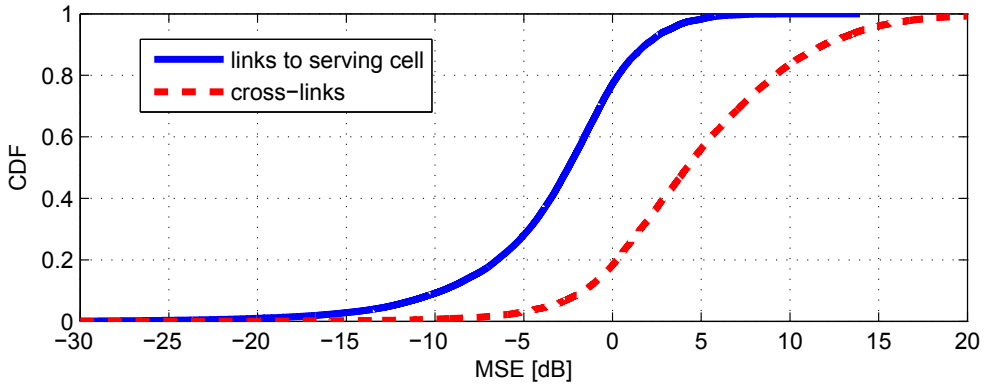

Figure 3.17.: Path loss estimation: Distribution of normalized mean squared error (3.79), using stage 1 output with $L_1 = 11$, $N_{\text{supp}} = 4$ and WINNER II C2 Channel model.

As stage 1 and 2 contain window sizes, which have to be chosen appropriately to the expected channel coherence bandwidth, in a next step the impact of those window sizes is investigated, shown by Fig. 3.18. The usual performance metric from previous sections, post-combining SINR, is chosen. A stage 1 window size of $L_1 = 11$ performs slightly better than $L_1 = 7$. Looking over the entire range of possible stage 2 window sizes L_2, one can observe that the maximum performance for high delay spread propagation channels (C3) occurs (for "perf. A/B", thus perfect stage 2 input parameters) around $L_2 = 13$, while for normal delay spread channels (C2) occurs at $L_2 = 19$. But: Over a very large range (from roughly $13 \leq L_2 \leq 25$, the performance does not vary much, just approximately 0.1 dB. The reason is that a trade-off occurs: A larger window offers strong interference suppression capabilities and more usable pilots for the estimation. On the other hand, the small-window approximation (3.59) only holds when the window size stays below the channel coherence bandwidth. The trade-off thus causes the effect that when increasing the window size beyond coherence bandwidth the small-window approximation does not hold and thus causes increased estimation errors while at the same time the increase of the window decreases the errors as more pilot symbols are exploited for one channel estimate. This stabilizes the performance and makes it in a certain range almost independent of the actual window size. This effect is very desirable for our multi-stage receiver, as it means that we do not have to fine-tune the window sizes, depending on propagation channel delay spreads. One value, e.g. $L_2 = 25$ is fair enough to suit all purposes. The losses for stage 2 compared to optimally adapted L_2 values are around 0.1 dB.

With the result from the window size tuning, an overall performance comparison is done, shown by Fig. 3.19, where the number of supporting cells N_{supp} is varied from 1 to 6. The performance gap between the different algorithms increases with increasing CoMP cluster size.

The top curve in solid brown gives the IRC performance in conjunction with perfect channel state information, thus an achievable upper limit in the absence of channel estimation errors. The bottom dash-dotted cyan curve represents the stage 2 output performance with estimated parameters. One can observe a difference of almost 2 dB for large CoMP clusters with $N_{supp} = 6$. In case the parameters noise and effective path loss are perfectly known ("perfect A/B", referring to the points in Fig. 3.14) the dashed green curve shows that the stage 2 output would benefit

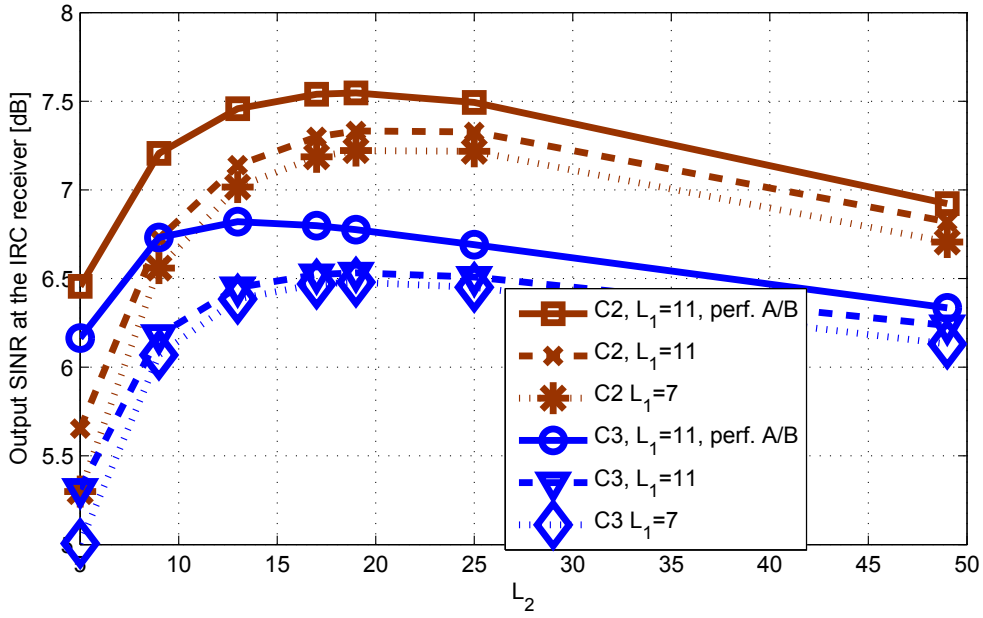

Figure 3.18.: Impact of window size L_2 on stage 2 performance: Output SINR at IRC combiner over L_2 for different L_1 and WINNER II Channel models C2 and C3 with $N_{supp} = 4$. "perf. A/B" means perfect knowledge at reference points (A) and (B) in Fig. 3.14.

by around half a dB compared to the estimated parameter case. As an upper limit for multi-user MMSE channel estimation by the full MWF, the blue curve shows the result when all parameters are perfectly known[34]. For $N_{supp} = 6$ this perfect full MWF is about 0.3 dB worse than with perfect CSI. As we have only a finite observation time and need to use a large window instead of the full window, the curve below in dashed orange, which is about 0.2 dB worse, gives the performance in case the channel covariance is determined based on the (infeasible) observation of perfect channel realizations (denoted "perfect A/C"). Thus the channel realizations are observed to obtain the channel covariance using (3.76) with N_t equaling an entire drop of perfect CSI as (non-causal) input. This provides a theoretical upper bound for our multi-stage approach. (But collecting an entire drop would introduce an intolerable processing delay, here 20 ms, which is not tolerable to do in practice, because of latency reasons.)

Our stage 3 output performance with estimated parameters, shown in dash-dotted magenta, is roughly "half way inbetween" the stage 2 performance and the full MWF with perfect parameter knowledge. In order to assess the impact of error propagation throughout the multiple stages, the stage 2 is fed with perfect parameter knowledge and used to generate the channel covariances for stage 3. This case is denoted "perfect A/B" and is shown in dotted black. We see that the error propagation just causes a loss of 0.1 dB for large cluster sizes $N_{supp} = 6$, vanishing for smaller cluster sizes.

Finally, Fig. 3.20 shows the impact of "warm-up time", thus the available amount of past ob-

[34]Building upon section 2.1.3, (3.20) provides the channel covariance based on perfect knowledge of the power-delay profile and the temporal fading auto-correlation.

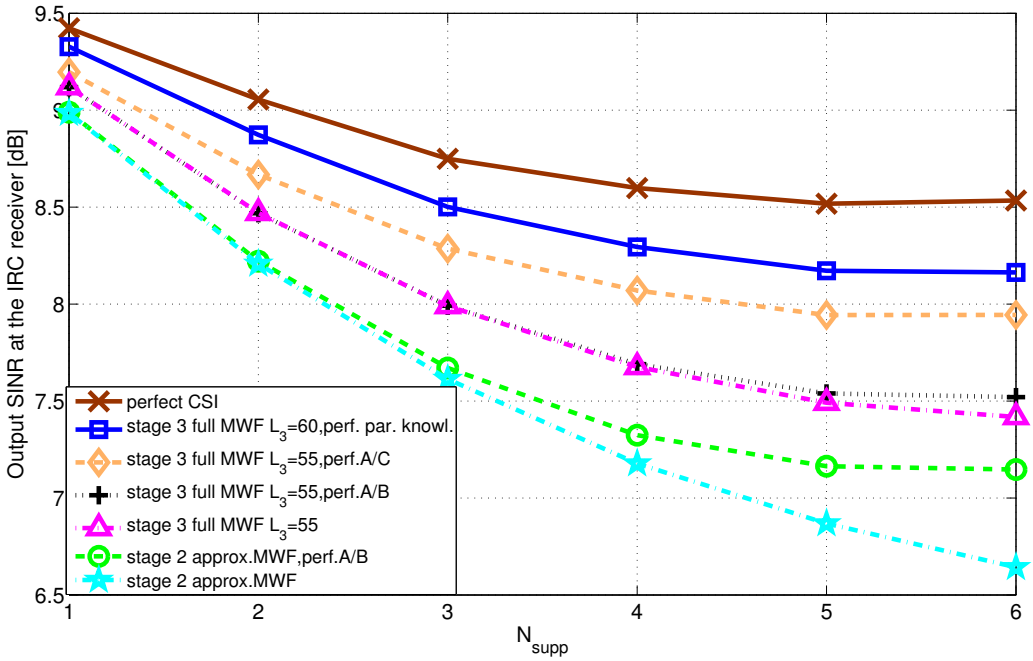

Figure 3.19.: Impact of coordination set sizes on estimator stages 2 and 3 (compared to perfect CSI and MWF with perfect parameter knowledge): SINR at the output of an IRC combiner over N_{supp}. WINNER II C2 Channel model, $L_0 = 55$, $L_1 = 11$, $L_2 = 25$, $L_3 = 55$, "warm-up" 20 TTIs.

servations (in number of subframes / TTIs), as well as the impact of the window size for the large window MWF. The curves for perfect channel knowledge or perfect parameter knowledge of course remain constant over warm-up time. The large-window MWF from stage 3 with estimated parameters demonstrates a very short warm-up time here. This means with only a few subframes of past observations, the channel estimator can already play out its full performance potential. This is a very important feature to be used in conjunction with a dynamic scheduler and dynamic (non-full-buffer) data traffic.

The warm-up times are comparatively shorter than the ones in the single-user case in section 3.2.3, Fig. 3.7. This comes from the fact that we have chosen for the multi-user setting an interference-limited SNR operation point, instead of the noise-limited operation point which is mainly adressed in 3.2.3. Low SNR will increase the time required for collecting statistics, while low SIR can be dealt with by appropriate estimators/receivers, introduced here in sections 3.3 and 3.4. Additionally, in the multi-user case we have exploited particular properties, discussed in section 3.4.3 and 3.4.4.4, for further bringing down the warm up time.

The impact of different window sizes is shown as well in Fig. 3.20. For the range of $50 \leq L_3 \leq 60$ no large differences are observed. $L_3 = 55$ performs best. $L_3 = 60$ comes close, especially with a large number of samples. $L_3 = 40$ is clearly weaker; here the reduced number of pilots for estimating the channel of a single data resource element costs performance.

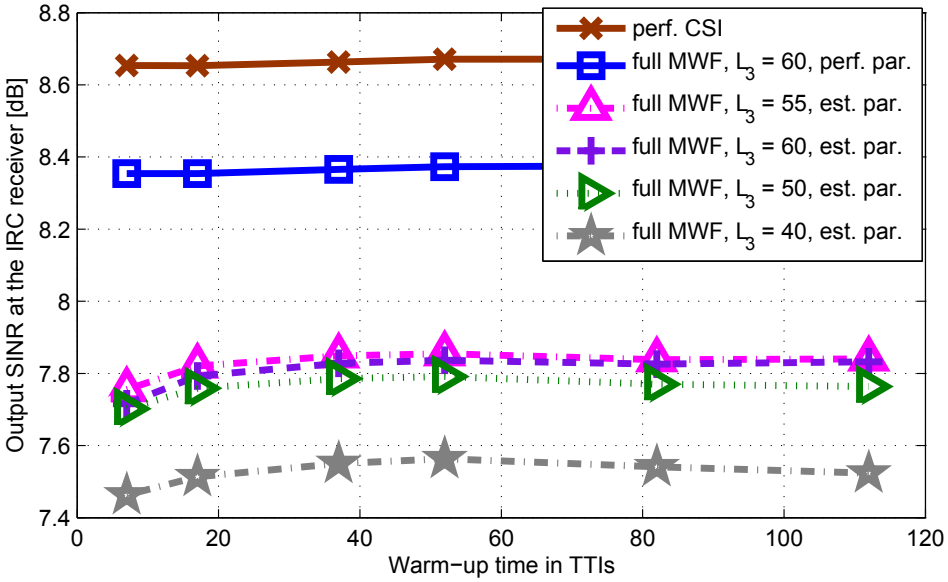

Figure 3.20.: Impact of "warm-up time" in number of TTIs/subframes on estimator stage 3: SINR at the output of an IRC combiner over number of subframe observations for $N_{supp} = 4$, $L_0 = 55$, $L_1 = 11$, $L_2 = 25$, $L_3 = 55$.

As Fig. 3.20 has shown, the warm-up times are very short with the multi-stage approach. As an outlook, an approach for further potential improvements is described in the patent [WRN13], including data-aided techniques and information exhcange via backhaul.

3.5. A Simplified Channel- and Estimator-Model for System Simulations

In the previous sections 3.2, 3.3 and 3.4 the channel estimation has been fully implemented on link level and assessed. In order to obtain the overall cellular system performance, system simulations are required. Typical assumptions are hexagonal playgrounds with e.g. 3 rings of in total 19 tri-sectored sites, thus $3 \cdot 19$ cells, where the number of Monte-Carlo-dropped users can easily exceed 1000, e.g. with 20 active users per cell, we have $3 \cdot 19 \cdot 20 = 1140$ users to consider within the simulation setting. For the sake of complexity, it is not tractable to explicitly implement the channel estimation functionality within system simulators, as the processing times simply get too long in commercially available workstations or personal computers. So an abstraction step is required, demanding a simple solution, which should also be easy to implement. This section deals with the issue of abstracting the channel estimator by a simple model.

Fortunately, due to the encouraging results of section 3.4, we have seen that the access to required parameter knowledge, even for time-variant, frequency-selective channels under the sometimes weak links for CoMP, is feasible and does not cost too much performance and time to collect

statistics. Thus, in our abstraction step, it appears to be tolerable to use an idealized MMSE channel estimator based on block fading channels.

The approach taken in this section is as follows: First the mean squared error of an idealized MMSE channel estimator is derived in a block fading channel. Then the properties of this MMSE estimator in terms of de-weighting unreliable sources are identified and conserved in a model where the error is just randomly generated. This feature will make the model for channel knowledge at the receiver side (CSIR) compatible with arbitrary receive combining algorithms. Supporting details of the MSE derivation and the model verification can be found in appendix E.

As this model shall also incorporate the possibility of emulating the channel state information at the transmitter side (CSIT), required in the downlink of CoMP / MIMO transmission, it is extended to incorporate feedback quantization and delay in a very simple manner. Reasonable parameters are discussed in appendix F.

Application examples of the model are in the following publications [Wil11a] [Art11] [ASWW12]; furthermore the model is used in the subsequent chapters 4 and 5 for downlink multi-cell multi-user CoMP simulations.

3.5.1. Simplified System Model

We start with the channel model introduced in section 3.1, (3.4): $\mathbf{y}_i = \sum_{\forall j \in \mathcal{J}} \sqrt{a_{ij}} \mathbf{S}_j \tilde{\mathbf{h}}_{ij}^{[\mathrm{P}]} + \mathbf{n}_i$.

The channel is assumed to be block-flat and thus, according to the discussion in section 2.2 invariant over the considered resource elements extending over time and frequency. With this simplification of a block flat (BF) channel, a single scalar channel coefficient $h_{ij}^{[BF]}$ is sufficient to replace the channel vectors $\tilde{\mathbf{h}}_{ij}^{[\mathrm{P}]}$, with transmit power ρ_j, path gain $\mathrm{E}\left\{|h_{ij}^{[BF]}|^2\right\} = \alpha_{ij}$ and effective path gain $a_{ij} = \rho_j \alpha_{ij}$. The length P pilot sequence will be used here in its vector form $\mathbf{s}_j$ instead of the diagonal matrix $\mathbf{S}_j$ for the ease of notation. In the following, we first focus on a single transmitter j. In section 3.5.5 we will discuss how the multi transmitter case will be handled. We can write our $P \times 1$ pilot receive vector as

$$\mathbf{y}_i = \sqrt{\rho_j} h_{ij}^{[BF]} \mathbf{s}_j + \mathbf{z}_i. \tag{3.80}$$

The noise vector $\mathbf{n}_i$ has been replaced by $\mathbf{z}_i$, which comprises both noise and "noise-like" interference from other transmitters, which are not explicitly modeled. $\mathbf{z}_i$ is a random vector with independent identically distributed elements[35] with variance σ_i^2. Note that the parameters interference plus noise power σ_i^2 as well as average channel gain α_{ij} depend on the respective receive antenna element i of the distributed antenna system. The average receive SINR is thus simply

$$\gamma_{ij} = \frac{\rho_j \cdot \alpha_{ij}}{\sigma_i^2} \tag{3.81}$$

[35] Because of the central limit theorem with many interfering sources it coarsely can be assumed as Gaussian.

3.5.2. Properties of the appropriate MMSE Estimator

As we have a linear signal model (3.80), the estimator which minimizes the mean squared error is the linear MMSE estimator. Starting, according to (3.18) with $\hat{h}_i^{[BF]} = \mathrm{E}\left\{h_{ij}^{[BF]}\mathbf{y}_i^H\right\}\mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\}^{-1}\mathbf{y}_i$ with further computations, shown in appendix E leads to

$$\hat{h}_{ij}^{[BF]} = \underbrace{\frac{\gamma_{ij}P}{1+\gamma_{ij}P}}_{\beta_{ij}}\left(h_{ij}^{[BF]} + \frac{\mathbf{s}_j^H\mathbf{z}_i}{\sqrt{\rho}\left\|\mathbf{s}_j\right\|^2}\right) \tag{3.82}$$

It is important to notice that the MMSE estimator inherently de-weights unreliable sources due to downscaling by the factor β_{ij}, using the available statistical knowledge. Obviously, this downscaling depends on the knowledge of the SINR γ_{ij} as well as on the number of pilot symbols P, with $\left\|\mathbf{s}_j\right\|^2 = P$ for unit power pilot sequences. The more pilots are available for the estimator and the higher the SINR, the more reliable is the result and thus the higher is β_{ij}. Otherwise a very low SINR leads to a β_{ij} tending towards zero:

$$\lim_{\gamma_{ij}P\to\infty}\beta_{ij} = 1, \quad \lim_{\gamma_{ij}P\to 0}\beta_{ij} = 0 \tag{3.83}$$

The MSE $\mathrm{E}\left\{\left|\hat{h}_{ij}^{[BF]} - h_{ij}^{[BF]}\right|^2\right\}$ of the MMSE estimator can be computed, according to appendix E (E.5), as

$$\mathrm{MSE}(\hat{h}_{ij}^{[BF]}) = \frac{\alpha_{ij}}{1+\gamma_{ij}P} \tag{3.84}$$

and the variance of the term after the de-weighting factor β_{ij} is

$$\sigma^2_{[e],ij} = \mathrm{Var}\left\{h_{ij}^{[BF]} + \frac{\mathbf{s}_j^H\mathbf{z}_i}{\sqrt{\rho}\left\|\mathbf{s}_j\right\|^2}\right\} = \frac{\alpha_{ij}}{\gamma_{ij}P} \tag{3.85}$$

3.5.3. Simplified Estimation Error Model for CSIR Inaccuracy

The objective of the error model in this thesis is to provide the same mean squared error as an actual MMSE estimator. Additionally, the model shall also preserve the de-weighting factor β_{ij}. This is because the model should be useable with different precoding and receive combining algorithms, impacted by this de-weighting[36].

A nice property of the MMSE estimator is the orthogonality principle [Kay93][Say08], which is used several times in this thesis for derivation of the estimator. This principle states that in order to be able to minimize the MSE, the error is statistically orthogonal to the estimated variable. Exploiting this principle for the model, it appears to be natural to generate a random error which

[36] E.g. the IRC receiver (3.48), containing matrix inversions is impacted by the actual scaling of imperfect channel coefficients, e.g. when different distributed antenna elements have very different reliabilities. So preservation of the scaling is required, as done in the model of this section.

is independent of the actual channel realization, where the properties of this error just depend on the statistical parameters. This leads us to an error model in the form

$$\hat{h}_{ij}^{(e,r)} = \beta_{ij} \left(h_{ij}^{[BF]} + \Theta_{ij} \right) \tag{3.86}$$

where we simply generate an error Θ_{ij} in the form of a complex i.i.d. Gaussian random variable with variance $\sigma^2_{[r],ij}$. In the model we use $\sigma^2_{[r],ij} = \sigma^2_{[e],ij} = \frac{\alpha_{ij}}{\gamma_{ij} P}$ from (3.85). Furthermore, as de-weighting we use $\beta_{ij} = \frac{\gamma_{ij} P}{1+\gamma_{ij} P}$ from (3.82).

Appendix E proofs that this model (3.86) leads to exactly the same MSE as the analytically computed MSE (3.84) of the estimator.

Let us have a look at the de-weighting of unreliable input of the MMSE estimator, caused by (3.83), which is reflected in the model. For comparison, the least squares estimator in the block-flat channel case (3.14) is used. The result is shown in Fig. 3.21. This result is very important for CoMP, as in the distributed antenna scenario, due to different path loss conditions, also a strong imbalance in terms of SNR between the different antennas may occur. Here β_{ij} resulting from the MMSE algorithm limits the occuring MSE, while for the least squares method this is not the case.

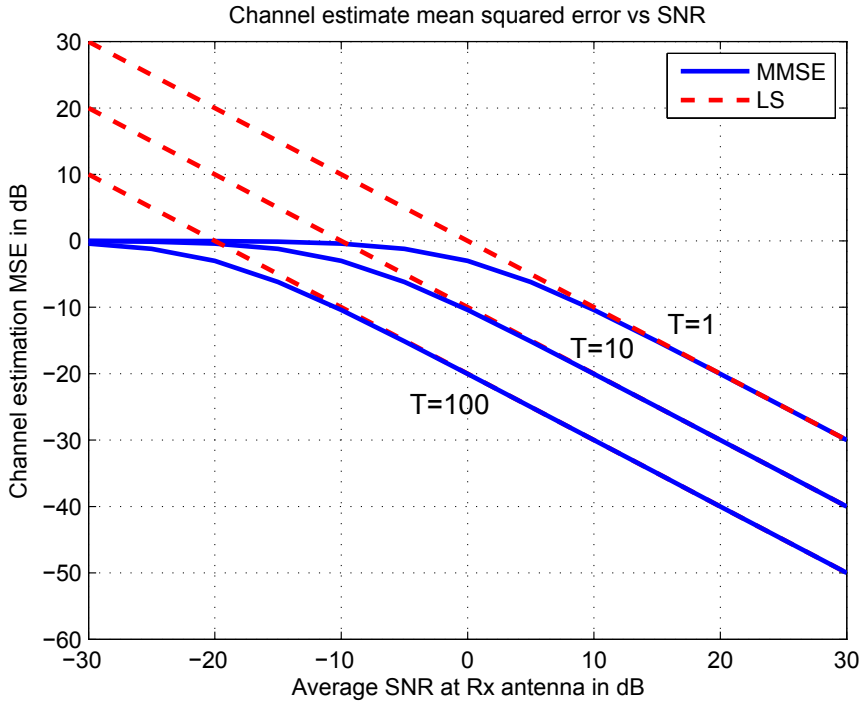

Figure 3.21.: Channel estimation MSE as a function of SNR, for the simplified block-flat fading model, for different numbers of pilot symbols $P = T$.

