DISSERTATION

Limited Feedback Transceiver Design for Downlink MIMO OFDM Cellular Networks

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften

Dipl.-Ing. Stefan Schwarz

geboren am 03. Juli 1984

Matrikelnummer: 0425136

Olbersdorf 138, 2842 Thomasberg, Österreich

November, 2013

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

Begutachter:

Univ. Prof. Dr.-Ing. Markus Rupp

Institute of Telecommunications
Technische Universität Wien
Österreich

Prof. Ph.D. Robert W. Heath Jr.

Department of Electrical and Computer Engineering
The University of Texas at Austin
USA

Abstract

Feedback in wireless communications is tied to a long-standing and successful history, facilitating robust and spectrally efficient transmission over the uncertain wireless medium. Since the application of multiple antennas at both ends of the communication link, enabling multiple-input multiple-output (MIMO) transmission, the importance of feedback information to achieve the highest performance is even more pronounced. Especially when multiple antennas are employed by the transmitter to handle the interference between multiple users, channel state information (CSI) is a fundamental prerequisite. The corresponding multi-user MIMO, interference alignment and coordination techniques are considered as a central part of future cellular networks to cope with the growing inter-cell-interference, caused by the unavoidable densification of base stations to support the exponentially increasing demand on network capacities. However, this vision can only be implemented with efficient feedback algorithms that provide accurate CSI at the transmitter without overloading the uplink channel.

In this dissertation, channel state information feedback algorithms for the downlink of MIMO OFDM cellular networks are proposed and evaluated. After developing the basic mathematical description of the data transmission to a user, the first part of the thesis is concerned with single-user MIMO (SU-MIMO) transmission, i.e., spatial multiplexing of multiple data streams to a single user. SU-MIMO is already incorporated and standardized in current cellular technology, such as 3GPP Long Term Evolution (LTE). The SU-MIMO feedback selection algorithms developed in this thesis are hence designed to optimize the achievable throughput of the transmission while keeping standard-compliance and the corresponding architectural constraints of the physical layer in mind. The performance of the proposed methods is evaluated with extensive link level simulations and by comparison to theoretical bounds on the achievable throughput.

In the second part of this dissertation, multi-user MIMO (MU-MIMO) is considered. Here only rudimentary specifications are provided in standards, giving a much larger design freedom. The focus is put on block-diagonalization based MU-MIMO. CSI feedback algorithms are proposed that are based on memoryless and predictive quantization on the Grassmann manifold. With predictive quantization, the temporal correlation of the wireless channel can be exploited to enable high fidelity quantization with a reasonable feedback overhead. A subspace quantization based antenna combiner is proposed for the case that the receivers are equipped with excess antennas. With this combiner a significant improvement in the CSI feedback accuracy can be achieved, reducing the residual multi-user interference. The performance of this combiner is analytically analyzed and contrasted to a strategy that maximizes the channel gain of a single user without considering interference.

Finally, the performance of SU- and MU-MIMO is investigated by means of simulations, revealing significant improvements with MU-MIMO provided sufficiently accurate CSI is available at the transmitter. Also, the area spectral efficiency of networking architectures using remote radio units and small cells is compared to the classical macro cellular network, demonstrating valuable capacity gains with remote radio units, which enable joint transmission over the distributed antenna system.

Kurzfassung

In der drahtlosen Kommunikation wird die Informationsrückführung vom Empfänger zum Sender, das sogenannte Feedback, als ein zentrales Mittel gesehen, um die Robustheit und Effizienz der Übertragung über den unsteten Kanal zu verbessern. Mit der Einführung von Mehrantennensystemen (MIMO) hat die Bedeutung des Feedbacks noch weiter zugenommen. Speziell dann, wenn Mehrantennensysteme eingesetzt werden, um die Interferenz zwischen Benutzern zu reduzieren, ist genaue Kanalinformation am Sender unumgänglich, um Konzepte wie räumliches Multiplexing mehrerer Benutzer (Multi-user (MU-)MIMO) und Interference Alignment zu ermöglichen. Solche Methoden sollen in zukünftigen zellulären Netzen, die der stark anwachsenden Nachfrage nach Übertragungskapazität nur durch eine Verdichtung der Netzstruktur begegnen können, eingesetzt werden, um die damit verbundene Zunahme der Interferenz zu bewältigen. Jedoch kann diese Vision nur mit effizientem Feedback Realität werden, um nicht im Gegenzug die Aufwärtsstrecke zu überlasten.

In der vorliegenden Dissertation wird dementsprechend die Entwicklung effizienter Feedback Algorithmen für zelluläre Netzwerke behandelt, deren physikalische Schicht auf MIMO und OFDM basiert. Zunächst wird die mathematische Modellierung der Datenübertragung beschrieben, um darauf aufbauend sogenannte Single-User MIMO (SU-MIMO) Systeme zu erörtern, bei denen ein einzelner Benutzer über mehrere Datenströme im räumlichen Multiplex bedient wird. Solche Konzepte sind bereits in standardisierten Kommunikations-Technologien, wie 3GPP Long Term Evolution (LTE), implementiert. Deswegen wird die Standard-Konformität bei der Herleitung der Feedback Methoden als eine entscheidende Nebenbedingung zur Optimierung der Übertragungsrate betrachtet. Die Leistungsfähigkeit der entwickelten Algorithmen wird mittels umfangreicher Simulationen und durch Vergleich mit theoretischen Durchsatzgrenzen analysiert.

Im zweiten Teil dieser Dissertation wird MU-MIMO Übertragung behandelt. Hierfür sind in aktuellen Kommunikations-Standards noch kaum Vorgaben gemacht, wodurch ein großer Gestaltungsfreiraum offen ist. Am Sender wird das Block-Diagonalization Verfahren angewendet, um die Datenströme der Benutzer zu trennen. Gedächtnislose und prädiktive Quantisierung von Elementen der Grassmann Mannigfaltigkeit wird entwickelt, um eine effiziente Rückführung der Kanalinformation zu erzielen. Für den Fall, dass die Empfänger über Überschussantennen verfügen, wird ein Verfahren zur Kombination der Antennenausgänge abgeleitet, das eine zusätzliche Reduktion des Quantisierungsfehlers erzielt und dadurch die verbleibende Interferenz zwischen den Benutzern minimiert. Das Potenzial dieser Methode zur Reduzierung der Rückführinformation wird mathematisch analysiert und mit einer Alternative verglichen, die Quantisierungseffekte vernachlässigt.

Ein Vergleich von SU- und MU-MIMO wird im letzten Abschnitt dieser Arbeit durchgeführt. Mittels Simulationen wird das Potenzial von MU-MIMO gegenüber SU-MIMO aufgezeigt, sofern ausreichend genaue Kanalinformation am Sender zur Verfügung steht. Des Weiteren werden verschiedene Netzstrukturen, basierend auf verteilten Antennen sowie Mikrozellen, gegenübergestellt und die Leistungsfähigkeit kohärenter Übertragung in verteilten Antennensystemen dargelegt.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Markus Rupp for sharing his profound scientific and technical knowledge and for introducing me to the amazing field of wireless communications. Prof. Rupp's guidance during the work on this thesis, but also the provided freedom to pursue own ideas and objectives contributed significantly to the success of this dissertation.

Many thanks to the great team of students and researches that has contributed to and evolved around the Vienna LTE simulators, especially to my colleagues Josep Colom Ikuno, Michal Šimko and Martin Taranetz. Valuable discussions, constructive team work and a good deal of fun made my life as a doctoral candidate more interesting and exciting.

My gratitude goes to Prof. Robert W. Heath Jr. for the time and effort spent as an examiner of this thesis and as a co-author of several conference papers and journal articles constituting an important part of this dissertation. Most of my work on distributed antenna systems was inspired by discussions with Prof. Heath during my three month research stay at the University of Texas at Austin.

The financial support of the Institute of Telecommunications, Vienna University of Technology, and of A1 Telekom Austria AG is gratefully appreciated.

Meinen Eltern Maria und Siegfried möchte ich von ganzem Herzen danken. Eure Unterstützung, euer Vertrauen und eure Geduld gaben mir den nötigen Rückhalt zur Erstellung dieser Arbeit. Zu guter Letzt, danke an meine Geschwister Claudia und Gerald und ihre Familien, Gerald, Silvy und Lena sowie Nicole und Laurenz, für zahllose schöne Stunden abseits des Alltages.

Contents

List of Figures

Chapter 1.

Introduction

> Concern for man himself and his fate must always be the chief interest of all technical endeavors, in order that the creations of our mind shall be a blessing and not a curse to mankind. Never forget this in the midst of your diagrams and equations.
>
> *(Albert Einstein)*

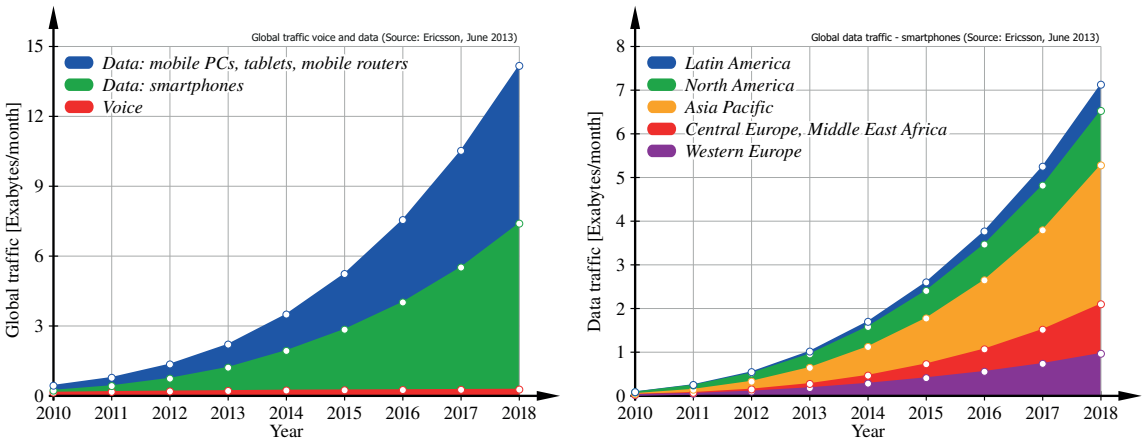
1.1. Trends and Development of Cellular Networks

Cellular networks are currently experiencing a tremendous growth of data traffic. A ten to seventeen-fold increase of global mobile data traffic between 2012 and 2017 is predicted by several studies [1–3]. A few notable statements of Cisco's Visual Networking Index [3], illustrating these trends, are:

- Mobile data traffic in 2012 was nearly twelve times the size of the entire Internet in 2000.
- Average smartphone usage grew 81 percent in 2012.
- Smartphones represented only 18 percent of total global handsets in use in 2012, but represented 92 percent of total global handset traffic.

The predicted trends in global mobile traffic according to Ericsson [2] are shown in Figure 1.1. The expected exponential growth in traffic due to data services (video, web-browsing, email, etc.) is illustrated in the figure, while the contribution of voice traffic stays approximately constant. The driving forces behind the expected traffic explosion are hence new applications enabled by modern devices such as smartphones and tablet-computers, while the classical role of the mobile as a phone becomes less and less significant. Similar tendencies are observed globally as shown in Figure 1.1b. Such observations substantiate a demand for larger network capacities in future cellular networks, but also for improved per-user data rates to support the requirements of novel applications.

The evolution in the mobile phone usage behavior is very well reflected in the historical development of cellular networks. The first fully digital mobile communication system, the Global System for Mobile Communications (GSM), was commercially launched in 1991. This second generation (2G) mobile communication system was developed by the European Telecommunications Standard



(a) Global voice and data traffic.

(b) Partitioning of smartphone traffic by regions.

Figure 1.1.: Estimated growth of the global mobile data and voice traffic and partitioning of the smartphone data traffic by geographic regions. (Source: Ericsson traffic exploration tool, June 2013 [4]; note: 1 Exabyte $= 10^{18}$ bytes).

Institute (ETSI) to replace first generation (1G) analog cellular networks. It was originally designed as a circuit-switched network optimized for telephony, to enable similar quality of experience as in fixed telephone networks. Soon, a need to support packet-switched data traffic for Internet applications was recognized and GSM was extended by General Packet Radio Service (GPRS) and Enhanced Data Rates for GSM Evolution (EDGE) to achieve data rates of up to 236 kbit/s over four time slots [5]. New features (e.g., video telephony, web-browsing) and further improvements in transmission rates were enabled with the launch of the third generation (3G) Universal Mobile Telecommunications System (UMTS) in 1999, originally standardized by the ETSI. The system was adopted by many countries and became a quasi world-standard. Due to this global interest, UMTS is nowadays kept and improved by the Third Generation Partnership Project (3GPP), a global collaboration of telecommunication associations and standardization bodies. 3GPP constantly evolves the cellular networking technology, providing major standard releases every few years. In that way, the maximum downlink data rates have improved from 384 kbit/s of the first UMTS release (Rel'99 [6]) to 330 Mbit/s achieved with High Speed Packet Access (HSPA) in Rel'11 [7]. Such data rates are enabled by the wideband code division multiple access (WCDMA) physical layer (PHY) technology in combination with 4×4 multiple-input multiple-output (MIMO) and aggregation of four 5 MHz carriers. In parallel, an entirely new air interface based on orthogonal frequency division multiple access (OFDMA) was released in 2009 (Rel'8 [8]) under the designation Long Term Evolution (LTE), enabling downlink data rates of up to 300 Mbit/s and setting the basis for an International Mobile Telecommunications-Advanced (IMT-A) capable fourth generation (4G) radio access technology [9]. In 2010, LTE advanced (LTE-A), i.e., LTE Rel'10 [10], breaking the 1 Gbit/s barrier, was officially announced as a 4G technology by the International Telecommunications Union (ITU). This technology allows for location-independent broadband Internet access and

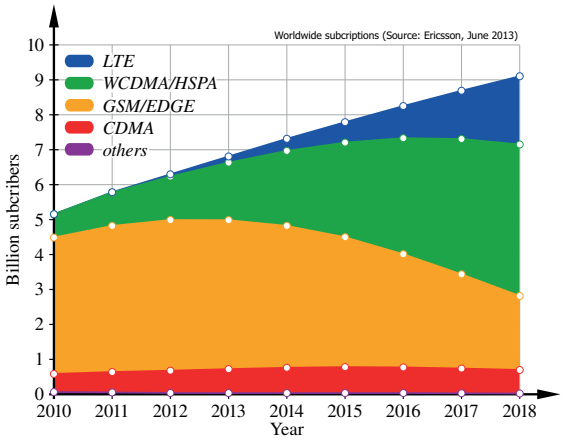


Figure 1.2.: Estimated number of subscribers worldwide partitioned by wireless access technology. (Source: Ericsson traffic exploration tool, June 2013 [4]).

facilitates video and multimedia streaming applications. Notice that the mentioned data rates are only achieved under ideal conditions; typical values are far below these numbers [11].

Figure 1.2 visualizes the current and future number of subscribers of the aforementioned wireless access technologies according to a study conducted by Ericsson [2]. Currently and over the next few years, GSM and its enhancements still dominate the market, while HSPA and even more LTE just start to gain momentum. Still, research today has to focus on the time period beyond the horizon shown in the figure. Hence, in this dissertation the OFDMA based PHY of LTE is considered as the basis technology for proposing novel algorithms and techniques.[1]

1.2. Motivation for this Dissertation and Scope of Work

Cellular networks play a central role in today's global networking and communication infrastructure and will become even more dominant in future, because people expect broadband wireless access everywhere. Coping with the expected exponential growth in mobile data traffic, as shown in Figure 1.1, is a considerable research challenge that many academic and industrial research groups worldwide are trying to solve currently. To achieve the performance of current cellular networks, three approaches have mostly been applied:

- increasing the amount of available spectrum,
- densifying the network,
- improving the PHY.

[1] The same concepts are mostly applicable to similar OFDMA systems as well, e.g., WiMAX [12] and WiFi [13].

Comparing GSM to LTE, the transmission bandwidth per carrier has increased from 200 kHz to 20 MHz, explaining already a large part of the performance improvement achieved between these two systems. LTE-A further enables aggregation of up to five LTE component carriers amounting in a total bandwidth of 100 MHz [14]. It seems straightforward to carry on with this strategy, but there is the problem that the available spectrum at the carrier frequencies enabled by current technology is limited [15]. A research report from 2010 showed that many European network operators do not even have 100 MHz of bandwidth at their disposal [16]. By utilizing the available spectrum more efficiently, e.g., by employing cognitive radio techniques and ultra wideband transmission [17, 18], and by refarming some of the old spectrum and utilizing so called spectrum white spaces [19], it is still possible to harvest additional bandwidth, but it is questionable whether such approaches can sustain the expected exponential growth in demand. An alternative, which is currently highly promoted, is to go for much higher carrier frequencies, so called millimeter-waves [20–22], where several GHz of spectrum are available for communication. But this technology is still in an early stage of research; it is considered as a potential long-term solution for 5G cellular, and it lies outside the scope of this dissertation. In the near future of LTE networks significant bandwidth expansions cannot be expected. Hence, alternative approaches to improve the network capacity are required.

In cellular networks, the same radio spectrum is reused in different cells, in order to improve the network capacity [23]. Several authors have considered the effect of increasing the density of base stations in the network, and improvements in the network capacity with increasing base station density have been revealed, e.g., [24–26]. Hence, network densification is considered as an important means to overcome spectrum shortcomings. This is achieved by augmenting the already existing macro-cellular base stations with comparatively cheap and easy to deploy radio equipment such as remote radio units (RRUs) and small cells (micro, pico, femto cells). Hereby obtained networking architectures are denoted as distributed antenna systems (DASs) [27] and heterogeneous networks (HetNets), respectively. RRUs are remotely controlled antenna elements that are connected to the macro base station via low-latency high-bandwidth dedicated connections using, e.g., micro-wave links or radio-over-fiber technology [28]. The purpose of the RRUs is to reduce the distance between users and the antenna ports of the base station, by distributing antennas over the cell area. Several publications have established the theoretical potential of DASs for improving the indoor coverage [29], reducing the outage probability of the network [30], increasing the network capacity [31], providing a more uniform coverage of the served area [32] and improving the area spectral efficiency (ASE) of the system [32]. Small cells, on the other hand, are autonomous base stations that are mostly deployed in user hot-spots to off-load traffic from the macro base station. They can be either open to all users of the network (open access) or closed to a specific group of users (closed access), leading to distinct throughput performance of the macro and small cell layers [33, 34]. The specifics of DASs, i.e., the pathloss differences with respect to different RRUs, are explicitly treated in the proposed transceivers and feedback algorithms of this dissertation. Small cells, however, do not require a specific treatment in that sense, because they behave like macro base stations.

Improvements of the PHY have increased the peak spectral efficiency of cellular networks from 1.92 bit/s/Hz, achieved with EDGE, to the 30 bit/s/Hz of LTE-A. These gains were obtained to some extent by improving the modulation and coding schemes; EDGE is limited to 8 phase shift keying (PSK), while LTE-A supports up to 64 quadratur amplitude modulation (QAM). The supported spectral efficiency of the transmission channel varies strongly in wireless communication systems, due to fading, pathloss and interference effects [23]. In LTE this is taken into account by supporting several modulation and coding schemes (MCSs), covering an efficiency range from 0.15 bit/modulated symbol to 5.55 bit/modulated symbol, which corresponds to a receive signal to noise ratio (SNR) range of approximately -5 dB to 20 dB [11]. Utilizing feedback information from the receivers, adaptive modulation and coding (AMC) is employed by the base station to match the transmission rate to the current channel conditions. Additionally, cases in which the selected transmission rate was too high, leading to decoding errors [11] at the receiver, are covered by the hybrid automatic repeat request (HARQ) PHY protocol. The optimal calculation of the feedback information for transmission rate adaptation is an important part of this dissertation.

More significant physical layer gains were obtained by the introduction of single-user MIMO (SU-MIMO), improving the robustness (diversity) and capacity of the transmission channel [35–37]. Channel state information (CSI) is useful for achieving the highest performance in multiple antenna wireless communications. In SU-MIMO, CSI is employed to match the spatial signature of the transmit signal to the channel, and to separate the spatially multiplexed data streams at the receiver [38]. CSI at the receiver is commonly acquired using channel estimation [39], while CSI at the transmitter (CSIT) for the downlink channel of the currently most prevalent frequency division duplex (FDD) systems can only be obtained from explicit feedback from the receiver. Codebook based precoding is an efficient way for attaining close to optimal SU-MIMO performance with a reasonable CSI feedback overhead [40]. This strategy was therefore incorporated into the LTE specifications, and Chapter 3 of this dissertation is devoted to the calculation of the optimal SU-MIMO feedback information.

There are a few weak points of SU-MIMO transmission. Firstly, the potential spatial multiplexing gain of the system is limited by the minimum of the number of transmit antennas at the base station and the number of receive antennas at the users. While it is mostly not a significant problem to put a large amount of antennas at the transmitter array, the number of receive antennas is often limited by hardware complexity and available space, especially in hand held devices such as mobile phones or tablet-computers. Additionally, a strong reduction of the channel capacity can be caused by antenna correlation, e.g., due to closely spaced antenna elements [41, 42]. For that reasons, multi-user MIMO (MU-MIMO) became increasingly popular in the research literature over the last few years [43, 44]. With MU-MIMO, the spatial multiplexing capabilities of the transmitter can be fully exploited by serving several users in parallel over the same time-frequency resources. Provided the transmitter has accurate CSI, multi-user interference can be pre-canceled by the transmitter allowing the users to be equipped with far less antennas than the transmitter. When applied in cellular

networks, the complexity burden is hence offloaded from the mobile devices to the stationary base station. In contrast to SU-MIMO, where CSIT is not essential but a welcome extra, MU-MIMO with less receive antennas per user than the total number of spatially multiplexed data streams suffers significantly from imperfect channel knowledge at the transmitter [45, 46]. Therefore accurate and efficient CSI feedback for MU-MIMO systems is an important research area. Chapter 4 of this dissertation is hence devoted to limited feedback MU-MIMO systems, presenting my contributions to predictive quantization on the Grassmann manifold.

Significant gains in the throughput performance of cellular networks are expected from coordinated multi-point (CoMP) transmission [47–49]. CoMP was introduced to cope with the increased inter-cell-interference encountered in dense wireless networks. Through coordination of neighboring cells, the interference can be reduced or even exploited to improve the throughput of the system. There are many approaches for solving the interference problematic, which can basically be classified as coordinated scheduling, coordinated beamforming or joint processing techniques, e.g., [50–52]. As in MU-MIMO, accurate CSIT is central for coordinated beamforming and joint processing techniques to achieve interference coordination. Although CoMP is not a central topic in this dissertation, the proposed limited feedback algorithms can be applied to obtain the necessary CSIT for many coordination techniques, e.g., interference alignment [53] and signal to leakage and noise ratio beamforming [54]. Also, coordinated transmission in a DAS, which is recognized as an important CoMP use case for LTE-A [55], is considered in Chapter 5 as an application scenario of the proposed transceivers and feedback algorithms.

1.3. Outline and Contributions

The focus of this dissertation is on OFDMA based single- and multi-user MIMO multi-cell wireless communication systems, as outlined in the previous section. Limited feedback algorithms for both MIMO concepts are proposed under the assumption of linear transceiver architectures. Additionally, a novel joint CSI feedback selection and antenna combining strategy for block diagonalization (BD)-based MU-MIMO with excess antennas at the receivers is proposed and analytically analyzed. Proportional fairness based multi-user scheduling algorithms are considered for distributing the time, space and frequency resources among the users in an optimal way [56]. For SU-MIMO, well-known scheduling algorithms can be applied [57], while for BD-based MU-MIMO an extension of the semi-orthogonal user selection (SUS) algorithm [58] is proposed. The performance of the proposed algorithms and techniques is analyzed by means of extensive link level simulations employing an LTE-A compliant simulator [59]. In Chapter 2, the basic system model employed throughout the thesis is derived. The main contributions of this dissertation are detailed in Chapters 3 to 5. The thesis is concluded in Chapter 6 with a discussion of the contributions and implications of this work and an outlook on possible future research. Background information as well as detailed proofs and derivations, which are not essential for a basic understanding of my contributions, are provided

in the appendices. The employed abbreviations and the mathematical notation are summarized in Appendices A and B. Below, a description of the contributions of the main chapters and listings of the corresponding publications are given.

Chapter 2: Mathematical Description of Downlink MIMO OFDM Transmission
In this chapter, the basic mathematical description of an OFDMA based single- and multi-user MIMO multi-cell wireless communication network is developed, building the basis for the contributions presented in subsequent chapters. Centralized antenna systems (CASs) as well as DASs are encompassed by the derived system model. A realistic model of the PHY is essential to accurately capture effects observed in wireless networks. The distinctive feature of DASs compared to classical CASs is the introduction of pathloss differences between spatially separated transmission points, improving the macro-diversity of the transmission against shadow fading and reducing the average access-distance to the base station. These pathloss differences are explicitly incorporated in the mathematical description of the system, facilitating their consideration in the proposed algorithms and techniques of later chapters to achieve an improved performance. Notice that in some designated derivations, simplifications are required to facilitate analytical tractability.