3.5.4. Simplified Estimation Error Model for CSIT Inaccuracy

When the transmitter uses a precoder[37], at the transmitter side, channel knowledge (CSIT) is required. In a closed-loop system, the channel is estimated first on the receiver side, providing $\hat{h}_i^{(e,r)}$. Then this channel coefficient is quantized and fed back via a feedback control signaling channel. From feedback delay and quantization, additional errors occur. A simple model now again treats those errors as additional distortion:

$$\begin{aligned} \hat{h}_{ij}^{(e,t)} &= \hat{h}_{ij}^{(e,r)} + \Theta_{ij}^{(t)} \\ &= \beta_{ij} \left(h_{ij} + \Theta_{ij} \right) + \Theta_{ij}^{(t)} \end{aligned} \tag{3.87}$$

As in the CSIR case, we model those additional errors $\Theta_{ij}^{(t)}$ as a complex Gaussian i.i.d. random variable with variance $\sigma_{[t],ij}^2$. A challenge will be to find reasonable values for $\sigma_{[t],ij}^2$. This is discussed in detail in appendix F.

3.5.5. Perspective Application of Simplified Estimator Model

While the previous sections 3.1 - 3.4 focused on realistic and detailed channel estimation algorithms in the uplink, the simplified estimator model of this section 3.5 can be used in both directions uplink and downlink. As chapters 4 and 5 are downlink-focused, the particularities for the model in that case are discussed here.

Note that the signal model (3.80) for this section so far discussed only a single link. For CoMP, we will observe large differences in path gains α_{ij} and thus the de-weighting factors β_{ij} will be clearly different, reducing the errors occuring for estimating weaker signals. The channels for those signals only can be estimated with inherently less accuracy and here the β_{ij} take into account statistical reliability knowledge in order to suppress estimation errors.

Our lessons learned from previous sections show that in CoMP a multi-user (i.e. joint) channel estimation across contributions from multiple cells is highly desirable. Note that the LTE standard allows to shift pilot sequences in their frequency position by the parameter ν_{shift} in [3GPc], so that pilot symbols of interest fall onto resource element positions of neighbor cell data symbols, which generates a noise-like behaviour of interference on pilot positions[38]. This would make pilot-based multi-user channel estimation impossible and so for CoMP it is not recommended to use ν_{shift}, presuming that the UEs are equipped with multi-user channel estimators (so clearly beyond basic LTE release 8).

While not every resource element will be explicitly modeled in system simulators, the Physical Resource Block (PRB) as the basic LTE allocation unit has to be modeled. Coarsely speaking, the PRB has about the size of coherence bandwidth and time for non-extreme propagation channels. Thus it is appealing to compute each realization and parameter of this channel estimation error model in a PRB-wise fashion. Assuming further that the UE channel estimator relies on an

[37] This case is especially relevant for the downlink, discussed in chapters 4 and 5

[38] LTE was initially not designed for CoMP, so its pilot structure was created for single cell channel estimation support.

observation window of a PRB size, we have $P \approx 10$. The above discussed multi-user channel estimator leads us to the approximation that the noise parameter σ_i^2 in the input SINRs (3.81) models interference from outside the CoMP cluster, but excludes intra-CoMP-cluster interference, which is an upper bound for the multi-user channel estimation inside the CoMP cluster. This approximation can be justified by looking at the results in Fig. 3.11 for the full MWF. The gap to perfect CSI for $N_{\mathrm{Supp}} = 7$ is about 0.3 dB, which is almost as small as the gap we observe in the single user case for the same SNR (10 dB) depicted in Fig. 3.6 on the right side. The observation motivates that multi-user channel estimation performance with separation of users by proper pilot sequences combined with multiuser channel estimation algorithms is not far from a single user setting under the same SNR and thus the exclusion of intra-cluster interference may serve as a simple upper bound approximation.

Those parameter settings will be used in chapters 4 and 5.

3.6. Summary of CoMP Channel Estimation

This chapter has discussed channel estimation for CoMP. First the single-user multi-cell scenario, which is noise-limited, has been discussed. The least squares estimator, in contrast to typical single cell settings, is a poor choice for CoMP, unable to cope with path loss imbalances between the different distributed antennas. The MMSE approach is a better choice but faces difficulties as well, because MMSE channel estimation in time-variant, frequency selective channels, requires extensive statistical parameter knowledge. Again, some weak links face unfavorable low SNR for that task. The consequence with state-of-the-art approaches, based on the sample covariance matrix, is that a large number of samples is required, in the order of a thousand subframes or more. This is undesirable, as the statistical parameters might already have changed, especially when the mobile is not continuously active. Furthermore, a long "warm-up time" before being able to effectively use CoMP is undesired as well. In this thesis, this problem was addressed by using the shrinkage technique for covariance estimation, clearly reducing the required "warm-up time". As a result the performance with shrinkage using statistical input from 200 subframes is equal to non-shrinkage with around 1000 subframes in a noise-limited environment .

Secondly, the multi-user setting was investigated. Here the assignment of pilot sequences to the different mobiles across different cells at first glance plays an important role. A potential modification of the LTE-A standard pilot assignment was investigated. Due to reduced cross-correlation, aiming towards orthogonality, this pilot assignment showed benefits especially for the simple algorithms, such as least squares. Interestingly, a full multi-user MMSE channel estimation, learning and exploiting the time-frequency statistical characteristics of the channel has almost identical performance with correlated sequences as with orthogonal ones. This allows to decrease the burden on the system as strict orthogonality in the choice and assignment of pilot sequences across the different cells is not required.

Again the parameter knowledge in the multi-user setting plays an important role. In this thesis, a multi-stage channel estimator was introduced, which allows to extract the statistics and the actual channel knowledge. The solution was designed such that only a few subframes for gaining statistical knowledge are enough to achieve its performance maximum which is even in large

CoMP clusters with 7 supporting cells just 1dB worse in terms of post-combining SINR than with perfect channel knowledge.

Finally, a simplified model has been introduced in order to simplify the abstraction of channel knowledge accuracy on the system level, avoiding the full channel estimation computation, which makes an assessment feasible within large system simulators. The model will be applied in the subsequent chapters.

4. Downlink Coordinated Multi-Point with Imperfect CSI

The focus in this chapter is on downlink precoding for a distributed multi-antenna system with imperfect channel knowledge. A comparison is made between different coordination strategies, where different precoding algorithms are used at the BS in conjunction with different receive combining approaches at the MS. A generalized Eigenbeamforming approach is formulated which can be used both for interference avoidance and joint transmission. The impact of path gain imbalances is shown to be a very important parameter, which is especially carved out in this investigation. Furthermore, the impact of number of transmit and receive antennas, as well as varying number of available pilot symbols for channel estimation are assessed throughout this chapter. To get a clear view on the dependencies of the important parameters, a two-user, two-cell scenario is selected.

In order to model the imperfect channel knowledge in a simple and efficient way, the achievements from section 3.5 are used. In chapter 4, a key question is, whether CoMP or non-CoMP behaves differently under imperfect channel knowledge. A further key question is to identify performance trends for the different coordination strategies in conjunction with appropriate multi-antenna processing. Parts of the achievements of this chapter have been published in [Wil11a].

4.1. System Model

The coordination cluster, being denoted as set $\mathcal{T} = \{t_1, t_2, \ldots, t_T\}$ of base station cells, with $|\mathcal{T}| = T$, is transmitting at a particular time instant k to the set of users $\mathcal{U}$. Each cell is equipped with N antennas, each mobile has M antennas. This means, for an OFDM resource element indexed k, l, the radio channel from BS v to MS u can be described by the MIMO channel matrix $\mathbf{H}_{uv} \in \mathbb{C}^{M \times N}$. The time-frequency dependency represented by k, l is omitted in the notation for the remaining part of this section for the sake of readability.

For the composite multi-cell channel matrix $\mathbf{H}_{u,\mathcal{T}} \in \mathbb{C}^{M \times NT}$ of the entire transmit set $\mathcal{T}$, the following notation is used

$$\mathbf{H}_{u,\mathcal{T}} = [\mathbf{H}_{u,t_1} | \mathbf{H}_{u,t_2} | \ldots | \mathbf{H}_{u,t_T}] \tag{4.1}$$

in order to stack all MIMO matrices of the coordinated set of transmitting cells. The path gain between the transmitting cell t_1 and the MS u is $\alpha_{u,t_1} = \mathrm{E}\{||\mathbf{H}_{u,t_1}||_2^2\}$. The cell transmits with power ρ_{t_1}. In order to extract the potentially strong receive power level differences between signals from different cells within the same CoMP cluster, a single parameter is introduced,

which will be varied in the simulation studies within this chapter. This parameter is called here *cell isolation*. For the two cell scenario with cell indices t_1, t_2, it is defined in dB as

$$\Delta_{t_1,t_2} = 10\log(\alpha_{u,t_1}\rho_{t_1}) - 10\log(\alpha_{u,t_2}\rho_{t_2}) \tag{4.2}$$

Let us now look at the entire transceiver system model. Fig. 4.1 illustrates this system model in the general coordinated case for a 2-cell, 2-user example.

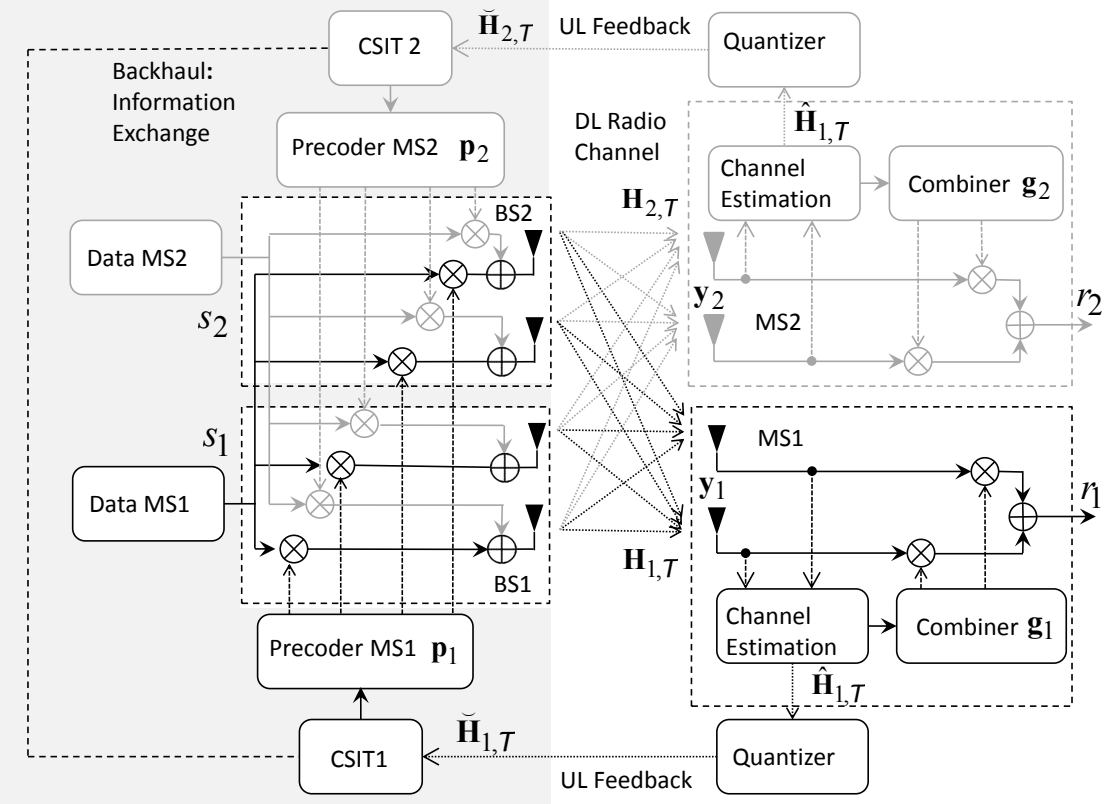

Figure 4.1.: Closed-loop joint transmission, 2 users, 2 cells; BS1 is the serving cell of MS1, BS2 serves MS2. "Data MS1" and "Data MS2" refers to data intended for MS1 and MS2, which act as data-sink. (Additional noise and interference at antenna input is omitted for simplicity.)

At a particular time-frequency resource element, the data symbol s_u is transmitted to user u. The transmission uses linear precoding across the coordinated antennas with weight vector $\mathbf{p}_u \in \mathbb{C}^{NT\times 1}$. The multi-antenna receive vector $\mathbf{y}_u \in \mathbb{C}^{M\times 1}$ thus can be written as

$$\mathbf{y}_u = \mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u s_u + \sum_{\forall \mu \in \mathcal{U}\setminus\{u\}} \mathbf{H}_{u,\mathcal{T}}\mathbf{p}_\mu s_\mu + \mathbf{z}_u. \tag{4.3}$$

The noise vector $\mathbf{z}_u \in \mathbb{C}^{M\times 1}$ contains thermal noise, as well as interference outside of the explicitly considered cluster of cooperating BS. Here the covariance is modeled as "spatially white"

$\mathrm{E}\left\{\mathbf{z}_u\mathbf{z}_u^H\right\} = \sigma_z^2\mathbf{I}$ which implies a large number of independent interfering signals according to the Central Limit Theorem. In chapter 5 also explicit neighbor cluster interfering signals are considered, being spatially non-white.

Note that (4.3) is more general than representing only the case that all BS in the coordination cluster contribute to joint transmission. The weight vector $\mathbf{p}_u$ may have zero elements for all antennas of a particular base station and thus this BS is excluded from joint transmission.

As the mobiles are equipped with multiple antennas, they can apply receive combining, which improves the useful signal output and/or suppresses interference. Within this thesis, for the sake of complexity, the receive combiner is restricted to a linear one, using the weights $\mathbf{g}_u \in \mathbb{C}^{M\times 1}$. In order to compute the SINR at the combiner output, we first decompose the post-combining receive signal in a similar way as done in the uplink (3.49) in section 3.3.1.2:

$$r_u = \mathbf{g}_u^H\mathbf{y}_u = \underbrace{\mathbf{g}_u^H\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u s_u}_{r_u^{[U]}} + \underbrace{\mathbf{g}_u^H\left(\sum_{\forall\mu\in\mathcal{U}\setminus\{u\}}\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_\mu s_\mu + \mathbf{z}_u\right)}_{r_u^{[I+N]}} \tag{4.4}$$

The SINR $\gamma_u^{[Out]}$ at the receive combiner output, can be computed as the useful signal power at the combiner output $\mathrm{E}\left\{|r_u^{[U]}|^2\right\}$ over the noise and interference contributions passing the receive combiner weights $\mathrm{E}\left\{|r_u^{[I+N]}|^2\right\}$. The expectation is done w.r.t. transmitted symbols and noise realizations, assuming unit norm data symbols, which are uncorrelated between users. This results into:

$$\gamma_u^{[Out]} = \frac{\left|\mathbf{g}_u^H\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u\right|^2}{\sum_{\forall\mu\in\mathcal{U}\setminus\{u\}}\left|\mathbf{g}_u^H\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_\mu\right|^2 + \left\|\mathbf{g}_u^H\right\|^2\sigma_n^2} \tag{4.5}$$

The achievable mean sum rate[1] over all served users within the considered set of cells can be computed according to Shannon as

$$R = \sum_{\forall u\in\mathcal{U}}\log_2\left(1+\gamma_u^{[out]}\right). \tag{4.6}$$

4.2. Considered Coordination Types

Let us first recap the different coordination types introduced in section 2.1.4.2, depicted by Fig. 2.7 and their implications on the system model introduced in chapter 4. Existing cellular systems are *non-coordinated* in terms of spatial processing. Each cell only has user data and channel knowledge (CSIT) available for the users which it is serving. Each transmission happens independent of what is done in other cells. An illustration is shown in Fig. 4.2. In this case the "backhaul information exchange" excludes CSIT, data and scheduling information. Simpler techniques, like inter-cell interference coordination (ICIC) [AL07] are not discussed in this chapter. For the system model (4.3) this means that the weight vector $\mathbf{p}_u$ actually only has

[1] In case of complex Gaussian distributed signals, interference, noise and channels.

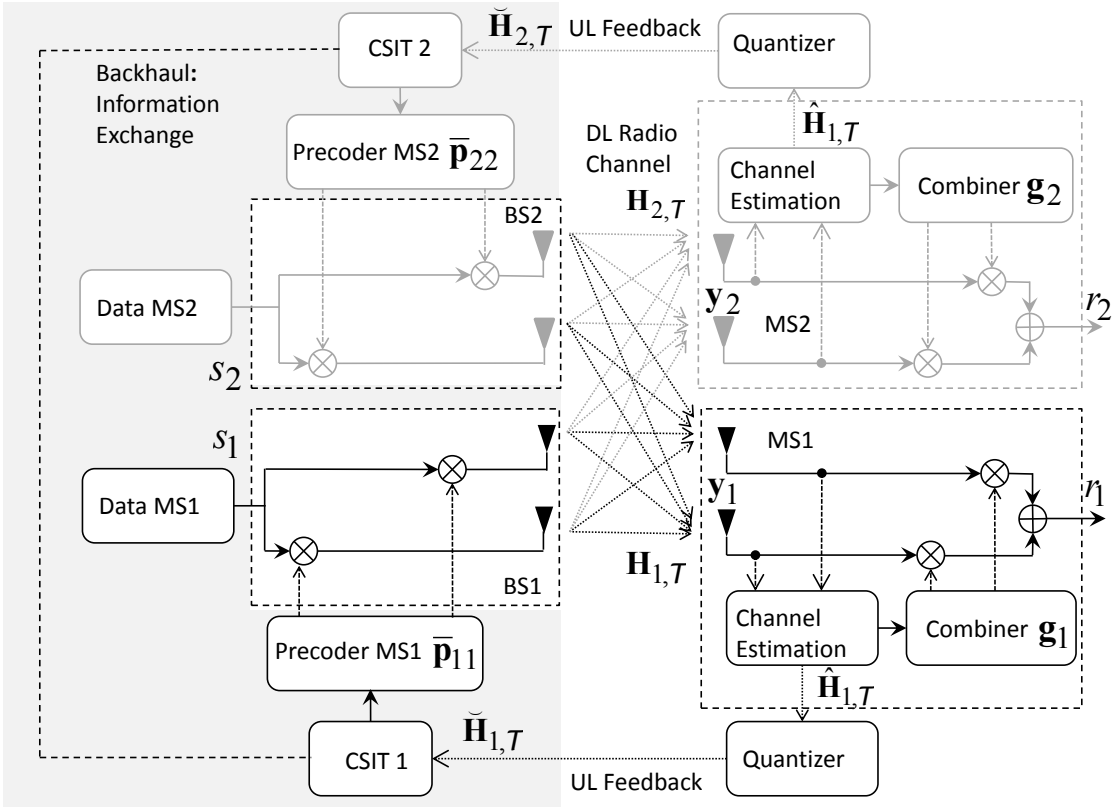

Figure 4.2.: Closed-loop coordinated scheduling/beamforming (CS/CB), 2 users, 2 cells. BS1 is the serving cell of MS1, BS2 serves MS2. "Data MS1" and "Data MS2" refers to data intended for MS1 and MS2, which act as data-sink. In absence of CSIT sharing via backhaul, this represents non-coordinated transmission. (Additional noise and interference at antenna input is omitted for simplicity.)

N non-zero elements - the number of antennas of the serving cell. The other $N(T-1)$ elements are zero. In this case we write $\bar{\mathbf{p}}_{uv} \in \mathbb{C}^{N\times 1}$ for the antenna weights of base station cell v serving user u. Note that the system model (4.3) still includes this case as $\bar{\mathbf{p}}_{uv}$ contains just the non-zero elements of $\mathbf{p}_u$, corresponding to the antenna elements of serving cell v. Nevertheless, rewriting (4.3) for transmission of only serving cell antennas (either coordinated or non-coordinated) to their served users yields

$$\mathbf{y}_u = \mathbf{H}_{u,v}\bar{\mathbf{p}}_{uv}s_u + \sum_{\forall \mu \in \mathcal{U}\setminus\{u\}} \mathbf{H}_{u,\bar{\mu}}\bar{\mathbf{p}}_{u\mu}s_\mu + \mathbf{z}_u, \tag{4.7}$$

where $\bar{\mu}$ represents the serving cell index corresponding to user μ. In case the different cells share CSIT and scheduler information, this is known as *coordinated beamforming* (or coordinated scheduling/beamforming - CS/CB) and covered as well in Fig. 4.2. The precoders are still per-cell with only N non-zero elements, represented by $\bar{\mathbf{p}}_{uv}$, as above. But in contrast to the non-coordinated case, the computation of $\bar{\mathbf{p}}_{uv}$ uses CSIT from neighbor cells (plus the information which MS is served). This generates the capability to suppress interference towards mobiles in

neighbor cells. One example for this CS/CB approach, already supported by the LTE-A standard is the multi precoding matrix indicator (PMI) reporting. (See also section 2.1.4.2 and similar "worst companion" PMI reporting [WH10].)

The most demanding scheme in terms of complexity and backhaul bandwidth consumption is *Joint Transmission*, illustrated by Fig. 4.1. All NT weight vector elements are used, generally with non-zeros, as all the distributed antennas of a coordination cluster may be involved in the transmission to a particular user within the coordinated cell set. Available knowledge on backhaul aspects is dicusssed in section 2.1.4.3. In chapter 4, backhaul rate constraints are not explicitely considered[2].

4.3. Considered Precoding and Receive Combining Strategies

4.3.1. Basic Assumptions

In the notation of the system model, only one data stream per served user is considered. In principle, especially if the MS is equipped with multiple receive antennas, multiple data streams (spatial multiplexing) could be considered as well. Based on findings in the literature, such as [BKBO13] , multi-user transmission with just one stream per mobile pays off compared to spatial multiplexing under realistic conditions, so this restriction to "*rank-1 transmission*" per user does not cause any notable disadvantages.

Precoders could also be non-linear, applying the principle of Costa's Dirty paper coding [Cos83], which can be realized by Tomlinson-Harashima precoding [SSF+10]. In the sections 2.1.4.1 and 2.1.4.2 it is discussed why the CoMP community focuses on *linear precoding* strategies, due to complexity and practical feasibility. This restriction to linear precoding is done as well in the chapters 4 and 5 in this thesis.