Chapter 3: Linear Transmission and CSI Feedback for Single-User MIMO
In MIMO communications, achieving the highest performance requires CSI at both, the transmitter and the receiver. In most wireless communication systems CSIT is acquired over dedicated feedback links from the users. Due to the limited capacity of these feedback links, the CSI is quantized by the users prior to signaling to the base station. Theoretical investigations have shown that a large part of the single-user channel capacity [38] can be achieved with a reasonable feedback overhead employing a codebook based precoding scheme, in which the precoders applied at the transmitter are restricted to a pre-defined codebook of matrices [40]. Provided the receiver has knowledge of the precoder codebook, the preferred precoder for the current channel conditions can be selected and signaled to the transmitter. Different precoder selection criteria have been considered in literature, e.g., maximizing the mutual information between channel input and output, minimizing the bit-error ratio (BER), and so on, mostly for frequency flat single-carrier transmission. Extensions to frequency selective multi-carrier orthogonal frequency division multiplexing (OFDM) have also been proposed, based on clustered feedback and pilot feedback interpolation approaches; see [60] for an overview of this work. Due to the effectiveness of this approach, codebook based precoding is incorporated into the specifications of LTE [8, 10] in a transmission mode denoted as closed loop spatial multiplexing (CLSM). In addition to precoder selection, transmission rank (i.e., the number of spatially multiplexed data streams) and transmission rate adaptation are also considered in CLSM.

In Chapter 3 of this dissertation, feedback selection algorithms for precoder, rank and transmission rate adaptation, applicable to LTE, are proposed. Directly applying the ideas presented in [60], i.e., maximizing the channel capacity or mutual information, or minimizing the theoretical BER, is not

effective for LTE, because the performance of this practical system is not realistically represented by such theoretical measures [11]. Hence, practical constraints (e.g., limited number of MCSs) and the PHY architecture imposed by LTE are considered in the proposed feedback selection algorithms to optimize the achieved throughput of the data transmission. In detail, the following contributions are provided:

- Joint selection of the transmission rate, rank and precoder to maximize the data throughput, while targeting a pre-specified maximum average block-error ratio (BLER). This target BLER is commonly dictated by the application.
- The frequency domain granularity of the feedback indicators can be adapted to match the frequency selectivity of the wireless channel. A trade-off between the feedback overhead and the potential multi-user frequency diversity gain, enabled by frequency selective multi-user scheduling, can hence be obtained.
- Several simplifications are considered to reduce the computational complexity of the proposed algorithms.
- An antenna subset selection algorithm is proposed to select the preferred transmit antenna subset in DASs.
- The performance of the proposed algorithms is investigated by means of simulations, and by comparison to theoretical performance bounds, revealing starting points for potential improvements of the LTE system.

Notice that the proposed algorithms are limited to linear transceiver architectures, for which a post-equalization signal to interference and noise ratio (SINR) can be defined. The contributions of this chapter are published in the following papers:

[i] S. Schwarz, M. Wrulich, and M. Rupp, "Mutual information based calculation of the precoding matrix indicator for 3GPP UMTS/LTE," in *International ITG Workshop on Smart Antennas*, Bremen, Germany, Feb. 2010, pp. 52–58

[ii] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *6th Conf. on Wireless Advanced*, London, UK, June 2010, pp. 1–6

[iii] S. Schwarz, M. Simko, and M. Rupp, "On performance bounds for MIMO OFDM based wireless communication systems," in *IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, San Francisco, CA, June 2011, pp. 311–315

[iv] S. Schwarz and M. Rupp, "Throughput maximizing feedback for MIMO OFDM based wireless communication systems," in *IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, San Francisco, CA, June 2011, pp. 316–320

Chapter 4: Linear Transmission and CSI Feedback for Multi-User MIMO
In this chapter, spatial multiplexing of several users is considered. The focus is put on scenarios in which the total number of spatially multiplexed data streams is larger than the number of receive antennas per user. In that case, the interference between the data streams of different users cannot be eliminated by the users, but must be pre-canceled by the precoder applied at the base station. It is known that the sum-rate capacity of such a MU-MIMO broadcast channel can be achieved by dirty paper coding (DPC) [65–67], but the complexity of this nonlinear precoding approach is currently considered as too high for practical implementations. Of greater practical interest are low-complexity linear precoding techniques such as zero forcing (ZF) beamforming [68] and BD precoding [69], which are therefore also assumed in this dissertation. Notice that BD precoding is a generalization of ZF beamforming, enabling multiple data-streams per user. Under the premise of perfect CSIT, interference between users is perfectly canceled with these techniques, by projecting the signals of interfering users onto the null spaces of each others' channel matrices. With BD and ZF precoding, the maximum number of degrees of freedom (DoF) [70] of the broadcast channel can be achieved, but an SNR loss is encountered with respect to DPC [71]. For the calculation of the precoders at the base station explicit CSI is required. Hence, codebook based precoding and feedback, as applied in SU-MIMO, are not useful in this case. Instead, the channel matrix is directly quantized by the users, employing a quantization procedure and a quantization codebook that is known by the base station. There also exists a codebook based MU-MIMO scheme denoted as per-user unitary beamforming and rate control [72]. According to [49, 73], this method is outperformed by ZF beamforming, even with quantized CSIT, and it is hence not considered in this dissertation. Notice that non-codebook based precoding with up to eight parallel spatial streams is supported in LTE by transmission mode 9, introduced in Rel'10.

When the precoders for BD precoding are calculated from quantized CSI, residual interference between users cannot be avoided, which significantly impacts the throughput performance of the transmission. It has been shown that the number of feedback bits must be scaled linearly with the logarithmic SNR (the SNR in [dB]) to maintain a bounded rate-gap with respect to perfect CSIT [45, 46]. Notice that very similar behavior is observed with interference alignment as well [53, 74]. Accurate CSIT is therefore an essential pre-requisite for such interference cancellation techniques. With a reasonable feedback overhead, the CSI accuracy requirements can often not be met using memoryless quantization. The interest of researchers has thus focused on the exploitation of the temporal channel correlation to improve the quantization accuracy. In addition to capacity limits, the feedback links in wireless communication systems also exhibit delay limits, due to the temporal variation of the wireless channel. It has been shown that outdated CSI degrades the throughput in a similar way as a quantization error [75]. Therefore, vector quantization over multiple time instants [76] is not an option to exploit the temporal channel correlation. Instead, differential [77, 78], predictive [79–82] and progressive refinement [83] quantization algorithms have been proposed to improve the quantization accuracy. Despite the success of [77–83], the algorithms are not as sophisticated as predictive coding techniques in other areas, e.g., speech coding [84]. Also, the

quantizers in [78, 79, 81–83] are restricted to quantization of vector channels for beamforming systems, whereas only single-user spatial multiplexing is considered by the channel matrix quantizers of [77, 80].

Mostly in parallel with this work, the development of a predictive CSI quantizer for MU-MIMO was pursued by myself as well. The description and derivation of this quantizer is a central topic of Chapter 4. CSI feedback for ZF and BD based MU-MIMO systems can be considered as a quantization problem on the Grassmann manifold of n-dimensional subspaces in an underlying m-dimensional Euclidean space ($n \leq m$). Background information on the Grassmannian is provided in Appendix C. The main contributions of this dissertation to Grassmannian quantization are:

- Efficient predictive quantization of temporally correlated n-dimensional subspaces in the m-dimensional Euclidean space. Despite $n \leq m$, no restrictions are imposed on the dimensionality of the quantization problem.
- Prediction of the current subspace from previously quantized observations. The prediction is achieved by translating the Grassmannian prediction problem to the tangent space associated with the manifold.
- Derivation of an adaptive Grassmannian quantization codebook to match the temporal evolution of the Grassmannian source. A local codebook is generated on the Grassmann manifold that covers a certain volume around the prediction of the current subspace. The size of the volume that needs to be covered is determined by the prediction accuracy, which depends on the strength of the temporal correlation of the source.
- Evaluation of the quantization error by means of Monte-Carlo simulations, and comparison to alternative quantizers.

When the number of data streams ℓ per user is less than the number of receive antennas, interference-free transmission is ensured by the BD precoder only over ℓ-dimensional subspaces of the users' channel matrices. An antenna combiner is applied at each user to separate the interference-free signal-space from the interference-contaminated space. To reduce the CSI feedback overhead, a selfish pre-selection of an ℓ-dimensional subspace by the users as CSI feedback is proposed, instead of quantizing the full channel matrix. Within this area, the following contributions are made:

- Proposal of subspace selection strategies to achieve either a maximal channel gain or a minimal CSI quantization error. The trade-off between these two approaches is investigated.
- Construction of the corresponding antenna combiners to filter out the signal-space. Specifically, the quantization based combining (QBC) method of [85] is extended to multiple data-streams.
- Analytic performance investigation of the proposed subspace selection and antenna combining strategies. Upper bounds on the throughput loss compared to perfect CSIT are derived, and the necessary number of feedback bits to achieve a given rate loss is determined.

Finally, the CSI quantizer is extended to frequency-selective multi-carrier OFDM systems by considering two approaches. With feedback pilot interpolation, the frequency-domain correlation of the subspaces on neighboring OFDM subcarriers is exploited by providing CSI feedback only for a subset of subcarriers, so-called CSI pilots, and interpolating the quantized CSI at the base station. This approach has already been investigated in [86] for one-dimensional subspaces. A high density of CSI pilots is required to achieve a satisfying interpolation performance. Alternatively, clustered CSI feedback is considered. Here, a subspace is determined by each user that provides the best representation of a block of subcarriers, avoiding the need for an interpolator at the base station.[2] It is shown, by means of simulations, that this approach can outperform CSI interpolation if the density of CSI pilots is small compared to the coherence bandwidth of the channel. The contributions of this chapter are published in the following conference papers and journal articles:

[v] S. Schwarz and M. Rupp, "Adaptive channel direction quantization based on spherical prediction," in *IEEE International Conference on Communications*, Ottawa, Canada, June 2012, pp. 3757 – 3762

[vi] S. Schwarz and M. Rupp, "Adaptive channel direction quantization – enabling multi user MIMO gains in practice," in *IEEE International Conference on Communications*, Ottawa, Canada, June 2012, pp. 6947 – 6952

[vii] S. Schwarz and M. Rupp, "Adaptive channel direction quantization for frequency selective channels," in *20th European Signal Processing Conference*, Bucarest, Romania, Aug. 2012, pp. 2536–2540

[viii] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems," in *38th International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013

[ix] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on a Grassmann-manifold for limited feedback beamforming systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4450–4462, 2013

[x] S. Schwarz and M. Rupp, "Antenna combiners for block-diagonalization based multi-user MIMO with limited feedback," in *IEEE International Conference on Communications, Workshop: Beyond LTE-A*, Budapest, Hungary, June 2013

[xi] S. Schwarz and M. Rupp, "Subspace quantization based combining for limited feedback block-diagonalization," *IEEE Trans. on Wireless Communications*, vol. PP, no. 99, pp. 1–12, Oct. 2013

[xii] S. Schwarz and M. Rupp, "Distributed downlink multi-user MIMO-OFDM with limited feedback," Oct. 2013, submitted to IEEE Transactions on Wireless Communications

[2] The respective definition of "best" is given in Chapter 4.

Chapter 5: Performance Evaluation and Application Scenarios
In this chapter, the performance of the proposed limited feedback algorithms and antenna combining techniques of Chapters 3 and 4 is compared and evaluated by means of Monte-Carlo simulations. The simulations are conducted with an LTE-A compliant link level simulator [59]. For complexity reasons only a small part of the cellular network, consisting of a few cells, can be simulated. A statistical model of the interference from more distant base stations is used to emulate the effects of this out-of-cell interference.The results are given in terms of the aggregated sum ASE of the cell.

One contribution of this chapter is to provide a cross-comparison of SU-MIMO and MU-MIMO with limited feedback and perfect CSIT, respectively. It is demonstrated that MU-MIMO can exploit the multiplexing capabilities of the base station, provided the CSIT is sufficiently accurate. On the other hand, if the CSI accuracy is low, MU-MIMO is outperformed by SU-MIMO, due to the residual multi-user interference caused by the CSI quantization error.

A performance comparison of different networking architectures is the second contribution of this chapter, applying ZF beamforming and BD precoding based MU-MIMO as transmit strategies. In these simulations a DAS, in which the transmit antennas are distributed over the cell area, is compared to a classical CAS, having all transmit antennas at the central base station, and to a macro-small cell overlay network, in which the macro network is augmented with additional autonomous small cell access points. The potential of the DAS for improving the ASE of the cellular network, compared to the other two networking architectures, is revealed in these investigations.

Similar results to those presented in this chapter, obtained with the same simulation methodology but restricted to single-stream transmission per user, are published in:

[xiii] S. Schwarz, R. Heath, Jr., and M. Rupp, "Multiuser MIMO in distributed antenna systems with limited feedback," in *IEEE 4th Int. Workshop on Heterogeneous and Small Cell Networks, GLOBECOM*, Anaheim, CA, Dec. 2012

[xiv] S. Schwarz, R. Heath Jr., and M. Rupp, "Single-user MIMO versus multi-user MIMO in distributed antenna systems with limited feedback," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, p. 54, 2013

[xv] S. Schwarz, J. Ikuno, M. Simko, M. Taranetz, Q. Wang, and M. Rupp, "Pushing the limits of LTE: A survey on research enhancing the standard," *IEEE Access*, vol. 1, pp. 51–62, 2013

Chapter 2.

Mathematical Description of Downlink MIMO OFDM Transmission

> If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.
>
> *(John von Neumann)*

In this chapter, the input-output relationship describing the downlink transmission between a base station of a cellular network and an attached user is presented and the post-equalization signal to interference and noise ratio (SINR) experienced by the user is derived. The broadband frequency selective wireless channel is converted into a set of non-interfering frequency flat subcarriers by the OFDM based physical layer (PHY) of the considered wireless technology [97, 98]. The cyclic-prefix applied by the transmitter is assumed to be long enough to cover the multipath delay of the wireless channel such that inter-symbol-interference is avoided. Also, inter-carrier-interference, e.g., due to Doppler-shifts or a carrier frequency offset [99–101], is supposed to be negligible. Multiple users are multiplexed in the time-frequency domain by means of OFDMA and/or in the spatial domain using multi-user MIMO (MU-MIMO) techniques. Coordination of the scheduling decisions among multiple base stations (i.e., coordinated scheduling) is not considered in this dissertation. Each base station can be equipped with remote radio units (RRUs) to form a distributed antenna system (DAS). All results in this dissertation are derived considering the presence of interference from other cells of the network. Hence, such out-of-cell interference is explicitly treated in the presented system model. A segment of an exemplary cellular network is shown in Figure 2.1. Each of the three autonomous cells shown in the figure is equipped with a 120° sectorized antenna array at the central base station and two RRUs. In the simulations of Chapter 5 such a network is investigated in detail.

This chapter is organized as follows: In Section 2.1, the input-output relationship of a user of the downlink MIMO OFDM based cellular network is developed, considering in-cell interference between the transmissions to different users of the same cell and out-of-cell interference caused by other cells of the network. The instantaneous post-equalization SINR of the user is derived in Section 2.2. In Section 2.3, the mathematical model employed to describe the effect of distributing

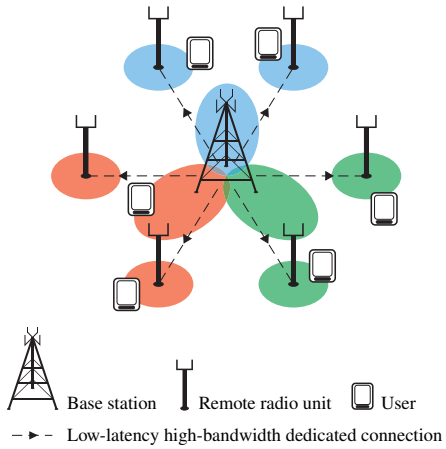


Figure 2.1.: Segment of an exemplary cellular network as considered in this dissertation.

transmit antennas over the cell area is introduced and in Section 2.4 the channel models employed to generate channel realizations for the simulations presented throughout the thesis are detailed.

2.1. Input-Output Relationship of a Cellular User

A cellular network consisting of $I+1$ cells is under consideration in this dissertation. In each cell $i \in \{0, \ldots, I\}$, the attached users are served via a central antenna array that is located at the macro base station and consists of $N_{0,i}$ antenna elements. Additionally, the cell is supplied with R_i RRUs that are distributed over the cell area, as illustrated in Figure 2.1. RRU $r \in \{1, \ldots, R_i\}$ is equipped with an antenna array comprising $N_{r,i}$ antenna elements. The total number of transmit antennas available in cell i is denoted $N_i = \sum_{r=0}^{R_i} N_{r,i}$.

In cell i, a total number of U_i users is served. The receive antenna array of user $u \in \mathcal{U}_i$, $\mathcal{U}_i = \{1, \ldots, U_i\}$ is composed of $M_{u,i}$ antennas. The channel between user u in cell i and RRU r in cell j at OFDM subcarrier n and symbol-time k is described by an $M_{u,i} \times N_{r,j}$ dimensional complex-valued channel matrix $\boldsymbol{\Xi}_{u,i}^{(r,j)}[n,k]$, employing the equivalent complex baseband representation of the OFDM wireless communication system, see, e.g., [102–104]. The channels to the central antenna array of cell j as well as all R_j RRUs are combined in the matrix

$$\boldsymbol{\Xi}_{u,i}^{(j)}[n,k] = \left[\boldsymbol{\Xi}_{u,i}^{(0,j)}[n,k], \ldots, \boldsymbol{\Xi}_{u,i}^{(R_j,j)}[n,k]\right] \in \mathbb{C}^{M_{u,i} \times N_j}. \tag{2.1}$$

Instead of representing the channel with $\boldsymbol{\Xi}_{u,i}^{(j)}[n,k]$, the $N_j \times M_{u,i}$ dimensional cell channel matrix $\mathbf{H}_{u,i}^{(j)}[n,k]$, defined by

$$\mathbf{H}_{u,i}^{(j)}[n,k] = \boldsymbol{\Xi}_{u,i}^{(j)}[n,k]^{\mathrm{H}}, \tag{2.2}$$

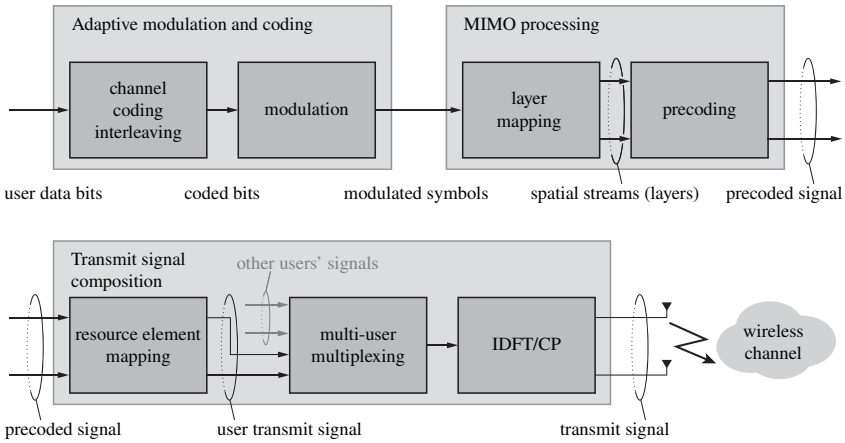


Figure 2.2.: MIMO OFDM transceiver architecture as considered in this dissertation.

is employed in this thesis, to simplify and shorten derivations and descriptions of later chapters. To simplify notation, the superscript $^{(j)}$ is omitted whenever in-cell channels, $j = i$, are considered. Also, if the description of an algorithm or method is independent of the subcarrier index n and/or the symbol-time index k, the respective index is dropped.[1] Following Long Term Evolution (LTE) notations, the subcarrier and symbol-time index pair $[n, k]$ is denoted as a resource element (RE).

The set of users that is selected by the scheduling algorithm of cell i to be served on a given RE is denoted $\mathcal{S}_i[n, k] \subseteq \mathcal{U}_i$. The employed single-user MIMO (SU-MIMO) and MU-MIMO schedulers are described in detail in Chapters 3 and 4, respectively. The number of served users in cell i on RE $[n, k]$ is given by $S_i[n, k] = |\mathcal{S}_i[n, k]|$. The transmit symbol vector intended for user $u \in \mathcal{S}_i[n, k]$ is written as $\mathbf{x}_{u,i}[n, k] \in \mathbb{C}^{\ell_{u,i}[k]\times 1}$, with $\ell_{u,i}[k] \leq M_{u,i}$ being the number of data-streams spatially multiplexed to user u. In general, the number of streams per user $\ell_{u,i}[k]$ could potentially change over subcarriers. With the considered LTE compliant transceiver architecture shown in Figure 2.2 this is not possible, because the mapping of the user data onto spatial streams (layer mapping) is performed before the mapping onto REs. Hence, $\ell_{u,i}[k]$ can at most change from one OFDM symbol to the next, provided that each OFDM symbol is coded individually.[2]

The total number of streams of cell i transmitted on RE $[n, k]$ is denoted $\ell_i[n, k] = \sum_{u\in\mathcal{S}_i[n,k]} \ell_{u,i}[k]$. To account for the maximum possible spatial multiplexing capabilities of the base station in cell i, the number of streams is constrained as $\ell_i[n, k] \leq N_i$. Although $\ell_{u,i}[k]$ is independent of the subcarrier index, the total number of streams depends on n due to the frequency selective multi-user scheduling. The transmit symbol vector is normalized as

$$\mathbb{E}\left(\mathbf{x}_{u,i}[n, k]\,\mathbf{x}_{u,i}[n, k]^{\mathrm{H}}\right) = \mathbf{I}_{\ell_{u,i}[k]}. \tag{2.3}$$

[1] Such situations are particularly indicated in the text.

[2] In LTE the number of streams per user is constant for the duration of one slot consisting of seven OFDM symbols. It is assumed that such restrictions are handled by the scheduler.

Prior to transmission over the wireless channel, the user symbol vector $\mathbf{x}_{u,i}[n,k]$ is precoded with a precoding matrix $\mathbf{F}_{u,i}[n,k] \in \mathbb{C}^{N_i \times \ell_{u,i}[k]}$, mapping the $\ell_{u,i}[k]$-dimensional transmit symbol vector onto the N_i transmit antennas. Notice that the allocation of the available transmit power $P_i[n,k]$ among users and spatial streams is considered in the precoding matrices, as detailed in Chapters 3 and 4. The precoders are obtained such that $P_i[n,k]$ is conserved, irrespective of the number of users and spatial streams. In the simulation results presented in this dissertation, power loading over REs is not considered: $P_i[n,k] = P_i,\ \forall n,k$. Better performance can be achieved by allocating the available transmit power over subcarriers following a water-filling power allocation policy [105], provided the required channel state information (CSI) at the transmitter (CSIT) is available.

With this notation, the $M_{u,i}$-dimensional received signal vector of user u in cell i at RE $[n,k]$ is

$$\mathbf{r}_{u,i}[n,k] = \mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k]\mathbf{x}_{u,i}[n,k] + \mathbf{H}_{u,i}[n,k]^{\mathrm{H}} \sum_{\substack{s \in \mathcal{S}_i[n,k] \\ s \neq u}} \mathbf{F}_{s,i}[n,k]\mathbf{x}_{s,i}[n,k]$$

$$+ \underbrace{\sum_{j=0, j\neq i}^{I} \mathbf{H}_{u,i}^{(j)}[n,k]^{\mathrm{H}} \sum_{s \in \mathcal{S}_j[n,k]} \mathbf{F}_{s,j}[n,k]\mathbf{x}_{s,j}[n,k] + \mathbf{z}_{u,i}[n,k]}_{\tilde{\mathbf{z}}_{u,i}[n,k]}, \tag{2.4}$$

where the additive white Gaussian noise (AWGN) added at the receiver is denoted $\mathbf{z}_{u,i}[n,k] \in \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma_z^2\, \mathbf{I}_{M_{u,i}}\right)$. The intended signal of user u is represented by the first summand in this equation. In-cell interference between the streams spatially multiplexed to several users in the same cell i is captured in the second summand. Out-of-cell interference from other cells $j \neq i$ of the cellular network is taken into account in the third term on the right hand side of Equation (2.4). The sum of out-of-cell interference and receiver noise is called the effective noise vector $\tilde{\mathbf{z}}_{u,i}[n,k]$.

The users are assumed to employ linear receive filters to equalize their respective channels and to separate the spatially multiplexed data-streams from each other and from the interference caused by the transmission to other users. The $\ell_{u,i}[k] \times M_{u,i}$ dimensional receive filtering matrix applied by user u in cell i is written as $\mathbf{G}_{u,i}[n,k]^{\mathrm{H}}$. Applying this matrix to the received signal vector, the estimated symbol vector is obtained as

$$\mathbf{y}_{u,i}[n,k] = \mathbf{G}_{u,i}[n,k]^{\mathrm{H}}\, \mathbf{r}_{u,i}[n,k] = \left(\mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}[n,k]\right)^{\mathrm{H}} \mathbf{F}_{u,i}[n,k]\mathbf{x}_{u,i}[n,k]$$

$$+ \left(\mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}[n,k]\right)^{\mathrm{H}} \sum_{\substack{s \in \mathcal{S}_i[n,k] \\ s \neq u}} \mathbf{F}_{s,i}[n,k]\mathbf{x}_{s,i}[n,k] + \mathbf{G}_{u,i}[n,k]^{\mathrm{H}}\, \tilde{\mathbf{z}}_{u,i}[n,k]. \tag{2.5}$$

The product of channel matrix and receive filter is referred to as effective channel matrix

$$\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] = \mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}[n,k]. \tag{2.6}$$

Considering the transceiver architecture of Figure 2.2, all signal processing steps starting from the spatial streams at the output of the layer mapper up to the received signal, obtained at the receiver after the inverse OFDM processing and the equalization, are incorporated in the input-output relationship of Equation (2.5). The non-linear mappings involved in the adaptive modulation and coding (AMC) stage are not covered by the system model, but are considered in more detail in Chapter 3.