A further important point to discuss is the issue of the *power constraint*. Existing power amplifier hardware imposes a peak power constraint per antenna element. Intelligent clipping preprocessing mechanisms are used to feed the power amplifier with a low peak-to-average power ratio (PAPR) in order to achieve a good power efficiency. In the literature, usually a sum power constraint over all antennas is assumed, for CoMP sometimes also a power constraint per base station. All those approaches are thus somewhat unrealistic as they imply that spent power could be shared among power amplifiers. The good news is that there is a simple practical way to live with the sum power constraint. The used system model (4.3) is from the perspective of a narrowband channel. In an OFDM system with different user allocations over the larger number of subcarriers in the whole frequency band, there is some statistical balancing of the spent power per antenna element. (This can be further enforced by intelligent scheduling to even out power imbalances across antenna elements.) Then the entire transmit signal is downscaled in order to

[2] In case the available backhaul bandwidth is small, we can implicitly take this into account by considering the supported coordination strategy. Thus, instead of Joint Transmission, only coordinated beamforming is practically possible in this case.

satisfy peak power constraints per antenna elements, as done e.g. in [KHWH13]. Having those solutions in mind, the precoding algorithms discussed below are simply fulfilling a sum power constraint over all considered antenna elements.

Finally, for the precoder computation, *explicit CSIT feedback* is used, which has been successfully demonstrated in [KHWH13] [WHHK13] and is discussed further in appendix F. Note that this explicit CSI allows for much more accuracy than existing release 11 multi-PMI or worst companion signaling in terms of amplitude and phase resolution of the channel.

4.3.2. Basic Precoding Algorithms

Eigenbeamforming for Non-Coord. Transmission, Zero Forcing for Joint Transmission
In case the transmitter only has channel knowledge for the user of interest, it can aim at maximizing the signal strength at the receive antennas. In a MISO configuration, the choice of finding the appropriate transmit weights can be formulated equivalently as problem of finding the optimal receive weights in a SIMO configuration - which is known as maximum ratio combining (MRC), discussed in section 3.2.3. The corresponding transmit precoding or beamforming algorithm is known as *maximum ratio transmission*. In the full MIMO (or distributed MIMO) configuration, maximizing the instantaneous power of the receive signal $||\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u||_2^2$ by appropriate choice of the precoding weights $\mathbf{p}_u \in \mathbb{C}^{NT\times 1}$, subject to $||\mathbf{p}_u||_2^2 = 1$, can be formulated as

$$\arg\max_{\mathbf{p}_u} \mathbf{p}_u^H \mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}} \mathbf{p}_u \tag{4.8}$$

In linear algebra, this expression is a quadratic form [PP12] which can be maximized by finding the Eigenvector $\mathbf{x}_{\max}$ corresponding to the largest Eigenvalue of the matrix $\mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}}$. Here, we write

$$\mathbf{p}_u \propto \mathbf{x}_{\max}\left\{\mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}}\right\}. \tag{4.9}$$

In the literature, this choice of transmit antenna weights is known as *Eigenbeamforming* (Ebf) [HBD00]. We use it in conjunction with non-coordinated cells, when no channel and scheduling knowledge from neighbor cells is available and thus maximization of receive signal strength is the right strategy from the perspective of one cell (serving one user per time-frequency resource). With the setup of Fig. 4.2, when restricting the transmission purely to the respective serving cells v, the maximization over $\bar{\mathbf{p}}_{uv} \in \mathbb{C}^{N\times 1}$, the N non-zero elements of $\mathbf{p}_u \in \mathbb{C}^{NT\times 1}$, can be expressed, using (4.7), as

$$\arg\max_{\bar{\mathbf{p}}_{uv}} \bar{\mathbf{p}}_{uv}^H \mathbf{H}_{u,v}^H \mathbf{H}_{u,v} \bar{\mathbf{p}}_{uv}, \tag{4.10}$$

which leads to

$$\bar{\mathbf{p}}_{uv} \propto \mathbf{x}_{\max}\left\{\mathbf{H}_{u,v}^H \mathbf{H}_{u,v}\right\}. \tag{4.11}$$

Another popular precoding strategy is *Zero Forcing* (ZF). When channel knowledge $\mathbf{H}_{\mu,\mathcal{T}}$ for all potential interference victims, indexed $\mu \neq u$, is available, the transmit weights for joint transmission according to Fig. 4.1 can be designed such that the outcome lies in the "nullspace of all other users" meaning simply

$$||\mathbf{H}_{\mu,\mathcal{T}}\mathbf{p}_u||_2^2 = 0, \forall \mu \neq u \tag{4.12}$$

This is probably the most used precoding approach in the literature. But it also has some drawbacks. As the weight computation criterion just forces the known interference to zero at the receive side, the resulting useful signal power may be poor. Especially when a certain amount of noise is present, this will cost performance in terms of receive SINR. Furthermore, ZF is in disadvantage when uncoordinated interference is present within the system, e.g. coming from independently operating neighbor cell clusters. In this case, a precoding algorithm which also cares for the achieved useful signal power would be preferable over ZF. Furthermore, ZF causes some restrictions for the system configuration in terms of the number of transmit and receive antennas, see e.g. [STS07][3]. Again, the "hard nulling" of ZF criterion seems unnecessary and preferably should be replaced by "something soft". This leads us to the considered approach of the next section.

4.3.3. Multi-Cell Generalized Eigenbeamforming

Joint Transmission Based on Generalized Eigenbeamforming
The SINR at the receive combiner output for JT, as depicted by Fig. 4.1 is given by (4.5). In general, an optimum transceiver strategy would aim at maximizing this expression. Unfortunately, this objective is not practical. The choice of a weight vector of one particular user effects the denominator of (4.5) for all other users. This means, in a maximization of (4.5), all the weights are coupled between all users of interest. There is no closed form solution for this problem. The problem is generally non-convex and the typical dimensions in terms of number of antennas and users are prohibitive for search strategies.

A different criterion for picking the weights in a multi-user MIMO setting was considered in [STS07]: Signal-over-leakage-plus-noise-ratio (SLNR). The term "leakage" refers to the sum interference caused to the user set of interest $\mathcal{U}$.

We now apply the SLNR criterion to the CoMP scenario. Hence, the aim is to maximize the ratio of the receive signal power $||\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u||_2^2$ over leakage power $\sum_{\forall\mu\in\mathcal{U}\setminus\{u\}}||\mathbf{H}_{\mu,\mathcal{T}}\mathbf{p}_u||_2^2$ plus summed power of $\mathbf{z}_u$, thus $M\sigma_z^2$,by appropriate choice of the precoding weights $\mathbf{p}_u$, subject to $||\mathbf{p}_u||_2^2 = 1$. This optimization task can be formulated as

$$\arg\max_{\mathbf{p}_u} \frac{\mathbf{p}_u^H \mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}} \mathbf{p}_u}{M\sigma_z^2 + \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}}||\mathbf{H}_{\mu,\mathcal{T}}\mathbf{p}_u||_2^2} \tag{4.13}$$

In linear algebra, this expression leads to a generalized Eigenvalue problem [Mey00] which can be solved by finding the generalized Eigenvector $\mathbf{x}_{\max}$ corresponding to the largest generalized Eigenvalue of the matrices $\mathbf{A}_{\mathrm{SLNR}} = \mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}}$ and $\mathbf{B}_{\mathrm{SLNR}} = M\sigma_z^2\mathbf{I} + \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}} \mathbf{H}_{\mu,\mathcal{T}}^H \mathbf{H}_{\mu,\mathcal{T}}$. As $\mathbf{B}_{\mathrm{SLNR}}$, constructed by a square of a matrix and a real-valued diagonal matrix, is positive semi-definite, it is invertible in the presence of some $\sigma_z^2 > 0$. Hence the generalized Eigenvalue problem can be translated to a simple Eigenvalue problem with matrix $\mathbf{B}_{\mathrm{SLNR}}^{-1}\mathbf{A}_{\mathrm{SLNR}}$. Thus for

[3] Roughly speaking, when the sum of receive antennas of allocated users in the cluster is more than the sum of BS transmit antennas, ZF cannot support this anymore. This limits the set of supported users $\mathcal{U}$ with ZF which is a drawback especially for CoMP, as $\mathcal{U}$ extends over several cells in contrast to typical single cell MU-MIMO.

applying SLNR to CoMP, for the corresponding antenna weights, we need to find the Eigenvector, corresponding to the largest Eigenvalue $\mathbf{x}_{\max}$ of $\mathbf{B}_{\mathrm{SLNR}}^{-1}\mathbf{A}_{\mathrm{SLNR}}$:

$$\mathbf{p}_u \propto \mathbf{x}_{\max}\left\{\left(M\sigma_z^2\mathbf{I} + \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}} \mathbf{H}_{\mu,\mathcal{T}}^H\mathbf{H}_{\mu,\mathcal{T}}\right)^{-1}\mathbf{H}_{u,\mathcal{T}}^H\mathbf{H}_{u,\mathcal{T}}\right\}. \tag{4.14}$$

In this thesis, this approach is denoted Generalized Eigenbeamforming (GEbf). Compared to ZF precoding, the following advantages are inherently in the antenna weight computation:

- The optimization criterion explicitly includes the useful signal power. For CoMP, this improves robustness against out-of-cluster interference. (Further mechanisms will be discussed in chapter 5.)
- The set of considered users $\mathcal{U}$ can be arbitrarily large.

In order to leverage its interference suppression potential, GEbf requires channel knowledge from the links of the CoMP cluster cells to all MS, which shall benefit from interference avoidance. As those links may be potentially weak in CoMP, causing channel estimation inaccuracies, the channel knowledge has to be considered critically in the investigations. Additional impairments by feedback quantization and delay have to be taken into account. These important aspects will be included in this chapter's investigations for all precoding and receive combining algorithms.

Coordinated Scheduling / Beamforming Based on Generalized Eigenbeamforming
While the result (4.14) covers the joint transmission case, we now formulate the same SLNR criterion for CS/CB. The optimization task now focuses on the non-zero elements $\bar{\mathbf{p}}_{uv} \in \mathbb{C}^{N\times 1}$ (out of $\mathbf{p}_u \in \mathbb{C}^{NT\times 1}$) belonging to cell v. Thus the useful signal part is $||\mathbf{H}_{u,v}\bar{\mathbf{p}}_{uv}||_2^2$ and the leakage is only considered from the perspective of cell v as $||\mathbf{H}_{\mu,v}\bar{\mathbf{p}}_{uv}||_2^2$, leading to

$$\arg\max_{\bar{\mathbf{p}}_{uv}} \frac{\bar{\mathbf{p}}_{uv}^H\mathbf{H}_{u,v}^H\mathbf{H}_{u,v}\bar{\mathbf{p}}_{uv}}{M\sigma_z^2 + \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}}||\mathbf{H}_{\mu,v}\bar{\mathbf{p}}_{uv}||_2^2}. \tag{4.15}$$

Likewise, the antenna weights are

$$\bar{\mathbf{p}}_{uv} \propto \mathbf{x}_{\max}\left\{\left(M\sigma_n^2\mathbf{I} + \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}} \mathbf{H}_{\mu,v}^H\mathbf{H}_{\mu,v}\right)^{-1}\mathbf{H}_{u,v}^H\mathbf{H}_{u,v}\right\}. \tag{4.16}$$

Thus here, the GEbf approach has been applied to both CoMP cases: JT as well as CS/CB. Its SLNR-based optimization approach is general enough for that purpose, leading to a very broad precoding principle, handled in this thesis and extended also to the out-of-cluster-leakage case, discussed in chapter 5.

4.3.4. Receive Combining Algorithms

In chapter 3, for uplink CoMP, receive combining algorithms were already discussed. MRC (3.41) is the optimal choice when no knowledge on interference is available. For the downlink, when the effective precoded channel is taken into account at the combiner, the combiner vector $\mathbf{g}_u \in \mathbb{C}^{M\times 1}$ is given by

$$\mathbf{g}_u^H \propto (\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u)^H. \tag{4.17}$$

IRC finds the optimal compromise between suppressing interference and enhancing the signal at the combiner output in an MMSE sense. It is derived in (3.48). For the downlink, this derivation can be extended in a straightforward manner, when the effective precoded channel is taken into account, by

$$\mathbf{g}_u^H \propto \left(\left(\sigma_z^2 \mathbf{I} + \sum_{\forall \mu \in \mathcal{U}\setminus\{u\}} \mathbf{H}_{\mu,\mathcal{T}}\mathbf{p}_u\mathbf{p}_u^H\mathbf{H}_{\mu,\mathcal{T}}^H \right)^{-1} \mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u \right)^H . \tag{4.18}$$

In order to leverage its interference suppression potential, IRC (similar as GEbf) requires channel knowledge from the links of neighbor cells to the MS. As those links may be potentially weak in CoMP, the quality of the channel knowledge has to be considered seriously in the following investigations.

4.4. Modeling Imperfect CSI

Chapter 3 discusses the channel estimation algorithms in detail. In section 3.5, a model has been introduced for imperfect CSIR and CSIT. The model is based on the mean squared error (MSE) and characterizes distortions as Gaussian noise. The de-weighting of unreliable links is preserved in the model (3.86) by β_{ij}. This means that in precoder and combiner weight computations (4.14),(4.18) where matrix inversions occur, the scalings of the channel coefficients due to β_{ij} with mismatched weights are preserved when computing the $\mathbf{p}_u$, which realistically models the impact of MMSE channel estimation. The feedback quantization and delay for modeling the CSIT is included in a similar manner in (3.87). Parameters and assumptions for this are discussed in appendix F.

Let us now discuss the downlink-specific properties of the chosen model parameters. In section 3.4, we have shown that with the multi-stage estimator, we are able to get rather close to perfect channel knowledge, presuming that the user has a typical allocation size of 5 PRBs, so that enough pilot symbols are available for the estimation. Appendix F shows that we can bring down the errors from feedback quantization to a very small value. If we want to pay the price for the feedback rate we have to make a trade-off between uplink resource consumption and downlink performance. Additionally, the channel aging effects can be handled by prediction [WHHK13] up to a certain extend. In case the system provides all this effort for CoMP, perfect CSIT and CSIR are almost present and thus will be considered in the simulations of section 4.5 as an upper bound. This is the first case.

In the second case, we consider restrictions in terms of MS channel estimation complexity and less elaborate feedback, supported by the standard. For this, we provide a more pessimistic parameter set for the model of section 3.5. Let us assume, the MS is equipped with a channel estimator, according to section 3.3.2.2, using the small window approximated MWF. Thus for the channel estimate of a particular resource element, the MS relies only on a few neighboring (in time or frequency) pilot symbols, where the channel approximately has not changed. In the simulations in section 4.5, we typically use a number of $P = 10$ pilot symbols, which e.g. is approximately the number of pilot symbols in one downlink PRB, see Fig. 2.8. Furthermore, the impact when a varying number of pilot symbols P is used in the estimator is shown as well.

Additionally, we assume in this second case for the CSIT, that the system has no counter-measure against channel aging, like prediction. For a feedback delay of 5 ms with a reporting interval of 10 ms, the MSE for pedestrian velocitites (3 km/h) becomes -12 dB, as shown in appendix F. This means even for the low mobility scenario, which is the most important for CoMP, the MSE due to channel aging is clearly non-negligible. Thus, it also makes no sense for the feedback quantization to be much more accurate than this MSE of -12 dB. This second case thus can be seen as a practical lower bound for a system which is intended for CoMP.

4.5. Simulation Results

A remark on the level of abstraction in modeling: The modeling in chapter 3 is closely aligned to LTE-A as the algorithms can already be used with current LTE-A. The modeling in chapter 4 is more abstract, as the considered precoding techniques demand feedback signaling beyond current LTE-A and thus strict alignment to LTE-A is not possible and not beneficial.

The different precoding and receive combining approaches are now compared in a 2-user, 2-cell scenario, introduced in section 4.1. This includes different levels of coordination:

- Ebf: Eigenbeamforming per cell (4.9), where different cells are not coordinated between each other. This is a non-CoMP baseline system.
- GEbf-CS: Generalized Eigenbeamforming (4.16) is used per-cell in conjunction with coordinated scheduling. Each BS serves only its own associated users but takes into account scheduling information and CSIT of neighbor cell users in order to spatially avoid interference. This demands small amount of backhaul bandwidth for CoMP.
- GEbf-JT: Generalized Eigenbeamforming (4.14) with joint transmission. The precoders are applied jointly across all cells of the coordination cluster. Besides CSIT and scheduling information, this also requires exchange of user data across all involved cells and thus consumes much more backhaul bandwidth.

These transmission strategies are combined with either MRC or IRC. The considered radio propagation channels $\mathbf{H}_{u,\mathcal{T}}$ within the narrowband system model of (4.3) are simply modeled as complex Gaussian i.i.d. Rayleigh fading. The parameters for CSIT/R are as discussed in section 4.4: Ideal CSIR/T provides an upper bound, while impaired CSIR/T, represents a more pessimistic parameter setting, when the MS channel estimator effort and the feedback mechanisms are less advanced.

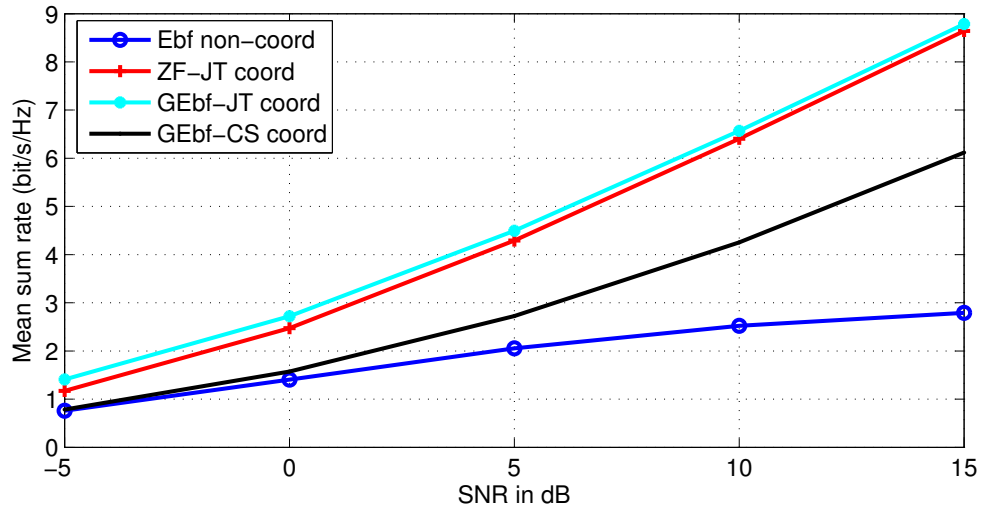

Figure 4.3.: Mean sum rate vs SNR: Comparison of ZF and GEbf for isolation $\Delta_{1,2} = 0\,\text{dB}$, $P = 10$, 2 Tx antennas per cell, 1 Rx antenna per MS.

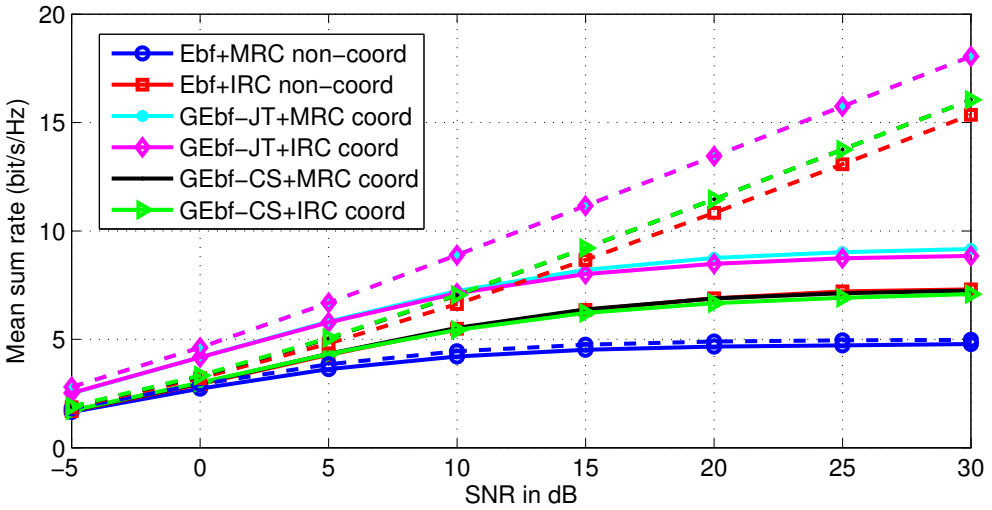

Figure 4.4.: Mean sum rate vs SNR: Comparison of different precoding and coordination strategies for isolation $\Delta_{1,2} = 0\,\text{dB}$, $P = 10$, 4 Tx antennas per cell, 2 Rx antennas per MS. Impaired CSI in solid lines, perfect CSI in dotted lines

In Fig. 4.3 it can be seen for JT that GEbf slightly outperforms ZF precoding, while at the same time not having the dimension restrictions of ZF. In lower SNR ranges[4], the resulting sum-rate of GEbf is around 10% higher than for ZF.

Fig. 4.4 depicts the sum-rate of the 2-cell system as a function of SNR, for the case that the users are both located at the cell edge (isolation 0dB). Each cell uses 4 Tx antennas, each MS 2 Rx antennas. For non-coordinated single-cell transmission (Ebf), one can observe that the sum-rate saturates early with increasing SNR and MRC receivers. The loss due to impaired channel

[4]These lower SNR ranges also model the case when out-of-cluster interference from additional cells is present.

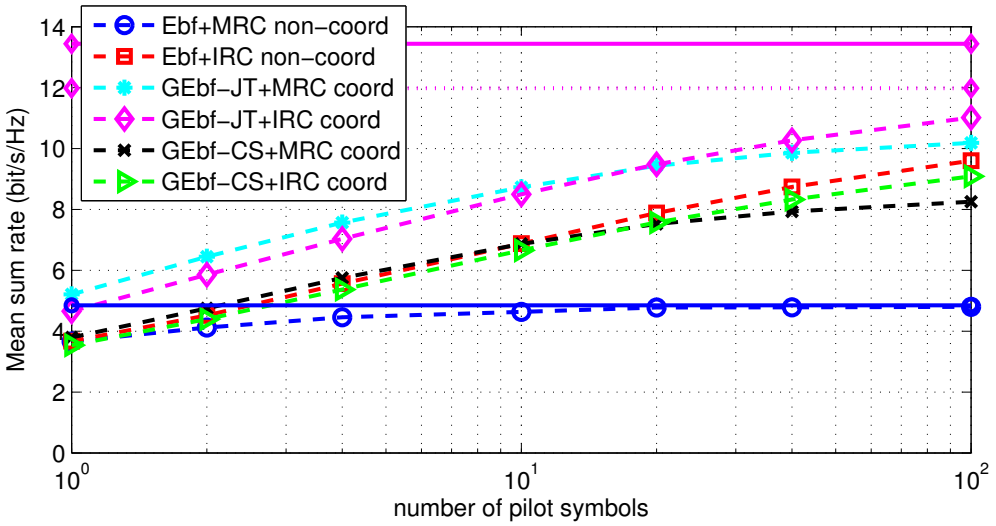

Figure 4.5.: Mean sum rate vs number of used pilot symbols P. Dotted line: Impaired CSIR+CSIT. Dashed: perfect CSIR and impaired CSIT. Solid:perfect CSIR+CSIT. SNR = 20 dB, isolation $\Delta_{1,2} = 0$ dB, 4 Tx antennas per cell, 2 Rx antennas per MS.

knowledge is almost negligible, thus this system is robust to those inacurracies but exhibits a low performance. When the MSs are using an IRC receiver to suppress inter-cell interference, the performance drastically improves, even in the non-coordinated case. At an SNR of 20 dB, the improvements over MRC in terms of sum-rate are more than 50% with impaired CSI. Under perfect CSI, those gains are even above 140%. To get closer to this point, larger allocation sizes ($P \gg 10$) are required in conjunction with advanced channel estimation techniques.