2.2. Instantaneous Post-Equalization SINR

The supported transmission rate of communication channels is frequently subject to significant fluctuations over time and frequency. Such fading effects are especially pronounced in cellular communication systems, where multipath propagation caused by reflections and refractions, shadowing of the radio signal due to obstacles, and movement of the users and/or obstacles can result in variations of the signal strength in the order of tens of decibels [106]. A common method to respond to these fading effects is the application of transmission rate adaptation, e.g., by means of AMC as employed in LTE, such as to match the current data rate to the channel conditions. As detailed throughout Chapter 3, rate adaptation in practical systems can be based on the instantaneous per-stream SINR experienced after the receive filter. Considering the input-output relationship of Equation (2.5), the instantaneous post-equalization SINR of stream $\nu \in \{1, \ldots, \ell_{u,i}[k]\}$ of user u in cell i is obtained as

$$\beta_{\nu,u,i}[n,k] = \frac{S_{\nu,u,i}[n,k]}{Z_{\nu,u,i}[n,k] + I^{(\text{self})}_{\nu,u,i}[n,k] + I^{(\text{in})}_{\nu,u,i}[n,k] + I^{(\text{out})}_{\nu,u,i}[n,k]}, \tag{2.7}$$

$$S_{\nu,u,i}[n,k] = \left|\mathbf{g}_{\nu,u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{f}_{\nu,u,i}[n,k]\right|^2,$$

$$Z_{\nu,u,i}[n,k] = \sigma_z^2 \left\|\mathbf{g}_{\nu,u,i}[n,k]\right\|^2,$$

$$I^{(\text{self})}_{\nu,u,i}[n,k] = \sum_{\mu=1,\mu\neq\nu}^{\ell_{u,i}[k]} \left|\mathbf{g}_{\nu,u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{f}_{\mu,u,i}[n,k]\right|^2,$$

$$I^{(\text{in})}_{\nu,u,i}[n,k] = \sum_{\substack{s\in\mathcal{S}_i[n,k]\\ s\neq u}} \left\|\mathbf{g}_{\nu,u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{s,i}[n,k]\right\|^2,$$

$$I^{(\text{out})}_{\nu,u,i}[n,k] = \sum_{j=0,j\neq i}^{I} \sum_{s\in\mathcal{S}_j[n,k]} \left\|\mathbf{g}_{\nu,u,i}[n,k]^{\mathrm{H}}\mathbf{H}^{(j)}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{s,j}[n,k]\right\|^2,$$

where $\mathbf{g}_{\nu,u,i}[n,k]$ and $\mathbf{f}_{\nu,u,i}[n,k]$ denote the ν-th column of $\mathbf{G}_{u,i}[n,k]$ and $\mathbf{F}_{u,i}[n,k]$, respectively. Notice that the statistical independence of the data symbols of the streams of a user according to Equation (2.3) has been exploited to obtain Equation (2.7), and that the symbols of different users

are too assumed as statistically independent

$$\mathbb{E}\left(\mathbf{x}_{u,i}[n,k]\mathbf{x}_{s,j}[n,k]^{\mathrm{H}}\right) = \mathbf{0}_{\ell_{u,i}[k]\times\ell_{s,j}[k]},\ \text{whenever } u \neq s \vee i \neq j. \tag{2.8}$$

In Equation (2.7), the useful signal power of stream ν is represented by the term $S_{\nu,u,i}[n,k]$ in the numerator. The residual interference power between the streams of the user after equalization is given by $I^{(\text{self})}_{\nu,u,i}[n,k]$. The in-cell interference power on stream ν from other users that are served in parallel in the same cell i is denoted $I^{(\text{in})}_{\nu,u,i}[n,k]$, and the out-of-cell interference power from other cells that operate at the same frequency is captured in the term $I^{(\text{out})}_{\nu,u,i}[n,k]$. Notice that depending on the considered transmission and reception strategy, some of these terms are equal to zero, e.g., for SU-MIMO there is no in-cell interference. In Chapters 3 and 4 this general SINR expression is further specialized to account for the considered transceiver architectures.

2.3. DAS Modeling

The most important effect that is caused by distributing antenna arrays over the cell area is the introduction of pathloss differences with respect to the spatially separated antenna arrays, improving the macro-diversity against shadow fading and reducing the average access-distance to the base station [31]. To account for these large-scale effects, the cell channel matrix is decomposed as

$$\mathbf{H}^{(j)}_{u,i}[n,k] = {\mathbf{C}^{(j)}_{u,i}}^{1/2}\bar{\mathbf{H}}^{(j)}_{u,i}[n,k], \tag{2.9}$$

where the channel gain matrix $\mathbf{C}^{(j)}_{u,i}$ is diagonal and characterizes pathloss and large-scale shadow fading, and the matrix $\bar{\mathbf{H}}^{(j)}_{u,i}[n,k]$ captures the time-frequency selective small-scale fading, with

$$\mathbb{E}\left(\left|\left[\bar{\mathbf{H}}^{(j)}_{u,i}[n,k]\right]_{l,m}\right|^2\right) = 1,\ \forall l,m. \tag{2.10}$$

The channel gain matrix is assumed as constant over the time-frequency span of interest. For the bandwidths supported by current cellular systems ($\approx 100\,\text{MHz}$), the frequency independence of the pathloss is satisfied with a high level of accuracy [107], although ultra-wideband systems may require the consideration of the frequency dependence of $\mathbf{C}^{(j)}_{u,i}$. Similarly in the time-domain, accounting for the time dependence of $\mathbf{C}^{(j)}_{u,i}$ may be necessary in high-mobility scenarios, but the focus of this dissertation lies on low to moderate mobility situations, in which updating $\mathbf{C}^{(j)}_{u,i}$ on a time-scale much larger than the duration of the data transmission is sufficient. The channel gain matrix is written as

$$\mathbf{C}^{(j)}_{u,i} = \operatorname{diag}\left(\gamma^{(1,j)}_{u,i}, \gamma^{(2,j)}_{u,i}, \ldots, \gamma^{(N_j,j)}_{u,i}\right), \tag{2.11}$$

where the large-scale channel gain between all receive antennas of user u in cell i and the transmit antenna k in cell j is denoted $\gamma^{(k,j)}_{u,i}$. Notice, if multiple antenna elements are collocated at one RRU,

then the corresponding large-scale channel gains are equal. The channel gains are determined by the pathloss due to the distance between the transmitting and receiving antennas, by a possible shadow fading loss and also by the transmit antenna gain if directional antennas are considered.

2.4. Simulation Relevant Channel Modeling

The time-frequency variation of the wireless channel as well as the spatial correlation properties are captured in the small-scale fading channel matrices $\bar{\mathbf{H}}_{u,i}^{(j)}[n,k]$. If not mentioned otherwise, no specific assumptions are made about $\bar{\mathbf{H}}_{u,i}^{(j)}[n,k]$ during the derivations of the proposed algorithms and techniques. In the conducted Monte-Carlo simulations, however, well defined models are employed to generate channel matrices that represent realistic scenarios, which are introduced in the following.

The matrices $\bar{\mathbf{H}}_{u,i}^{(j)}[n,k]$ are representations of the transmission channel incorporating the OFDM processing at transmitter and receiver. They are hence obtained by sampling the channel transfer function, see [104]. Physical models of the wireless channel, as obtained from measurement campaigns, however, exist for the time-domain representation of the channel. In the considered simulation models, an L-tap finite impulse response (FIR) filter, as specified in [108, 109], is employed to represent the multipath time-domain channel. The L channel matrices corresponding to the multipath components are denoted $\mathbf{T}_{u,i,1}^{(j)}(t),\ldots,\mathbf{T}_{u,i,L}^{(j)}(t)$, where (t) indicates their dependence on time. Below, important statistical properties of the channel are summarized that are employed throughout the thesis to characterize the simulation scenarios.

Frequency correlation: The correlation of the elements of the matrices $\bar{\mathbf{H}}_{u,i}^{(j)}[n,k]$ in the frequency domain, i.e., over subcarriers n, is determined by the delay-spread of the channel impulse response in the time-domain. Assuming a root mean square (RMS) delay spread of τ_{RMS}, the 50% coherence bandwidth of the channel is defined as

$$B_C = \frac{1}{\kappa\,\tau_{\text{RMS}}}, \tag{2.12}$$

with κ denoting a constant that depends on the environment [110]. The 50% coherence bandwidth indicates the bandwidth over which the frequency correlation function of the elements of the matrices $\bar{\mathbf{H}}_{u,i}^{(j)}[n,k]$ is above 0.5. Throughout this thesis, the value $\kappa = 5$ specified in [111] is employed to obtain an order of magnitude estimate, facilitating comparison of different channel models.

Temporal correlation: The correlation of the elements of the matrices $\bar{\mathbf{H}}_{u,i}^{(j)}[n,k]$ in the time domain, i.e., over OFDM symbols k, is determined by the Doppler-spread of the taps of the channel

impulse response. Similarly to the definition of the coherence bandwidth above, the 50% coherence time of the channel, assuming Clarke's model [112], is defined as [111]

$$T_C = \frac{9}{16\pi f_d}, \tag{2.13}$$

where f_d is the maximum channel Doppler frequency. Again, T_C should be interpreted as an order of magnitude estimate. The constant multiplying the factor $\frac{1}{f_d}$ is determined by the shape of the Doppler spectrum. In the considered simulations, a channel model following Jakes' sum-of-sinusoids approach [113] is employed to determine the temporal evolution of the wireless channel [114]. The given constant $\frac{9}{16\pi}$ in (2.13) is valid for Jakes' Doppler spectrum [113].

Spatial correlation: The spatial correlation between the elements of the matrices $\mathbf{T}^{(j)}_{u,i,\ell}(t)$ is determined in the presented simulations by a Kronecker correlation model [115]

$$\mathbf{T}^{(j)}_{u,i,\ell}(t) = \mathbf{R}_{\text{TX}}^{1/2} \bar{\mathbf{T}}^{(j)}_{u,i,\ell}(t) \mathbf{R}_{\text{RX}}^{1/2}, \tag{2.14}$$

$$\mathbf{R}_{\text{TX}} = \mathbb{E}\left(\mathbf{T}^{(j)}_{u,i,\ell}(t) \mathbf{T}^{(j)}_{u,i,\ell}(t)^{\text{H}}\right),$$

$$\mathbf{R}_{\text{RX}} = \mathbb{E}\left(\mathbf{T}^{(j)}_{u,i,\ell}(t)^{\text{H}} \mathbf{T}^{(j)}_{u,i,\ell}(t)\right).$$

Similarly to the channel gain matrix, the matrices $\mathbf{R}_{\text{TX}}$ and $\mathbf{R}_{\text{RX}}$ are assumed independent of time and also independent of the FIR tap index ℓ. The matrix $\bar{\mathbf{T}}^{(j)}_{u,i,\ell}(t)$ consists of independent and identically distributed (i.i.d.) elements

$$\mathbb{E}\left(\text{vec}\left(\bar{\mathbf{T}}^{(j)}_{u,i,\ell}(t)\right) \text{vec}\left(\bar{\mathbf{T}}^{(j)}_{u,i,\ell}(t)\right)^{\text{H}}\right) = \mathbf{I}_{N_j \cdot M_{u,i}}. \tag{2.15}$$

Then the spatial correlation of $\mathbf{T}^{(j)}_{u,i,\ell}(t)$ is obtained as

$$\mathbb{E}\left(\text{vec}\left(\mathbf{T}^{(j)}_{u,i,\ell}(t)\right) \text{vec}\left(\mathbf{T}^{(j)}_{u,i,\ell}(t)\right)^{\text{H}}\right) = \mathbf{R}_{\text{RX}} \otimes \mathbf{R}_{\text{TX}}. \tag{2.16}$$

The correlation matrices are generated according to the Third Generation Partnership Project (3GPP) document [116, Appendix B.2.3] assuming uniform linear arrays; e.g., the user-side correlation matrices for two and four receive antennas are

$$\mathbf{R}_{\text{RX}} = \begin{bmatrix} 1 & \alpha_{\text{corr}} \\ \alpha_{\text{corr}} & 1 \end{bmatrix}, \; \mathbf{R}_{\text{RX}} = \begin{bmatrix} 1 & \alpha_{\text{corr}}^{\frac{1}{9}} & \alpha_{\text{corr}}^{\frac{4}{9}} & \alpha_{\text{corr}} \\ \alpha_{\text{corr}}^{\frac{1}{9}} & 1 & \alpha_{\text{corr}}^{\frac{1}{9}} & \alpha_{\text{corr}}^{\frac{4}{9}} \\ \alpha_{\text{corr}}^{\frac{4}{9}} & \alpha_{\text{corr}}^{\frac{1}{9}} & 1 & \alpha_{\text{corr}}^{\frac{1}{9}} \\ \alpha_{\text{corr}} & \alpha_{\text{corr}}^{\frac{4}{9}} & \alpha_{\text{corr}}^{\frac{1}{9}} & 1 \end{bmatrix}, \tag{2.17}$$

with $\alpha_{\text{corr}} \in [0, 1]$ determining the strength of the correlation. The transmit correlation matrix is always considered as an identity matrix, assuming transmit antennas at the base station that are spaced sufficiently far apart from each other to achieve negligible correlation.

Chapter 3.

Linear Transmission and CSI Feedback for Single-User MIMO

> The worthwhile problems are the ones you can really solve or help solve, the ones you can really contribute something to.
>
> *(Richard P. Feynman)*

Since the introduction of single-user spatial multiplexing in radio communications in the early nineties of the last century by A. Paulraj and T. Kailath [117], single-user MIMO (SU-MIMO) is heavily adopted in commercial systems to leverage the theoretically established advantages of multiple antennas at the transmitter and the receiver for improving the transmission rate (spatial multiplexing) and the reliability (diversity) of the communication channel [35–37]. The highest performance in MIMO communications is achieved if instantaneous channel state information (CSI) is available at both, the transmitter and the receiver. A comprehensive overview of results on the Shannon capacity of SU-MIMO with different assumptions about the availability of CSI at the transmitter and the receiver is provided in [118]. The focus of this dissertation is hence on the acquisition of instantaneous CSI. It is assumed that the receivers obtain CSI autonomously using channel estimation, while the base station relies on finite rate feedback links from the users to obtain instantaneous CSI at the transmitter (CSIT).

In this chapter, the most commonly implemented variant of SU-MIMO in commercial wireless communications, that is, codebook based linear precoding [40], is considered. With codebook based precoding efficient CSI feedback is facilitated, enabling reasonably close to optimal performance with an acceptable feedback overhead [119]. The capacity of the point-to-point AWGN MIMO channel with perfect instantaneous CSI at the transmitter and the receiver can be achieved by singular value decomposition (SVD) based unitary precoding and reception along with power loading across the eigenmodes of the channel [106]. In the considered wireless communication system, motivated by the approach standardized in LTE, a coarse approximation of SVD based precoding is implemented. In this approach the unitary precoders are restricted to a pre-specified set (codebook) of matrices and the continuous power loading is replaced with an on-off switching of spatial data

streams, effectively leading to a codebook of semi-unitary precoders.[1] Based on this codebook, the preferred precoder that optimizes a given performance criterion is determined by the users and this information is efficiently conveyed to the base station by signaling the codebook index together with the number of spatial streams that are switched on, i.e., the transmission rank. This idea of implicit CSI feedback can be straightforwardly extended to MIMO OFDM, by providing feedback for each resource element (RE) separately. To reduce the CSI feedback overhead, the correlation of the wireless channel in the time-frequency domain can be exploited by means of clustered feedback [120] or feedback interpolation [121]. Again motivated by the approach standardized in LTE, clustered feedback is considered in this dissertation. Here, the optimal precoder is determined for a set of REs, denoted as resource block (RB), effectively scaling down the feedback overhead by the size of the set in the time domain and in the frequency domain. The feedback clustering considered in this chapter is detailed in Section 3.2.

In addition to CSI feedback for MIMO precoding, feedback for transmission rate adaptation is another important topic of this chapter. In this dissertation, transmission rate adaptation is achieved by means of adaptive modulation and coding (AMC), supporting a set of pre-specified combinations of modulation alphabets and coding rates of the forward error correction code (FEC), denoted as modulation and coding schemes (MCSs), to cover the projected operating regime of the cellular network. The binary channel coder and the modulation mapper are joined over a bit-interleaver, forming a bit-interleaved coded modulation (BICM) architecture. BICM is employed by many of today's waveform communication systems, due to its high flexibility in allowing to combine virtually any binary code with any modulation format, while providing performance close to Shannon capacity [122, 123]. With the considered system architecture of Figure 2.2, the same MCS is applied on all REs that are assigned to a user. Hence, an average channel quality measure must be defined that realistically represents the achievable transmission rate over a multitude of OFDM subcarriers. To this end, effective signal to interference and noise ratio (SINR) averaging, a technique commonly employed for link to system level abstraction [11], is used in this dissertation to determine the average channel quality, which is utilized as CSI feedback for transmission rate adaptation.

If multiple users are served in a cell, CSI about the achievable transmission rate can additionally be exploited during multi-user scheduling to achieve a multi-user diversity gain. It has been shown that the sum capacity over Rayleigh fading channels with opportunistic scheduling scales double logarithmically with the number of users, due to the statistically independent fading of the different users; see, e.g., [124]. Leveraging the multi-user diversity in OFDMA requires time-frequency selective feedback from the users. Similarly to the precoder feedback, clustered feedback is useful to reduce the CSI feedback overhead of the achievable rate feedback as well. The trade-off between the feedback granularity and the achieved multi-user diversity gain is investigated in a simulation-based study at the end of this chapter.

[1] A $p \times q$ matrix $\mathbf{U}$ with $q \leq p$ is called semi-unitary if $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}_q$.

This chapter is organized as follows: In Section 3.1, the concepts of AMC are described in more detail and the principles of codebook based precoding and capacity achieving SVD based precoding are reviewed, considering a system architecture in accordance with the LTE standard. Taking into account the constraints mentioned above, a discrete joint-optimization problem for finding the optimal feedback indicators for the achievable transmission rate, the precoders and the transmission rank is defined in Section 3.2. To reduce the complexity of an exhaustive search, a suboptimal sequential solution of the optimization problem is derived. Also, an antenna subset selection algorithm is proposed that is useful for improving the performance of codebook based precoding in distributed antenna systems (DASs). In Section 3.3, the throughput and block-error ratio (BLER) performance of the proposed feedback selection algorithms is investigated by means of Monte-Carlo simulations. Furthermore, the performance is compared to theoretical throughput bounds that take into account the constraints imposed by the considered technology. The details of these bounds are provided in Appendix D. The methods presented in this chapter are published in [61–64, 96].

3.1. Principles of Link Adaptation and Linear Precoding

3.1.1. Link Adaptation

In cellular communications, the quality of the transmission link is subject to significant variations over time and frequency, due to macroscopic (pathloss, shadowing) and microscopic (multipath interference) fading and due to interference from neighboring cells [106]. A common approach to counteract these effects and to improve the reliability and data rate of the transmission is link adaptation, where the transmission parameters of the communication link are adapted to account for the current channel conditions. The preferred link adaptation method of early code division multiple access (CDMA) systems is fast power control, where variations in the channel gain are compensated by the transmit power of the base station. More recently, e.g., in High Speed Packet Access (HSPA) and LTE, link adaptation is implemented through AMC, improving the network capacity by exploiting the fading conditions of the wireless channel [125]. In all cases, link adaptation is based on CSIT.

In this dissertation, link adaptation by means of AMC in accordance with the LTE standard [10, 126] is considered. The corresponding BICM system architecture of LTE is illustrated in Figure 3.1, presenting a more detailed block-diagram of the AMC part of Figure 2.2. The user input data is processed according to the following steps:

- Mapping of payload bits onto codewords. In LTE, the number of codewords $C_{\ell_{u,i}[k]}$ of a user is restricted as $C_{\ell_{u,i}[k]} \leq 2$ and is implicitly determined by the transmission rank, i.e., the number of spatial data streams $\ell_{u,i}[k]$.

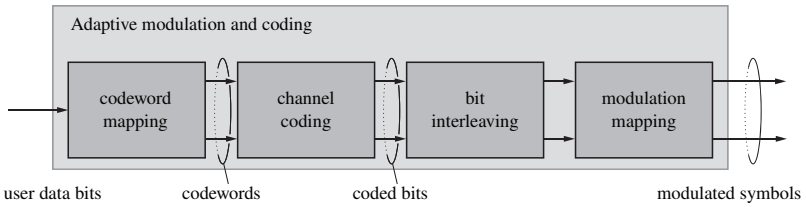


Figure 3.1.: AMC based on a BICM architecture according to the LTE specifications.

- Channel coding of the codewords. The coding rate of each codeword can be chosen independently in LTE, accounting for the varying channel quality over spatial streams.
- Bit interleaving of the coded bits. Independent interleaving is applied to each codeword in LTE. Bit interleaving is employed to reduce the impact of error bursts due to channel fading [127], improving the performance of the FEC.
- Mapping of the coded bits onto modulated symbols. In AMC, a set of modulation alphabets $\mathcal{A}$ is supported to enable robust transmission in case of bad channel conditions and high spectral efficiency in case of good channel quality. In LTE, 4 quadratur amplitude modulation (QAM), 16 QAM and 64 QAM are implemented.

Following the AMC stage, the modulated symbols corresponding to the codewords are mapped onto the spatial streams by the layer mapping shown in Figure 2.2. In case that $\ell_{u,i}[k] > C_{\ell_{u,i}[k]}$, a codeword is divided onto multiple spatial streams [10].

The capacity of BICM is derived for memoryless channels in [122]. In Figure 3.2, a comparison between the capacity achieved with the BICM architecture and the Shannon capacity is shown for a single-input single-output (SISO) AWGN channel. The three BICM curves correspond to the performance of 4/16/64 QAM. The saturation of the BICM capacity at high signal to noise ratio (SNR) occurs due to the finite number of symbols in the modulation alphabets. It can be seen that the theoretical spectral efficiency obtained with BICM is close to the Shannon capacity (implying Gaussian signaling) over a wide SNR range, explaining the popularity of BICM for commercial implementations. Notice though that powerful channel codes and complex detection algorithms are required in practice to achieve close to optimal performance with BICM, involving soft-information exchange between the symbol demapper and the channel decoder and possibly iterative detection [123, 128]. Also shown in the figure is the performance achieved with the MCSs defined in the LTE specifications [126] as obtained from link level simulations employing the Vienna LTE link level simulator [59]. No iterations between the soft-output symbol demapper and the soft-input channel decoder are considered in these simulations and perfect CSI at the receiver is assumed for data detection. The LTE efficiency curve is obtained as the maximum over the 15 MCSs of LTE. The efficiency is calculated only from frequency bins that are used for data transmission, i.e., the additional loss caused by the overhead for reference symbols is not taken into account. The values at the saddle points of this curve are equal to the peak spectral efficiencies of the individual

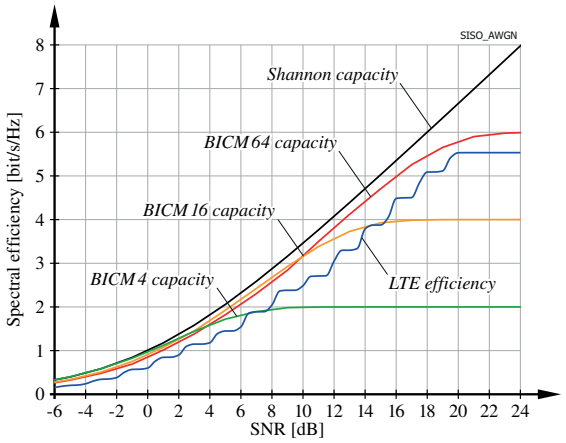


Figure 3.2.: Shannon capacity versus BICM capacity and the efficiency of LTE over an AWGN channel.

MCSs. A loss of approximately 2 dB is encountered by LTE compared to the BICM capacity. Over a SISO AWGN channel the SNR range from -10 dB to 20 dB can effectively be covered by the link adaptation of LTE. The incurred loss of LTE compared to the BICM capacity can be attributed to the imperfect operation of the channel code.

3.1.2. Linear Precoding for SU-MIMO

In SU-MIMO transmission, only one user is served per cell on a given time-frequency resource. Hence, the in-cell interference term $I^{(\text{in})}_{\nu,u,i}[n,k]$ in the SINR expression (2.7) is equal to zero. In this chapter it is assumed that the out-of-cell interference is treated as additional Gaussian noise by the transmitter and the receiver, and that the users are able to estimate the power of the effective noise defined in Equation (2.4). Notice that Gaussianity of the out-of-cell interference may not be fulfilled in all cases, e.g., if there are only a few dominant interferers. Then better performance can be achieved with receivers that estimate the interference statistics [129]. However, Gaussianity of the out-of-cell interference can be justified by the central limit theorem, considering the increasing density of cellular networks. With this simplifying assumption, the general input-output relationship of user u in cell i according to Section 2.1 is reduced to a point-to-point AWGN MIMO channel, possibly with distributed transmit antennas

$$\mathbf{y}_{u,i}[n,k] = \left(\mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}[n,k]\right)^{\mathrm{H}} \mathbf{F}_{u,i}[n,k]\mathbf{x}_{u,i}[n,k] + \mathbf{G}_{u,i}[n,k]^{\mathrm{H}}\,\tilde{\mathbf{z}}_{u,i}[n,k], \tag{3.1}$$

where $\tilde{\mathbf{z}}_{u,i}[n,k] \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \tilde{\sigma}_z^2\,\mathbf{I}_{M_{u,i}}\right)$.