When coordinated beamforming/scheduling is used (GEbf-CS), the performance in conjunction with MRC receivers is roughly the same as non-coordinated cells with IRC receivers for impaired CSI and slightly better with perfect CSI. When adding IRC receivers to CS/CB there are no additional gains, as the inter-cell interference was already handled by the transmit side. From this we can conclude that CS/CB helps to shift the complexity from the MS side to the BS side, but, in case the MS is already equipped with proper interference-suppression capabilities, CS/CB can only provide small sum-rate gains over non-coordinated systems. The best performance is obtained by joint transmission (GEbf-JT). As JT is able to exploit the coherent transmission across all the antennas of the CoMP cluster, additional antenna gains can be exploited. With perfect CSI, JT is factor 3 better than Ebf+MRC in terms of sum-rate at an SNR of 20 dB. With impaired CSI that gain is around 90%.

Fig. 4.5 shows the sum-rate performance as a function of used pilot symbols P for an SNR of 20 dB at the cell edge (isolation $\Delta_{1,2} = 0$ dB). In order to be able to use more pilot symbols for the LMMSE channel estimation, the allocation size needs to be sufficiently large in terms of allocated physical resource blocks (PRB). For $P = 100$, the user allocation has to be at least 13 PRBs. This means, that the user requires enough data to be transmitted, and the scheduler needs to be constrained in order to generate large enough user allocations. Furthermore, the channel estimator at the MS side needs to be able to operate beyond coherence bandwidth and time, which implies that it requires extensive channel statistics for all links to all BS in the CoMP cluster. So, an algorithm like the multi-stage channel estimation described in section 3.4 is

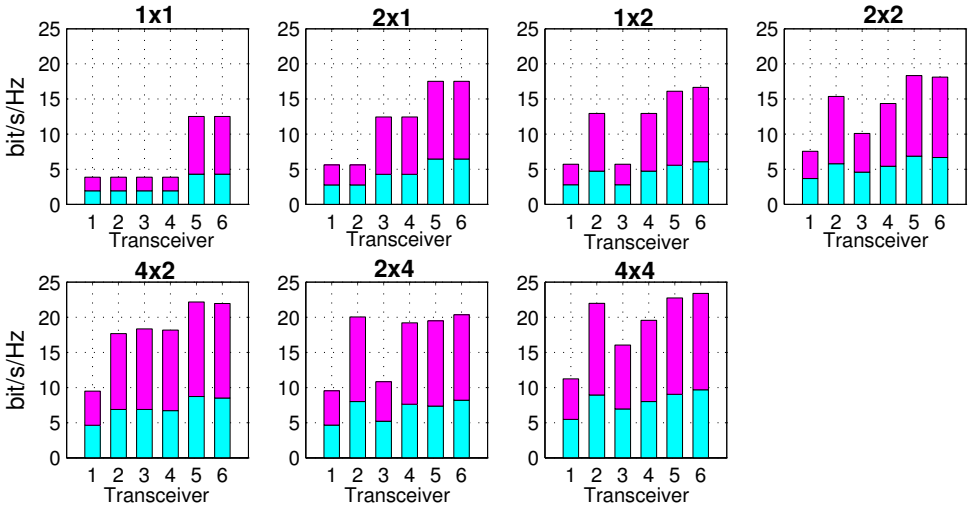

Figure 4.6.: Mean sum rate for different transceiver techniques, depending on the number of Tx and Rx antennas $M \times N$. Perfect CSI (pink) vs modeled CSI errors (cyan) for SNR = 20 dB, isolation $\Delta_{1,2} = 0$ dB, $P = 10$. The transceiver indices on the x-axis are (1) Ebf+MRC, (2) Ebf+IRC, (3) Gebf-CS+MRC, (4) Gebf-CS+IRC, (5) Gebf-JT+MRC, (6) Gebf-JT+IRC.

required at the MS side. Only with low mobility and low delay spreads, e.g. a pico or femto cell scenario, might still have such a large coherence bandwidth and -time that the channel variation is negligible and simpler channel estimators are beneficial enough. The results in Fig. 4.5 show that impaired CSIT (dashed line) generates a certain fixed rate gap over perfect CSI (solid line). Note that impaired CSIT means here perfect channel knowledge at the receive side (CSIR); the impairments stem only from feedback quantization and delay (channel aging). As there is no dependency on the number of pilots in those two types of curves, the resulting sum-rate over P is a constant straight line. Note that the gap between perfect CSI and imperfect CSIT is around 10% for GEbf-JT while it is negligible for Ebf+MRC. Again, as above, the CoMP case shows more sensitivity to inaccurate channel knowledge. When impairments for both CSIR and CSIT are taken into account, we can observe that with increasing P, the sum-rate clearly increases and approaches the dashed line (impaired CSIT, perfect CSIR). E.g. from $P = 10$ to $P = 100$ we can observe an almost 30% increase of sum-rate.

Fig. 4.6 shows the impact of number of antennas at transmit and receive side for SNR = 20 dB, isolation $\Delta_{1,2} = 0$ dB, $P = 10$. For 1×1, meaning $N = 1$ Tx per BS cell and $M = 1$ Rx per MS, only JT can bring any benefits. It is by factor of 3 better in sum-rate for perfect CSI and less than a factor of 2 with impaired CSI. In the 2×1 case, sum-rate gains over non-coordination by CS/CB are roughly half of the JT gains. With larger number of receive antennas, IRC with non-coordinated transmitters shows to be a very powerful technique.

Fig. 4.7 illustrates the effect when users are located at the cell edge (isolation $\Delta_{1,2} = 0$ dB) or are positioned at more inner cell positions (increasing isolation $\Delta_{1,2}$). As the perfect CSI case is considered: For inner cell positions, with e.g. $\Delta_{1,2} = 30$ dB, the performance of all considered schemes is converging. This is because interference becomes negligible, so algorithms which deal with interference provide no additional benefits. Also JT cannot bring notable antenna

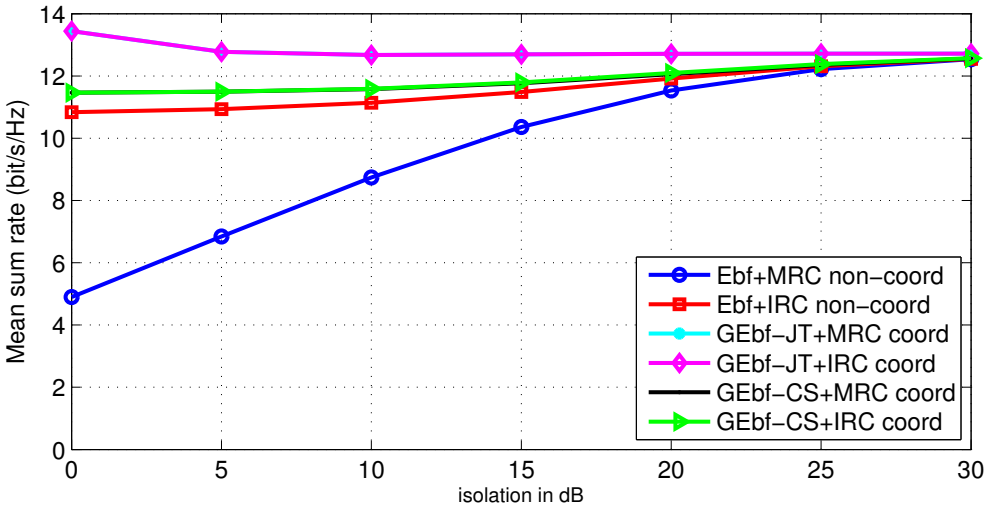

Figure 4.7.: Mean sum rate vs cell isolation $\Delta_{1,2}$: Comparison of transceiver algorithms with perfect CSI for SNR = 20 dB.

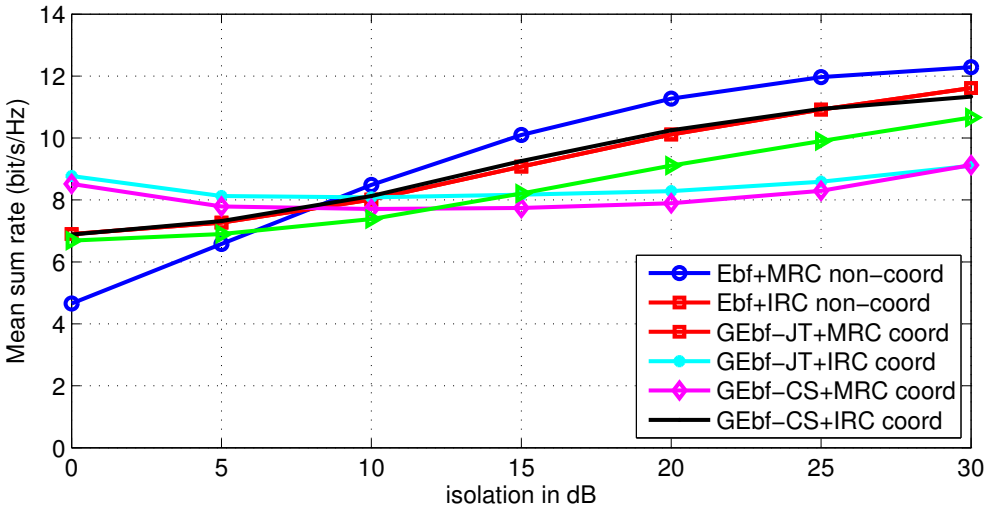

Figure 4.8.: Mean sum rate vs cell isolation $\Delta_{1,2}$: Comparison of transceiver algorithms with impaired CSI for SNR = 20 dB and $P = 10$.

gains from coherent precoding, as the link attenuations for the supporting cells are simply too strong to generate notable receive signal power boosts. The non-coordinated transmitters with MRC (Ebf+MRC) now show a strong mean sum rate decay for user positions getting closer to the cell edge[5]. GEbf-CS or Ebf+IRC show a minor performance loss at cell outer parts compared to cell inner parts. This is because precoder or combiner degrees of freedom are now used to suppress interference instead of boosting the useful signal level. JT even becomes better at the cell edge, as all antennas of the CoMP cluster become equally important, increasing the coherent superposition gains, or antenna gains, generated by the precoder.

[5]User position do not have to be geometrically exactly inbetween cells for this effect, the parameter $\Delta_{1,2} = 0$ dB simply results in equal average receive signal power levels, which depend on shadowing etc.

Fig. 4.8 represents the imperfect CSI case: here the behavior drastically changes. At the cell edge, the interference-aware transceiver techniques still are better, but with smaller relative mean sum rate gains. For mobiles closer to the cell inner part, at $\Delta_{1,2} \approx 8\,\text{dB}$ there is an intersection point. From this point on, the performance order of the transceivers, simply speaking, reverses. This means that Gebf-JT, which was best performing at the cell edge, now becomes the worst transceiver choice in terms of mean sum rate. And the simplest non-coordinated scheme Ebf+MRC, worst at cell edge, becomes the best one in the cell center. This is because the simplest transceivers, also from above results, have shown to be the most robust ones towards CSI impairments, while joint transmission was the most sensitive scheme.

4.6. Conclusions and Implications on System Design

The results from the previous section have clear implications on the system design. CS/CB is mainly an option when mobiles are only equipped with low complex receivers, lacking interference awareness. Joint transmission is powerful, especially at the cell edge. With impaired channel knowledge, though, it should not be used in the cell inner parts, e.g. with isolation $\Delta_{1,2} > 8\,\text{dB}$. The results of Fig. 4.8 suggests a *switching of techniques*, using Ebf+MRC for cell-inner mobiles and, if enough backhaul bandwidth available, GEbf+JT at the cell edge[6]. In case it can be afforded to have larger numbers of antennas at the MS receiver (e.g. 4 or even more), IRC receivers become a very powerful tool for suppressing interference not even requiring coordination at the base station side. So IRC receivers are a clear competitor to DL CoMP or can complement it, where available.

In order to boost performance in presence of impaired CSI, the system design should aim to operate at the right side of the x-axis of Fig. 4.5. This means that the scheduler should provide large-enough allocation sizes. The channel estimator should be able to exploit pilots beyond coherence bandwidth and -time, as discussed in sections 4.5 and 3.4. In case a simpler channel estimator requires almost constant channels, a deployment scenario with low delay spreads and/or low Doppler spreads are desirable, e.g. pico or femto cells. User allocations in LTE-A are indicated by control channel subframe-wise for a duration of 1 ms. The results in section 4.5 led in conclusion to the following potential extension of the standard: In [DW11] a patented solution for modified signaling is provided in order to improve channel estimation performance. Forward control signaling indicates that user data allocation at the same frequency position will be repeated in one or more subsequent subframes. The MS can now buffer the entire allocation, delay the processing for those one or more subframes and execute the channel estimation based on the overall pilot set of the subsequent subframes. The improved channel quality will result in a higher probability for correctly decoding the data and thus a higher system mean sum rate. E.g. with GEbf+IRC, when relying for the channel estimation on 4 subframes instead of one, e.g. with $P = 40$ instead of $P = 10$, the mean sum rate increases by more than 20%.

[6] Users in different coordination modes, e.g. JT or non-coordinated, could use different parts of the frequency band in order to realize this.

5. Coping with Out-of-Cluster Leakage for Downlink Transmission

Chapter 4 has introduced downlink CoMP techniques for dealing with inter-cell interference inside the coordination cluster. As discussed in section 2.3, the size of practical coordination clusters will always be limited, at least due to the introduced backhaul latencies (and backhaul bandwidth consumption). So, comparable to the inter-cell interference problem, the system performance is limited, which is now called the *inter cluster-interference problem* or *out-of-cluster leakage*.

The target of this chapter is to address this downlink problem of inter-cluster interference by appropriate precoding techniques[1]. Inside the clusters, joint transmission shall be made. Each scheduler shall operate independently per cluster. After scheduling decisions are done, between clusters some small amount of information is shared on allocation decision associated with CSI. Based on this information, the schedulers can be recomputed. The framework of this selected approach (with appropriate control signaling) was filed as patent in [Wil11c]. The research results of this section were published in [Wil11b].

Existing multi-user MIMO precoding techniques like ZF, block-diagonalization [SSH04], multi-user Eigenmode-transmission [BH07] as well as the selected Generalized Eigenbeamforming approach in chapter 4 can be applied on CoMP to compute the JT antenna weights for the users inside the cluster, but cannot directly adress additional users outside the cluster. When looking at existing approaches in the literature beyond this, a technique called soft interference nulling (SIN) [NH10] is addressing a precoding weight computation problem for JT inside clusters with interference avoidance between clusters. The drawback of SIN is that, in order to compute the precoding weights, a sequence of convex optimization problems has to be solved. Thus the resulting complexity appears to be prohibitive for fast real-time usage, while the solution itself is still not optimal, as it is an approximative approach.

This chapter aims at a closed-form solution for the precoder which can be easily used in practice. As we have seen from chapter 4 that the accuracy of channel knowledge plays an important role in judging the performance of transceiver algorithms. In the investigations of this chapter CSI impairments will be taken into account as well.

[1] In the uplink the processing cluster for receive combining could be centered individually per user - this means that the out-of-cluster leakage problem is less severe. In the downlink this is not applicable, as transmit precoding is associated with radiated power which cannot be arbitrarily increased per base station, and so JT involves a certain common cluster shared by all users.

5.1. System Model and Considered Scenario

We follow the notation of section 4.1 for stacking multiple MIMO channel matrices to a CoMP cluster channel (4.1) and use the definition of cell isolation from (4.2).

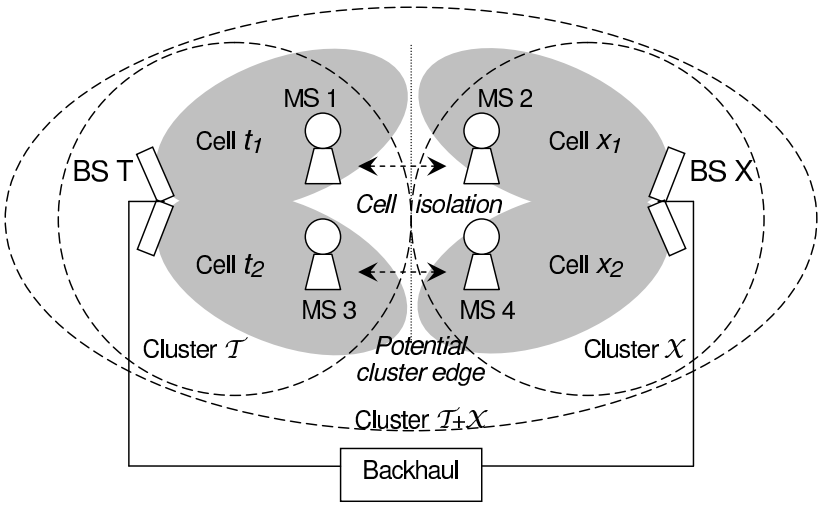

Figure 5.1.: Cluster-edge scenario example with two clusters, $\mathcal{T} = \{t_1, t_2\}$ and $\mathcal{X} = \{x_1, x_2\}$, each consisting of a BS, having two cells (here: sectors); and one MS per cell. For benchmarking purposes a supercluster $\mathcal{T} + \mathcal{X}$ over 4 cells is considered.

We expand the scenario now by taking into account an explicit set $\mathcal{X} = \{x_1, x_2, \ldots, x_X\}$ of $|\mathcal{X}| = X$ interfering cells outside the cluster :

$$\mathbf{H}_{u,\mathcal{X}} = [\mathbf{H}_{u,x_1} | \mathbf{H}_{u,x_2} | \ldots | \mathbf{H}_{u,x_X}] \tag{5.1}$$

The receive signal vector $\mathbf{y}_u \in \mathbb{C}^{M\times 1}$ of MS u for the cluster of interest, comprising cell set $\mathcal{T}$, corrupted by the the interfering cell set $\mathcal{X}$, where $\mathcal{X}$ at current time instant serves the user set $\mathcal{V}$ in cluster $\mathcal{X}$ can be expressed as

$$\mathbf{y}_u = \mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u s_u + \sum_{\forall \mu \in \mathcal{U}\setminus\{u\}} \mathbf{H}_{u,\mathcal{T}}\mathbf{p}_k s_k + \sum_{\forall \nu \in \mathcal{V}} \mathbf{H}_{u,\mathcal{X}}\mathbf{p}_\nu s_\nu + \mathbf{z}_u. \tag{5.2}$$

Note that $\mathbf{H}_{u,\mathcal{X}}$ covers composite CoMP-channels of the set of out-of-cluster cells $\mathcal{X}$ to the MS indexed u. The first summand in (5.2) represents the useful signal part. The second summand is the intra-cluster interference, which shall be handled by joint transmission. The third summand is inter-cluster interference. The fourth contains noise and not explicitly modeled additional interference contributions, its covariance - as in chapter 4 - is modeled as "spatially white" $\mathrm{E}\left\{\mathbf{z}_u \mathbf{z}_u^H\right\} = \sigma_z^2 \mathbf{I}$.

Our particular focus now lies on the additional third summand. This out-of-cluster leakage has to be considered by the precoder of BS in cluster $\mathcal{T}$ which has to be designed within this chapter. Additionally, the receive combiner $\mathbf{g}_u \in \mathbb{C}^{M\times 1}$ of user u can take this into account as well. Like in chapter 4 we consider linear precoders and combiners.

In order to compute the SINR, a similar approach as in chapter 4 can be used: The signal is decomposed like in (4.4), leading to the following SINR $\gamma_u^{[Out]}$ at the receive combiner output

$$\gamma_u^{[Out]} = \frac{\left|\mathbf{g}_u^H \mathbf{H}_{u,\mathcal{T}} \mathbf{p}_u\right|^2}{\sum_{\forall \mu \in \mathcal{U} \setminus \{u\}} \left|\mathbf{g}_u^H \mathbf{H}_{u,\mathcal{T}} \mathbf{p}_k\right|^2 + \sum_{\forall \nu \in \mathcal{V}} \left|\mathbf{g}_u^H \mathbf{H}_{\nu,\mathcal{X}} \mathbf{p}_\nu\right|^2 + \left\|\mathbf{g}_u^H\right\|^2 \sigma_z^2} \tag{5.3}$$

where the expectation is done w.r.t. transmitted symbols and noise realizations, assuming unit norm data symbols, which are uncorrelated between users. Likewise, the achievable mean sum rate over all served users within the considered set of cells can be computed according to Shannon as

$$R = \sum_{\forall u \in \mathcal{U}} \log_2 \left(1 + \gamma_u^{[out]}\right). \tag{5.4}$$

Fig. 5.1 depicts a scenario for the example of $|\mathcal{T}| = |\mathcal{U}| = |\mathcal{X}| = |\mathcal{V}| = 2$. This means, we consider two clusters with two cells each. Within each cell, one user is served, sharing the same time-frequency resources as all the other cells.

In section 4.5, a single cell isolation parameter $\Delta_{1,2}$ was used in order to describe the 2-cell scenario of chapter 4. The setting of Fig. 5.1 targets two intra-site clusters where the two cells each are the sectors of the respective site. Inside each cluster, the most relevant point for CoMP is the case where $\Delta_{t1,t2} = \Delta_{x1,x2} = 0\,\text{dB}$ - thus users having equally strong links for cells t1 and t2 in cluster $\mathcal{T}$ (or in cluster $\mathcal{X}$: equally strong links to cells x1 and x2). The worst case for inter cluster interference is when $\Delta_{t1,x1} = \Delta_{t2,x2} = 0\,\text{dB}$. Thus the users of interest are in an area where they receive signals of the four cells equally strong. We will address this operation point in the simulation results in section 5.3, as well as situations where the users are located more in the cluster inner parts. In order to simplify the investigation, a single cell isolation parameter Δ_C is used to modify the receive power differences between cells belonging to different clusters. The above described example for receiving 4 cells equally strong is created by $\Delta_C = 0\,\text{dB}$, while 4 users located each at a very cell inner point means e.g. $\Delta_C = 30\,\text{dB}$, with $\Delta_C = \Delta_{t1,x1} = \Delta_{t1,x2} = \Delta_{t2,x2} = \Delta_{t2,x1}$. (Note that $\Delta_{t1,t2} = \Delta_{x1,x2} = \text{const.} = 0\,\text{dB}$.)

5.2. Transceiver Design for Joint Transmission with Out-of-Cluster Leakage Awareness

5.2.1. Discussion of Precoder Options

From the perspective of one user u it will be beneficial if its associated BS precoder $\mathbf{p}_u$ maximizes its SINR (5.3) which results in maximizing its mean sum rate (5.4) as well. The problem is, like in chapter 4, that the choice of precoder weights for one user impacts the SINR for all other users, leading to a coupled optimization problem. In chapter 4 it was already explained why this is not tractable in a practical system. The problem in this chapter is even more complex, as we want to consider neighbor cells outside the cluster as well.