SVD based precoding and equalization: The capacity of the point-to-point AWGN MIMO channel with perfect CSI at the transmitter and the receiver can be achieved by SVD based precoding

and equalization [106]. Although perfect CSI is unrealistic in practice, the method still provides a good benchmark for comparing the performance of limited feedback techniques and it is hence briefly reviewed.

The compact-form SVD of the channel matrix $\mathbf{H}_{u,i}[n,k]$ at RE $[n,k]$ can be written as

$$\mathbf{H}_{u,i}[n,k] = \mathbf{U}_{u,i}[n,k]\mathbf{\Sigma}_{u,i}[n,k]\mathbf{V}_{u,i}[n,k]^{\mathrm{H}}, \tag{3.2}$$

$$\mathbf{\Sigma}_{u,i}[n,k] = \operatorname{diag}\left(\sigma_{u,i}^{(1)}[n,k], \ldots, \sigma_{u,i}^{(\ell_{\max})}[n,k]\right),$$

where the semi-unitary matrices $\mathbf{U}_{u,i}[n,k] \in \mathbb{C}^{N_i \times \ell_{\max}}$ and $\mathbf{V}_{u,i}[n,k] \in \mathbb{C}^{M_{u,i} \times \ell_{\max}}$ denote the matrices of left singular vectors and right singular vectors, respectively. The ν-th diagonal element $\sigma_{u,i}^{(\nu)}[n,k]$ of the singular value matrix $\mathbf{\Sigma}_{u,i}[n,k]$ is equal to the ν-th largest singular value of $\mathbf{H}_{u,i}[n,k]$. Assuming a full-rank channel, the maximum number of streams is $\ell_{\max} = \min(N_i, M_{u,i})$. By setting the precoder and the receive filter according to

$$\mathbf{F}_{u,i}[n,k] = \mathbf{U}_{u,i}[n,k]\mathbf{P}_{u,i}[n,k]^{1/2}, \tag{3.3}$$

$$\mathbf{G}_{u,i}[n,k] = \mathbf{V}_{u,i}[n,k], \tag{3.4}$$

with $\mathbf{P}_{u,i}[n,k] = \operatorname{diag}(p_1[n,k], \ldots, p_{\ell_{\max}}[n,k])$ being a diagonal power loading matrix, the effective channel is decomposed into $\ell_{\max}$ parallel non-interfering AWGN SISO channels, denoted as spatial modes. The input-output relationship simplifies to

$$[\mathbf{y}_{u,i}[n,k]]_\nu = \sqrt{p_\nu[n,k]}\sigma_{u,i}^{(\nu)}[n,k]\,[\mathbf{x}_{u,i}[n,k]]_\nu + \tilde{z}_{u,i}[n,k], \quad \nu \in \{1, \ldots, \ell_{\max}\}, \tag{3.5}$$

with $\tilde{z}_{u,i}[n,k] \sim \mathcal{N}_{\mathbb{C}}(0, \tilde{\sigma}_z^2)$ because the receive filter is semi-unitary. To achieve the highest transmission rate, the diagonal elements of $\mathbf{P}_{u,i}[n,k]$ have to be set according to the water-filling power allocation policy over the squared singular values normalized by the effective noise variance. The water level is equal to the transmit power P_i [106]. Assuming equal and spatially uncorrelated out-of-cell interference on all receive antennas, the effective noise variance is calculated as

$$\tilde{\sigma}_z^2 = \sigma_z^2 + \frac{1}{M_{u,i}} \sum_{j=0, j\neq i}^{I} \left\| (\mathbf{H}_{u,i}^{(j)}[n,k])^{\mathrm{H}} \mathbf{F}_j[n,k] \right\|^2. \tag{3.6}$$

Notice that knowledge of the out-of-cell channel matrices and the precoders applied in the other cells is not required in practice to estimate the effective noise variance [130]. Due to the applied water level, the total power constraint

$$\|\mathbf{F}_{u,i}[n,k]\|^2 = P_i \tag{3.7}$$

is satisfied by the precoder. In this derivation independent per RE power constraints are assumed. If power loading over REs is considered, the water-filling power allocation is calculated over spatial streams and REs; see Appendix D.

Codebook based precoding with linear equalization: In the previous paragraph it is shown that the capacity achieving transmit strategy of the point-to-point AWGN MIMO channel involves linear precoding and reception employing semi-unitary matrices. To calculate the precoder at the base station, knowledge of the left singular matrix together with the SNR obtained on each spatial mode is required. This suggests that with limited feedback these values should be quantized and fed back by the users [131].

Nonetheless, with codebook based precoding a different approach is taken, which is shown to outperform direct quantization of the channel matrix in [40]. Instead of quantizing the channel, the optimal precoder is determined by the user from a given precoder quantization codebook. In [40], criteria for the selection of the optimal precoder and corresponding quantization codebook constructions are proposed for single carrier systems. It is shown that the optimal quantization codebooks for the transmission of ℓ spatial streams are maximally spaced subspace packings in the Grassmann manifold of ℓ-dimensional subspaces in the N_i-dimensional Euclidean space $\mathcal{G}\left(N_i, \ell\right)$; see Appendix C for a short introduction of the Grassmann manifold. Depending on the considered performance criterion, different distance metrics are employed for the construction of the Grassmannian codebooks, e.g., the projection two-norm, the Fubini-Study distance or the chordal distance [40]. The corresponding codebooks can be represented by sets of semi-unitary matrices, i.e., orthonormal bases, spanning the ℓ-dimensional subspaces. Notice that this approach can also be interpreted as a high SNR approximation of direct channel quantization. At high SNR it is known that the water-filling power allocation converges to equal power allocation over all spatial modes [132]. Therefore, quantization of the left singular matrix only is sufficient, which can be efficiently achieved using a Grassmannian codebook. Alternative codebook designs are based on vector quantization [133], discrete Fourier-transform matrices [134], QAM [135] and other concepts [136].

Motivated by the gains promised by these investigations, codebook based precoding has been implemented in commercial cellular technology, e.g., in LTE's closed loop spatial multiplexing (CLSM) transmission mode [137]. Hence, the precoder codebook employed for SU-MIMO transmission is assumed as given in this dissertation. Specifically, the precoder codebook for the transmission of ℓ streams over N_i transmit antennas, consisting of semi-unitary precoders, is denoted

$$\mathcal{Q}_{\ell}^{(N_i)} \subset \left\{ \mathbf{Q} \in \mathbb{C}^{N_i \times \ell} \middle| \mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{I}_{\ell} \right\}. \tag{3.8}$$

Such precoder codebooks are defined for all possible numbers of data streams $\ell \in \{1, \ldots, N_i\}$. Assuming that precoder $\mathbf{Q}_{u,i}[n,k] \in \mathcal{Q}_{\ell_{u,i}[k]}^{(N_i)}$ is selected as the preferred precoder of user u on RE $[n,k]$, the precoder applied for transmission in cell i is

$$\mathbf{F}_{u,i}[n,k] = \sqrt{\frac{P_i}{\ell_{u,i}[k]}}\mathbf{Q}_{u,i}[n,k], \tag{3.9}$$

accounting for the power constraint P_i of the base station. The power is equally distributed over all $\ell_{u,i}[k]$ spatial streams. Power allocation among spatial streams is restricted to an on-off switching of

spatial modes, by selecting the preferred transmission rank $\ell_{u,i}[k]$. Hence, the continuous trade-off between MIMO beamforming and spatial multiplexing, which is achieved with the water-filling power allocation employed with SVD based precoding, is coarsely approximated in this case. The capacity loss of SU-MIMO with limited feedback is known to decrease exponentially in the number of feedback bits b [131]. The number of feedback bits b is related to the codebook size as

$$b = \log_2\left(Q_\ell^{(N_i)}\right), \quad Q_\ell^{(N_i)} = \left|\mathcal{Q}_\ell^{(N_i)}\right|. \tag{3.10}$$

To separate the spatial data streams at the receiver, a linear equalizer filter $\mathbf{G}_{u,i}[n,k]$ is employed in this dissertation. Specifically, in the presented simulations zero forcing (ZF) and minimum mean-squared error (MMSE) equalization are considered, with the corresponding receive filtering matrices

$$\mathbf{G}_{u,i}^{(\mathrm{ZF})}[n,k] = \mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k]\Big(\mathbf{F}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k]\Big)^{-1}, \tag{3.11}$$

$$\mathbf{G}_{u,i}^{(\mathrm{MMSE})}[n,k] = \mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k]\Big(\mathbf{F}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k] + \tilde{\sigma}_z^2\mathbf{I}_{\ell_{u,i}[k]}\Big)^{-1}. \tag{3.12}$$

In the presented simulations, the precoder codebooks proposed in the LTE specification [10] are employed. Instead of applying maximally spaced Grassmannian subspace packings, these codebooks are designed having computational and implementation complexity in mind, e.g., by minimizing the amount of complex multiplications involved in precoding or by changing only the phase of the transmit signal and not its amplitude, reducing the requirements posed on amplifier linearity [138]. In case of two transmit antennas, only seven possible precoders are defined by the LTE standard. With four and eight transmit antennas, 64 and 621 precoders are supported, respectively [10]. Other transmit antenna configurations are currently not considered by the standard.

3.2. Implicit CSI Feedback Algorithms

The aim of the CSI feedback selection algorithms proposed in this section is to maximize the instantaneous user throughput, given the CSI available at the receiver as obtained from channel estimation using, e.g., training symbols [39]. Naturally, instantaneous CSI feedback is reasonable only if the current channel estimate at the receiver, employed for calculating the feedback, is representative for the time when the feedback is utilized to adapt the data transmission. Hence, the delay experienced in the feedback path must be sufficiently small compared to the coherence time of the channel to ensure similar channel conditions; see Section 3.3 for a more detailed investigation on the impact of a feedback delay on the proposed algorithms. If this cannot be guaranteed, statistical (or long-term) CSI feedback should be employed instead, as proposed, e.g., in [139].

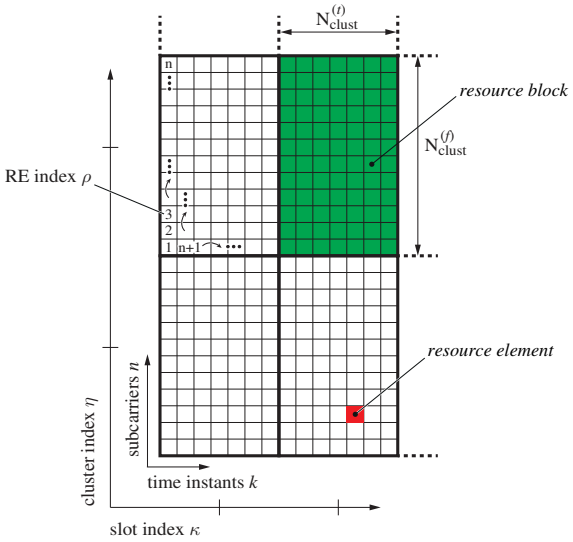


Figure 3.3.: OFDM time-frequency resource grid and clustering of REs into RBs.

3.2.1. Feedback Clustering

Before going into the details of the proposed feedback selection algorithms, some useful notation is introduced in this section to capture the concept of feedback clustering. As mentioned above, feedback clustering is a technique for reducing the CSI feedback overhead, by exploiting the correlation of the wireless channel in the time-frequency domain. With feedback clustering, the optimal CSI feedback is determined not for each RE individually, but for a set of consecutive time instants and subcarriers. The intuition behind this approach is that the best precoder stays constant over a specific time-frequency interval, due to the restricted size of the precoder codebook $\mathcal{Q}_\ell^{(N_i)}$ and the correlation of the channel. Therefore, the size of this interval is determined by the coherence time and bandwidth of the channel, but also by the size of the codebook; see [140] for an overview of different clustering approaches for precoder feedback. Similarly, due to the restricted set of supported MCSs, the achievable transmission rate does not change arbitrarily fast. But even if this assumption is not fulfilled and significant variations of the channel occur over the cluster size, optimal feedback indicators can still be determined using appropriate averaging as explained in Section 3.2.2, though for the price of a reduced throughput as investigated in Section 3.3.3.

The idea of clustering is illustrated in Figure 3.3. This figure shows the time-frequency resource grid spanned by the OFDM subcarriers and symbols. The pair $[n, k]$ of subcarrier index and symbol-time index is denoted as RE, as already introduced before. With clustering, a set of $N_{\text{clust}}^{(f)}$ subcarriers and $N_{\text{clust}}^{(t)}$ time instants is combined to a so called RB.[2] The number of REs per cluster

[2]Notice the slight abuse of LTE notation; the RBs considered in this thesis can have variable size, while those of LTE consist of exactly 12 subcarriers and 7 OFDM symbols.

is $N_{\text{RE}} = N_{\text{clust}}^{(f)} N_{\text{clust}}^{(t)}$. The time axis is divided into slots by the time domain clustering. The feedback is calculated for each slot κ individually. Similarly, the frequency axis is divided into clusters by the frequency domain clustering. A specific RB is indexed with the pair $[\eta, \kappa]$ of cluster index and slot index. The number of RBs per time slot is $N_{\text{RB}} = \frac{N_{\text{tot}}}{N_{\text{clust}}^{(f)}}$, with N_{tot} denoting the total number of subcarriers. To index the REs of a specific RB $[\eta, \kappa]$, the single RB specific RE index $\rho \in \{1, \ldots, N_{\text{RE}}\}$ is employed, which relates to the index pair $[n, k]$ as

$$k = (\kappa - 1) N_{\text{clust}}^{(t)} + \left\lceil \frac{\rho}{N_{\text{clust}}^{(f)}} \right\rceil, \; n = (\eta - 1) N_{\text{clust}}^{(f)} + \rho - (k - 1) N_{\text{clust}}^{(f)}, \tag{3.13}$$

$$\eta = \left\lceil \frac{n}{N_{\text{clust}}^{(f)}} \right\rceil, \; \kappa = \left\lceil \frac{k}{N_{\text{clust}}^{(t)}} \right\rceil, \tag{3.14}$$

$$\rho = \left(n - N_{\text{clust}}^{(f)} (\eta - 1) \right) + \left(k - N_{\text{clust}}^{(t)} (\kappa - 1) - 1 \right) N_{\text{clust}}^{(f)}. \tag{3.15}$$

This is equivalent to indexing the REs within an RB along subcarriers one time instant after the other, as illustrated in Figure 3.3.

With clustering, the same precoder is applied for all REs within an RB. Hence, the RB index $[\eta, \kappa]$ is employed in that case to index the precoders instead of the RE index $[n, k]$, i.e, $\mathbf{Q}_{u,i}[n, k]$ is replaced with $\mathbf{Q}_{u,i}[\eta, \kappa]$. Similarly, the transmission rank is indexed with the time slot index $[\kappa]$ instead of the OFDM symbol index $[k]$, i.e., $\ell_{u,i}[\kappa]$.

3.2.2. Feedback Selection Algorithm

With the support of AMC and codebook based precoding with a variable transmission rank, the assumed transceiver architecture is able to adapt the parameters of the SU-MIMO transmission to the current channel conditions such as to maximize the MIMO gain and to ensure a reliable data transmission. This is achieved with the provision of CSIT for each time slot κ using the following three feedback indicators:

- The *preferred MCSs* $\mathbf{m}_{u,i}[\eta, \kappa] \in \mathcal{M}^{C_\ell}, \; \forall \eta \in \{1, \ldots, N_{\text{RB}}\}$ are signaled with the *channel quality indicators (CQIs)*. The set of supported MCSs of the considered technology is denoted $\mathcal{M}$. As the MCS determines the spectral efficiency of the data transmission, the CQI is equivalent to the achievable transmission rate over RB $[\eta, \kappa]$. RB and codeword specific CQI feedback is considered to enable exploitation of the multi-user diversity during scheduling.
- The *preferred transmission rank* $\ell_{u,i}[\kappa]$ is signaled with the *rank indicator (RI)*. With rank adaptive transmission, a trade-off between the SNR gain provided by beamforming and the spatial multiplexing gain of the MIMO system is enabled, adjusting the transmission to the current SINR experienced by a user.

- The *precoding matrix indicator (PMI)* is employed for MIMO precoding, selecting the *favored precoders* $\mathbf{Q}_{u,i}[\eta,\kappa],\ \forall\eta\in\{1,\ldots,N_{\mathrm{RB}}\}$ from the quantization codebooks $\mathcal{Q}_\ell^{(N_i)},\ \ell\in\{1,\ldots,\ell_{\max}\}$. The maximum number of spatial streams is $\ell_{\max}=\min(N_i,M_{u,i})$.

The selection of the preferred feedback indicators is based on maximizing the achievable throughput. Due to the finite block-length of the codewords $c_{u,i}[\kappa]\in\{1,\ldots,C_{\ell_{u,i}[\kappa]}\}$ and other imperfections of the channel code, a vanishing BLER is in general not achieved. The system is rather designed to operate below a target BLER $P_b^{(t)}$ that is commonly determined by the application driving the data transmission. This target BLER is considered in the proposed feedback selection algorithm as an implicit constraint of the optimization problem.

In the following, feedback clustering with a cluster size of $N_{\mathrm{clust}}^{(f)}$ in the frequency domain and $N_{\mathrm{clust}}^{(t)}$ in the time domain is assumed. The same cluster size is applied for CQI and PMI feedback to simplify the exposition, although different cluster sizes can be accommodated with an extended notation; see [64]. The RI is a wide band feedback value that is valid for all REs of a time slot κ. The CSI feedback is determined for each slot κ individually; hence, κ is assumed as fixed in the following.

Assuming that precoder $\mathbf{Q}[\eta]\in\mathcal{Q}_\ell^{(N_i)}$ is employed during transmission on RB $[\eta,\kappa]$, the post-equalization SINR of data stream $\nu\in\{1,\ldots,\ell\}$ on RE $[n,k]$ equals

$$\beta_{\nu,u,i}[n,k]=\frac{\frac{P_i}{\ell}\left|\mathbf{g}_{\nu,u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{q}_\nu[\eta]\right|^2}{\frac{P_i}{\ell}\sum_{\mu=1,\mu\neq\nu}^{\ell}\left|\mathbf{g}_{\nu,u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{q}_\mu[\eta]\right|^2+\tilde{\sigma}_z^2\left\|\mathbf{g}_{\nu,u,i}[n,k]\right\|^2},\tag{3.16}$$

with $\mathbf{g}_{\nu,u,i}[n,k]=[\mathbf{G}_{u,i}[n,k]]_{:,\nu}$, $\mathbf{q}_\nu[\eta]=[\mathbf{Q}[\eta]]_{:,\nu}$ and $\tilde{\sigma}_z^2$ from Equation (3.6). Due to SU-MIMO transmission, the in-cell interference term of the more general instantaneous SINR defined in (2.7) does not appear in Equation (3.16). Also, the semi-unitary precoder $\mathbf{Q}$ is employed in (3.16) instead of its scaled version $\mathbf{F}$ from (3.9) as in Equation (2.7), to express the dependency of the SINR on the precoder codebook more clearly.

In order to determine the achievable transmission rate over an RB comprising N_{RE} REs, mutual information effective SINR mapping (MIESM) is employed in this work to average the corresponding post-equalization SINRs [141]. The idea of MIESM is to identify the time-frequency selective channel experienced over the N_{RE} REs, with an AWGN channel that achieves the same average spectral efficiency in terms of the BICM capacity. Assuming transmission with MCS $m[\eta]\in\mathcal{M}$ on RB $[\eta,\kappa]$, this is achieved with the following averaging function

$$\beta_{\nu,u,i}^{(m[\eta])}[\eta,\kappa]=B_{m[\eta]}^{-1}\left(\frac{1}{N_{\mathrm{RE}}}\sum_{\rho=1}^{N_{\mathrm{RE}}}B_{m[\eta]}\left(\beta_{\nu,u,i}[\rho]\right)\right),\tag{3.17}$$

where the absolute RE index $[n, k]$ has been replaced with the implicit RB specific RE index $[\rho]$ using Equation (3.15). In Equation (3.17), the function $B_{m[\eta]}(\beta)$ is defined as

$$B_{m[\eta]}(\beta) = B_{A_{m[\eta]}}\left(\frac{\beta}{\phi_{m[\eta]}}\right), \tag{3.18}$$

with $B_{A_{m[\eta]}}(\beta)$ denoting the BICM capacity of the modulation alphabet $A_{m[\eta]} \in \mathcal{A}$ that is associated with the MCS $m[\eta]$. The scalar $\phi_{m[\eta]}$ is employed for calibration purposes, to adapt the MIESM averaging to the performance of the different MCSs; see Appendix E for more details on MIESM and the calibration of the method. The inverse of the bijective function $B_{m[\eta]}(\beta)$ is denoted $B^{-1}_{m[\eta]}(\cdot)$.

The RB specific AWGN equivalent effective SNR $\beta^{(m[\eta])}_{\nu,u,i}[\eta, \kappa]$ is dependent on the spatial stream index ν. As mentioned in Section 3.1.1, in LTE the data of multiple spatial streams can be jointly coded, that is, a codeword $c \in \{1, \ldots, C_\ell\}$ is mapped onto several streams $\nu \in \{1, \ldots, \ell\}$. If this is the case, it is sufficient to provide CQI feedback for each codeword only, instead of for each stream. To accommodate this case, MIESM averaging is applied to determine the average SNR not only over REs, but also over spatial streams. Denoting the set of spatial streams associated with codeword c as $\mathcal{L}_c \subseteq \{1, \ldots, \ell\}$, the codeword dependent effective SNR is defined as

$$\beta^{(m[\eta])}_{c,u,i}[\eta, \kappa] = B^{-1}_{m[\eta]}\left(\frac{1}{N_{\text{RE}}\,|\mathcal{L}_c|}\sum_{\nu\in\mathcal{L}_c}\sum_{\rho=1}^{N_{\text{RE}}} B_{m[\eta]}\left(\beta_{\nu,u,i}[\rho]\right)\right). \tag{3.19}$$

The AWGN equivalent SNR $\beta^{(m[\eta])}_{c,u,i}[\eta, \kappa]$ is employed to estimate the BLER achieved with MCS $m[\eta]$ on RB $[\eta, \kappa]$, using precomputed AWGN look-up tables that quantify the performance of the considered technology; see Figure E.1b in Appendix E for the AWGN BLERs attained with the MCSs of LTE, which are employed as look-up tables in the simulations presented throughout this thesis. Denoting the relationship between the SNR β of an AWGN channel and the corresponding BLER of MCS $m[\eta]$ as $g_{m[\eta]}(\beta)$, the estimated BLER of codeword c over RB $[\eta, \kappa]$ equals

$$P^{(m[\eta])}_{c,u,i}[\eta, \kappa] = g_{m[\eta]}\left(\beta^{(m[\eta])}_{c,u,i}[\eta, \kappa]\right). \tag{3.20}$$

To account for the target BLER $P^{(t)}_b$, a function $h_{m[\eta]}(P, P^{(t)}_b)$ is defined that outputs the spectral efficiency $e_{m[\eta]}$ of MCS $m[\eta]$ if the BLER P is less than the target BLER and zero otherwise

$$h_{m[\eta]}(P, P^{(t)}_b) = \begin{cases} e_{m[\eta]}, & P \le P^{(t)}_b \\ 0, & P > P^{(t)}_b \end{cases}. \tag{3.21}$$

With this function, the spectral efficiency of codeword c using MCS $m[\eta]$ for the transmission over RB $[\eta, \kappa]$ is estimated as

$$e^{(m[\eta])}_{c,u,i}[\eta, \kappa] = \left(1 - P^{(m[\eta])}_{c,u,i}[\eta, \kappa]\right) h_{m[\eta]}\left(P^{(m[\eta])}_{c,u,i}[\eta, \kappa], P^{(t)}_b\right). \tag{3.22}$$

In this equation, the success probability of the data transmission is taken into account in the first term, and the achieved spectral efficiency in case of successful transmission is quantified with the second term. Hence, $e_{c,u,i}^{(m[\eta])}[\eta,\kappa]$ corresponds to the expected value of the spectral efficiency. Notice though that zero spectral efficiency is output in case that the BLER target is not satisfied, preventing these MCSs to be selected by the optimization problem defined below. The target BLER is hence implicitly considered as a hard constraint of the optimization problem.