As mentioned above, a suboptimal technique called soft interference nulling (SIN) [NH10] is not taken into account in this thesis due to prohibitive complexity. The most commonly used

suboptimal technique, ZF, also cannot be used properly, as it soon faces its limits of supported number of users because of limited degrees of freedom by limited numbers of antennas, as discussed in the context of MU-MIMO e.g. in [STS07]. The reason for exceeding the manageable degrees of freedom for ZF in CoMP is that we consider also interference-victim-users $\mathcal{V}$ outside our own joint transmission clusters.

5.2.2. Leakage-Aware Precoder beyond Cluster Borders

The considered approach in this thesis builds upon the principle of maximizing signal-over-leakage-plus-noise ratio (SLNR). In chapter 4, this principle was used either for joint transmission or for interference avoidance by CS/CB. Now in this chapter 5 we combine these two approaches and design a precoder which does joint transmission inside the cluster in conjunction with interference avoidance between the clusters, based on CS/CB.

The optimization task (4.13) for the JT precoding weights $\mathbf{p}_u$ is used as a starting point and will be extended. The receive signal power $||\mathbf{H}_{u,\mathcal{T}}\mathbf{p}_u||_2^2$ and the intra-cluster leakage power $L_{\mathcal{U}} = \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}} ||\mathbf{H}_{\mu,\mathcal{T}}\mathbf{p}_u||_2^2$, as well as the summed power of $\mathbf{z}_u$, thus $M\sigma_z^2$, are the same as in (4.13). Additionally, we want to avoid leakage on a set of relevant users[2] $\mathcal{V}$ outside the cluster. This leakage can be expressed as $L_{\mathcal{V}} = \sum_{\forall\nu\in\bar{\mathcal{V}}} \|\mathbf{H}_{\nu,\mathcal{T}}\mathbf{p}_u\|^2$. Thus, the optimization task of chosing the precoding weights $\mathbf{p}_u$, subject to $||\mathbf{p}_u||_2^2 = 1$ can be formulated as

$$\arg\max_{\mathbf{p}_u} \frac{\mathbf{p}_u^H \mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}} \mathbf{p}_u}{M\sigma_z^2 + L_{\mathcal{U}} + L_{\mathcal{V}}} \tag{5.5}$$

As in chapter 4, this is a generalized Eigenvalue problem, which fortunately can be translated to a simple Eigenvalue problem[3] for the matrix $\mathbf{C}_{\mathrm{SLNR}}^{-1}\mathbf{A}_{\mathrm{SLNR}}$, where $\mathbf{A}_{\mathrm{SLNR}} = \mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}}$ and

$$\mathbf{C}_{\mathrm{SLNR}} = M\sigma_z^2\mathbf{I} + \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}} \mathbf{H}_{\mu,\mathcal{T}}^H \mathbf{H}_{\mu,\mathcal{T}} + \sum_{\forall\nu\in\mathcal{V}} \mathbf{H}_{\nu,\mathcal{T}}^H \mathbf{H}_{\nu,\mathcal{T}}. \tag{5.6}$$

The solution is the Eigenvector, corresponding to the largest Eigenvalue $\mathbf{x}_{\max}$ of $\mathbf{C}_{\mathrm{SLNR}}^{-1}\mathbf{A}_{\mathrm{SLNR}}$:

$$\mathbf{p}_u \propto \mathbf{x}_{\max}\left\{\left(M\sigma_z^2\mathbf{I} + \sum_{\forall\mu\in\mathcal{U}\setminus\{u\}} \mathbf{H}_{\mu,\mathcal{T}}^H \mathbf{H}_{\mu,\mathcal{T}} + \sum_{\forall\nu\in\mathcal{V}} \mathbf{H}_{\nu,\mathcal{T}}^H \mathbf{H}_{\nu,\mathcal{T}}\right)^{-1} \mathbf{H}_{u,\mathcal{T}}^H \mathbf{H}_{u,\mathcal{T}}\right\}. \tag{5.7}$$

This precoder design algorithm offers the following advantages:

- It can be described in closed-form.
- The weights can be computed individually per user. The optimization is fully decoupled between different user's weights.

[2] Relevant means that the interference has non-negligible power clearly impacting the throughput performance of this respective user.

[3] $\mathbf{C}_{\mathrm{SLNR}}$ is invertible as it is positive definite for $\sigma_z^2 > 0$, see explanation in chapter 4 for matrix $\mathbf{B}_{\mathrm{SLNR}}$, which still holds for $\mathbf{C}_{\mathrm{SLNR}}$.

- It has no hard degrees of freedom limitation and can work with arbitrary numbers of considered users.
- It has the same complexity order as ZF[4].

5.2.3. System Architecture and Control Signaling Aspects

The selected approach takes care that each cluster can be run by independent schedulers. The precoders act per-cluster as the ones in Fig. 4.1, besides a potential recomputation, based on neighbor cluster information. For dealing with the out-of-cluster leakage, a minimal information exchange is required between different clusters. As this information exchange happens via the backhaul, the two major bottlenecks there are latency and data rate. The considered approach [Wil11c] takes care of this by just requiring one message per cluster-to-cluster connection, instead of several steps of request, response etc., and one post-scheduling precoder adaptation.

In the scenario depicted by Fig. 5.1, the following steps are taken to make out-of-cluster leakage suppression work:

- All MSs measure the channels to the different cells. Also outside the cluster, first a coarse average receive power level is required, which can e.g. be obtained from handover measurements existing in LTE / LTE-A systems.
- In case the cell of the neighbor cluster is within a certain range[5] of the cells in the own serving cluster in terms of receive power level, then a channel estimation is required.
- All MSs quantize the CSI, e.g. using methods from appendix F, of the own serving cluster and relevant selected cells from neighbor clusters.
- With this CSIT, the processing unit of the cluster also has knowledge about potential interference victims outside this cluster.
- The role of being an interference victim depends on average power levels, which only changes on a slow timescale. Thus the clusters update their neighbor clusters on a slow time scale on the list of potential out-of-cluster-interference victim mobiles.
- Now usual scheduling happens within the cluster. Scheduling typically involves an estimation of the supported rates of the radio link (in practical systems for the choice of the modulation and coding scheme). This typically requires also a computation of the precoder weights. Note that those weights will be refined later due to the subsequent steps for out-of-cluster leakage reduction.
- In case the cluster processing unit schedules a MS from the list of out-of-cluster-interference victim mobiles, it informs the neighbor cluster associated to this MS on the corresponding

[4]This is because the basic algorithmic principle still follows [STS07], where the complexity of ZF and SLNR was both assessed for the MU-MIMO case as cubic to the number of transmit antennas which is in our CoMP case $\mathcal{O}\left\{(NT)^3\right\}$

[5]A few numbers can be deduced from the simulation results in section 5.3.

time-frequency resources (PRBs) where this MS is allocated. Additionally the respective CSIT for these resources is added which was obtained from the previous report of the MS[6]

- This scheduling information is timed such that the neighbor cluster has enough time to recompute its precoding weights based on this information. In case the backhaul latency is T_ν and the expected precoding computation time is T_C, the cluster signals its single unidirectional message time $T_\nu + T_C + T_\Delta$ ahead of the beginning of the respective subframe, where T_Δ accounts for uncertainities, e.g. latency jitter or computation time variation.
- The recomputed weights lead to reduced inter-cluster interference, while fulfilling the demand that the schedulers can operate independently.

It is important to notice that this approach requires no exchange of user data between clusters, just CSIT and scheduling information. Hence the resulting backhaul bandwidth consumption is very moderate.

5.2.4. IRC Receiver beyond Cluster Borders

In chapter 4, the IRC algorithm was formulated inside the cluster (4.18). Clearly, the interference covariance matrix, which is used to suppress unwanted signals can be expanded to include signals coming from neighbor clusters. This leads to the following receive weights:

$$\mathbf{g}_u^H \propto \left(\left(\sigma_z^2 \mathbf{I} + \sum_{\forall \mu \in \mathcal{U} \setminus \{u\}} \mathbf{H}_{\mu,\mathcal{T}} \mathbf{p}_u \mathbf{p}_u^H \mathbf{H}_{\mu,\mathcal{T}}^H + \sum_{\forall \nu \in \mathcal{V}} \mathbf{H}_{\nu,\mathcal{X}} \mathbf{p}_\nu \mathbf{p}_\nu^H \mathbf{H}_{\nu,\mathcal{X}}^H \right)^{-1} \mathbf{H}_{u,\mathcal{T}} \mathbf{p}_u \right)^H . \quad (5.8)$$

This kind of IRC receiver now has out-of-cluster leakage awareness, without requiring base station coordination. Its potential limitation of interference suppression capabilities is due to the fact that MSs are usually not equipped with a large number of receive antennas.

5.3. Simulation Results

Fig. 5.2 shows the simulation results with perfect CSI for different SNR points, where the noise reflects AWGN, as well as interference from outside the considered clusters. The baseline transmit precoder approach is non-coordinated transmission, using Eigenbeamforming (black), as described in section 4.3.2. One can observe that its performance saturates already for low SNR at around 2.4 bit/s/Hz with MRC receivers (solid line). The IRC receivers (dashed line) can improve the situation, but even awareness of neighbor cluster interference cannot prevent the strong rate limitation, which is 30% above the MRC results. This is due to the limited number of receive antennas (2 Rx in this case) which do not provide enough degrees of freedom in order to remove the inter-cell and inter-cluster interference.

[6] Alternatively, channel prediction can address potential channel aging effects, see appendix F.

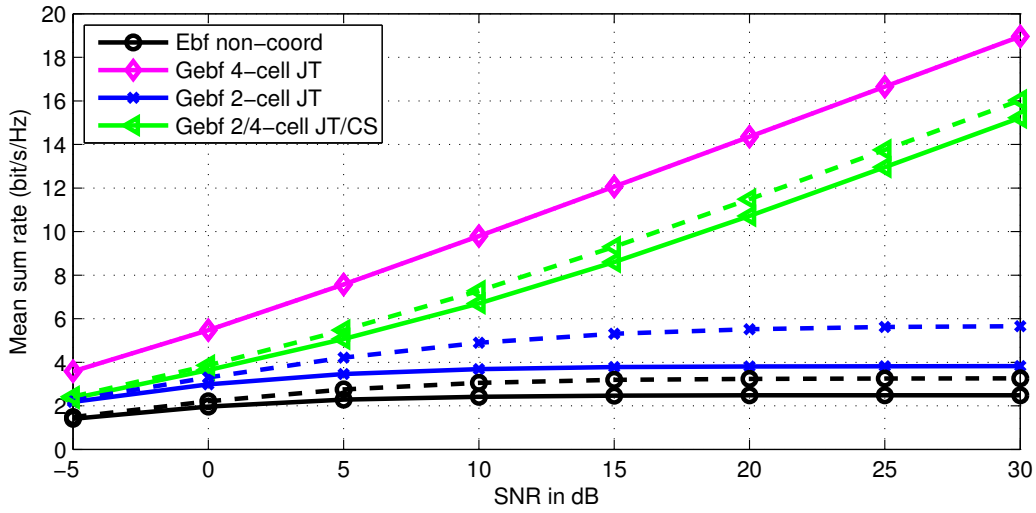

Figure 5.2.: Mean sum rate vs SNR: Comparison of different transceiver strategies for perfect CSI, isolation $\Delta_C = 0\,\text{dB}$, $N = 4$ Tx antennas per cell, $M = 2$ Rx antennas per MS. MRC is depicted in solid lines, IRC in dashed lines.

As an upper bound the 4-cell joint transmission case is provided ("Gebf 4-cell JT"), where generalized Eigenbeamforming, as in section 4.3.3 is used. This represents the case where a "supercluster" is formed out of the two clusters. As this does not fulfill the target of independently operating clusters, it is just a benchmark for achievable performance. The mean sum rate is growing with increasing SNR, as the precoder can avoid all inter-cell and inter-cluster interference, and e.g. factor 6 higher than Ebf+MRC is achieved. 4-cell GEbf performs identical with MRC and IRC as the transmit side already handles the inter-cell interference.

Let us now have a look at the 2-cell joint transmission case ("Gebf 2-cell JT"), when inter-cluster interference is not adressed. Clearly it is superior to non-coordinated transmission, but comes along with a large performance gap to 4-cell JT. The approach (5.7) described in section 5.2.2 enables 2-cell joint transmission while taking care of the leakage for all 4 cells. Thus it can be seen as coordinated scheduling/beamforming towards the neighbor cluster. It is depicted by "Gebf 2/4-cell JT/CS". Due to the improved precoder design, the gap between "Gebf 2-cell JT" and "Gebf 4-cell JT" now clearly can be reduced by using "Gebf 2/4-cell JT/CS".

Fig. 5.3 includes impaired CSI into the performance, based on the modeling parameters in section 4.4. With these conservative parameters for number of usable pilots P, feedback quantization and delay losses (MSE of -12 dB), the ranking of the algorithms still remains unchanged but the performance differences become smaller and also for the 4-cell JT case the mean sum rate saturates over SNR.

Fig. 5.4 depicts the case when the isolation parameter Δ_C is varied. Thus, the left side of the x-axis represents the situation when the mobile stations of interest are located at the cluster edge, in terms of receive power levels. The right side of the x-axis represents cluster inner locations. The black curve shows that Ebf+MRC is intra-cluster interference limited. These problems can be reduced when using an IRC receiver (dashed line). In the cluster inner part (high Δ_C), no degrees of freedom need to be spend on neighbor cluster interference. Thus

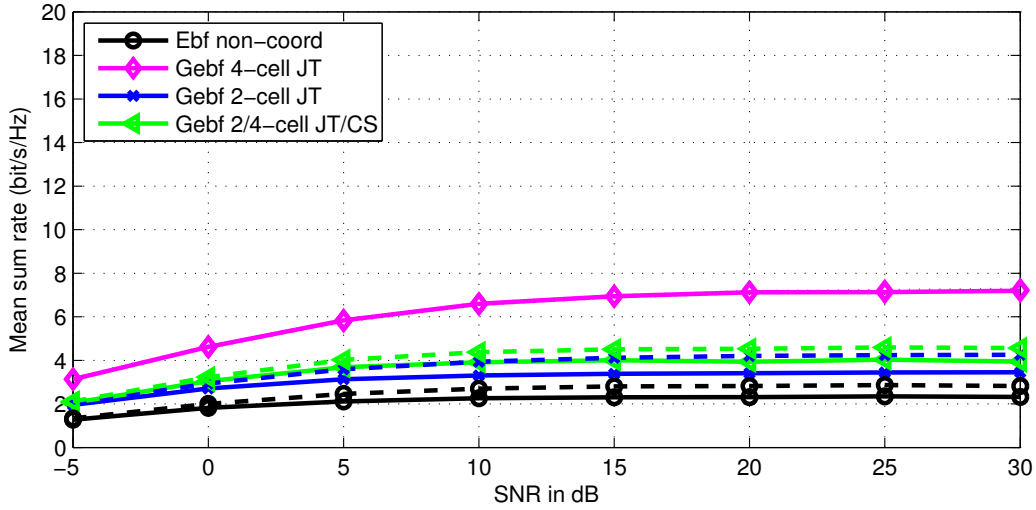

Figure 5.3.: Mean sum rate vs SNR: Comparison of different transceiver strategies for impaired CSI, $P = 10$, isolation $\Delta_C = 0$ dB, $N = 4$ Tx antennas per cell, $M = 2$ Rx antennas per MS. MRC is depicted in solid lines, IRC in dashed lines.

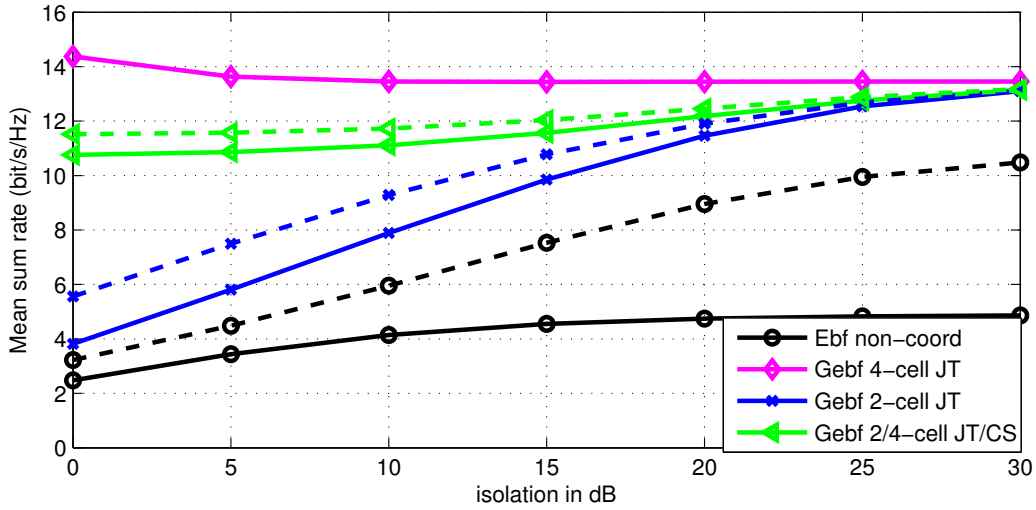

Figure 5.4.: Mean sum rate vs Isolation Δ_C: Comparison of different transceiver strategies for perfect CSI, SNR 20 dB, $N = 4$ Tx antennas per cell, $M = 2$ Rx antennas per MS. MRC is depicted in solid lines, IRC in dashed lines.

already 2 Rx antennas provide a good interference suppression performance in the inner parts of a 2-cell cluster.

2-cell JT (blue) works well inside the cluster but shows its weakness at the cluster edge (for low values of Δ_C). Here the IRC receiver (dashed blue) can help, but brings only limited performance gain due to only 2 Rx antennas. The proposed precoding algorithm for coping with out-of-cluster leakage (green) (5.7) is much closer to the 4-cell JT case than to the 2-cell JT case.

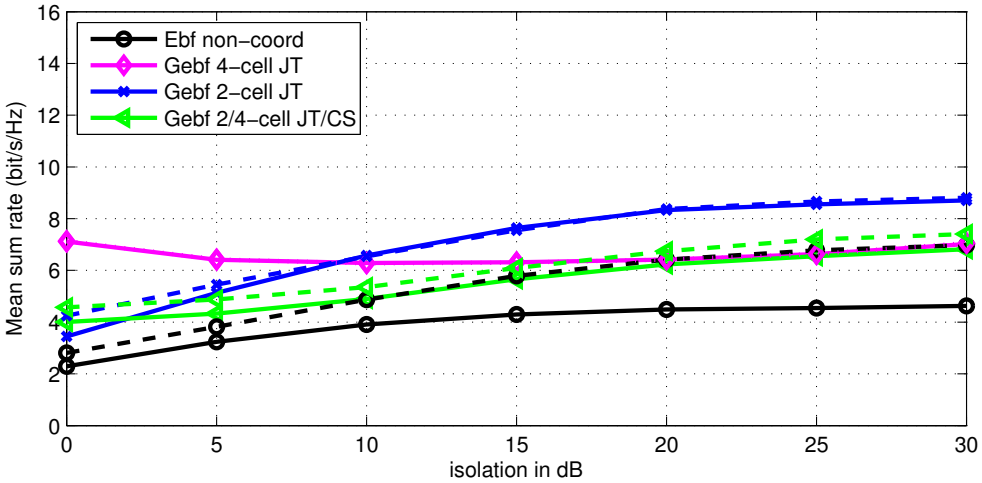

Figure 5.5.: Mean sum rate vs Isolation Δ_C: Comparison of different transceiver strategies for impaired CSI, $P = 10$, SNR 20 dB, $N = 4$ Tx antennas per cell, $M = 2$ Rx antennas per MS. MRC is depicted in solid lines, IRC in dashed lines.

Fig. 5.5 looks at the same situation when the CSI is impaired. Here the gains of the leakage-aware precoding (5.7) get much smaller, e.g. around 20% at the cluster edge for MRC receivers and vanish for $\Delta_C > 2\,\text{dB}$. They may become even negative. This is because of inaccurate interference suppression, being more harmful than helpful with poor CSIT. Note that the best strategy then is to use pure intra-cluster JT. These results provide a hint for a switching point of different precoding strategies, which was mentioned in section 5.2.3: When the receive power levels of the neighbor cell clusters are less than e.g. 2 dB below the own cell cluster, the out-of-cluster leakage suppression scheme (5.7) should be used. For users which are "further inside" their own clusters, no extra information exchange is required. Note that this number depends on the exact receiver and channel estimation capabilities of the MS.

5.4. Conclusion

This chapter has discussed the weaknesses of CoMP at cluster edges. New precoding algorithms have been designed to deal with this problem. For MS inside clusters, joint transmission is made, while between different clusters interference avoiding precoding is used. The approach uses the criterium of maximizing SLNR. The signaling concept is described, supporting a "backhaul-friendly" approach in terms of required rate consumption and supported latency.

With impaired channel knowledge, the performance gains from this leakage-aware approach are significantly smaller, but still benefits are there for MS which are directly at the cluster edge. Inner cluster mobiles from neighbor clusters should be ignored when taking into account the leakage to neighbor clusters. This saves backhaul bandwidth and avoids losses from inaccurate channel knowledge.

6. Outlook to CoMP in Next Generation Wireless Systems

Finally a brief outlook on future trends is given: Today, we are at the dawn of 5th generation wireless systems [WJK+14]. Projects like METIS [MET] or 5GNOW [5GN], where the author of this thesis is involved, are driving collaborative research towards 5G. This future generation has a broad range of targets, like better supporting the Internet of Things, lower latency and efficient coexistence of small packet services and ultra high bit rate traffic. Building upon the lessons learned from 4G systems, 5G systems can start integrating CoMP right from the beginning into its design. The trends to cell densification to achieve higher and higher spectral efficiencies per area intensifies the inter-cell interference problem and thus will create high synergies for using CoMP in conjunction with heterogeneous networks. This prevails for both current 4G as well as future generation systems.

In 5GNOW, the focus lies on new waveforms, like Filter-bank Multicarrier (FBMC) [FB11] and Generalized Frequency Division Multiplexing (GFDM) [FKB09]. Very recently, a new waveform contender, which is called Universal Filtered Multicarrier (UFMC), also known as Universal Filtered (UF)-OFDM, has entered the discussion [WSC14][SW14][WS13]. The benefits in CoMP are that the improved spectral properties of the waveform allow for relaxing requirements, such as local oscillator accuracy of the devices, due to better robustness against CFO [VWS+13]. Research in this area is ongoing, helping to realize a broad range of device classes in 5G.

Another emerging vision for the future are *user-centric*, "cell-less" mobile systems [BAD+12]. They will go very well together with CoMP. Additionally, data centers and *cloud processing* are more and more upcoming, e.g. to save costs for operators. This centralization of processing is a natural fit for using CoMP and the centralized architecture can directly support the joint channel estimation techniques investigated in this thesis.

All those trends show that CoMP will play an important role also in the future.

7. Overall Conclusion

This work derived practical algorithms which cope with realization challenges of coordinated multi-point (CoMP) transmission and reception, thus distributed antenna systems were enabled. A key point is to obtain the channel knowledge in order to realize potential CoMP gains. The focus of this thesis is on uplink pilot-based, linear channel estimation, close to the 3GPP LTE-A standard. Chapter 3 addresses the issue of multi-cell channel estimation under the demanding particularities of CoMP, like low SINR and multiple users. This thesis also investigates the often neglected issue of obtaining the required parameters for advanced channel estimation. It is shown that conventional methods require very long observation times for collecting the statistical parameters, which is often prohibitive in real wireless systems. An improved covariance estimation approach called *shrinkage*, known from financial mathematics, is applied here in the CoMP context to deal with this problem. It was shown in section 3.2 that shrinkage could significantly reduce the required observation times for multi-cell single-user channel estimation.

Regarding multi-user, multi-cell channel estimation, it has been shown that conventional basic channel estimation approaches, e.g. based on least squares methods, fail under the difficult conditions of CoMP. Thus a multi-user matrix Wiener filter (MWF) has been formulated in section 3.3 of this thesis, including a simplification option, which is close to perfect channel knowledge, when the statistical parameters are known.

Standard-compliant pilot sequence assignments to users located across multiple cells are compared to potential orthogonal alternatives beyond the current standard. Separate (thus per-user) channel estimation, suitable e.g. for distributed processing implementation, is compared to joint channel estimation (for all users). The separate channel estimators clearly benefit from orthogonal sequences, but the joint MWF can handle both non-orthogonal and orthogonal sequences almost equally well. As a result, a modification of the standard is in this case not required, provided that the base stations are using an advanced channel estimation approach.

Then section 3.4 deals with the parameter estimation challenge for multi-user, multi-cell channel estimation. This problem is difficult, as an advanced channel estimator, like the MWF, requires extensive statistical knowledge to separate the mutually interfering user signals well enough. As a challenging prerequisite, user-individual parameter collection is required, as in practical cellular systems, due to dynamic scheduling and non-contiguous data traffic patterns, the sets of active users are strongly varying over time and frequency. The elaborated solution for this problem uses a multi-stage channel estimator. In a first stage, no-prior knowledge is required. Simple estimators are gathering information on path loss. The noise variance is extracted using subspace-based methods. Subspace properties, such as the frequency-shift-invariant nature of the channel covariance, are exploited to reduce the required observation time. In a second stage, a simplified multi-user channel estimator, using a small processing window within coherence bandwidth, provides a first channel estimate, which allows to compute or track the channel

covariance matrices of the users individually. Again, methods are introduced to shorten the required observation time. This statistical knowledge now enables the third stage: The usage of a MWF, exploiting almost the entire number of pilots of the user allocation, beyond typical coherence bandwidth and -time. The multi-stage channel estimation method has been shown to deliver a performance close to perfect channel knowledge. Even for large CoMP cluster sizes of 7 cells, the SINR at the output of an IRC combiner is with 1 dB close to perfect channel knowledge.

In order to abstract the assessment of channel estimation, a model has been derived in section 3.5. This model can be used for both uplink and downlink. For downlink transmission, feedback quantization and delays of the channel information have to be taken into account. This was done based on work discussed in appendix F, which was already implemented and successfully shown in a demonstrator. The introduction of a channel estimation error model is very helpful for dealing with computational-intensive system simulations.

Chapter 4 now applies this CSI model for downlink closed-loop investigations on different transceiver and coordination strategies. A precoder for CoMP is formulated, based on the signal-over-leakage-plus-noise criterion known from MU-MIMO, which outperforms zero forcing precoding and provides more flexibility for using it. The investigations show that pure IRC receivers without coordination at the transmit side can be powerful competitors to CoMP method. When impaired CSI is considered, the degradation for CoMP is more severe than for non-CoMP, due to the poor accuracy of channel knowledge at the cross-links. This leads to the fact that schemes like joint transmission are only beneficial for cell edge users and should not be used, when the receive power levels for different cells are larger than around 8 dB. Coordinated beamforming (CB) is helpful for shifting complexity from MS to BS, but once the MS has an IRC receiver, CB does not add much improvement.

Chapter 5 discusses the downlink cluster edge problem. Inter-cluster interference limits the performance of CoMP systems. A novel precoder is introduced which enables joint transmission within clusters and avoids interference between clusters. Its performance is assessed, showing very large gains for perfect CSI. Once impaired CSI is considered, this approach should only be used for users which have receive power levels from neighbor cluster cell of up to 2 dB within the serving cluster cells.

As lessons learned from chapter 4 and 5 we summarize “CoMP is sensitive to channel knowledge”. As a consequence CoMP should operate with sufficiently large user allocations in time and frequency to provide more pilots for channel estimation, which can be exploited by advanced channel estimators, addressed in chapter 3. Potential signaling extensions to the standard [DW11] could provide additional knowledge. E.g. the allocation will be repeated at the same frequency position also in the next subframe, which boosts channel estimation performance. For precoding in CoMP, appropriate advanced feedback reporting schemes are required (see appendix F), which are not yet supported by the standard.

To sum up, this thesis has investigated realization problems of CoMP transmission and reception in the framework of 3GPP LTE-A and beyond, and provides algorithmic solutions to make this technology feasible to operate in networks with inter-cell interference. As a result the proposed solutions lead to a much more fair and balanced user experience over the entire cell area for mobile communication systems.

A. Uplink LTE-A Signal Structure

The investigations on uplink channel estimation in chapter 3 are building upon the 3GPP LTE-A standard. Its physical channels and modulation schemes are described in [3GPc]. This appendix contains some details (similarly summarized in [NRWtB11]) on the LTE-A uplink signal formats, which form the background of the work in chapter 3. Fig. A.1 depicts the basic time-frequency resource: The physical resource block (PRB). The forward error coding for one transmission time interval (TTI) lasts over one so-called subframe, which has a duration of 1 ms. The subframe is subdivided into two slots. Each slot contains 7 OFDM symbols (if the default cyclic prefix is used[1]). Thus each OFDM symbol lasts $1/14$ ms. The subcarrier spacing is 15 kHz, leading to a symbol body length of $1/15$ ms and the remaining 4.6 μs form the cyclic prefix[2]. Those 4.6 μs create a protection against delay spreads and timing offsets corresponding to a length of 1.5 km of light speed propagation, which is well suited for macro cells in urban environments. Here, typical inter-site distances are around 500 m.

The signal format in the LTE/LTE-A uplink is called single carrier frequency division multiple access (SC-FDMA). The basis is formed by OFDM, but symbols are spread by DFT-precoding across the user allocation [3GPc] for the sake of PAPR reduction. As this does not impact the average SINR of the allocation and is also not used for the pilot symbols, the DFT-spreading of the data is not explicitly handled in the signal model of this thesis, e.g. in (3.1). Instead, the symbols s_{jkl} represent the superposition of multiple DFT-spread data symbols.

One PRB covers a width of 12 subcarriers, thus a bandwidth of 180 kHz. One PRB represents the minimum allocation size in LTE/LTE-A. For 20 MHz bandwidth, a 2048 FFT is used with a sample rate of 30.72 MSamples/s. The 1200 used subcarriers cover 18 MHz bandwidth, the band edges remain unused for fulfilling spectral mask requirments; the DC carrier is unused as well.

Uplink demodulation reference symbols (DMRS), thus the pilots, are in the center of each slot, covering the entire frequency range of the allocation. Small allocation sizes of up to 3 PRBs use computer-generated QPSK-sequences out of a table in [3GPc]. For all other medium to large allocation sizes, Zadoff-Chu sequences are used. Those sequences have the *constant amplitude zero autocorrelation* (CAZAC) property. This means the sequence amplitude is constant, the autocorrelation of a (prime-number-length) Zadoff-Chu sequence with its cyclically shifted version is zero and the cross-correlation of two different (prime-number-)length M sequences is

[1] There are 6 OFDM symbols per slot with the long CP. This parameter set is not used in the thesis as it causes a loss of spectral efficiency due to increased CP overhead.

[2] The first OFDM symbol of a subframe has a few samples longer CP in order to match to a multiple of the UMTS chip rate.

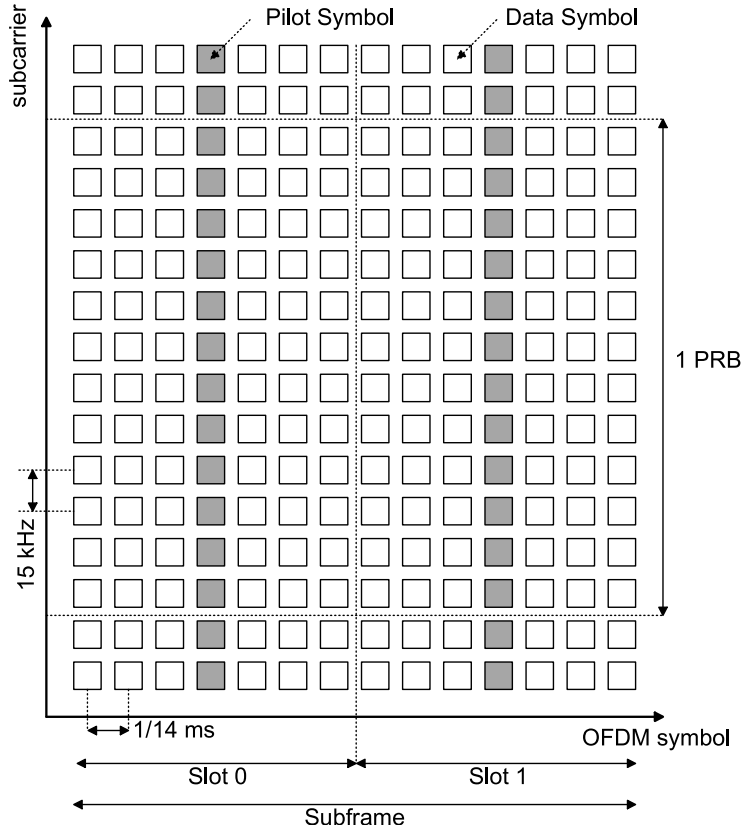

Figure A.1.: Time-frequency structure of an LTE/LTE-A uplink physical resource block (PRB); each square represents one OFDM resource element.

well defined as $1/\sqrt{M}$. Zadoff-Chu sequences' m-th element of the q-th root can be described as

$$\bar{s}_q(m) = \exp\left(-j\frac{\pi q m(m+1)}{M}\right),\ 0 \leq m < M,\ 0 < q < M \wedge gcd(M,q) = 1 \qquad \text{(A.1)}$$

3GPP uses 30 sequence groups and one or two base sequences in order to construct the roots. In order to fill the non-prime number allocation length, the sequence is extended by a modulo repetition, forming the base sequences $\mathbf{s}$. As all possible LTE allocation sizes are a multiple of 12, they are no prime number, which slightly impacts the sequence properties. The diagonal pilot matrices $\mathbf{S}_j$ of (3.4) are constructed as

$$\mathbf{S}_j = diag(\mathbf{s}) \qquad \text{(A.2)}$$

For the discussion of CAZAC sequence properties see e.g. [BLM03].

In an uplink coordinated multi-point system with multiple users, the CAZAC sequences of the users will interfere with each other. Different propagation channels and sequences stretching over frequency beyond coherence bandwidth will strongly impact the cross-correlation properties at the receive side, which is discussed in section 3.3.1.3 and depicted in Fig. 3.9.

The usage of different Zadoff-Chu sequence roots in different cells is intended in the 3GPP LTE/LTE-A standard to more or less randomize pilot sequence interference, as the original LTE design was made for single cell reception and thus single cell channel estimation as well. For MU-MIMO, the standard also offers the possibility to orthogonalize pilot sequence of different users by introducing different sequence shifts of the same root sequence for different users. With the α_{shift} parameter, the cyclically-shifted sequence is

$$s_{[\alpha_{\text{shift}}]}(m) = \exp\left(j\alpha_{\text{shift}}m\right) s(m) \tag{A.3}$$

Here, the m-th element gets a continuous multiplication, which happens due to the sequence placement of Fig. A.1 in the frequency domain. This corresponds to a cyclic shift in the time domain, still preserving the CAZAC properties. Sequences with the same roots, but different shifts are orthogonal.

A discussion on the usage of this approach over multiple cells for CoMP is done in section 3.3.1.3.

B. Terminology and Non-Zero-Mean Channels

Throughout the literature, the terminology on covariance and correlation is inconsistent. Furthermore, the terminology of a linear estimator is used inconsistently. The purpose of this chapter is to clarify the used terminology in this thesis. Additionally, for the case of non-zero mean channels, potential modifications towards affine MMSE estimators are discussed.

Terminology of Covariance and Correlation
In strict mathematical terminology the term *covariance* of random variables x and y is defined as $\mathrm{cov}(x, y) = \mathrm{E}\left\{(x - \mathrm{E}\{x\})(y - \mathrm{E}\{y\})^*\right\}$. The term *correlation* in mathematical terminology is a normalized form of the covariance, where the random variables are "standardized" by their standard deviations σ_x, σ_y; so the correlation results in $\mathrm{corr}(x, y) = \mathrm{E}\left\{\frac{x-\mathrm{E}\{x\}}{\sigma_x}\frac{(y-\mathrm{E}\{y\})^*}{\sigma_y}\right\}$.

In contrast, in the literature of signal processing, correlation is often used in a different definition, without zero-centering and normalization (or standardization), as $\mathrm{E}\left\{xy^*\right\}$. In a further variant in terminology, $\mathrm{E}\left\{xy^*\right\}$ in different signal processing literature is also defined as covariance, for instance in notable channel estimation literature, like [HKR97].

In this thesis the mathematical definition of correlation is used. For the term covariance this work follows [HKR97] and uses for the cross-covariance matrix of random vectors $\mathbf{x}$ and $\mathbf{y}$ the non-zero-centered variant $\mathrm{E}\left\{\mathbf{x}\mathbf{y}^H\right\}$. Likewise, the term auto-covariance matrix for $\mathbf{y}$ is defined as $\mathrm{E}\left\{\mathbf{y}\mathbf{y}^H\right\}$.

For zero-mean channels (the main focus of this thesis) our used term covariance coincides with the strict mathematical terminology for covariance. Furthermore it coincides with the (non-zero-centered, non-standardized) term correlation used in notable parts of the signal processing literature.

Terminology of Linear and Affine MMSE Estimators
We have constrained the class of used estimators to strictly linear ones in (3.6) and (3.7). Again, terminology in literature varies here: A notable fraction, e.g. [Kay93], uses the term linear in a wider sense, which is actually *affine*. For (3.6) this would mean a modification $\hat{h}_{ijkl} = \mathbf{w}_{ijkl}^H \mathbf{y}_i + c_{ijkl}$ including a constant c_{ijkl}. Note that this also creates a confusion in the definition of the Wiener filter, whether it is a *linear* MMSE estimator, or an *affine* one. In this thesis we have sticked to the linear variant. For zero mean channels both affine and linear variants coincide, as $c_{ijkl} = 0$, and, furthermore, affine and linear MMSE coincide with MMSE, as we use a generalized linear model, according to (3.5).

Modification Steps for Extension towards an Affine Estimator
An extension to an affine MMSE estimator can be derived in the following way: According to

[Say08] equation (4.13), non-zero mean h_{ijkl} and $\mathbf{y}_i$ can be replaced by the centered variables $h_{ijkl} - \mathrm{E}\{h_{ijkl}\}$ and $\mathbf{y}_i - \mathrm{E}\{\mathbf{y}_i\}$. Then the linear estimation problem is solved with those centered variables. Note that this additionally requires knowledge of $\mathrm{E}\{h_{ijkl}\}$. If $\mathrm{E}\{h_{ijkl}\}$ is estimated accurately enough in a further extension of the parameter estimation process, we can expect better performance from the affine solution compared to the linear one, as the estimator has additional degrees of freedom.

The main focus in this thesis lies on zero-mean channels. The derived linear estimators are also applicable to non-zero mean channels. However, in that setting affine estimators have the potential to outperform linear ones. In order to achieve full optimality in the MMSE sense, for non-zero mean channels, an affine estimator is even required. This is because, according to [Say08] Theorem 1.2, for optimality in an MMSE sense, the estimator has to be unbiased, thus $\mathrm{E}\left\{\hat{h}_{ijkl}\right\} = \mathrm{E}\{h_{ijkl}\}$. With a linear estimator $\mathrm{E}\left\{\hat{h}_{ijkl}\right\} = \mathrm{E}\left\{\mathbf{w}_{ijkl}^H \mathbf{y}_i\right\}$ it becomes clear from the system model (3.4) that for zero mean channels the mean estimate is zero (assuming zero mean noise): The linear estimator is unbiased. For non-zero mean channels this is not generally the case. The necessity of an affine MMSE estimator for full MMSE optimality can also be seen by e.g. observing [Kay93], equation (14.6).

C. Analytic 2-Cell SINR Computation

This chapter analytically computes the expected value of the SINR γ_{av} at the IRC combiner output for a CoMP set of 2 cells, containing one user each, in case of perfect channel knowledge at the receiver. So it guides the simulations of sections 3.3 and 3.4.

The following simplifications are made in order to enable the analytic treatment: This appendix focuses on narrowband channels with i.i.d. Rayleigh fading, thus the channel coefficients with time indices k and frequency indices l are complex Gaussian distributed[1]. The effective path gain factor a_{jj} for each user j to its serving cell is normalized to one, while the cross-links to supporting cells are parametrized by $\alpha = a_{ij}, i \neq j$.

We start with the signal model from section 3.3, using (3.47) to describe the receive vector in receive antenna dimension for two users (with user indices $j = 1, 2$):

$$\mathbf{y}_{kl}^{[\mathrm{B}]} = \sum_{j=1}^{2} \mathbf{h}_{jkl}^{[\mathrm{B}]} s_{jkl} + \mathbf{n}_{kl}^{[\mathrm{B}]}. \tag{C.1}$$

Note that, according to (3.3), the channel vector $\mathbf{h}_{jkl}^{[\mathrm{B}]}$ already contains path gains and transmit powers, captured in the diagonal matrix $\mathbf{A}_{\mathbf{j}}^{[\mathrm{B}]} = \mathrm{diag}\left([a_{1j}, a_{2j}]\right)$.

In (3.48) the IRC receiver is derived with $\mathbf{g}_{jkl} = \mathbf{Z}_{jkl}^{-1}\mathbf{h}_{jkl}^{[\mathrm{B}]}$, resulting in a post-combining SINR given in (3.50). In our setting with perfect channel knowledge we now compute the average SINR as a function of the effective path gain coefficients a_{ij} as

$$\gamma_{\mathrm{av}} = \mathrm{E}\left\{ \frac{\mathbf{g}_{jkl}^H \mathbf{h}_{jkl}^{[\mathrm{B}]} \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{g}_{jkl}}{\mathbf{g}_{jkl}^H \mathbf{Z}_{jkl} \mathbf{g}_{jkl}} \right\} = \mathrm{E}\left\{ \frac{\mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_{jkl}^{-H} \mathbf{h}_{jkl}^{[\mathrm{B}]} \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_{jkl}^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}}{\mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_{jkl}^{-H} \mathbf{Z}_{jkl} \mathbf{Z}_{jkl}^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}} \right\} = \mathrm{E}\left\{ \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_{jkl}^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} \right\}. \tag{C.2}$$

The expectation is done over time k and frequency l.

Let us continue this notation by focusing on the SINR of user 1; the SINR of user 2 is identical due to fully identical statistical parameters. Furthermore, for the sake of improved readability, the indices k and l are dropped, as well as the superscript $[\mathrm{B}]$. We write the channel vectors as

$$\mathbf{h}_1 = \left[h_{1,1}, h_{1,2}\right]^T;\ \mathbf{h}_2 = \left[h_{2,1}, h_{2,2}\right]^T. \tag{C.3}$$

$\mathbf{Z}_1$ contains the properties of noise and interference for user 1 perspective:

$$\mathbf{Z}_1 = \sigma_n^2 \mathbf{I} + \mathbf{h}_2 \mathbf{h}_2^H. \tag{C.4}$$

[1] The more detailed spatial channel model [WIM06] is intended for computer simulations and cannot be used reasonably for analytic computations.

In order to obtain the inverse $\mathbf{Z}_1^{-1}$ of (C.4) , required in (C.2), we use the Sherman-Morrison formula (C.13). This leads to

$$\mathbf{Z}_1^{-1} = \sigma_n^{-2}\mathbf{I} - \frac{\sigma_n^{-2}\mathbf{I}\mathbf{h}_2\mathbf{h}_2^H\sigma_n^{-2}\mathbf{I}}{1+\mathbf{h}_2^H\sigma_n^{-2}\mathbf{I}\mathbf{h}_2} = \sigma_n^{-2}\left(\mathbf{I} - \frac{\mathbf{h}_2\mathbf{h}_2^H}{\sigma_n^2+\mathbf{h}_2^H\mathbf{h}_2}\right). \tag{C.5}$$

Inserting (C.5) into (C.2) gives

$$\gamma_{\text{av}} = \mathrm{E}\left\{\sigma_n^{-2}\begin{pmatrix} h_{1,1}^* & h_{1,2}^* \end{pmatrix}\begin{pmatrix} 1-\frac{|h_{2,1}|^2}{\sigma_n^2+|h_{2,1}|^2+|h_{2,2}|^2} & -\frac{h_{2,1}h_{2,2}^*}{\sigma_n^2+|h_{2,1}|^2+|h_{2,2}|^2} \\ -\frac{h_{2,2}h_{2,1}^*}{\sigma_n^2+|h_{2,1}|^2+h_{2,2}|^2} & 1-\frac{|h_{2,2}|^2}{\sigma_n^2+|h_{2,1}|^2+|h_{2,2}|^2}\end{pmatrix}\begin{pmatrix} h_{1,1} \\ h_{1,2}\end{pmatrix}\right\} \tag{C.6}$$

By linearity of the expectation operator, we can write (C.6) as

$$\begin{aligned}\gamma_{\text{av}} \quad = \quad & \sigma_n^{-2}\,\mathrm{E}\left\{|h_{1,1}|^2\left(1-\frac{|h_{2,1}|^2}{\sigma_n^2+|h_{2,1}|^2+|h_{2,2}|^2}\right)\right\} \\ & -\sigma_n^{-2}\underbrace{\mathrm{E}\left\{-\frac{h_{1,1}^*h_{1,2}h_{2,2}^*h_{2,1}}{\sigma_n^2+|h_{2,1}|^2+|h_{2,2}|^2}\right\}}_{=0} - \sigma_n^{-2}\underbrace{\mathrm{E}\left\{-\frac{h_{1,1}h_{1,2}^*h_{2,2}h_{2,1}^*}{\sigma_n^2+|h_{2,1}|^2+|h_{2,2}|^2}\right\}}_{=0} \\ & +\sigma_n^{-2}\,\mathrm{E}\left\{|h_{1,2}|^2\left(1-\frac{|h_{2,2}|^2}{\sigma_n^2+|h_{2,1}|^2+|h_{2,2}|^2}\right)\right\}\end{aligned} \tag{C.7}$$

Here the second and third summand become zero due to the statistical independence and zero mean assumption of the channel coefficients.

Due to Rayleigh fading, the amplitudes are Rayleigh distributed with parameter σ_R with probability densities $f(x;\sigma_R) = \frac{x}{\sigma_R^2}\exp\left(-\frac{x^2}{2\sigma_R^2}\right)$. For the direct links $i=j$ with normalized path gain $\mathrm{E}\{|h_{i,j}|^2\}=1$, the coefficients $h_{1,1}$ and $h_{1,2}$ have the parameter $\sigma_R = \frac{1}{\sqrt{2}}$.[2] The cross-links $h_{1,2}$ and $h_{2,1}$ have $\sigma_R = \sqrt{\frac{\alpha}{2}}$.