The optimal feedback values at time slot κ are jointly determined from the following optimization problem maximizing the achievable throughput, i.e., the sum spectral efficiency over all RBs and codewords

$$\{\ell_{u,i}[\kappa], \mathbf{Q}_{u,i}[\eta,\kappa], \mathbf{m}_{u,i}[\eta,\kappa]\} = \underset{\ell,\mathbf{Q}[\eta],\mathbf{m}[\eta]}{\operatorname{argmax}} \sum_{\eta=1}^{N_{\mathrm{RB}}} \sum_{c=1}^{C_\ell} e_{c,u,i}^{(m_c[\eta])}[\eta,\kappa] \tag{3.23}$$

$$\text{subject to: } \ell \leq \min\left(N_i, M_{u,i}\right),$$

$$\mathbf{Q}[\eta] \in \mathcal{Q}_\ell^{(N_i)},\ \forall \eta \in \{1,\ldots,N_{\mathrm{RB}}\}$$

$$\mathbf{m}[\eta] = [m_1[\eta],\ldots,m_{C_\ell}[\eta]]^{\mathrm{T}} \in \mathcal{M}^{C_\ell},\ \forall \eta \in \{1,\ldots,N_{\mathrm{RB}}\}.$$

In general, a joint optimization over all N_{RB} RBs is required to solve this problem, because the individual summands of the sum over η are coupled over the transmission rank ℓ. Only if ℓ is fixed, the summands in (3.23) decouple and the optimal precoder and MCSs for each RB can be determined independently. Based on this observation, the most efficient way for solving the joint optimization problem (3.23) is

- Fix the transmission rank ℓ and solve the decoupled optimization problem for each η

$$\left\{\mathbf{Q}_{u,i}^{(\ell)}[\eta,\kappa], \mathbf{m}_{u,i}^{(\ell)}[\eta,\kappa], e_{u,i}^{(\ell)}[\eta,\kappa]\right\} = \underset{\mathbf{Q}[\eta],\mathbf{m}[\eta]}{\operatorname{argmax}} \sum_{c=1}^{C_\ell} e_{c,u,i}^{(m_c[\eta])}[\eta,\kappa] \tag{3.24}$$

$$\text{subject to: } \mathbf{Q}[\eta] \in \mathcal{Q}_\ell^{(N_i)},$$

$$\mathbf{m}[\eta] \in \mathcal{M}^{C_\ell}.$$

  Here, the optimized efficiency on RB $[\eta,\kappa]$ when transmitting ℓ streams is denoted $e_{u,i}^{(\ell)}[\eta,\kappa]$.
- Maximize the sum of the optimized efficiencies with respect to the transmission rank ℓ

$$\ell_{u,i}[\kappa] = \underset{\ell\in\{1,\ldots,\ell_{\max}\}}{\operatorname{argmax}} \sum_{\eta=1}^{N_{\mathrm{RB}}} e_{u,i}^{(\ell)}[\eta,\kappa], \tag{3.25}$$

$$\mathbf{Q}_{u,i}[\eta,\kappa] = \mathbf{Q}_{u,i}^{(\ell_{u,i}[\kappa])}[\eta,\kappa],$$

$$\mathbf{m}_{u,i}[\eta,\kappa] = \mathbf{m}_{u,i}^{(\ell_{u,i}[\kappa])}[\eta,\kappa].$$

A closed-form solution of this optimization problem is not possible, because the variables are confined to pre-specified codebooks and are hence discrete. As soon as the optimization problem is solved, the feedback indicators are obtained as the codebook indices corresponding to the optimal solutions, i.e., the PMIs are the indices of the precoders $\mathbf{Q}_{u,i}[\eta,\kappa]$ in the codebook $\mathcal{Q}_{\ell_{u,i}[\kappa]}^{(N_i)}$, the RI is equal to the transmission rank $\ell_{u,i}[\kappa]$ and the CQIs are determined by the indices of $\mathbf{m}_{u,i}[\eta,\kappa]$ within $\mathcal{M}$.

3.2.3. Approximate Sequential Solution

The complexity involved in the exhaustive search required to obtain the optimal solution of Equation (3.23) may often be too high for practical implementations. For example, finding the optimum of only one RB in an LTE compliant system with eight transmit antennas and four receive antennas already requires a search over almost $10\,000$ options, which may not be feasible within the strict delay requirements of instantaneous CSI feedback.

To reduce the complexity of the exhaustive search, an approximate sequential solution is proposed, selecting the precoders independently of the MCSs by employing a coarser estimation of the achievable throughput. This estimation is obtained from the BICM capacity. To make the calculation independent of the modulation alphabet $A \in \mathcal{A}$ and hence of the MCS, the *BICM system capacity* is defined as the envelope of the modulation alphabet dependent capacities

$$B\left(\beta\right) = \max_{A\in\mathcal{A}} B_A\left(\beta\right). \tag{3.26}$$

The RB specific estimated achievable throughput using precoder $\mathbf{Q}[\eta] \in \mathcal{Q}_\ell^{(N_i)}$ is obtained as

$$B_{u,i}^{(\mathbf{Q}[\eta])}[\eta,\kappa] = \sum_{\nu=1}^{\ell}\sum_{\rho=1}^{N_{\mathrm{RE}}} B\left(\beta_{\nu,u,i}[\rho]\right), \tag{3.27}$$

with $\beta_{\nu,u,i}[\rho]$ from Equation (3.16). Maximizing this value with respect to the precoders, the transmission rank dependent optimal precoders are obtained

$$\left\{\mathbf{Q}_{u,i}^{(\ell)}[\eta,\kappa], B_{u,i}^{(\ell)}[\eta,\kappa]\right\} = \underset{\mathbf{Q}[\eta]\in\mathcal{Q}_\ell^{(N_i)}}{\operatorname{argmax}} B_{u,i}^{(\mathbf{Q}[\eta])}[\eta,\kappa], \tag{3.28}$$

with $B_{u,i}^{(\ell)}[\eta,\kappa]$ denoting the transmission rank dependent optimized achievable throughput. The solutions for the precoders can then be employed as a-priori knowledge in Equation (3.23) to obtain the corresponding transmission rank and MCSs.

If this is still too complex, the transmission rank ℓ can also be selected independently of the MCSs

by maximizing the sum of $B_{u,i}^{(\ell)}[\eta,\kappa]$ over the RBs

$$\ell_{u,i}[\kappa] = \underset{\ell\in\{1,\ldots,\ell_{\max}\}}{\operatorname{argmax}} \sum_{\eta=1}^{N_{\mathrm{RB}}} B_{u,i}^{(\ell)}[\eta,\kappa], \tag{3.29}$$

$$\mathbf{Q}_{u,i}[\eta,\kappa] = \mathbf{Q}_{u,i}^{(\ell_{u,i}[\kappa])}[\eta,\kappa].$$

Utilizing these solutions as a-priori knowledge in Equation (3.23) the corresponding MCSs are obtained. The optimization of the MCSs cannot be further simplified to ensure that the upper bound on the BLER specified by $P_b^{(t)}$ is satisfied.

The most expensive step involved in solving these optimization problems is the calculation of the post-equalization SINR in (3.16) for every precoder, because it involves matrix inversions to determine the equalizer $\mathbf{G}_{u,i}[n,k]$. This can be avoided by employing the *pre-equalization mutual information* achieved with Gaussian signaling to estimate the achievable throughput according to [142]

$$I_{u,i}[n,k] = \log_2 \det\left(\mathbf{I}_{M_i} + \frac{P_i}{\tilde{\sigma}_z^2 \ell}\,\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{Q}[\eta]\mathbf{Q}[\eta]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]\right). \tag{3.30}$$

Replacing $\sum_{\nu=1}^{\ell} B\left(\beta_{\nu,u,i}[\rho]\right)$ in Equation (3.27) with $I_{u,i}[n,k]$, the same optimization steps as in Equations (3.28) and (3.29) can again be performed to calculate suboptimal solutions for the precoders and the transmission rank. Then, the post-equalization SINR hast to be calculated only once, for the optimization of the MCSs. The precoder selection based on the mutual information is also proposed in [40]. The performance of the different methods is compared in Section 3.3.2.

3.2.4. Antenna Subset Selection for DASs

The available transmit power P_i of cell i is distributed uniformly over all N_i transmit antennas by the precoders of the considered semi-unitary codebook (3.8). In case of distributed transmit antennas, when pathloss differences between the remote radio units (RRUs) become significant, it is beneficial to concentrate the transmit power on those antenna arrays that experience good channel quality, thus employing only a subset of the antennas for transmission. To transmit the data from $\tilde{N}_i \leq N_i$ antennas, the reduced size codebook $\mathcal{Q}_\ell^{(\tilde{N}_i)}$ is employed to distribute the signal onto the $\tilde{N}_i$ activated antennas, while the other antenna elements are deactivated.

The preferred antenna subset can be determined by solving the optimization problem (3.23) for all possible antenna subset choices, amounting in

$$\sum_{\tilde{N}_i=1}^{N_i} \binom{N_i}{\tilde{N}_i} = 2^{N_i} - 1 \tag{3.31}$$

possibilities and thus in an additional feedback overhead of roughly N_i bits.

To enable efficient transmit antenna subset selection, with minimal extra CSI feedback overhead, it is assumed that the in-cell channel gain matrix $\mathbf{C}_{u,i}$ from (2.9) is known by both, the base station and the user. This is a reasonable, often made assumption in DASs, because the large-scale statistics change only very slowly over time and frequency, and can thus be learned either from the uplink [143], provided the duplex distance is not too large, or via a very low rate feedback link (e.g., during connection setup). The CSI feedback overhead can then be reduced by considering only the antenna subsets with the largest channel gains as possible candidates for antenna subset selection. Thus, e.g., if $\tilde{N}_i = 4$, the optimization problem (3.23) is only solved for the four strongest distributed antennas, reducing the additional CSI feedback overhead to $\log_2(N_i)$ bits. The performance-investigation of this method is postponed until Chapter 5, when DASs are evaluated.

As an alternative to antenna subset selection, power loading over transmit antennas can be employed [144], which requires accurate knowledge of the instantaneous channel gain at the transmitter, thus increasing the CSI feedback overhead significantly.

3.3. Performance Investigation

In this section, the performance of the proposed CSI feedback selection algorithms is investigated by means of Monte-Carlo simulations. Due to the involved non-linear functions (MIESM, BICM capacity) and the discrete nature of the proposed optimization problem, an analytic performance investigation was not successful. Instead, the theoretical throughput bounds developed in Appendix D are applied in Section 3.3.1 to evaluate the performance of LTE using the proposed CSI feedback algorithms, and to identify the dominant sources of the observed throughput loss with respect to the Shannon capacity. In Section 3.3.2, the performance of the individual feedback indicators, i.e., the RI, CQI and PMI, is scrutinized and the impact of a delay in the feedback path is evaluated. Finally, in Section 3.3.3, the impact of the CSI feedback granularity applied with feedback clustering on the achieved throughput is examined.

The simulations presented in this dissertation are mostly obtained with the Vienna LTE-A link level simulator [49, 59]. This open source *MATLAB* based LTE compliant link level simulator is publicly available for download, pursuing the goal of facilitating reproducibility in wireless communications research. The link level simulator is augmented with a system level simulator enabling the simulation of large cellular networks with reasonable complexity. Single cell simulations are considered in this section; multiple cells causing out-of-cell interference are treated in Chapter 5. The results are presented in dependence of the equivalent average transmit SNR defined as

$$\text{SNR} = \frac{P_i}{\sigma_z^2} = \frac{E_i}{N_0}, \tag{3.32}$$

with E_i denoting the symbol energy and N_0 being the noise power spectral density.

Table 3.1.: Simulation parameters of the SU-MIMO performance investigation.

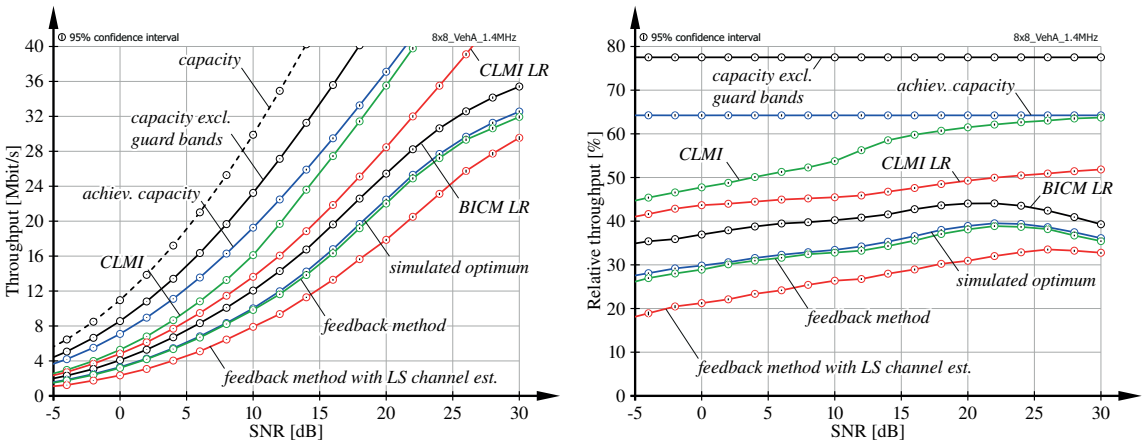
Parameter	Value
Carrier bandwidth	$\{1.4, 10\}$ MHz
Carrier frequency	$f_c = 2$ GHz
Signal to noise ratio	SNR $\in [-10, 40]$ dB
Number of subcarriers	$N_{\text{tot}} \in \{72, 600\}$
Channel models	power delay profile based:
	vehicular A, pedestrian B [108]
	typical urban [109]
Number of receive antennas	$M_{u,i} \in \{4, 8\}$
Number of transmit antennas	$N_i \in \{4, 8\}$
Number of users	$U_i \in \{1, 2, 4, \ldots, 64\}$
Spatial correlation parameter	$\alpha_{\text{corr}} \in \{0, 0.9\}$
Maximum channel Doppler frequency	$f_d \in [10, 500]$ Hz
CSI feedback delay	$\tau_d \in [0, 8]$ TTIs
Multi-user scheduling	proportional fair
MIMO receiver	ZF

Centralized antenna systems, without RRUs, are studied in the simulations presented in this section; see Chapter 5 for results on DASs. Pathloss and shadow fading are not considered, the instantaneous SNR observed by a user is rather determined by the microscopic channel fading realization and the variance of the receiver noise. In case of multi-user simulations, the average SNR of all users is equal. A block fading channel model is assumed with a temporally constant channel for the duration of one LTE transmission time interval (TTI), denoted as subframe (1 ms or 14 OFDM symbols). The channel realizations of consecutive subframes are either statistically independent or correlated according to Clarke's model; see Section 2.4 for details. CSI feedback is provided for each TTI. In most simulations a feedback delay of zero is assumed, meaning that the CSIT is acausally available before the transmission. The impact of a feedback delay is separately investigated. Proportional fair multi-user scheduling is applied [56] utilizing the LTE optimized scheduling framework of [145]. Important simulation parameters are summarized in Table 3.1.[3]

3.3.1. Comparison of LTE to Theoretical Throughput Bounds

In this section, a single user scenario employing the largest LTE compliant antenna configuration of $N_i = 8$ transmit antennas and $M_{u,i} = 8$ receive antennas is assumed, transmitting over a bandwidth of 1.4 MHz (72 subcarriers). The spatial correlation parameter of the receive antenna array is optimistically assumed as $\alpha_{\text{corr}} = 0$. The performance of LTE's CLSM transmission mode, utilizing PMI, RI and CQI feedback, is compared to the throughput bounds derived in Appendix D. CSI feedback clustering is applied with $N_{\text{clust}}^{(t)} = 14$ and $N_{\text{clust}}^{(f)} = 12$. The performance with perfect

[3] Parameter sets are specified by curly brackets $\{\cdot\}$; intervals are defined by square brackets $[\cdot]$.



(a) Absolute throughput.

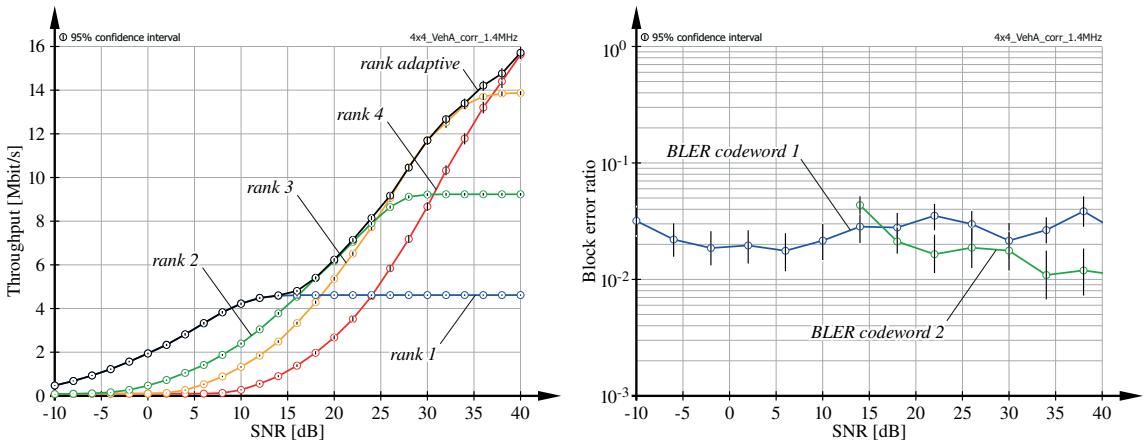
(b) Relative throughput.

Figure 3.4.: Comparison of the achieved throughput to channel capacity and the proposed throughput bounds with $N_i \times M_{u,i} = 8 \times 8$.

and estimated channel knowledge at the receiver is simulated, employing a least-squares channel estimator [39]. The additional noise caused by the channel estimator has to be considered in the post-equalization SINR $\beta_{\nu,u,i}[n,k]$ used for the calculation of the feedback indicators to ensure an accurate CQI estimation [62]. This is achieved by employing the first-order Taylor approximation proposed in [146] together with the analytic expression of the channel estimation error variance of the least-squares channel estimator derived in [147].

The results of the investigation are shown in Figure 3.4 in terms of the absolute throughput and the relative throughput with respect to channel capacity. As detailed in Appendix D, the proposed throughput upper bounds shown in the figure account for the throughput loss due to the system overhead for guard bands and reference symbols (achievable capacity), the restriction to codebook based precoding (closed loop mutual information (CLMI)), the application of a linear receiver (CLMI with linear receiver (CLMI LR)) and the BICM architecture (BICM with linear receiver (BICM LR)). It can be seen in Figure 3.4b that each of these factors entails a throughput loss in the order of $10 - 20\,\%$ of channel capacity, amounting in an achievable throughput of approximately $35 - 45\,\%$ of channel capacity (BICM LR). The simulated optimum performance of LTE with perfect channel knowledge at the receiver, obtained by exhaustive search, is found another $3 - 7\,\%$ below the BICM LR bound. With the proposed feedback selection algorithms close to optimal performance is attained (feedback method). The loss of approximately $1\,\%$ can be explained by the non-zero BLER achieved with the proposed CQI selection. With imperfect channel knowledge at the receiver a further throughput reduction to $20 - 30\,\%$ of channel capacity is experienced (feedback method with LS channel est.).

Notice that the performance of LTE is pessimistically represented in these results due to some assumptions of the simulation. With all other LTE compliant carrier bandwidths ($3/5/10/20\,$MHz),



(a) Rank adaptation versus fixed rank transmission.

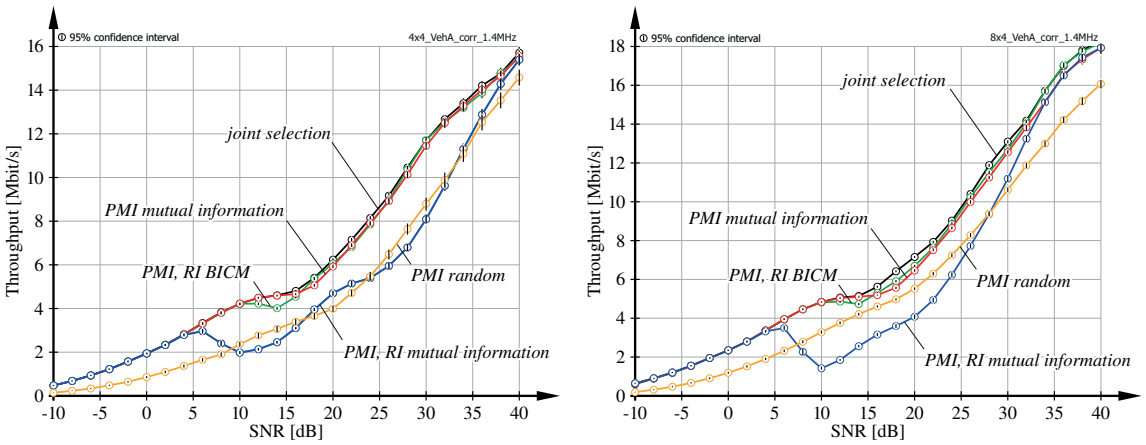
(b) BLER achieved with the proposed feedback algorithms.

Figure 3.5.: Comparison of the proposed rank adaptive scheme to fixed rank transmission and simulated BLER achieved with the proposed feedback algorithms.

the guard band overhead is reduced from 23 % to 10 %. The bandwidth of 1.4 MHz was chosen for complexity reasons to be able to determine the optimal performance of LTE with an exhaustive search. The channel estimator performance can be improved, e.g., by means of linear MMSE channel estimation providing close to perfect performance [39]. Also, significantly improved MIMO detection is possible, e.g., with successive interference cancellation [106]. Still, it is questionable whether such highly complex algorithms are feasible in practice in the near future, particularly in mobile phones.

3.3.2. Evaluation of the Feedback Algorithms

The performance of the individual feedback indicators is evaluated in the following in more detail. In the first scenario, an $N_i \times M_{u,i} = 4 \times 4$ antenna configuration is investigated assuming strongly correlated receive antennas, i.e., $\alpha_{\text{corr}} = 0.9$ in Equation (2.17). Other simulation parameters are set as in the previous section considering perfect channel estimation at the receiver. The corresponding throughput versus SNR is shown in Figure 3.5a, comparing the performance of fixed rank transmission ($\ell_{u,i} \in \{1, 2, 3, 4\}$) to rank adaptation by means of the proposed joint feedback selection algorithm of Equation (3.23). These results show that the proposed feedback selection algorithm is able to determine the optimal transmission rank in dependency of the channel conditions, and automatically trades-off the MIMO beamforming gain for the MIMO multiplexing gain to achieve the best performance. The corresponding BLER of the two codewords is shown in Figure 3.5b demonstrating that the target BLER $P_b^{(t)} = 0.1$ is observed by the proposed CQI selection. Due to the limited set of only 15 MCSs supported in LTE the achieved BLER is markedly below the constraint $P_b^{(t)} = 0.1$. Notice that the second codeword is activated just in case $\ell_{u,i} \geq 2$; hence the BLER curve starts only at 14 dB SNR.



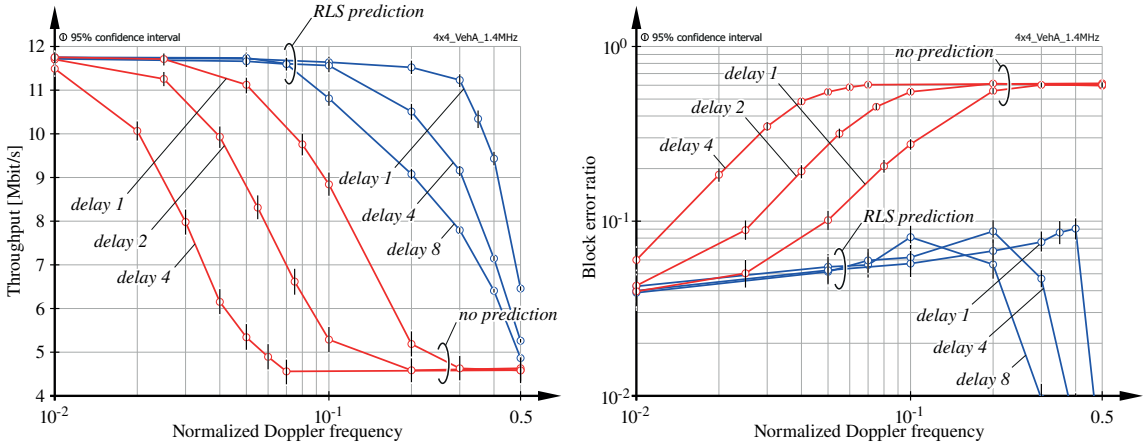
(a) $N_i \times M_{u,i} = 4 \times 4$. **(b)** $N_i \times M_{u,i} = 8 \times 4$.

Figure 3.6.: Performance of approximate sequential CSI feedback selection schemes compared to the joint optimization.

The same scenario is considered in the results presented in Figure 3.6a to investigate the performance of the approximate sequential solutions (3.28) and (3.29) of the joint CSI feedback selection problem (3.23). The figure shows that the best performance is achieved with the accurate rate estimation considered in the joint selection algorithm. If the PMI and RI are selected from the BICM capacity as in Section 3.2.3 almost equally good performance is achieved, with a complexity that is approximately a factor of ten below the joint selection in case of LTE. Similar throughput is attained if the pre-equalization mutual information of Equation (3.30) is employed instead of the BICM capacity for the PMI selection to avoid the calculation of the post-equalization SINR. Notice though that the RI selection based on the pre-equalization mutual information leads to a significant throughput reduction, because the performance of higher rank transmission is obviously overestimated. For comparison the achieved transmission rate with random precoder selection is also shown in Figure 3.6a. It can be seen that a gain of approximately $5 - 10\,\text{dB}$ in SNR is obtained with optimal precoder selection. Similar behavior is observed in Figure 3.6b for the case of eight transmit antennas. Notice that the LTE precoder codebook for eight antennas is four times larger than the four antenna codebook.

Next, the impact of a delay in the feedback path on the transmission rate obtained with the proposed instantaneous CSI feedback algorithms is investigated. Again, the $N_i \times M_{u,i} = 4 \times 4$ configuration is assumed, this time with uncorrelated receive antennas, i.e., $\alpha_{\text{corr}} = 0$. The effect of the feedback delay is determined by the strength of the temporal correlation of the wireless channel. Hence, the performance of the system is evaluated in terms of the maximum channel Doppler frequency f_d, defining the channel coherence time according to Equation (2.13). The results are plotted in dependency of the normalized channel Doppler frequency

$$\nu_d = f_d T_s, \tag{3.33}$$



(a) Throughput versus channel Doppler frequency.

(b) Block error ratio versus channel Doppler frequency.