Using Wolfram Mathematica [Wol], inserting the distributions in (C.7), and fixing the parameter σ_n^2 to a constant (here 0.5 for an SNR of 3 dB), we obtain:

$$\gamma_{\text{av}}(\alpha)|_{\sigma_n^2=0.5} \approx \frac{2-2.92291\alpha-1.07709\alpha^2+2\alpha^3+e^{0.5/\alpha}\left(1-3.\alpha+2\alpha^2\right)\mathrm{Ei}\left(-\frac{0.5}{\alpha}\right)}{(-1+\alpha)^2} \tag{C.8}$$

Here, $\mathrm{Ei}(x)$ refers to the exponential integral function $\mathrm{Ei}(x) = \int_{-\infty}^{x}\frac{e^t}{t}\,\mathrm{d}t$.

With $\sigma_n^2 = 0.1$ (for an SNR of 10 dB), we get:

$$\gamma_{\text{av}}(\alpha)|_{\sigma_n^2=0.1} \approx \frac{10-28.1318\alpha+8.13178\alpha^2+10.\alpha^3+e^{0.1/\alpha}\left(1-11.\alpha+10.\alpha^2\right)\mathrm{Ei}\left(-\frac{0.1}{\alpha}\right)}{(-1+\alpha)^2} \tag{C.9}$$

This relationship (C.9) of γ_{av} as a function of path gain parameter α is depicted in Fig. C.1. We observe a trade-off: Increasing α increases the interference caused by the unwanted user's signal

[2] The channel coefficients are drawn from a complex Gaussian distribution with unit variance, resulting in variance $1/\sqrt{2}$ for both real and imaginary part which corresponds to σ_R

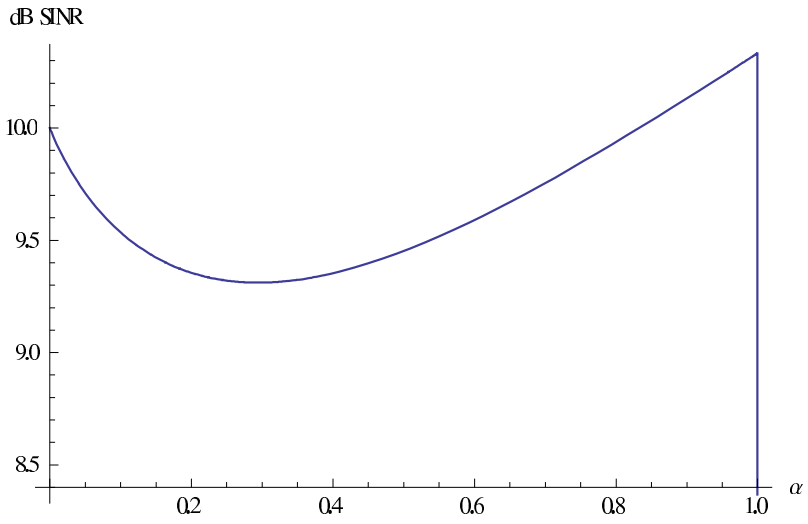

Figure C.1.: Analytic result of SINR over α at SNR=10dB.

contribution. On the other hand, the useful signal contribution caught by the supporting cell increases. The receive combiner SINR output, finding the best trade-off of interference suppression and useful signal enhancement shows a clear non-monotonic behaviour. The case $\alpha = 0$ represents the situation of an isolated cell. This is either caused by path loss, penetration loss and shadowing effects or by multiple access resource separation (ICIC "inter-cell interference coordination") between cells, like TDMA, which leads to halving the number of radio resources available per cell. Clearly, the SINR in case of $\alpha = 0$ is identical to the SNR, as the system can only make use of one receive antenna, but observes no interference. The other extreme case is $\alpha = 1$, where the signal contributions from serving and supporting cell are at same average strength for both users. This has the highest interference, with an SINR at each antenna input below 0 dB, but also the strongest useful signal combining gains. We see that the SINR at the receive combiner output is about 0.5 dB higher than the SNR of a single antenna, even in this highest interference case. This shows the performance potential of plain linear receive combining and strongly justifies CoMP, as both cells can make full use of all radio resources. Furthermore, due to antenna combining gains, both cells have higher achievable rates than if they were fully isolated from interference.

Of course, in order to uphold these gains, perfect CSI is required at the base station. Chapter 3 fully deals with the question of enabling a channel knowledge which is as close to perfect CSI as possible.

C.1. Remark on Extension to more than 2 Users

The question is: How to extend this analytical SINR computation to more than 2 cells and users? The suggested approach is described here.

The noise and interference matrix $\mathbf{Z}_1$ is now extended by one additional interferer (growing the dimensions of vectors and matrices accordingly):

$$\mathbf{Z}_1 = \sigma_n^2\mathbf{I} + \mathbf{h}_2\mathbf{h}_2^H + \mathbf{h}_3\mathbf{h}_3^H. \tag{C.10}$$

For obtaining an analytical result for the inverse, it is suggested to treat the impact of $1 + N_{\text{supp}}$ users successively by N_{supp} rank-1 updates. Start the inversion with the intermediate result for the first interferer $j = 2$ with a first rank-1 update, using Sherman-Morrison formula like in (C.5):

$$\bar{\mathbf{Z}}_1^{[1]\,-1} = \sigma_n^{-2}\mathbf{I} - \frac{\sigma_n^{-2}\mathbf{I}\mathbf{h}_2\mathbf{h}_2^H\sigma_n^{-2}\mathbf{I}}{1 + \mathbf{h}_2^H\sigma_n^{-2}\mathbf{I}\mathbf{h}_2} = \sigma_n^{-2}\left(\mathbf{I} - \frac{\mathbf{h}_2\mathbf{h}_2^H}{\sigma_n^2 + \mathbf{h}_2^H\mathbf{h}_2}\right). \tag{C.11}$$

Then the second interferer $j = 3$ can be included into the inverse by computing:

$$\bar{\mathbf{Z}}_1^{[2]\,-1} = \bar{\mathbf{Z}}_1^{[1]\,-1} - \frac{\bar{\mathbf{Z}}_1^{[1]\,-1}\mathbf{h}_3\mathbf{h}_3^H\bar{\mathbf{Z}}_1^{[1]\,-1}}{1 + \mathbf{h}_3^H\bar{\mathbf{Z}}_1^{[1]\,-1}\mathbf{h}_3}. \tag{C.12}$$

Further potential interferers could be added in the same step-by-step treatment. But it is clear that the number of terms to treat will "explode" and will not be reasonably manageable by analytic computations. Thus we suggest computer simulations, like used in chapter 3. This will allow to account for time-variant frequency selective channels obtained from realistic spatial channel model and to reasonably include the different types of channel estimation algorithms.

C.2. Sherman Morrison Formula

$$\left(\mathbf{X} + \mathbf{u}\mathbf{v}^H\right)^{-1} = \mathbf{X}^{-1} - \frac{\mathbf{X}^{-1}\mathbf{u}\mathbf{v}^H\mathbf{X}^{-1}}{1 + \mathbf{v}^H\mathbf{X}^{-1}\mathbf{u}} \tag{C.13}$$

D. Proof of IRC Variants

The purpose of this chapter is to proof in (3.48) the step from the second-last to the last row. In essence, it shows that the receive covariance in an IRC receiver can be written *with and without* the user term of interest.

Let's define the interference-plus-noise covariance excluding the user of interest, indexed j, as

$$\mathbf{Z}_j = \sum_{\forall \mu \in \mathcal{J}, \mu \neq j} \mathbf{h}_{\mu kl}^{[\mathrm{B}]} \mathbf{h}_{\mu kl}^{[\mathrm{B}]\,H} + \sigma_n^2 \mathbf{I} \tag{D.1}$$

The second-last row of (3.48) was derived from the orthogonality principle and it contains the user of interest as well. Plugging in (D.1) gives

$$\mathbf{g}_{jkl} = \left(\sum_{\forall \mu \in \mathcal{J}} \mathbf{h}_{\mu kl}^{[\mathrm{B}]} \mathbf{h}_{\mu kl}^{[\mathrm{B}]\,H} + \sigma_n^2 \mathbf{I} \right)^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} = \left(\mathbf{Z}_j + \mathbf{h}_{jkl}^{[\mathrm{B}]} \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \right)^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}. \tag{D.2}$$

As the matrix inversion including the user of interest can be interpreted as the inverse of a rank-1 matrix update, we apply the Sherman-Morrison formula (C.13), resulting in

$$\begin{aligned}
\mathbf{g}_{jkl} &= \left(\mathbf{Z}_j + \mathbf{h}_{jkl}^{[\mathrm{B}]} \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \right)^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} \\
&= \left(\mathbf{Z}_j^{-1} - \frac{\mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_j^{-1}}{1 + \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}} \right) \mathbf{h}_{jkl}^{[\mathrm{B}]} \\
&= \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} - \frac{\mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}}{1 + \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}} \\
&= \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]} \overbrace{\left(1 - \underbrace{\frac{\mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}}{1 + \mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}}}_{\Delta} \right)}^{\xi}.
\end{aligned} \tag{D.3}$$

The quadratic form $\mathbf{h}_{jkl}^{[\mathrm{B}]\,H} \mathbf{Z}_j^{-1} \mathbf{h}_{jkl}^{[\mathrm{B}]}$ is real valued and positive, as $\mathbf{Z}_j$ is a covariance matrix and thus is hermitian and positive semidefinite (and its inverse as well). This means that Δ is real-valued as well and that ξ is just a real-valued constant, a scaling factor.

As a conclusion, formulating the IRC receiver with and without the user term of interest j is just a change in normalization by the scaling factor ξ and hence does not impact the receiver output SINR.

E. Derivation of Simplified Channel Estimation Model

This appendix chapter contains the details of the derivation for the channel estimation error model introduced in section 3.5. The system model (3.80) of 3.5.1 is used.

E.1. Derivation of the Appropriate MMSE Estimator

As (3.80) is a linear model, the corresponding estimator minimizing the mean squared error is the linear MMSE estimator or Wiener filter. Its weights can be deduced using (3.18) and thus

$$\hat{h}_{ij}^{[BF]} = \mathrm{E}\left\{h_{ij}^{[BF]}\mathbf{y}_i^H\right\}\mathrm{E}\left\{\mathbf{y}_i\mathbf{y}_i^H\right\}^{-1}\mathbf{y}_i. \tag{E.1}$$

Inserting (3.80) into (E.1) leads to

$$\hat{h}_{ij}^{[BF]} = \sqrt{\rho_j}\alpha_{ij}\mathbf{s}^H(\sigma_i^2\mathbf{I} + \rho_j\alpha_{ij}\mathbf{s}_j\mathbf{s}_j^H)^{-1}(\sqrt{\rho_j}h_{ij}^{[BF]}\mathbf{s}_j + \mathbf{z}_i) \tag{E.2}$$

using the statistical independence of channel and noise. For computing the second factor, the Sherman-Morrison formula (C.13) is used.

$$\begin{aligned}(\sigma_i^2\mathbf{I} + \rho_j\alpha_{ij}\mathbf{s}_j\mathbf{s}_j^H)^{-1} &= \frac{1}{\sigma_i^2}\mathbf{I} - \frac{\frac{1}{\sigma_i^2}\mathbf{I}(\rho_j\alpha_{ij}\mathbf{s}_j\mathbf{s}_j^H)(\frac{1}{\sigma_i^2}\mathbf{I})}{1+\mathbf{s}_j^H(\frac{1}{\sigma_i^2}\mathbf{I})(\rho_j\alpha_{ij}\mathbf{s}_j)} \\ &= \frac{\frac{1}{\sigma_i^2}(1+\gamma_{ij}\left\|\mathbf{s}_j\right\|^2)\mathbf{I} - \frac{\gamma_{ij}}{\sigma_i^2}\mathbf{s}_j\mathbf{s}_j^H}{1+\gamma_{ij}\left\|\mathbf{s}_j\right\|^2}\end{aligned} \tag{E.3}$$

Using this result (E.3) into (E.2) yields

$$\begin{aligned}\hat{h}_{ij}^{[BF]} &= \alpha_{ij}\sqrt{\rho_j}\mathbf{s}_j^H\frac{\frac{1}{\sigma_i^2}(1+\gamma_{ij}\left\|\mathbf{s}_j\right\|^2)\mathbf{I} - \frac{\gamma_{ij}}{\sigma_i^2}\mathbf{s}_j\mathbf{s}_j^H}{1+\gamma_{ij}\left\|\mathbf{s}_j\right\|^2}(\sqrt{\rho_j}h_{ij}^{[BF]}\mathbf{s}_j + \mathbf{z}_i) \\ &= \frac{\alpha_{ij}\sqrt{\rho_j}}{1+\gamma_{ij}\left\|\mathbf{s}_j\right\|^2}\frac{1}{\sigma_i^2}\mathbf{s}_j^H(\sqrt{\rho_j}h_{ij}^{[BF]}\mathbf{s}_j + \mathbf{z}_i) \\ &= \underbrace{\frac{\gamma_{ij}\left\|\mathbf{s}_j\right\|^2}{1+\gamma_{ij}\left\|\mathbf{s}_j\right\|^2}}_{\beta_{ij}}(h_{ij}^{[BF]} + \frac{\mathbf{s}_j^H\mathbf{z}_i}{\sqrt{\rho_j}\left\|\mathbf{s}_j\right\|^2})\end{aligned} \tag{E.4}$$

Now, the MSE of the MMSE estimator is computed as

$$\begin{aligned}
\mathrm{MSE}(\hat{h}_{ij}^{[BF]}) &= \mathrm{E}\left\{\left|\hat{h}_{ij}^{[BF]} - h_{ij}^{[BF]}\right|^2\right\} = \mathrm{E}\left\{\left|\frac{\gamma_{ij}\left\|\mathbf{s}_j\right\|^2}{1+\gamma_{ij}\left\|\mathbf{s}_j\right\|^2}(h_{ij}^{[BF]} + \frac{\mathbf{s}_j^H\mathbf{z}_i}{\sqrt{\rho}\left\|\mathbf{s}_j\right\|^2}) - h_{ij}^{[BF]}\right|^2\right\} \\
&= \mathrm{E}\left\{\left|\frac{1}{1+\gamma_{ij}P}h_{ij}^{[BF]} - \frac{\gamma_{ij}P\mathbf{s}_j^H\mathbf{z}_i}{(1+\gamma_{ij}P)\sqrt{\rho_j}P}\right|^2\right\} \\
&= \frac{1}{(1+\gamma_{ij}P)^2}\alpha_{ij} + \frac{(\gamma_{ij}P)^2\sigma_i^2}{(1+\gamma_{ij}P)^2\rho_j P} \\
&= \frac{\alpha_{ij}}{1+\gamma_{ij}P}. \qquad \text{(E.5)}
\end{aligned}$$

E.2. Computations for the Error Model

Checking the overall MSE of the model (3.86) leads to:

$$\begin{aligned}
\mathrm{MSE}(\hat{h}_{ij}^{(e,r)}) &= \mathrm{E}\left\{\left|\hat{h}_{ij}^{(e,r)} - h_{ij}^{[BF]}\right|^2\right\} = \mathrm{E}\left\{\left|\beta_{ij}h_{ij}^{[BF]} + \beta_{ij}\Theta_i - h_{ij}^{[BF]}\right|^2\right\} \\
&= (\beta-1)^2\,\mathrm{E}\left\{\beta_{ij}h_{ij}^{[BF]}\beta_{ij}h_{ij}^{[BF]\,*}\right\} + \beta_{ij}^2\,\mathrm{E}\left\{\Theta_i\Theta_i^*\right\} = \alpha_{ij}\left((\beta_{ij}-1)^2 + \frac{\beta_{ij}^2}{\gamma_{ij}P}\right) \\
&= \alpha_{ij}\left(\frac{\gamma_{ij}^2P^2}{(1+\gamma_{ij}P)^2} - \frac{2\gamma_{ij}P}{1+\gamma_{ij}P} + 1 + \frac{\gamma_{ij}^2P^2}{(1+\gamma_{ij}P)^2\gamma_{ij}P}\right) \\
&= \alpha_{ij}\frac{\gamma_{ij}^2P^2 - 2\gamma_{ij}P - 2\gamma_{ij}^2P^2 + 1 + 2\gamma_{ij}P + \gamma_{ij}^2P^2 + \gamma_{ij}P}{(1+\gamma_{ij}P)^2} \\
&= \alpha_{ij}\frac{1+\gamma_{ij}P}{(1+\gamma_{ij}P)^2} \\
&= \frac{\alpha_{ij}}{1+\gamma_{ij}P} \qquad \text{(E.6)}
\end{aligned}$$

This proofs that the model has identical MSE (E.6) as the actual MMSE estimator, computed in (E.5).

F. Remarks on CSIT Feedback

F.1. Discussion of Background for Explicit CSIT Feedback

Limited feedback [HLR+08] is an ongoing research topic, discussed in the research community. In LTE, codebooks are used, signaling so-called preferred matrix indicators [3GPc]. In the LTE-A release 11 standard, limited feedback for CoMP has found its way as so-called "multi PMI reporting", meaning that for links to multiple cells a PMI can be signaled to describe the channel. T. Wild et. al proposed a variant of this scheme already for release 10 under the name "worst companion" PMI reporting [WH10], [AL08] [AL10] [GAB+11], see also section 2.1.4.2.

Note that the standardized approaches are based on limited feedback, describing narrowband channels. This generates problems in today's systems, as they are broadband, e.g. having a bandwidth of 20 MHz. The radio channels of different OFDM subcarriers are correlated. Frequency-domain based feedback, as used in current LTE-A releases has the problem that it either drops relevant information or inhibits redundancy, by granularizing the feedback into certain subband sizes.

The time-domain representation of the channel is a better choice for dealing with such correlations, as can be seen from (2.12). Theoretical considerations for such a time-domain-based broadband feedback were made in [CSM08]. A practical reporting scheme was suggested in [Wil10], called "rake compression", as the feedback concentrated on the dominant taps of the radio channel, like "rake fingers" known from CDMA. Rake compression was practically demonstrated [KHWH13], showing that with a feedback rate of 20 kbit/s (with a reporting interval of 10 ms), in a 4-pico cell scenario (each cell equipped with 2 transmit antennas) a receive SINR of 20 dB could be achieved with JT ZF precoding. With this SINR value, the largest modulation and coding scheme for LTE-A can be achieved. This means the limited feedback is accurate enough for maximum LTE-A performance, using 64-QAM.

In [WHHK13] it was shown that this time-domain feedback reporting approach is very well compatible with channel prediction to combat the derogatory effects of feedback delays. This will improve the supported MS velocities. Additionally, for finding the time domain channel taps, their sparsity was exploited by appropriate algorithms and reverse waterfilling bit loading, motivated from information theory, was used for reducing required reporting rates for weaker channel taps.

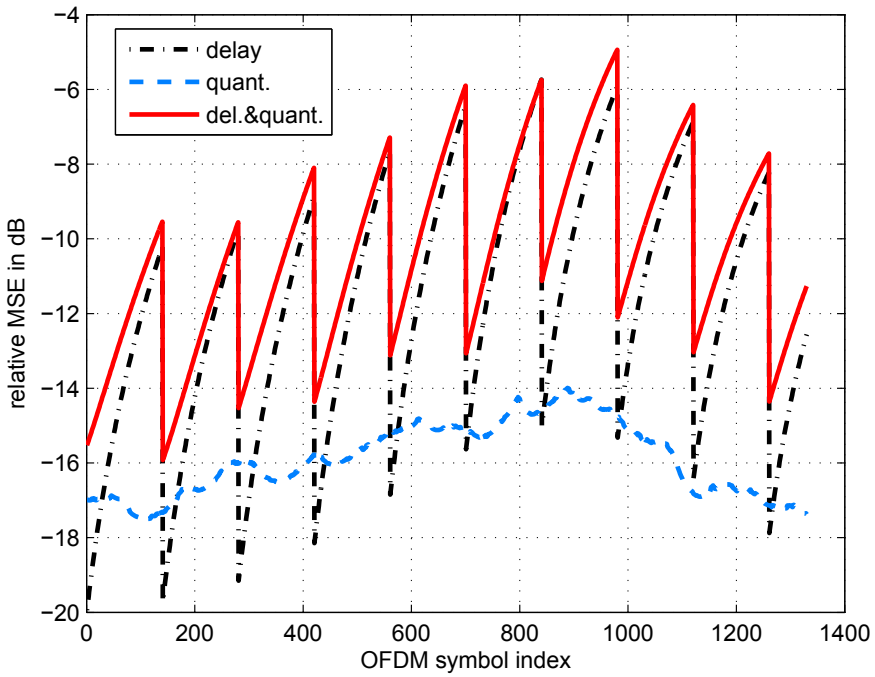

Figure F.1.: Relative MSE over time due to feedback quantization and feedback reporting delay. SCME urban macro channel, 3 km/h. Reporting delay 5ms, reporting interval 10 ms.

F.2. Parameters for Feedback Delay and Quantization

For chapters 4 and 5, the feedback delay and quantization is handled by the CSIT model (3.87) $\hat{h}_{ij}^{(e,t)} = \hat{h}_{ij}^{(e,r)} + \Theta_{ij}^{(t)}$.

The errors from quantization and channel aging $\Theta_{ij}^{(t)}$ are modeled as a complex Gaussian i.i.d. random variable with variance $\sigma_{[t],ij}^2$. The challenge is now to find reasonable values for $\sigma_{[t],ij}^2$.

For this purpose, we focus on the following scenario: An urban macro channel with low mobility, which is the most representative and sensible setting for CoMP. Feedback reporting delays in LTE systems are a multiple of the subframe duration of 1 ms. Typical values are 5 ms or above, resulting from the total required time of the following steps: The downlink pilots need to be received, the channel needs to be estimated and quantized and the next suitable uplink subframe has to be chosen for transmission over the uplink. Again, the BS has to receive and decode the feedback and has to take it into account in the scheduler for the next subsequent downlink subframe. Reasonable feedback reporting intervals for LTE systems are 10 ms.

The MSE over time due to channel aging and feedback quantization is shown in Fig. F.1, a result from [Wil10].

Each time, a new feedback report is available at the transmitter, its information is most up-to-date, only 5 ms old. In Fig. F.1 we can see this by the dips of the MSE, especially at the black

curve (delay effects only), but also at the red curve (delay plus quantization). Over time, this information gets outdated resulting in a rising MSE, until 10 ms later, thus 140 LTE OFDM symbols, a new feedback report is available, generating the next MSE dip.

The quantization errors depend on the number of invested feedback bits, according to Shannon's rate distortion theory. The relative MSE due to quantization distortion varies with the channel fading over time. The combined effect of quantization and delay still has a "sawtooth" characteristic, due to the reporting intervals, but the dips are not as deep as in the delay-only case without quantization, as the lower MSE values cannot be reached anymore due to quantization noise.

While Fig. F.1 provides a very detailed characteristic of the MSE behavior, we simplify this by providing one average MSE value for the model: In SCME urban macro, 3 km/h, the MSE is at -12 dB, resulting in $\sigma^2_{[t],ij} \approx 0.063$.

Bibliography

[3GPa] 3GPP TR 25.996, Technical Specification Group RAN; Spatial channel model for MIMO simulations. Technical Report V6.1.0, 3GPP.

[3GPb] 3GPP TS 36.141, Technical Specification Group RAN; E-UTRA; Base Station conformance testing (Release 11). Technical Report V11.0.0, 3GPP.