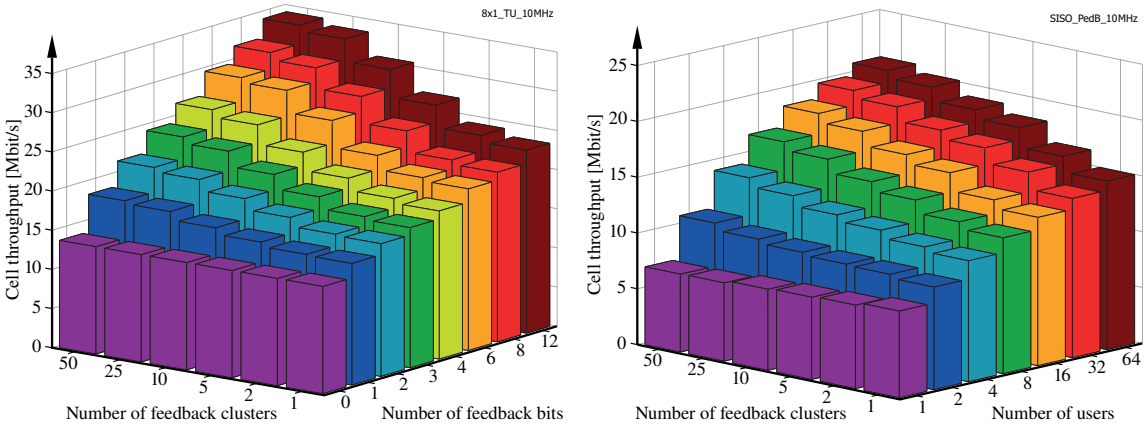
Figure 3.7.: Sensitivity of the proposed instantaneous CSI feedback algorithms with respect to a delay in the feedback path.

with $T_s = 1\,\text{ms}$ being the temporal sampling rate of the channel, i.e., the CSI feedback interval. The throughput degradation due to the feedback delay is shown in Figure 3.7a. With a delay of 1 TTI, i.e., $\tau_d = 1\,\text{ms}$, optimal performance is achieved up to a normalized Doppler frequency of $\nu_d^{(c)} = 0.025$, corresponding to a user speed of

$$v = \frac{\nu_d^{(c)}}{T_s\, f_c}\, c \approx 15\,\text{km/h}, \tag{3.34}$$

with $f_c = 2\,\text{GHz}$ denoting the carrier frequency and c being the speed of light. With $\tau_d = 4\,\text{ms}$ a further reduction to $5\,\text{km/h}$ is incurred. Hence, without any delay-compensation, the proposed algorithms are suitable for quasi-stationary situations. The reason for the performance degradation can be seen in Figure 3.7b. Due to the mismatch between the channel observed during CSI feedback calculation and the channel for which the outdated CSIT is employed for rate adaptation, the BLER is severely increased and the target BLER is not satisfied. A simple means to overcome this problem is the inclusion of an adaptive link margin to account for the channel mismatch [148].

Alternatively, channel prediction can be employed to compensate for the feedback delay. To demonstrate the capabilities of this approach, finite impulse response (FIR) filter based channel prediction is implemented. For simplicity, each column of the channel matrix $\mathbf{H}_{u,i}[n,k]$ is independently predicted. Also, each RB $[\eta,\kappa]$ is represented by the single channel matrix in the center of the RB for CSI feedback calculation, utilizing the channel sub sampling approach of [61]. The FIR filter coefficients are trained by means of the recursive least squares (RLS) algorithm [149]. The corresponding results, labeled with RLS prediction, are shown in Figure 3.7. The delay sensitivity is significantly reduced with channel prediction, enabling optimal performance up to user speeds of $100\,\text{km/h}$ and $40\,\text{km/h}$ at delays of 1 TTI and 8 TTIs, respectively. Interestingly, the throughput reduction at high Doppler frequencies is not caused by an increased BLER. It is rather a consequence



(a) Variation of PMI feedback granularity and codebook size.

(b) Variation of CQI feedback granularity and user number.

Figure 3.8.: Cell throughput with proportional fair scheduling versus CSI feedback granularity, codebook size and number of users.

of the reduction of the gain of the predicted channel matrix, due to diminishing magnitudes of the FIR filter coefficients with decreasing channel correlation. Hence, the SINR is automatically reduced with decreasing channel correlation, entailing a more conservative CQI estimation.

3.3.3. Impact of CSI Feedback Granularity

The SU-MIMO CSI feedback indicators investigated in this chapter can be exploited in a twofold way to improve the downlink data rate of cellular networks. Firstly, the single user throughput is increased through the optimized selection of the precoders, transmission rank and MCSs enabled by the proposed feedback selection algorithm of Equation (3.23). Secondly, the CQI feedback is useful in multi-user scheduling to achieve a multi-user diversity gain and hence to improve the sum data rate in a multi-user scenario [145, 150]. In both cases, the potential gain is larger with a smaller CSI feedback cluster size. The downside of reducing the cluster size is an increase in the CSI feedback overhead. This trade-off is investigated below by means of a simulation based study.

In the first scenario the improvement of the single user throughput with reducing PMI feedback cluster size and with increasing precoder codebook size is evaluated. To enable a variable precoder codebook size, which is not considered in LTE, a random codebook of independent and isotropically distributed semi-unitary matrices is generated [46]. Such random codebooks are known to be asymptotically optimal in the case of independent and identically distributed (i.i.d.) channel matrices; hence α_{corr} is set equal to zero. The results are averaged over random codebook realizations. An $N_i \times M_{u,i} = 8 \times 1$ antenna configuration is considered. The typical urban channel model [109] is employed, having a $50\,\%$ coherence bandwidth of $B_C = 400\,\text{kHz}$. A carrier bandwidth of $10\,\text{MHz}$ is assumed. The precoder codebook size is varied from one element to 2^{12} elements corresponding to a feedback

overhead of zero to 12 bits per cluster. The number of feedback clusters per TTI is varied from one to 50. The results of the simulation are shown in Figure 3.8a. A significant throughput improvement from 14 Mbit/s without precoder feedback to 34 Mbit/s with a feedback overhead of $12 \cdot 50$ bit/ms $= 600$ kbit/s is obtained. Such investigations are also useful to determine the optimal trade-off between cluster size and codebook size for a fixed amount of CSI feedback bits. For example, the best performance with 10 bits per TTI is obtained with a single cluster using a precoder codebook of size 2^{10}, amounting in 22 Mbit/s downlink throughput. On the contrary, with 10 clusters each using a codebook of size two only 17 Mbit/s are achieved. Naturally, these results depend on the channel coherence bandwidth.

In the second scenario the impact of the CQI cluster size on the achievable multi-user scheduling gain is investigated. A 10 MHz SISO system is assumed to eliminate the effect of the precoder feedback. The pedestrian B channel model [108] with a coherence bandwidth of $B_C = 267$ kHz is employed. The number of users U_i is varied from one to 64 and the number of CSI feedback clusters is increased from one to 50. The simulation results are plotted in Figure 3.8b, visualizing the improvement of the multi-user gain with an increasing number of feedback clusters. With one feedback cluster, only temporal multi-user diversity can be exploited, leading to a throughput improvement by a factor of two when increasing the number of users form one to 64. With 50 feedback clusters a threefold throughput gain is obtained because the frequency diversity can additionally be exploited during scheduling. With 64 users the cell throughput grows by 6 Mbit/s when increasing the number of clusters from one to 50. Notice though that this downlink improvement is outweighed by the increased CQI feedback overhead, scaling from $4 \cdot 64$ bit/ms $= 256$ kbit/s up to $4 \cdot 64 \cdot 50$ bit/ms $= 12.8$ Mbit/s, if 4 bits per CQI are employed as in LTE. Solutions for this problem of increasing feedback overhead with a growing number of users have been published in literature, e.g., K-significant scheduling [151] and thresholding, i.e., providing feedback only in case a threshold is exceeded [152].

3.4. Summary

In this chapter, limited feedback SU-MIMO transmission in wireless communications is investigated. Following the approach that is currently implemented in most commercial systems, the transmission parameters of the single user link are adapted by means of AMC and codebook based precoding to ensure reliable and spectrally efficient data transmission. This is achieved by the provision of CSIT using feedback indicators for the selection of the optimal transmission rate, transmission rank and precoders. The best performance is achieved with instantaneous CSIT enabling the exploitation of the space, time, frequency and multi-user diversity inherent in wireless communication systems.

By deriving an accurate estimate of the achievable user data rate, based on the post-equalization SINR, the optimization problem for the joint-selection of the optimal CSI feedback indicators is derived in this chapter. The performance of this method is investigated by means of simulations,

demonstrating that close to optimal throughput is achieved with the proposed algorithm. Several simplifying approximations are considered to reduce the complexity of the exhaustive search. A promising trade-off between complexity and performance is obtained by selecting the precoders based on the pre-equalization mutual information and considering the post-equalization SINR only for transmission rank and rate adaptation, significantly reducing the amount of matrix inversion.

The sensitivity of instantaneous CSI feedback with respect to a delay in the feedback path is investigated. The obtained results show that this kind of feedback is restricted to quasi-stationary situations if no delay compensation is incorporated. The scope of the algorithms can be extended to moderate mobility (~ 50 km/h) if delay compensation by means of channel prediction is implemented. The achievable throughput with limited feedback is strongly dependent on the granularity of the feedback indicators in the OFDM time-frequency grid, requiring a careful choice of the feedback parameters to achieve a good trade-off between the uplink overhead and the downlink throughput.

Chapter 4.

Linear Transmission and CSI Feedback for Multi-User MIMO

> When you see something that is technically sweet, you go ahead and do it and you argue about what to do about it only after you have had your technical success.
>
> *(J. Robert Oppenheimer)*

Spatial multiplexing of multiple users, i.e., multi-user MIMO (MU-MIMO), is considered as a promising technique in multi-antenna broadcast systems to achieve high spectral efficiencies by serving multiple users in parallel over the same time-frequency resources [44, 67]. In contrast to single-user MIMO (SU-MIMO), the potential multiplexing gain of MU-MIMO is only confined by the capabilities of the transmitter. Hence, with MU-MIMO the need for multiple antennas at the users is eliminated, facilitating the development of small and cheap user equipments. Another important advantage of MU-MIMO over SU-MIMO is the reduced sensitivity to channel impairments such as line-of-sight propagation and antenna correlation, which cause an increase in the singular value imbalance of the single-user channel and hence effectively limit the transmission rank of SU-MIMO [153, 154]. Unfortunately, these advantages come at the cost of significantly increased susceptibility of the achievable multi-user throughput with respect to the accuracy of the channel state information (CSI) at the transmitter (CSIT), due to residual multi-user interference incurred with imperfect channel knowledge [45, 46, 155, 156].

The literature on MU-MIMO precoding can be coarsely partitioned into two main concepts, that is, linear precoding and non-linear precoding. While non-linear techniques based on Tomlinson-Harashima precoding [155] and vector perturbation [156, 157] provide an advantage in terms of the achievable throughput, linear strategies such as zero forcing (ZF) beamforming [68] and block diagonalization (BD) precoding [69] are considered as practically important for complexity reasons. In addition to their complexity advantage, ZF and BD precoding have the benefit that partial CSIT, specifically channel subspace information, is sufficient for the calculation of the precoders, reducing the burden on the feedback channel. Notice that BD precoding can be viewed as a generalization

of ZF beamforming to multiple data stream transmission per user; hence in the sequel mostly BD precoding is used to refer to both methods.

For the reasons mentioned above, MU-MIMO transmission based on BD precoding is investigated in this dissertation. Although codebook based precoding as in SU-MIMO is an option for MU-MIMO as well [72], the precoders employed with BD precoding are not confined to a codebook but are calculated during operation from explicit channel knowledge at the base station. Correspondingly, the implicit CSI feedback of SU-MIMO is replaced with direct quantization of the channel matrix in MU-MIMO. To this end, memoryless and predictive CSI quantization algorithms are proposed in this chapter, building upon the concepts of Grassmannian quantization [158, 159]. The temporal correlation of the wireless channel (see Section 2.4) can be exploited by the proposed predictive quantizer to achieve high fidelity quantization in low to moderate mobility scenarios, i.e., when the channel coherence time is large. In case of distributed antennas, knowledge of the channel gain matrix (2.11) is utilized during quantization to further improve the accuracy.

Similar to SU-MIMO, channel quality indicator (CQI) feedback is employed in MU-MIMO as well, for transmission rate adaptation and multi-user scheduling. In contrast to single-user transmission, the achievable rate with MU-MIMO cannot be estimated accurately by the users ahead of scheduling due to the unknown precoders. Exploiting the precoder construction of BD precoding, a lower bound on the expected SINR of a user with imperfect CSIT is derived, facilitating estimation of the achievable transmission rate. This bound is proposed as CQI feedback for multi-user scheduling.

In case the number of data streams per user is less than the number of receive antennas, interference cancellation by means of BD is only achieved over a subspace of a users' channel matrix. Ideally, joint user subspace selection and scheduling must be performed at the transmitter to maximize the achievable throughput, requiring full channel knowledge for all users at the base station. To reduce the CSI feedback overhead, it is, however, proposed to perform a pre-selection of the preferred channel subspace by the users based on selfish arguments, effectively decreasing the dimensionality of the Grassmannian quantization problem. To filter-out the interference-free subspace at the users, semi-unitary antenna combiners are employed. It is shown how the degrees of freedom (DoF) provided by the excess antennas can be utilized to trade-off the residual multi-user interference, due to imperfect CSI, for the effective channel gain experienced by a user.

The proposed CSI feedback algorithms are extended to frequency selective OFDM systems employing the same concepts as in SU-MIMO, i.e., by means of CSI feedback clustering and interpolation. The performance of these two methods is evaluated in this chapter using simulations, with the result that interpolation is only reasonable if the distance between CSI pilots is small compared to the channel coherence bandwidth, challenging its value for practical implementations.

This chapter is organized as follows: In Section 4.1, the ZF beamforming and BD precoding constructions are introduced, providing the motivation for the proposed Grassmannian quantization algorithms. The memoryless and predictive CSI quantizers are detailed in Sections 4.2.1 and 4.2.2,

respectively. In Section 4.3, the proposed antenna combiners are derived. Finally, in Section 4.4, the feedback methods are extended to frequency selective systems by means of interpolation and clustering, and the lower bound on the SINR for CQI feedback is presented in Section 4.4.2. The performance of the methods and algorithms is investigated in Section 4.4.3 through extensive Monte-Carlo simulations. Relevant background information on the Grassmann manifold, useful for the understanding of the proposed Grassmannian quantizer, is provided in Appendix C. Proofs and derivations related to the antenna combiners and the SINR lower bound proposed in this chapter are presented in Appendices F and G. The MIMO minimum mean-squared error (MMSE) equalizer that is employed in some of the simulations presented in Section 4.4.3 is derived in Appendix H. The impact of the residual multi-user interference due to CSIT inaccuracies is effectively reduced by this equalizer by exploiting the BD construction to estimate the multi-user interference. The material presented in this chapter is published in parts in [87–96].

4.1. Zero-Forcing and Block-Diagonalization Precoding

In this section, an overview of ZF and BD precoding is provided and the relevant CSIT for precoder calculation is identified. It is shown that channel subspace information is sufficient if uniform power allocation over spatial streams and users is considered, which is known to be a asymptotically optimal in the limit of high SNR [71]. Although not considered in this dissertation, optimal power allocation by means of water-filling can be facilitated with additional channel magnitude feedback [69].

4.1.1. Transmit Strategy

The input-output relationship of user u in cell i on resource element (RE) $[n, k]$ assuming MU-MIMO transmission is

$$\begin{aligned}\mathbf{y}_{u,i}[n,k] &= \mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k]\mathbf{x}_{u,i}[n,k] \\ &+ \mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}} \sum_{\substack{s\in\mathcal{S}_i[n,k] \\ s\neq u}} \mathbf{F}_{s,i}[n,k]\mathbf{x}_{s,i}[n,k] + \mathbf{G}_{u,i}[n,k]^{\mathrm{H}}\,\tilde{\mathbf{z}}_{u,i}[n,k],\end{aligned} \tag{4.1}$$

with the effective channel $\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] \in \mathbb{C}^{N_i \times \ell_{u,i}[k]}$ being specified in Equation (2.6) and $\tilde{\mathbf{z}}_{u,i}[n,k] \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \tilde{\sigma}_z^2\,\mathbf{I}_{M_{u,i}}\right)$, hence the out-of-cell interference is treated as additional Gaussian noise. To simplify notations, the user set $\mathcal{U}_i$ is re-ordered after scheduling such that the served users $\mathcal{S}_i[n,k]$ can be indexed as $\mathcal{S}_i[n,k] = \{1,\ldots,S_i[n,k]\}$. The goal of BD precoding is to calculate the precoders $\mathbf{F}_{s,i}[n,k], \forall s$ such that the interference at each user due to all other users is perfectly eliminated. BD was originally proposed for the case that each user is served over $\ell_{u,i}[k] = M_{u,i}$ streams. Then no antenna combiner $\mathbf{G}_{u,i}[n,k]$ is required at the users. If the number of data streams

$\ell_{u,i}[k]$ is less than the number of receive antennas $M_{u,i}$, the BD precoder is calculated for the effective channel $\mathbf{H}_{u,i}^{\text{eff}}[n,k]$. Hence the multi-user interference is only canceled in the $\ell_{u,i}[k]$-dimensional subspace of $\text{span}\left(\mathbf{H}_{u,i}[n,k]\right)$ that is spanned by $\mathbf{H}_{u,i}^{\text{eff}}[n,k]$. The precoders are obtained from the following conditions

$$\mathbf{H}_{u,i}^{\text{eff}}[n,k]^{\text{H}}\mathbf{F}_{s,i}[n,k] = \mathbf{0}_{\ell_{u,i}[k]\times\ell_{s,i}[k]},\ \forall u, s \in \mathcal{S}_i[n,k] \text{ and } s \neq u, \tag{4.2}$$

$$\text{rank}\left(\mathbf{H}_{u,i}^{\text{eff}}[n,k]^{\text{H}}\mathbf{F}_{u,i}[n,k]\right) = \ell_{u,i}[k],\ \forall u \in \mathcal{S}_i[n,k]. \tag{4.3}$$

Notice the similarity of this formulation of BD to interference alignment [160, 161], with the difference that there is only a single transmitter in BD that has access to the data of all users.[1] Certain feasibility conditions on the number of streams per user and the total number of spatial streams $\ell_i[n,k]$ have to be fulfilled to assure that a solution to this problem exists, such as $\ell_{u,i}[k] \leq M_{u,i}$ and $\ell_i[n,k] \leq N_i$ [69]. Provided these feasibility conditions are satisfied, the solution is obtained from

$$\mathbf{F}_{u,i}[n,k] \in \text{null}\left(\overline{\mathbf{H}}_{u,i}[n,k]\right),\ \text{rank}\left(\mathbf{F}_{u,i}[n,k]\right) = \ell_{u,i}[k],\ \forall u \in \mathcal{S}_i[n,k], \tag{4.4}$$

$$\overline{\mathbf{H}}_{u,i}[n,k] = \left[\tilde{\mathbf{H}}_{1,i}[n,k], \ldots, \tilde{\mathbf{H}}_{u-1,i}[n,k], \tilde{\mathbf{H}}_{u+1,i}[n,k], \ldots, \tilde{\mathbf{H}}_{S_i[n,k],i}[n,k]\right]^{\text{H}}, \tag{4.5}$$

$$\text{span}\left(\tilde{\mathbf{H}}_{s,i}[n,k]\right) = \text{span}\left(\mathbf{H}_{s,i}^{\text{eff}}[n,k]\right), \tag{4.6}$$

$$\tilde{\mathbf{H}}_{s,i}[n,k]^{\text{H}}\tilde{\mathbf{H}}_{s,i}[n,k] = \mathbf{I}_{\ell_{s,i}[k]},\ \tilde{\mathbf{H}}_{s,i}[n,k] \in \mathbb{C}^{N_i\times\ell_{s,i}[k]},\ \forall s \in \mathcal{S}_i[n,k]. \tag{4.7}$$

Here, the matrices $\tilde{\mathbf{H}}_{s,i}[n,k]$ form orthonormal bases for the respective effective channels $\mathbf{H}_{s,i}^{\text{eff}}[n,k]$ and can, e.g., be obtained from a QR or a singular value decomposition (SVD) of $\mathbf{H}_{s,i}^{\text{eff}}[n,k]$. The precoder of user u lies in the left null space of all other users' effective channels. Hence, after antenna combining, the transmission to user u does not interfere with the transmission to any of the other users. As the same holds true for every $u \in \mathcal{S}_i[n,k]$, interference-free transmission to all users is achieved. To satisfy the rank constraint (4.3), $\text{rank}\left(\mathbf{F}_{u,i}[n,k]\right) = \ell_{u,i}[k]$ must be fulfilled.

As mentioned above, to maximize the SINR of the users, a joint optimization of the precoders $\mathbf{F}_{s,i}[n,k]$ and the antenna combiners $\mathbf{G}_{s,i}[n,k]$ at the base station, based on perfect knowledge of $\mathbf{H}_{s,i}[n,k], \forall s$, is required; see, e.g., [162]. With the proposed approach, however, it is sufficient to feed back the subspace $\text{span}\left(\mathbf{H}_{s,i}^{\text{eff}}[n,k]\right), \forall s$, effectively reducing the feedback overhead, facilitating independent design of the transmit- and receive-filters and enabling the application of Grassmannian quantization, as detailed Section 4.1.2. Useful selections of $\mathbf{H}_{s,i}^{\text{eff}}[n,k]$ are proposed in Section 4.3.

In case of ZF beamforming, transmitting a single stream per user $\ell_{u,i}[k] = 1,\ \forall u$, the effective channel matrix $\mathbf{H}_{u,i}^{\text{eff}}[n,k]$ of user u is reduced to the effective channel vector $\mathbf{h}_{u,i}^{\text{eff}}[n,k]$ and the

[1] When applied for coordinated multi-point (CoMP) transmission, BD is hence classified as a joint transmission strategy.

precoding matrix $\mathbf{F}_{u,i}[n,k]$ is obtained as the column-vector $\mathbf{f}_{u,i}[n,k] \in \mathbb{C}^{N_i \times 1}$. Combining the normalized effective channel vectors (channel directions)

$$\tilde{\mathbf{h}}_{u,i}[n,k] = \frac{\mathbf{h}_{u,i}^{\mathrm{eff}}[n,k]}{\left\|\mathbf{h}_{u,i}^{\mathrm{eff}}[n,k]\right\|} \tag{4.8}$$

of all served users in the so-called cell channel direction matrix

$$\tilde{\mathbf{H}}_i[n,k] = \left[\tilde{\mathbf{h}}_{1,i}[n,k], \ldots, \tilde{\mathbf{h}}_{S_i[n,k],i}[n,k]\right]^{\mathrm{H}} \in \mathbb{C}^{S_i[n,k] \times N_i}, \tag{4.9}$$

a closed-form solution of the *cell ZF precoding matrix* $\mathbf{F}_i[n,k]$ of all users is obtained as

$$\mathbf{F}_i[n,k] = \tilde{\mathbf{H}}_i[n,k]^{\mathrm{H}} \left(\tilde{\mathbf{H}}_i[n,k]\tilde{\mathbf{H}}_i[n,k]^{\mathrm{H}}\right)^{-1} \mathbf{P}_i[n,k]^{1/2} = \tilde{\mathbf{F}}_i[n,k]\mathbf{P}_i[n,k]^{1/2}, \tag{4.10}$$

$$\mathbf{F}_i[n,k] = \left[\mathbf{f}_{1,i}[n,k], \ldots, \mathbf{f}_{S_i[n,k],i}[n,k]\right], \quad \mathbf{P}_i[n,k] = \mathrm{diag}\left(p_{1,i}[n,k], \ldots, p_{S_i[n,k],i}[n,k]\right),$$

where the precoding vectors $\mathbf{f}_{u,i}[n,k]$ are obtained as the columns of $\mathbf{F}_i[n,k]$. Considering equal power allocation, the powers $p_{u,i}[n,k]$ are calculated as

$$p_{u,i}[n,k] = \frac{P_i}{S_i[n,k]\left\|[\tilde{\mathbf{F}}_i[n,k]]_{:,u}\right\|^2}, \ \forall u \in \{1, \ldots, S_i[n,k]\}, \tag{4.11}$$

with P_i denoting the instantaneous total power constraint of the base station.

With the BD condition (4.2), the input-output relationship with perfect CSIT simplifies to

$$\mathbf{y}_{u,i}[n,k] = \mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\, \mathbf{F}_{u,i}[n,k]\, \mathbf{x}_{u,i}[n,k] + \mathbf{G}_{u,i}[n,k]^{\mathrm{H}}\, \tilde{\mathbf{z}}_{u,i}[n,k]. \tag{4.12}$$

The precoder $\mathbf{F}_{u,i}[n,k]$ obtained from (4.4) is unique only up to right-multiplication with any full-rank $\ell_{u,i}[k] \times \ell_{u,i}[k]$ matrix, as this multiplication has no impact on $\mathrm{span}\,(\mathbf{F}_{u,i}[n,k])$.