[3GPc] 3GPP TS 36.211, Technical Specification Group RAN1; E-UTRA Physical Channels and Modulation (Release 10). Technical Report V10.0.0, 3GPP.

[3GPd] 3GPP TS 36.213, Technical Specification Group RAN1; E-UTRA Physical Layer Procedures (Release 10). Technical Report V10.0.0, 3GPP.

[3GPe] 3GPP TS 36.521, Technical Specification Group RAN1; User Equipment (UE) conformance specification Radio transmission and reception Part 1: Conformance Testing (Release 10). Technical Report V10.0.0, 3GPP.

[3GPf] 3GPP TS 36.814, Technical Specification Group RAN1; Further Advancements for E-UTRA Physical Layer Aspects. Technical report, 3GPP.

[3GPg] 3GPP TS 36.819, Technical Specification Group RAN; Coordinated multi-point operation for LTE physical layer aspects (Release 11). Technical report, 3GPP.

[5GN] 5GNOW - 5th Generation Non-Orthogonal Waveforms for Asynchronous Signalling, http://www.5gnow.eu/.

[AL07] Alcatel-Lucent. 3GPP TSG-RAN1-071967: DL E-UTRA Performance Checkpoint. Technical report, Apr. 2007.

[AL08] Alcatel-Lucent. 3GPP TSG-RAN1-083759: UE PMI feedback signalling for user pairing/coordination. Technical report, Sept. 2008.

[AL10] Alcatel-Lucent. 3GPP TSG-RAN1-100944: Performance of coordinated beamforming with multiple PMI feedback. Technical report, Feb. 2010.

[Ala98] S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journ. on Selected Areas in Commun.*, 16(8):1451 –1458, Oct. 1998.

[Ala01] N.S. Alagha. Cramer-Rao bounds of SNR estimates for BPSK and QPSK modulated signals. *IEEE Commun. Letters*, 5(1):10–12, 2001.

[Ana81] F. Ananasso. Nulling performance of null-steering arrays with digital phase-only weights. *Electronics Letters*, 17(7):255 –257, 2 1981.

[Art11] ARTIST4G D2.3, Advanced link-to-system modelling. ARTIST4G deliverable, ICT ARTIST4G, Mar. 2011. https://ict-artist4g.eu/projet/deliverables.

[ASWW12] D. Aziz, S. Sentuerk, A. Weber, and T. Wild. A simple model for imperfect channel state information and its application for the assessment of Interference Alignment. In *Proc. IEEE Int. Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 507–511, 2012.

[BAD+12] F. Boccardi, O. Aydin, U. Doetsch, T. Fahldieck, and H.-P. Mayer. User-centric architectures: Enabling CoMP via hardware virtualization. In *Proc. IEEE Personal, Indoor and Mobile Radio Communications, PIMRC*, pages 191–196, 2012.

[BG96] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: turbo-codes. *IEEE Trans. on Commun.*, 44(10):1261 –1271, Oct. 1996.

[BH07] F. Boccardi and H. Huang. A Near-Optimum Technique using Linear Precoding for the MIMO Broadcast Channel. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process., ICASSP*, volume 3, pages III–17–III–20, Apr. 2007.

[BHP02] H. Bolcskei, Jr. Heath, R.W., and A.J. Paulraj. Blind channel identification and equalization in OFDM-based multiantenna systems. *IEEE Trans. on Signal Process.*, 50(1):96 – 109, Jan. 2002.

[BHS05] D.S. Baum, J. Hansen, and J. Salo. An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM). In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, volume 5, pages 3132 – 3136, May 2005.

[BKBO13] E. Björnson, M. Kountouris, M. Bengtsson, and B. E. Ottersten. Receive Combining vs. Multi-Stream Multiplexing in Downlink Systems with Multi-Antenna Users. *IEEE Trans. on Signal Process.*, 61(13):3431–3446, July 2013.

[BKKR11] K. Balachandran, J.H. Kang, K. Karakayali, and K.M. Rege. NICE: A Network Interference Cancellation Engine for Opportunistic Uplink Cooperation in Wireless Networks. *IEEE Trans. on Wireless Commun.*, 10(2):540 –549, Feb. 2011.

[BLM03] I. Barhumi, G. Leus, and M. Moonen. Optimal training design for MIMO OFDM systems in mobile wireless channels. *IEEE Trans. on Signal Process.*, 51(6):1615 – 1624, June 2003.

[Bre03] D.G. Brennan. Linear diversity combining techniques. *Proceedings of the IEEE*, 91(2):331 – 356, 2003.

[BS11] S.Z. Budišin and P. Spasojević. Björck sequence sets. *Electronics Letters*, 47(8):491–493, 2011.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[BZ71] C. A. Baird and C. L. Zahm. Performance criteria for narrowband array processing. In *Proc. IEEE Conf. Decision and Control*, volume 10, pages 564 –565, Dec. 1971.

[CFV00] D. Chizhik, G.J. Foschini, and R.A. Valenzuela. Capacities of multi-element transmit and receive antennas: Correlations and keyholes. *Electronics Letters*, 36(13):1099 –1100, June 2000.

[Cha66] R.W. Chang. Synthesis of band-limited orthogonal signals for multi-channel data transmission. *The Bell System Technical Journal*, 46(1):1775–1796, 1966.

[CJ08] V.R. Cadambe and S.A. Jafar. Interference Alignment and Degrees of Freedom of the K-User Interference Channel. *IEEE Trans. on Inform. Theory*, 54(8):3425 –3441, Aug. 2008.

[Cla68] R. H. Clarke. A statistical theory of mobile radio reception. *Bell Systems Technical Journal*, 47:957–1000, 1968.

[CMSH05] N. Czink, G. Matz, D. Seethaler, and F. Hlawatsch. Improved MMSE estimation of correlated MIMO channels using a structured correlation estimator. In *Proc. IEEE Signal Process. Advances in Wirel. Commun., SPAWC*, pages 595–599, 2005.

[Cos83] M. Costa. Writing on dirty paper (corresp.). *IEEE Trans. on Inform. Theory*, 29(3):439 – 441, May 1983.

[CPR11] Common Public Radio Interface (CPRI); Interface Specification. Technical Report V5.0, http://www.cpri.info, 2011.

[CSM08] G. Caire and H. Shirani-Mehr. Feedback schemes for multiuser MIMO-OFDM downlink. In *Information Theory and Applications Workshop*, 2008.

[CT06] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.

[CWEH10] Yilun Chen, A. Wiesel, Y.C. Eldar, and A.O. Hero. Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Trans. on Signal Process.*, 58(10):5016 – 5029, Oct. 2010.

[DDM+13] U. Doetsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier. Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE. *Bell Labs Technical Journal*, 18(1), 2013.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journ. of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[DS05] G. Dimic and N.D. Sidiropoulos. On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm. *IEEE Trans. on Signal Process.*, 53(10):3857 – 3868, Oct. 2005.

[DW11] U. Doetsch and T. Wild. EP 2 544420 A1 Method for transmitting data in a communication system, first network node and second network node thereof. European Patent Application, 2011.

[FB11] B. Farhang-Boroujeny. OFDM Versus Filter Bank Multicarrier. *IEEE Signal Process. Magazine*, 28(3):92–112, May 2011.

[FD11] T. Fahldieck and M. Doll. CoMP Backhaul Infrastructure Concepts. In *Coordinated Multi-Point in Mobile Communications*, chapter 12.3, pages 301–310. Cambridge University Press, 2011.

[FG98] G. J. Foschini and M. J. Gans. On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas. *Wirel. Pers. Commun.*, 6(3):311–335, 1998.

[FKB09] G. Fettweis, M. Krondorf, and S. Bittner. GFDM - Generalized Frequency Division Multiplexing. In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, pages 1–4, Apr. 2009.

[Fos96] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, 1(2):41–59, 1996.

[GAB+11] J. Giese, M. A. Amin, S. Brueck, M. Doll, T. Wild, L. Thiele, and M. Olbrich. Downlink Simulation Results. In *Coordinated Multi-Point in Mobile Communications*, chapter 14.4, pages 193–208. Cambridge University Press, 2011.

[Gal63] R. G. Gallager. *Low Density Parity Check Codes*. PhD thesis, M.I.T., 1963.

[Gap08] W. Gappmair. Cramer-Rao Lower Bound for Non-Data-Aided SNR Estimation of Linear Modulation Schemes. *IEEE Trans. on Commun.*, 56(5):689–693, 2008.

[GFH95] W.H. Gerstacker, R.F.H. Fischer, and J.B. Huber. Blind equalization for digital cable transmission with Tomlinson-Harashima precoding and shaping. In *Proc. IEEE Intern. Conf. on. Commun., ICC*, volume 1, pages 493 –497, June 1995.

[GHH+10] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu. Multi-Cell MIMO Cooperative Networks: A New Look at Interference. *IEEE Journ. on Selected Areas in Commun.*, 28(9):1380 –1408, Dec. 2010.

[Gol05] A. Goldsmith. Multicarrier modulation. In *Wireless Communications*, chapter 12, pages 374–402. Cambridge University Press, 2005.

[HBD00] J.S. Hammerschmidt, C. Brunner, and C. Drewes. Eigenbeamforming - a novel concept in array signal process. In *Proc. VDE/ITG European Wireless Conf.*, 2000.

[HFM11] C. Hoymann, L. Falconetti, and P. Marsch. Backhaul Requirements of Practical CoMP Schemes. In *Coordinated Multi-Point in Mobile Communications*, chapter 12.2, pages 291–301. Cambridge University Press, 2011.

[HKR97] P. Hoeher, S. Kaiser, and P. Robertson. Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process.. ICASSP*, volume 3, pages 1845 –1848, Apr. 1997.

[HLR+08] R.W. Heath, D.J. Love, B.D. Rao, V. Lau, D. Gesbert, and M. Andrews. Exploiting limited feedback in tomorrow's wireless communication networks. *IEEE Journ. on Selected Areas in Commun.*, 26(8):1337–1340, Oct. 2008.

[HTH+09] H.C. Huang, M. Trivellato, A. Hottinen, M. Shafi, P.J. Smith, and R.A. Valenzuela. Increasing downlink cellular throughput with limited network MIMO coordination. *IEEE Trans. on Wireless Commun.*, 8(6):2983–2989, 2009.

[ITU] ITU-R Recommendation M.1225: Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000, 1997. Technical report.

[JC94] W. C. Jakes and D. C. Cox, editors. *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.

[Kay93] S. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[KHWH13] J. Koppenborg, C. Hoek, T. Wild, and G. Herzog. Test of Downlink Joint Transmission with Distributed Antennas. In *Proc. Intern. ITG Workshop on Smart Antennas*, Mar. 2013.

[KSP+02] J.P. Kermoal, L. Schumacher, K.I. Pedersen, P.E. Mogensen, and F. Frederiksen. A stochastic MIMO radio channel model with experimental validation. *IEEE Journ. on Selected Areas in Commun.*, 20(6):1211 – 1226, Aug. 2002.

[KV96] H. Krim and M. Viberg. Two decades of array signal process. research: the parametric approach. *IEEE Signal Process. Magazine*, 13(4):67 –94, July 1996.

[LW04a] O. Ledoit and M. Wolf. Honey, I Shrunk the Sample Covariance Matrix. *Journ. of Portfolio Management*, 30(4), Dec. 2004.

[LW04b] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journ. of Multivariate Analysis*, 88(2):365–411, 2004.

[Mac79] V.H. MacDonald. The Cellular Concept. *The Bell System Technical Journal*, 58(1):15–41, 1979.

[Mac03] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[Mar10] P. Marsch. *Coordinated Multi-Point under a Constrained Backhaul and Imperfect Channel Knowledge*. PhD thesis, TU Dresden, 2010.

[MET] METIS 2020 Mobile and wireless communications Enablers for the Twenty-twenty Information Society 5G, https://www.metis2020.com/.

[Mey00] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Siam, 2000.

[MF11] P. Marsch and G. P. Fettweis, editors. *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*. Cambridge University Press, 2011.

[MFS10] A. Mueller, P. Frank, and J. Speidel. Performance of the LTE Uplink with Intra-Site Joint Detection and Joint Link Adaptation. In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, pages 1 –5, May 2010.

[MGL12] R. Mungara, G. George, and A. Lozano. System-level performance of distributed cooperation. In *Proc. Asilomar Conf. Signals, Systems and Computers*, pages 1561–1565, Nov. 2012.

[MS02] A.N. Mody and G.L. Stuber. Receiver implementation for a MIMO OFDM system. In *Proc. IEEE Global Telecommun. Conf., Globecom*, volume 1, pages 716–720, 2002.

[NH10] C.T.K. Ng and H. Huang. Linear Precoding in Cooperative MIMO Cellular Networks with Limited Coordination Clusters. *IEEE Journ. on Selected Areas in Commun.*, 28(9):1446 –1454, Dec. 2010.

[NRWtB11] L.-H. Nguyen, R. Rheinschmitt, T. Wild, and S. ten Brink. Limits of channel estimation and signal combining for multipoint cellular radio (CoMP). In *Proc. IEEE Intern. Sympos. on Wireless Commun. Systems, ISWCS*, pages 176 –180, Nov. 2011.

[OORW08] M. Olsson, A. Osseiran, F. Roemer, and T. Wild. Multi-antenna processing in the WINNER air interface. In *Proc. XXIXth General Assembly of the International Union of Radio Science (URSI)*, 2008.

[OSI] ITU-T Recommendation X.200 (1994) | ISO/IEC 7498-1: 1994, Information Technology - Open Systems Interconnection - Basic Reference Model. Technical report.

[PB00] D.R. Pauluzzi and N.C. Beaulieu. A comparison of SNR estimation techniques for the AWGN channel. *IEEE Trans. on Commun.*, 48(10):1681 – 1691, Oct. 2000.

[PNG03] A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[PP12] K. B. Petersen and M. S. Pedersen. The matrix cookbook, Nov. 2012. Version 20121115.

[PRK85] A. Paulraj, R. Roy, and T. Kailath. Estimation Of Signal Parameters Via Rotational Invariance Techniques- Esprit. In *Proc. Asilomar Conf. on Circuits, Systems and Computers*, pages 83 –89, Nov. 1985.

[Say08] Ali H. Sayed. *Adaptive Filters*. Wiley-IEEE Press, 2008.

[SB05] M. Schubert and H. Boche. Iterative multiuser uplink and downlink beamforming under SINR constraints. *IEEE Trans. on Signal Process.*, 53(7):2324 – 2334, July 2005.

[Sch86] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. on Antennas and Propagation*, 34(3):276 – 280, Mar. 1986.

[Sha48] Claude E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[SHC+10] A. Simonsson, B. Hagerman, J. Chistoffersson, L. Klockar, C. Koutsimanis, and P. Cosimini. LTE Downlink Inter-Cell Interference Assessment in an Existing GSM Metropolitan Deployment. In *Proc. IEEE Vehicular Technology Conf., VTC-Fall*, pages 1 –5, Sept. 2010.

[Skl97] B. Sklar. Rayleigh fading channels in mobile digital communication systems .i. characterization. *IEEE Commun. Magazine*, 35(7):90 –100, July 1997.

[Spe05] J. Speidel. *Multiple Input Multiple Output (MIMO): Drahtlose Nachrichtenübertragung hoher Bitrate und Qualität mit Mehrfachantennen*. Tele-Kommunikation aktuell. Verlag für Wiss. und Leben Heidecker, 2005.

[SS00] F. Sanzi and J. Speidel. An adaptive two-dimensional channel estimator for wireless OFDM with application to mobile DVB-T. *IEEE Trans. on Broadcasting*, 46(2):128 –133, June 2000.

[SSF+10] C. Stierstorfer, C. Siegl, R.F.H. Fischer, T. Wild, and C. Hoek. Network MIMO downlink transmission. In *Proc. Intern. OFDM Workshop (InOWo'10)*, Aug. 2010.

[SSH04] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt. Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. *IEEE Trans. on Signal Process.*, 52(2):461 – 471, Feb. 2004.

[STS07] M. Sadek, A. Tarighat, and A.H. Sayed. A leakage-based precoding scheme for downlink multi-user MIMO channels. *IEEE Trans. on Wireless Commun.*, 6(5):1711 –1721, May 2007.

[SW14] F. Schaich and T. Wild. Waveform contenders for 5G – suitability for short packet and low latency transmissions. In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, May 2014.

[SZ01] S. Shamai and B.M. Zaidel. Enhancing the cellular downlink capacity via co-processing at the transmitting end. In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, volume 3, pages 1745 –1749 vol.3, 2001.

[Tel99] I. Emre Telatar. Capacity of multi-antenna Gaussian channels. *European Trans. on Telecommun.*, 10:585–595, 1999.

[TS11] M. Tuechler and A.C. Singer. Turbo Equalization: An Overview. *IEEE Trans. on Inform. Theory*, 57(2):920 –952, Feb. 2011.

[TSS+08] L. Thiele, M. Schellmann, S. Schiffermuller, V. Jungnickel, and W. Zirwas. Multi-Cell Channel Estimation using Virtual Pilots. In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, pages 1211 –1215, May 2008.

[vB09] C. H. (Kees) van Berkel. Multi-core for mobile phones. In *Proc. Conf. on Design, Automation and Test in Europe*, DATE '09, pages 1260–1265, 2009.

[Ven07] S. Venkatesan. Coordinating base stations for greater uplink spectral efficiency in a cellular network. In *Proc. IEEE Personal, Indoor and Mobile Radio Communications, PIMRC*, pages 1 –5, Sept. 2007.

[Ver89] S. Verdu. The capacity region of the symbol-asynchronous Gaussian multiple-access channel. *IEEE Trans. on Inform. Theory*, 35(4):733 –751, July 1989.

[VJG03] S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Trans. on Inform. Theory*, 49(10):2658 – 2668, Oct. 2003.

[VT02] H. L. Van Trees. *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. Wiley-Interscience, 1 edition, March 2002.

[VWS+13] V. Vakilian, T. Wild, F. Schaich, S. ten Brink, and J.-F. Frigon. Universal-filtered multi-carrier technique for wireless systems beyond LTE. In *Proc. IEEE Globecom Workshops*, pages 223–228, Dec. 2013.

[WA06] J. Wang and K. Araki. Pilot Symbol Aided MAP-Based 3D Channel Estimation for Multi-User MIMO-OFDM Systems. *IEICE Trans.*, 89-B(3):801–808, 2006.

[WE71] S. Weinstein and P. Ebert. Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform. *IEEE Trans. on Commun. Technology*, 19(5):628 –634, Oct. 1971.

[WH10] T. Wild and C. Hoek. EP 2 166 807 A1 Method for building sets of mobile stations in SDMA systems, corresponding mobile station, base stations, operation and maintenance centre and radio communication network. European Patent Specification, 2010.

[WHHK13] T. Wild, C. Hoek, G. Herzog, and J. Koppenborg. Multi-antenna OFDM channel feedback compression exploiting sparsity. In *Proc. European Wireless Conf. (EW2013)*, pages 1–6, 2013.

[Wie49] N. Wiener. *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*. Wiley, New York, 1949.

[Wil06] T. Wild. Successive user insertion for long-term adaptive beams using short-term CQI. In *Proc. Intern. OFDM Workshop (InOWo'06)*, Aug. 2006.

[Wil10] T. Wild. A Rake-Finger Based Efficient Channel State Information Feedback Compression Scheme for the MIMO OFDM FDD Downlink. In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, pages 1–5, May 2010.

[Wil11a] T. Wild. Comparing downlink coordinated multi-point schemes with imperfect channel knowledge. In *Proc. IEEE Vehicular Technology Conf., VTC-Fall*, pages 1 –5, Sept. 2011.

[Wil11b] T. Wild. Coping with out-of-cluster leakage in downlink coordinated multi-point. In *Proc. IEEE Intern. Sympos. on Wireless Commun. Systems, ISWCS*, pages 710 –714, Nov. 2011.

[Wil11c] T. Wild. EP 2523360 A1 Method for interference reduction in a radio communication system, first radio access network node, second radio access network node and mobile station thereof. European Patent Application, 2011.

[WIM06] IST-4-027756 WINNER II D1.1.2 v1.2 WINNER II Channel Models. WINNER deliverable, IST WINNER, 2006.

[Win84] J.H. Winters. Optimum combining in digital mobile radio with co-channel interference. *IEEE Trans. on Vehicular Technology*, 33(3):144 – 155, Aug. 1984.

[WIN05] IST-2003-507581 WINNER D2.7 Assessment of Advanced Beamforming and MIMO Technologies. Technical report, 2005.

[WJK+14] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Yejian Chen, S. ten Brink, I Gaspar, N. Michailow, A Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, and F. Wiedmann. 5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications. *IEEE Commun. Magazine*, 52(2):97–105, Feb. 2014.

[WK85] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. on Acoustics, Speech and Signal Process.*, 33(2):387 – 392, Apr. 1985.

[WNtB12] T. Wild, L.-H. Nguyen, and S. ten Brink. Joint channel estimation across multiple cells in coordinated multi-point. In *Proc. IEEE Intern. Sympos. on Wireless Commun. Systems, ISWCS*, Aug. 2012.

[WNtB13] T. Wild, L.-H. Nguyen, and S. ten Brink. Multi-stage channel estimation across multiple cells in uplink joint reception. In *Proc. IEEE Vehicular Technology Conf., VTC-Spring*, 2013.

[Wol] Wolfram Mathematica Software, http://www.wolfram.com/mathematica/.

[WRN13] T. Wild, R. Rheinschmitt, and L.-H. Nguyen. EP2757750 A1 Apparatuses, Methods, and Computer Programs for a Channel Estimator and a Base Station Transceiver . European Patent Application, 2013.

[WS13] T. Wild and F. Schaich. EP2840749 Receiver Architectures For Universal Filtered Multi-Carrier. European Patent Application, 2013.

[WSC14] T. Wild, F. Schaich, and Y. Chen. 5G Air Interface Design based on Universal Filtered (UF-)OFDM. In *Proc. Intern. Conf. on Digital Signal Process., DSP*, Aug. 2014.

[WSS06] H. Weingarten, Y. Steinberg, and S. Shamai. The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel. *IEEE Trans. on Inform. Theory*, 52(9):3936 –3964, Sept. 2006.

[Wyn94] A.D. Wyner. Shannon-theoretic approach to a Gaussian cellular multiple-access channel. *IEEE Trans. on Inform. Theory*, 40(6):1713 –1727, Nov. 1994.

[XJY05] X. Xu, Y. Jing, and X. Yu. Subspace-based noise variance and SNR estimation for OFDM systems. In *Proc. IEEE Wirel. Commun. and Networking Conf., WCNC*, volume 1, pages 23 – 26, Mar. 2005.

[YG06] T. Yoo and A. Goldsmith. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *IEEE Journ. on Selected Areas in Commun.*, 24(3):528 – 541, Mar. 2006.

[ZMH10] J. Zhang, X. Mu, and L. Hanzo. Joint Channel Impulse Response and Noise-Variance Estimation for OFDM/SDMA Systems Based on Expectation Maximization. In *Proc. IEEE Vehicular Technology Conf., VTC-Fall*, pages 1 –5, Sept. 2010.

[ZTW+11] W. Zirwas, L. Thiele, T. Weber, N. Palleit, and V. Jungnickel. Channel Estimation for CoMP. In *Coordinated Multi-Point in Mobile Communications*, chapter 9.1, pages 193–208. Cambridge University Press, 2011.