In [69], this ambiguity is removed by treating the interference-free input-output relationship (4.12) as a SU-MIMO system (cf. Equation (3.1)), and additionally performing SVD based precoding with water-filling power allocation over the obtained single-user channel. Hence, if $\tilde{\mathbf{F}}_{u,i}[n,k]$ denotes an arbitrary semi-unitary solution of (4.4), the unambiguous precoder according to [69] is obtained as

$$\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\, \tilde{\mathbf{F}}_{u,i}[n,k] = \mathbf{U}_{u,i}[n,k]\mathbf{\Sigma}_{u,i}[n,k]\mathbf{V}_{u,i}[n,k]^{\mathrm{H}}, \tag{4.13}$$

$$\mathbf{F}_{u,i}[n,k] = \tilde{\mathbf{F}}_{u,i}[n,k]\mathbf{V}_{u,i}[n,k]\mathbf{P}_{u,i}[n,k]^{1/2}, \tag{4.14}$$

with (4.13) denoting an SVD of $\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\, \tilde{\mathbf{F}}_{u,i}[n,k]\right)$ and $\mathbf{P}_{u,i}[n,k]$ being the diagonal power allocation matrix obtained form the water-filling solution; see Section 3.1.2.

This, however, is not a reasonable approach with the subspace CSI feedback proposed below, because the feedback only provides an arbitrary orthonormal basis for $\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]$ that neither matches the

individual orientations of the eigenmodes of the effective channel nor contains information on their magnitude. Instead, semi-unitary precoding with equal power allocation is considered, enforcing the following additional precoder constraint

$$\mathbf{F}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k] = \frac{P_i}{S_i[n,k]\,\ell_{u,i}[k]}\tilde{\mathbf{F}}_{u,i}[n,k]^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}[n,k] = P_{u,i}[n,k]\mathbf{I}_{\ell_{u,i}[k]}. \tag{4.15}$$

The corresponding solution is unique up to right-multiplication with any unitary matrix, which has no impact on the *achievable per user transmission rate with perfect CSIT and equal power allocation*

$$R_{\mathrm{BD}} = \mathbb{E}\log_2\det\left(\mathbf{I}_{\ell_{u,i}[k]} + \rho\,\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}[n,k]\tilde{\mathbf{F}}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]\right),\ \rho = \frac{P_{u,i}[n,k]}{\tilde{\sigma}_z^2}. \tag{4.16}$$

4.1.2. Limited Feedback Model

With the assumption of equal power allocation, the BD precoders can be calculated at the base station if knowledge of the $\ell_{u,i}[k]$-dimensional subspace $\mathrm{span}\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]\right)$ is available for every $u \in \mathcal{S}_i[n,k]$. This subspace can be interpreted as a point on the Grassmann manifold of $\ell_{u,i}[k]$-dimensional subspaces in the N_i-dimensional Euclidean space, denoted as $\mathcal{G}\left(N_i, \ell_{u,i}[k]\right)$. In general it can be represented equivalently by any matrix $\mathbf{S}$ spanning the subspace

$$\mathbf{S} \equiv \mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] \Leftrightarrow \mathrm{span}\left(\mathbf{S}\right) = \mathrm{span}\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]\right). \tag{4.17}$$

In this dissertation the most common approach to represent points on the Grassmann manifold is employed, using orthonormal bases as in Equations (4.6) and (4.7). A more detailed introduction to the Grassmannian is given in Appendix C.

Perfect knowledge of $\tilde{\mathbf{H}}_{u,i}[n,k]$, $\forall u \in \mathcal{S}_i[n,k]$ from (4.6) is required at the base station to achieve zero multi-user interference. With limited feedback, however, a quantized version $\hat{\mathbf{H}}_{u,i}[n,k]$ of the subspace is fed back to the transmitter. To convey the CSI to the transmitter, a quantization codebook $\mathcal{Q}_{u,i}[n,k]$ of size 2^b is employed by user u, with b denoting the number of feedback bits

$$\mathcal{Q}_{u,i}[n,k] = \left\{\mathbf{Q}_j \in \mathbb{C}^{N_i \times \ell_{u,i}[k]} \middle| \mathbf{Q}_j^{\mathrm{H}}\mathbf{Q}_j = \mathbf{I}_{\ell_{u,i}[k]},\ j \in \{1,\ldots,2^b\}\right\}. \tag{4.18}$$

The codebook is known by the base station. The quantized channel subspace is obtained by minimizing the subspace chordal distance [46] between $\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]$ and the elements of the codebook

$$\hat{\mathbf{H}}_{u,i}[n,k] = \underset{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]}{\mathrm{argmin}}\ \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k], \mathbf{Q}_j\right), \tag{4.19}$$

$$\mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k], \mathbf{Q}_j\right) = \ell_{u,i}[k] - \mathrm{tr}\left(\tilde{\mathbf{H}}_{u,i}[n,k]^{\mathrm{H}}\mathbf{Q}_j\mathbf{Q}_j^{\mathrm{H}}\tilde{\mathbf{H}}_{u,i}[n,k]\right), \tag{4.20}$$

$$\mathrm{d}_{\mathrm{c,min}}^2 = \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k], \hat{\mathbf{H}}_{u,i}[n,k]\right) = \ell_{u,i}[k] - \sum_{\ell=1}^{\ell_{u,i}[k]}\cos\left(\varphi_\ell^{(q)}\right)^2. \tag{4.21}$$

In this equation, the $\ell_{u,i}[k]$ principal angles between $\mathrm{span}(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]) = \mathrm{span}(\tilde{\mathbf{H}}_{u,i}[n,k])$ and $\mathrm{span}(\hat{\mathbf{H}}_{u,i}[n,k])$ are denoted $\varphi_{\ell}^{(q)}$. When the chordal distance is employed as quantization metric the expected value of the residual multi-user interference, due to the CSI quantization error, is minimized. This is investigated in more detail in Sections 4.3 and 4.4.2, when considering the theoretical performance of BD precoding with limited feedback and deriving the lower bound on the expected SINR. The quantization problem (4.19) is well known as Grassmannian quantization on the Grassmann manifold $\mathcal{G}\left(N_i, \ell_{u,i}[k]\right)$ [158].

It is assumed that the base station treats the quantized subspaces $\hat{\mathbf{H}}_{u,i}[n,k]$ as the actual subspaces $\tilde{\mathbf{H}}_{u,i}[n,k]$, and calculates the precoders from (4.4), replacing $\overline{\mathbf{H}}_{u,i}[n,k]$ with

$$\hat{\overline{\mathbf{H}}}_{u,i}[n,k] = \left[\hat{\mathbf{H}}_{1,i}[n,k], \ldots, \hat{\mathbf{H}}_{u-1,i}[n,k], \hat{\mathbf{H}}_{u+1,i}[n,k], \ldots, \hat{\mathbf{H}}_{S_i[n,k],i}[n,k]\right]^{\mathrm{H}}. \tag{4.22}$$

The effective channel $\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]$ can be decomposed into its range space component $\mathbf{H}_{u,i}^{\mathrm{eff,r}}[n,k]$ and its left null space component $\mathbf{H}_{u,i}^{\mathrm{eff,n}}[n,k]$, with respect to the orthonormal basis $\hat{\mathbf{H}}_{u,i}[n,k]$

$$\begin{aligned}\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] &= \hat{\mathbf{H}}_{u,i}[n,k]\hat{\mathbf{H}}_{u,i}[n,k]^{\mathrm{H}}\,\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] + \left(\mathbf{I}_{N_i} - \hat{\mathbf{H}}_{u,i}[n,k]\hat{\mathbf{H}}_{u,i}[n,k]^{\mathrm{H}}\right)\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] \\ &= \underbrace{\hat{\mathbf{H}}_{u,i}[n,k]\hat{\mathbf{H}}_{u,i}[n,k]^{\mathrm{H}}\,\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]}_{\mathbf{H}_{u,i}^{\mathrm{eff,r}}[n,k]} + \underbrace{\hat{\mathbf{H}}_{u,i}^{\perp}[n,k]\hat{\mathbf{H}}_{u,i}^{\perp}[n,k]^{\mathrm{H}}\,\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]}_{\mathbf{H}_{u,i}^{\mathrm{eff,n}}[n,k]},\end{aligned} \tag{4.23}$$

with $\hat{\mathbf{H}}_{u,i}^{\perp}[n,k] \in \mathbb{C}^{N_i \times N_i - \ell_{u,i}[k]}$ being an orthonormal basis for the orthogonal complement of $\mathrm{span}(\hat{\mathbf{H}}_{u,i}[n,k])$. Using this notation and recalling that $\hat{\mathbf{H}}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{s,i}[n,k] = 0$, $\forall s \neq u$, due to the BD construction, the input-output relationship with quantized CSIT is

$$\begin{aligned}\mathbf{y}_{u,i} &= \mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k]\mathbf{x}_{u,i}[n,k] \\ &+ \mathbf{H}_{u,i}^{\mathrm{eff,n}}[n,k]^{\mathrm{H}} \sum_{s=1,s\neq u}^{S_i[n,k]} \mathbf{F}_{s,i}[n,k]\mathbf{x}_{s,i}[n,k] + \mathbf{G}_{u,i}[n,k]^{\mathrm{H}}\,\tilde{\mathbf{z}}_{u,i}[n,k].\end{aligned} \tag{4.24}$$

Thus, with quantized CSIT perfect interference cancellation is not achieved. The residual interference is determined by the quantization error of $\tilde{\mathbf{H}}_{u,i}[n,k]$, captured by the null space component $\mathbf{H}_{u,i}^{\mathrm{eff,n}}[n,k]$. The *per-user transmission rate with quantized CSIT* is [46]

$$\begin{aligned}R_{\text{BD-Quant}} &= \mathbb{E}\log_2\det\left(\mathbf{I}_{\ell_{u,i}[k]} + \sum_{s=1}^{S_i[n,k]} \mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\mathbf{F}_{s,i}[n,k]\mathbf{F}_{s,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]\right) \\ &- \mathbb{E}\log_2\det\left(\mathbf{I}_{\ell_{u,i}[k]} + \sum_{s=1,s\neq u}^{S_i[n,k]} \mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]^{\mathrm{H}}\mathbf{F}_{s,i}[n,k]\mathbf{F}_{s,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]\right),\end{aligned} \tag{4.25}$$

where the expectation is taken with respect to channel and corresponding precoder realizations.

4.2. Explicit CSI Feedback Algorithms

In interference limited multi-user precoding systems, such as MU-MIMO and interference alignment, strict requirements are placed on the CSIT accuracy, because CSI imperfections directly impact the residual multi-user interference and strongly deteriorate the throughput performance [45, 46, 74]. Therefore, efficient CSI quantization is central in such systems to obtain sufficiently accurate CSIT with a reasonable feedback overhead. As mentioned above, for the calculation of the BD precoders with equal power allocation channel subspace information is sufficient.[2] This information can efficiently be conveyed by the Grassmann manifold, putting Grassmannian quantization at the center of this section. In Section 4.2.1, memoryless quantization of channel subspaces is studied, where the CSI at each transmission time interval (TTI) is independently quantized, i.e., without considering the past. Efficient memoryless quantization in distributed antenna systems (DASs) is achieved by taking into account the channel gain matrix (2.11) during construction of the quantization codebook, to match the spatial distribution of the channel matrix. Still, often sufficient CSIT accuracy cannot be obtained with an acceptable feedback overhead employing memoryless techniques. To circumvent this problem, predictive quantization is proposed in Section 4.2.2, exploiting the temporal channel correlation to improve the quantizer's fidelity. The performance of the quantizers is investigated numerically by evaluating the quantization mean-squared error (MSE) in Section 4.2.3. The derivation of channel quality feedback for transmission rate adaptation and multi-user scheduling is postponed until Section 4.4.2 when frequency selective systems are treated.

The methods proposed in this section are designed for CSI quantization on a single frequency flat subcarrier. Efficient extensions to multi-carrier OFDM are proposed in Section 4.4, considering feedback clustering and interpolation. Also, antenna combining is not considered in Sections 4.2.1 and 4.2.2; hence, $\ell_{u,i}[k] = M_{u,i}$ and $\mathbf{G}_{u,i}[n,k] = \mathbf{I}_{M_{u,i}}$ is assumed. Excess receive antennas, i.e., the case $\ell_{u,i}[k] \leq M_{u,i}$, and antenna combining are covered in Section 4.3. Notice that it is possible to apply the predictive quantization proposed in Section 4.2.2 in the frequency domain as well, i.e., over OFDM subcarriers. This is not treated in this thesis, although it has been investigated by myself in [89] and by others, e.g., in [86]. With sufficiently dense placement of CSI pilots in the frequency domain, a substantial prediction gain can be achieved, especially if large system bandwidths are considered. The interpolation and clustering methods proposed in this dissertation, however, facilitate a significant reduction of the required CSI pilots in the frequency domain.

4.2.1. Memoryless Grassmannian Quantization

With memoryless quantization, the temporal correlation of the source to be quantized is neglected and the quantization is performed independently at each time instant k, employing a pre-specified

[2] The same holds true for other concepts as well, e.g., interference alignment [53] and signal to leakage and noise ratio beamforming [54].

quantization codebook $\mathcal{Q}_{u,i}[n,k]$. Possibly different codebooks are used by different users to avoid situations in which users feed back the same CSI, excluding them from being spatially multiplexed with each other. The average quantization distortion in terms of the chordal distance MSE

$$D = \mathbb{E}\left(\mathrm{d}_{\mathrm{c,min}}^2\right) \tag{4.26}$$

is critically determined by the construction of the quantization codebook.

Codebooks for isotropic subspaces If the channel subspace $\mathrm{span}\left(\tilde{\mathbf{H}}_{u,i}[n,k]\right)$ is uniformly distributed on the Grassmann manifold $\mathcal{G}\left(N_i, M_{u,i}\right)$, i.e., the channel matrix $\mathbf{H}_{u,i}[n,k]$ is isotropically distributed (e.g., independent and identically distributed (i.i.d.) Rayleigh fading), the minimum average quantization distortion is achieved by maximally spaced Grassmannian subspace packings with respect to the chordal distance [40]. These packings are essentially uniform on the Grassmann manifold in terms of the chordal distance. Unfortunately, finding good codebooks is in general hard except for special cases; an algorithm for obtaining good codebooks is provided, e.g., in [163].

More suitable for analytical investigations, however, are random quantization codebooks, consisting of elements $\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]$ that are chosen independently and uniformly distributed on $\mathcal{G}\left(N_i, M_{u,i}\right)$ [46]. The elements of such codebooks can be obtained as the compact left singular matrices of i.i.d. Gaussian matrices $\mathbf{H} \in \mathbb{C}^{N_i \times M_{u,i}},\ [\mathbf{H}]_{m,n} \sim \mathcal{N}_{\mathbb{C}}\left(0,1\right)$. Such a codebook construction, denoted as random vector quantization (RVQ), is applied in this dissertation whenever memoryless quantization is evaluated. In that case, the results are averaged over codebook realizations by employing independent quantization codebooks at each time instant k. Bounds on the average distortion achieved with random isotropically distributed Grassmannian quantization codebooks are derived in [159], and the codebooks are shown to perform asymptotically optimal in a number of applications, e.g., [164].

Codebooks for correlated subspaces In case the channel subspace to be quantized is not obtained from an isotropically distributed channel matrix, better quantization performance can be achieved if knowledge of the spatial channel correlation is exploited. An efficient heuristic codebook construction is proposed in [165] for single-user multiple-input single-output beamforming systems, which effectively "colors" a Grassmannian line packing according to the channel correlation matrix. This approach is extended here for Grassmannian subspace packings to exploit the pathloss differences experienced in DASs. Considering the DAS model of Equation (2.9), a random Grassmannian codebook that is statistically matched to the distribution of the channel subspace can be obtained as

$$\mathcal{Q}_{u,i}^{(\mathrm{corr})}[n,k] = \left\{ \mathbf{Q}_j^{(\mathrm{corr})} \middle| \mathbf{Q}_j^{(\mathrm{corr})}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}} = \mathbf{C}_{u,i}^{1/2}\bar{\mathbf{H}} \in \mathbb{C}^{N_i \times M_{u,i}},\ [\bar{\mathbf{H}}]_{m,n} \sim \mathcal{N}_{\mathbb{C}}\left(0,1\right) \right\}. \tag{4.27}$$

Here, $\mathbf{Q}_j^{(\mathrm{corr})}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ represents a compact SVD of the matrix $\mathbf{C}_{u,i}^{1/2}\bar{\mathbf{H}}$. The codebook construction is denoted as correlated RVQ. An analytic performance characterization of this construction was not yet

successful. Simulations demonstrating the performance gain achieved in DASs with this codebook are provided in Section 5.2 (see also [95, 96]). The method can be applied to account for the spatial correlation between the entries of the channel matrix as well. Alternative approaches to consider the pathloss differences in DASs or the spatial channel correlation exist in literature, e.g., in [166] a quantization codebook for ZF beamforming based on estimated channel statistics is proposed, and in [167] a feedback bit allocation scheme for DASs is derived.

4.2.2. Predictive Grassmannian Quantization

MU-MIMO techniques are mostly considered for situations in which users are either static or moving slowly through the cell. If this is the case, the temporal correlation of the wireless channel is often high, which can be exploited to improve the CSI quantization accuracy and to reduce the feedback overhead. A common approach in quantization and source coding for exploiting the correlation of the source signal is vector quantization [76], where multiple consecutive samples of the source signal are combined to a vector and are jointly coded/quantized. Although such techniques are rate-distortion optimal, they cannot be applied for CSI quantization in wireless communications, because the throughput performance of MU-MIMO with outdated CSI degrades in a similar way as with a quantization error [75]. Hence, the delay of vector quantization caused by gathering and processing the data cannot be tolerated. Instead, predictive quantization on the Grassmann manifold is proposed in this section to achieve efficient subspace quantization. Predictive quantizers are able to leverage the temporal correlation of the source signal at consecutive time instants to provide higher fidelity at a given quantization rate. The prediction of points on the Grassmannian is realized by translating the problem to the tangent space associated with the manifold. The statistics of the prediction error in the tangent space are approximated with a Gaussian distribution, which is valid for sufficiently small prediction errors. Based on this approximation, a Grassmannian quantization codebook is generated to quantize the prediction error. The performance of the proposed algorithm is evaluated by means of Monte Carlo simulations of the quantization MSE in Section 4.2.3. An application of the proposed quantizer for CSI quantization in limited feedback MU-MIMO is investigated in Section 4.4.3. Basic concepts and relationships associated with the Grassmann manifold that are exploited in the following derivation are detailed in Appendix C. To shorten notations, the RE index $[n, k]$ is partly omitted for intermediate steps and auxiliary variables of the derivation.

Predictive quantization model: The predictive Grassmannian quantizer derived below explicitly considers the effect of distributing antenna arrays over the cell area, by employing the decomposition of the channel matrix into a temporally constant channel gain matrix $\mathbf{C}_{u,i}$ and a small-scale fading matrix $\bar{\mathbf{H}}_{u,i}[n, k]$ according to Equation (2.9). The channel gain matrix $\mathbf{C}_{u,i}$ is supposed to be known by the transmitter and the receiver. The temporal variation of the channel due to the movement of

obstacles and/or users is hence captured in $\bar{\mathbf{H}}_{u,i}[n,k]$. In the derivation of the quantizer, a general model for the temporal evolution of $\bar{\mathbf{H}}_{u,i}[n,k]$ is considered

$$\bar{\mathbf{H}}_{u,i}[n,k] = G\left(\bar{\mathbf{H}}_{u,i}[n,k-1], \bar{\mathbf{H}}_{u,i}[n,k-2], \ldots\right) + \bar{\mathbf{J}}_{u,i}[n,k] = \bar{\mathbf{H}}_{u,i}^{(d)}[n,k] + \bar{\mathbf{J}}_{u,i}[n,k]. \tag{4.28}$$

Here, the deterministic dependence of the current channel on the past is described with $\bar{\mathbf{H}}_{u,i}^{(d)}[n,k]$ and the random innovation is captured in the matrix $\bar{\mathbf{J}}_{u,i}[n,k]$, which is assumed as i.i.d. Gaussian

$$\mathrm{vec}\left(\bar{\mathbf{J}}_{u,i}[n,k]\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma_j^2[n,k]\mathbf{I}_{N_i \cdot M_{u,i}}\right). \tag{4.29}$$

No specific assumptions are made about the nature of the deterministic function $G(\cdot)$ in (4.28), because the behavior of the wireless channel is strongly dependent on the surrounding environment [168]. In [169] it is shown that Rayleigh fading processes can be accurately modeled using auto-regressive relations, fitting well to the considered decomposition in (4.28).

In the proposed quantizer, a prediction algorithm is implemented to provide an estimate of the deterministic evolution $\bar{\mathbf{H}}_{u,i}^{(d)}[n,k]$, based on previously quantized channel observations. Assuming that a prediction $\bar{\mathbf{H}}_{u,i}^{(p)}[n,k]$ of $\bar{\mathbf{H}}_{u,i}^{(d)}[n,k]$ is available, the channel can be expressed as

$$\mathbf{H}_{u,i}[n,k] = \mathbf{C}_{u,i}^{1/2}\Big(\bar{\mathbf{H}}_{u,i}^{(p)}[n,k] + \underbrace{\bar{\mathbf{E}}_{u,i}^{(p)}[n,k] + \bar{\mathbf{J}}_{u,i}[n,k]}_{\bar{\mathbf{E}}_{u,i}[n,k]}\Big) = \mathbf{H}_{u,i}^{(p)}[n,k] + \mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k], \tag{4.30}$$

with $\bar{\mathbf{E}}_{u,i}^{(p)}[n,k]$ denoting the prediction error. The prediction error is assumed as independent of the innovation and i.i.d. Gaussian: $\mathrm{vec}(\bar{\mathbf{E}}_{u,i}^{(p)}[n,k]) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_p^2[n,k]\mathbf{I}_{N_i \cdot M_{u,i}})$. Thus, the total error $\bar{\mathbf{E}}_{u,i}[n,k]$ is distributed as

$$\mathrm{vec}\left(\bar{\mathbf{E}}_{u,i}[n,k]\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma_e^2[n,k]\mathbf{I}_{N_i \cdot M_{u,i}}\right), \quad \sigma_e^2[n,k] = \sigma_p^2[n,k] + \sigma_j^2[n,k]. \tag{4.31}$$

Given the prediction $\bar{\mathbf{H}}_{u,i}^{(p)}[n,k]$ at both ends of the CSI feedback link, the channel can be reproduced from knowledge of the error $\bar{\mathbf{E}}_{u,i}[n,k]$. With an accurate prediction ($\sigma_p^2[n,k] \ll 1$) and a smoothly varying channel ($\sigma_j^2[n,k] \ll 1$), the variance of the signal to be quantized can thus be reduced by quantizing $\bar{\mathbf{E}}_{u,i}[n,k]$ instead of $\bar{\mathbf{H}}_{u,i}[n,k]$.

By taking advantage of the CSIT invariances introduced in Section 4.1.2, the required feedback information can further be reduced. Using orthonormal bases $\tilde{\mathbf{H}}_{u,i}[n,k]$ and $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$ for the actual channel $\mathbf{H}_{u,i}[n,k]$ and the predicted channel $\mathbf{H}_{u,i}^{(p)}[n,k]$, as defined by Equation (4.7), the channel matrix can be written as

$$\mathbf{H}_{u,i}[n,k] = \tilde{\mathbf{H}}_{u,i}[n,k]\mathbf{D}_{u,i}[n,k] = \tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]\mathbf{D}_{u,i}^{(p)}[n,k] + \mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k], \tag{4.32}$$

$$\mathbf{D}_{u,i}[n,k] = \tilde{\mathbf{H}}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k], \quad \mathbf{D}_{u,i}^{(p)}[n,k] = \tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}^{(p)}[n,k]. \tag{4.33}$$

The channel subspace, represented by the orthonormal basis $\tilde{\mathbf{H}}_{u,i}[n,k]$, is thus obtained as

$$\tilde{\mathbf{H}}_{u,i}[n,k] = \left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]\mathbf{D}_{u,i}^{(p)}[n,k] + \mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k]\right)\mathbf{D}_{u,i}[n,k]^{-1}. \tag{4.34}$$

Exploiting the differential geometry associated with the Grassmannian, it is possible to describe the error between the predicted subspace $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$ and the observed subspace $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}[n,k])$ with the error tangent

$$\mathbf{T}_{u,i}[n,k] = T\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k], \tilde{\mathbf{H}}_{u,i}[n,k]\right) \in \mathcal{T}\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]\right) \subset \mathbb{C}^{N_i \times M_{u,i}}, \tag{4.35}$$

according to Equation (C.9), where $\mathcal{T}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$ denotes the tangent space associated with the prediction $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$. The shortest path on the manifold between the predicted subspace and the actual subspace, which is denoted as the geodesic $\Gamma(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k], \mathbf{T}_{u,i}[n,k], p)$, defined by Equation (C.11), is specified by the tangent $\mathbf{T}_{u,i}[n,k]$. Having knowledge of $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$ and $\mathbf{T}_{u,i}[n,k]$, it is hence possible to obtain an equivalent representation of $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}[n,k])$ from

$$\tilde{\mathbf{H}}_{u,i}[n,k] \equiv \tilde{\mathbf{H}}_{u,i}^{(e)}[n,k] = \Gamma\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k], \mathbf{T}_{u,i}[n,k], 1\right), \tag{4.36}$$

$$\mathrm{span}\left(\tilde{\mathbf{H}}_{u,i}[n,k]\right) = \mathrm{span}\left(\tilde{\mathbf{H}}_{u,i}^{(e)}[n,k]\right), \;\; \tilde{\mathbf{H}}_{u,i}^{(e)}[n,k]^{\mathrm{H}}\tilde{\mathbf{H}}_{u,i}^{(e)}[n,k] = \mathbf{I}_{M_{u,i}}. \tag{4.37}$$

In general, the two orthonormal bases $\tilde{\mathbf{H}}_{u,i}[n,k]$ and $\tilde{\mathbf{H}}_{u,i}^{(e)}[n,k]$ are not equal but equivalent representatives of the same point on the Grassmann manifold $\mathcal{G}(N_i, M_{u,i})$ in terms of the equivalence relationship (C.2). For the purpose of precoder calculation for BD MU-MIMO either matrix can be employed; an explicit distinction between $\tilde{\mathbf{H}}_{u,i}[n,k]$ and $\tilde{\mathbf{H}}_{u,i}^{(e)}[n,k]$ is hence omitted in the following. To calculate the subspace $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}[n,k])$ at the base station, it is sufficient to feed back the error tangent $\mathbf{T}_{u,i}[n,k]$. Notice that this tangent lies in the $(N_i - M_{u,i})$-dimensional tangent space associated with $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$. Therefore, compared to quantization of $\bar{\mathbf{E}}_{u,i}[n,k]$, the dimensionality of the quantization problem is reduced from N_i to $N_i - M_{u,i}$.

Approximation of the error tangent statistics: To derive an efficient quantizer for the error tangent knowledge about the statistics of $\mathbf{T}_{u,i}[n,k]$ is required. Due to the nonlinear operations involved in the calculation of the tangent, obtaining a closed-form statistical description is hard. Therefore, an approximation is employed in the following that is valid if the channel correlation is sufficiently large.

According to the definition in Equation (C.9), the tangent $\mathbf{T}_{u,i}[n,k]$ is obtained as

$$\mathbf{T}_{u,i}[n,k] = \mathbf{U}\mathbf{\Phi}\mathbf{V}^{\mathrm{H}}, \;\; \mathbf{U}\tan(\mathbf{\Phi})\mathbf{V}^{\mathrm{H}} = \mathbf{\Theta}_{u,i}[n,k], \tag{4.38}$$

$$\mathbf{\Theta}_{u,i}[n,k] = \Bigg(\underbrace{\mathbf{I}_{N_i} - \tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}}_{\mathbf{P}_{u,i}[n,k]}\Bigg)\tilde{\mathbf{H}}_{u,i}[n,k]\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\tilde{\mathbf{H}}_{u,i}[n,k]\right)^{-1}. \tag{4.39}$$

Here, the diagonal matrix $\mathbf{\Phi} = \mathrm{diag}(\varphi_1^{(p)}, \ldots, \varphi_{M_{u,i}}^{(p)})$ is composed of the principal angles between the subspaces spanned by $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$ and $\tilde{\mathbf{H}}_{u,i}[n,k]$, which are obtained from the arctangent of the singular values of $\mathbf{\Theta}_{u,i}[n,k]$. Under the assumption $\varphi_i^{(p)} \leq 0.35 \mathrel{\hat{=}} 20°$ an error that is below 4% is caused by the approximation

$$\mathrm{atan}\left(\varphi_i^{(p)}\right) \approx \varphi_i^{(p)}. \tag{4.40}$$

With this approximation, the tangent simplifies to

$$\mathbf{T}_{u,i}[n,k] \approx \mathbf{\Theta}_{u,i}[n,k] = \mathbf{P}_{u,i}[n,k]\tilde{\mathbf{H}}_{u,i}[n,k]\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\tilde{\mathbf{H}}_{u,i}[n,k]\right)^{-1}, \tag{4.41}$$

where the matrix $\mathbf{P}_{u,i}[n,k]$, which is defined in Equation (4.39), can be identified as a projection onto the orthogonal complement of $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$. Substituting $\tilde{\mathbf{H}}_{u,i}[n,k]$ from Equation (4.34) the approximated tangent can be written as

$$\mathbf{T}_{u,i}[n,k] \approx \mathbf{P}_{u,i}[n,k]\mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k]\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]\right)^{-1}, \tag{4.42}$$

where the first term from (4.34) vanishes due to the projection $\mathbf{P}_{u,i}[n,k]$ and the matrix $\mathbf{D}_{u,i}[n,k]^{-1}$ is pulled inside the brackets. The tangent matrix is hence obtained as the product of the left null space component of the channel matrix with respect to the predicted subspace $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$, and the inverse of the term $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]$ that is determined by the range space component. Applying an SVD to the term in brackets

$$\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k] = \mathbf{Y}_{u,i}[n,k]\mathbf{\Lambda}_{u,i}[n,k]\mathbf{W}_{u,i}[n,k]^{\mathrm{H}}, \tag{4.43}$$

$$\mathbf{Y}_{u,i}[n,k], \mathbf{\Lambda}_{u,i}[n,k], \mathbf{W}_{u,i}[n,k] \in \mathbb{C}^{M_{u,i} \times M_{u,i}},$$

the approximated tangent can be formulated as

$$\mathbf{T}_{u,i}[n,k] \approx \mathbf{P}_{u,i}[n,k]\mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k]\left(\mathbf{W}_{u,i}[n,k]\mathbf{\Lambda}_{u,i}[n,k]^{-1}\mathbf{Y}_{u,i}[n,k]^{\mathrm{H}}\right). \tag{4.44}$$

Assuming $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]$ as fixed/observed, the distribution of the tangent is determined by the null space component $\mathbf{P}_{u,i}[n,k]\mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k]$. This term is obtained as the projection of the zero-mean Gaussian error $\mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k]$ onto the orthogonal complement of $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$ and is therefore itself Gaussian distributed. The distribution of $\mathbf{T}_{u,i}[n,k]$ is hence obtained as

$$\mathrm{vec}\left(\mathbf{T}_{u,i}[n,k]\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \mathbf{C}_{u,i}^{(t)}[n,k]\right), \tag{4.45}$$

$$\mathbf{C}_{u,i}^{(t)}[n,k] = \sigma_e^2[n,k]\Big(\underbrace{\mathbf{Y}_{u,i}[n,k]\mathbf{\Lambda}_{u,i}[n,k]^{-2}\mathbf{Y}_{u,i}[n,k]^{\mathrm{H}}}_{\mathbf{C}_M \in \mathbb{C}^{M_{u,i} \times M_{u,i}}} \otimes \underbrace{\mathbf{P}_{u,i}[n,k]\mathbf{C}_{u,i}\mathbf{P}_{u,i}[n,k]}_{\mathbf{C}_N \in \mathbb{C}^{N_i \times N_i}}\Big). \tag{4.46}$$

Notice that the correlation matrix (4.46) is obtained from a Kronecker product. It is therefore possible to derive a statistically matched quantization codebook for the error tangent $\mathbf{T}_{u,i}[n,k]$ by correlating i.i.d. Gaussian matrices

$$\mathcal{Q}_{u,i}^{(t,\text{corr})}[n,k] = \left\{ \mathbf{C}_N^{1/2} \mathbf{Q}_j \mathbf{C}_M^{1/2} \;\middle|\; \mathbf{Q}_j \in \mathbb{C}^{N_i \times M_{u,i}},\, [\mathbf{Q}_j]_{l,m} \sim \mathcal{N}_{\mathbb{C}} \left(0, \sigma_e^2[n,k]\right) \right\}. \tag{4.47}$$

The correlation matrix $\mathbf{C}_N$ can be calculated by the transmitter and the receiver, because both have knowledge of the predicted subspace and of the channel gain matrix. Unfortunately, $\mathbf{C}_M$ is not known a-priori and can therefore not be exploited to improve the quantization efficiency. Hence, the codebook is designed under the assumption that $\mathbf{C}_M$ is a scaled identity, which is the case if $\mathbf{\Lambda}_{u,i}[n,k]$ in (4.46) is a scaled identity. The *error tangent correlation matrix* is then obtained as

$$\mathbf{C}_{u,i}^{(t)}[n,k] \approx \left(\frac{\sigma_e[n,k]}{\lambda_{u,i}[n,k]} \right)^2 \left(\mathbf{I}_{M_{u,i}} \otimes \mathbf{C}_N \right). \tag{4.48}$$

To further motivate this approach, two special cases are considered below that enable a more detailed investigation of the factor $\lambda_{u,i}[n,k]$.

Error tangent statistics with $M_{u,i} = 1$: When the users are equipped with single receive antennas, the channel matrix is reduced to the channel vector $\mathbf{h}_{u,i}[n,k]$ and $\mathbf{C}_M$ is obtained as

$$\mathbf{C}_M = \frac{1}{\|\mathbf{h}_{u,i}[n,k]\|^2 \cos\left(\varphi^{(p)}\right)^2}, \;\; \cos\left(\varphi^{(p)}\right)^2 = \left| \tilde{\mathbf{h}}_{u,i}^{(p)}[n,k]^{\mathrm{H}} \tilde{\mathbf{h}}_{u,i}[n,k] \right|^2, \tag{4.49}$$

with $0 \leq \varphi^{(p)} \leq \pi/2$ denoting the principal angle of the subspace prediction error. In this case, $\mathbf{C}_M$ is reduced to a scalar and has an impact only on the variance of the error tangent. Notice that the variance is increased in fading dips, i.e., when $\|\mathbf{h}_{u,i}[n,k]\|^2$ is small, and in case the subspace prediction error is large.

Error tangent statistics for i.i.d. Gaussian channels: When the channel matrix is i.i.d. Gaussian distributed $[\mathbf{H}_{u,i}[n,k]]_{l,m} \sim \mathcal{N}_{\mathbb{C}}(0, \gamma_{u,i})$, it is possible to determine the expected value of the matrix $\mathbf{C}_M$ with respect to the unknown range space component of $\mathbf{H}_{u,i}[n,k]$ within $\operatorname{span}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$. This is achieved by considering the inverse of $\mathbf{C}_M$

$$\mathbf{C}_M^{-1} = \mathbf{Y}_{u,i}[n,k] \mathbf{\Lambda}_{u,i}[n,k]^2 \mathbf{Y}_{u,i}[n,k]^{\mathrm{H}} = \tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}} \mathbf{H}_{u,i}[n,k] \mathbf{H}_{u,i}[n,k]^{\mathrm{H}} \tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]. \tag{4.50}$$

With a compact SVD of the channel matrix $\mathbf{H}_{u,i}[n,k] = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$, this product can be written as

$$\mathbf{C}_M^{-1} = \tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}} \mathbf{U} \mathbf{\Sigma}^2 \mathbf{U}^{\mathrm{H}} \tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]. \tag{4.51}$$

Notice that $\mathbf{U}$ is an isotropically distributed unitary matrix that is statistically independent of the singular value matrix $\mathbf{\Sigma}$, due to the assumption that the channel is i.i.d. Gaussian and hence also

isotropic [170, Theorem 1]. Applying an SVD to the product $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]^{\mathrm{H}}\mathbf{U} = \mathbf{Q}\cos(\mathbf{\Phi})\,\mathbf{W}^{\mathrm{H}}$, the inverse of $\mathbf{C}_M$ is obtained as

$$\mathbf{C}_M^{-1} = \mathbf{Q}\cos(\mathbf{\Phi})\left(\mathbf{W}^{\mathrm{H}}\mathbf{\Sigma}^2\mathbf{W}\right)\cos(\mathbf{\Phi})\,\mathbf{Q}^{\mathrm{H}}. \tag{4.52}$$

Due to the isotropy of $\mathbf{U}$, its projection onto $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$ is isotropically distributed within the subspace. Hence, the $M_{u,i} \times M_{u,i}$ matrices $\mathbf{Q}$, $\cos(\mathbf{\Phi})$ and $\mathbf{W}$ are statistically independent, and $\mathbf{Q}$ and $\mathbf{W}$ are isotropically distributed unitary matrices [170, Theorem 1]. Furthermore, $\cos(\mathbf{\Phi})$ is composed of the cosines of the principal angles between $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$ and $\mathrm{span}(\tilde{\mathbf{H}}_{u,i}[n,k])$

$$\cos(\mathbf{\Phi}) = \mathrm{diag}\left(\cos\left(\varphi_1^{(p)}\right),\ldots,\cos\left(\varphi_{M_{u,i}}^{(p)}\right)\right). \tag{4.53}$$

Because of the isotropy of $\mathbf{W}$ and its statistical independence of $\mathbf{\Sigma}$, it follows that

$$\mathbf{W}^{\mathrm{H}}\mathbf{\Sigma}^2\mathbf{W} \sim \mathcal{W}_{M_{u,i}}^{\mathbb{C}}\left(N_i, \gamma_{u,i}\,\mathbf{I}_{M_{u,i}}\right), \tag{4.54}$$

with $\mathcal{W}_{M_{u,i}}^{\mathbb{C}}\left(N_i, \gamma_{u,i}\,\mathbf{I}_{M_{u,i}}\right)$ denoting a central complex Wishart distribution of dimension $M_{u,i}$, having N_i DoF. With this notation, the correlation matrix $\mathbf{C}_M$ is obtained as

$$\mathbf{C}_M = \mathbf{Q}\cos(\mathbf{\Phi})^{-1}\left(\mathbf{W}^{\mathrm{H}}\mathbf{\Sigma}^2\mathbf{W}\right)^{-1}\cos(\mathbf{\Phi})^{-1}\,\mathbf{Q}^{\mathrm{H}}, \tag{4.55}$$

where the inverse matrix in the center is distributed according to the inverse Wishart distribution $\mathcal{W}_{M_{u,i}}^{-\mathbb{C}}\left(N_i, \gamma_{u,i}^{-1}\,\mathbf{I}_{M_{u,i}}\right)$. Correspondingly, the expected value of $\mathbf{C}_M$ can be calculated as [171]

$$\mathbb{E}(\mathbf{C}_M) = \frac{\mathbb{E}\left(\mathbf{Q}\cos(\mathbf{\Phi})^{-2}\,\mathbf{Q}^{\mathrm{H}}\right)}{\gamma_{u,i}(N_i - M_{u,i})} = \frac{1}{\gamma_{u,i}(N_i - M_{u,i})}\left(\frac{1}{M_{u,i}}\sum_{j=1}^{M_{u,i}}\frac{1}{\cos(\varphi_j^{(p)})^2}\right)\mathbf{I}_{M_{u,i}}, \tag{4.56}$$

where the second equality follows from the isotropy of $\mathbf{Q}$. Hence, averaging out the unknown orientation of the range space component of $\mathbf{H}_{u,i}[n,k]$ within $\mathrm{span}\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]\right)$ and the unknown singular values, a scaled identity matrix is obtained. The exact scaling is determined by the distribution of the chordal distance prediction error among the subspace dimensions. Notice that the expected value of (4.49) in case of i.i.d. Rayleigh fading is obtained from (4.56) by setting $M_{u,i} = 1$.

Quantization algorithm: Based on the observations of the previous paragraphs, a predictive Grassmannian quantizer is proposed that generates at each time instant k a quantization codebook, according to the tangent statistics derived in Equations (4.45) and (4.48). To determine the correlation matrix $\mathbf{C}_{u,i}^{(t)}[n,k]$ in (4.48), a subspace prediction $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$ is required at the CSI encoder and the decoder. In this paragraph the prediction is assumed to be given; subspace prediction algorithms are proposed in the next paragraph. In general the scaling $\left(\frac{\sigma_e[n,k]}{\lambda_{u,i}[n,k]}\right)^2$ of $\mathbf{C}_{u,i}^{(t)}[n,k]$ is time-dependent; a tracking algorithm is proposed to adapt the scaling of the quantization codebook correspondingly.

As described in Section 4.1.2, the chordal distance is employed as quantization metric. The chordal distance cannot be evaluated directly in the tangent space associated with the Grassmannian. It is therefore necessary to project the tangent codebook onto the manifold to obtain an equivalent Grassmannian codebook that enables the quantization. The quantization algorithm is described in the following steps:

-2. Initialize the codebook scale parameter $s_{u,i}[n,0]=0$ and the scale growth rate $g>1$. These parameters are used to track the scaling of $\mathbf{C}_{u,i}^{(t)}[n,k]$.

-1. Initialize a Gaussian quantization codebook $\mathcal{Q}_{u,i}^{(0)}$ of size 2^b for a standard normal random matrix of size $N_i \times M_{u,i}$. In this dissertation this codebook is randomly generated as

$$\mathcal{Q}_{u,i}^{(0)} = \left\{ \mathbf{Q}_j^{(0)} \middle| \ \left[\mathbf{Q}_j^{(0)}\right]_{l,m} \sim \mathcal{N}_{\mathbb{C}}\left(0,1\right) \right\}. \tag{4.57}$$

0. Correlate the elements of $\mathcal{Q}_{u,i}^{(0)}$ to match the correlation of the prediction error $\mathbf{C}_{u,i}^{1/2}\bar{\mathbf{E}}_{u,i}[n,k]$

$$\mathcal{Q}_{u,i}^{(\text{corr})} = \left\{ \mathbf{C}_{u,i}^{1/2}\mathbf{Q}_j^{(0)} \middle| \, \forall \mathbf{Q}_j^{(0)} \in \mathcal{Q}_{u,i}^{(0)} \right\}. \tag{4.58}$$

Repeat the following steps for each $k>0$:

1. Predict the current subspace and calculate the projection matrix $\mathbf{P}_{u,i}[n,k]$ in Equation (4.39).
2. Project the correlated codebook onto the orthogonal complement of $\operatorname{span}(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k])$ to determine the tangent codebook matching the structure of the error tangent correlation matrix $\mathbf{C}_{u,i}^{(t)}[n,k]$ defined in (4.48)

$$\mathcal{Q}_{u,i}^{(t)}[n,k] = \left\{ \mathbf{P}_{u,i}[n,k]\mathbf{Q}_j \middle| \, \forall \mathbf{Q}_j \in \mathcal{Q}_{u,i}^{(\text{corr})} \right\} \subset \mathcal{T}\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]\right). \tag{4.59}$$

3. Calculate two scaled codebooks to track the scaling of $\mathbf{C}_{u,i}^{(t)}[n,k]$

$$\mathcal{Q}_{-,u,i}^{(t)}[n,k] = \left\{ g^{s_-}\,\mathbf{Q}_j^{(t)} \middle| \, \forall \mathbf{Q}_j^{(t)} \in \mathcal{Q}_{u,i}^{(t)}[n,k] \right\}, \tag{4.60}$$

$$\mathcal{Q}_{+,u,i}^{(t)}[n,k] = \left\{ g^{s_+}\,\mathbf{Q}_j^{(t)} \middle| \, \forall \mathbf{Q}_j^{(t)} \in \mathcal{Q}_{u,i}^{(t)}[n,k] \right\}, \tag{4.61}$$

$$s_- = s_{u,i}[n,k-1]-1, \quad s_+ = \min\left(s_{u,i}[n,k-1]+1,0\right). \tag{4.62}$$

4. Project the obtained tangent codebooks onto the manifold to obtain two Grassmannian codebooks $\mathcal{Q}_{-,u,i}[n,k]$ and $\mathcal{Q}_{+,u,i}[n,k]$. This is achieved with the geodesic $\Gamma\left(\cdot\right)$ defined in (C.11)

$$\mathcal{Q}_{-,u,i}[n,k] = \left\{ \Gamma\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k], \mathbf{Q}_j^{(t)}, 1\right) \middle| \forall \mathbf{Q}_j^{(t)} \in \mathcal{Q}_{-,u,i}^{(t)}[n,k] \right\}, \tag{4.63}$$

$$\mathcal{Q}_{+,u,i}[n,k] = \left\{ \Gamma\left(\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k], \mathbf{Q}_j^{(t)}, 1\right) \middle| \forall \mathbf{Q}_j^{(t)} \in \mathcal{Q}_{+,u,i}^{(t)}[n,k] \right\}. \tag{4.64}$$

5. Quantize the observed subspace $\tilde{\mathbf{H}}_{u,i}[n,k]$ with respect to both Grassmannian codebooks

$$\mathbf{Q}_- = \underset{\mathbf{Q}_j \in \mathcal{Q}_{-,u,i}[n,k]}{\operatorname{argmin}} \mathrm{d}_\mathrm{c}^2\left(\tilde{\mathbf{H}}_{u,i}[n,k], \mathbf{Q}_j\right),\ d_- = \mathrm{d}_\mathrm{c}^2\left(\tilde{\mathbf{H}}_{u,i}[n,k], \mathbf{Q}_-\right), \tag{4.65}$$

$$\mathbf{Q}_+ = \underset{\mathbf{Q}_j \in \mathcal{Q}_{+,u,i}[n,k]}{\operatorname{argmin}} \mathrm{d}_\mathrm{c}^2\left(\tilde{\mathbf{H}}_{u,i}[n,k], \mathbf{Q}_j\right),\ d_+ = \mathrm{d}_\mathrm{c}^2\left(\tilde{\mathbf{H}}_{u,i}[n,k], \mathbf{Q}_+\right). \tag{4.66}$$

6. Determine the "winning" codebook, i.e., the one with minimum quantization error

$$w = \underset{i \in \{-,+\}}{\operatorname{argmin}}\, d_i. \tag{4.67}$$

7. Determine the quantized channel subspace and update the codebook scale parameter

$$\hat{\mathbf{H}}_{u,i}[n,k] = \mathbf{Q}_w \in \mathcal{Q}_{w,u,i}[n,k],\ \ s_{u,i}[n,k] = s_w. \tag{4.68}$$

8. Feedback the indices of $\mathbf{Q}_w$ and s_w.

The tangent codebook obtained in step 2 of the algorithm is equal in distribution to the error tangent specified by Equations (4.45) and (4.48), despite a scaling factor. The codebook scale parameter $s_{u,i}[n,k]$ is employed to track the scaling $\left(\frac{\sigma_e[n,k]}{\lambda_{u,i}[n,k]}\right)^2$ of $\mathbf{C}_{u,i}^{(t)}[n,k]$, by determining the appropriate up- or down-scaling of the codebook scale parameter in steps 3 and 6 of the algorithm. The ratio of the scaling between two consecutive time instants is given by

$$\frac{g^{s_{u,i}[n,k]}}{g^{s_{u,i}[n,k-1]}} = g^{\pm 1}. \tag{4.69}$$

Thus, the tracking speed of the algorithm is determined by the variance growth rate g; increasing g causes a faster tracking speed, but also a larger steady state tracking error. The optimal value of g is determined by means of simulations in my corresponding publication [91], where it is shown that the optimum depends on the subspace dimension, but the sensitivity of the algorithm to the calibration of g is not strongly pronounced. In this publication also the accuracy of the approximation of the error tangent statistics in (4.45) is evaluated, demonstrating that the estimated tangent variance fits well to the observed variance up to very high channel Doppler frequencies.

The feedback overhead of the proposed quantization algorithm is $b+1$ bits per quantization instant. The 1 bit overhead is caused by the codebook scale parameter that must be signaled to the decoder as additional side-information. The same prediction algorithm is employed at the encoder and at the decoder, such that both sides are able to calculate the two scaled codebooks $\mathcal{Q}_{-,u,i}[n,k]$ and $\mathcal{Q}_{+,u,i}[n,k]$. With the feedback information, i.e., the indices of $\mathbf{Q}_w$ and s_w, the quantized channel subspace $\hat{\mathbf{H}}_{u,i}[n,k]$ and the scale parameter $s_{u,i}[n,k]$ can therefore be reproduced by the decoder.

Prediction algorithm: The Grassmannian quantization algorithm proposed in the previous paragraph is based on the availability of a prediction $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$ of the current channel subspace, to calculate a quantization codebook that is matched to the temporal evolution of the subspace. The prediction algorithm must be based on quantized CSI to enable the prediction at the encoder and at the decoder. Trivially, the previously quantized subspace can be employed as a prediction $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k] = \hat{\mathbf{H}}_{u,i}[n,k-1]$, leading to a differential quantizer as proposed in [78]. Better performance can be achieved with a more sophisticated prediction. One possibility is to implement the prediction directly on the Grassmann manifold, e.g., by means of a linear predictor

$$\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k] = \sum_{j=1}^{N_p} \hat{\mathbf{H}}_{u,i}[n,k-j]\; a_j. \tag{4.70}$$

In general, the obtained prediction $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$ is not compatible with the considered semi-unitary matrix representation of points on $\mathcal{G}\left(N_i, M_{u,i}\right)$, which could be resolved by applying an SVD decomposition after linear prediction to determine the corresponding subspace. With this approach, however, an optimal solution for the filter coefficients is hard to obtain.

This difficulty can be avoided by predicting the tangent $T(\hat{\mathbf{H}}_{u,i}[n,k-1], \tilde{\mathbf{H}}_{u,i}[n,k])$ between the quantized subspace $\hat{\mathbf{H}}_{u,i}[n,k-1]$ at time instant $k-1$ and the observed subspace $\tilde{\mathbf{H}}_{u,i}[n,k]$ at time instant k. With the geodesic defined by this predicted tangent, a subspace prediction $\tilde{\mathbf{H}}_{u,i}^{(p)}[n,k]$ can indirectly be obtained. In the tangent space well known prediction algorithms from Euclidean geometry can be reused. Although statistical models for the temporal evolution of the channel matrix exist, there are no corresponding models for the temporal evolution of tangents available. Due to the highly nonlinear relationship between subspaces and tangents, it is difficult to derive such a model. Instead, predictors that are not based on an underlying channel model are proposed in this dissertation. For one-dimensional subspaces adaptive finite impulse response (FIR) filters are employed, while a regression model is considered for higher dimensional subspaces. The adaptive filtering approach is not used for higher dimensional subspaces, due to the large number of filter coefficients required to achieve an accurate prediction and the corresponding slow filter convergence speed.

The basic idea of regression based prediction is proposed by Zhang et.al. for one-dimensional subspaces in [82] under the name *robust Grassmannian prediction*. Its extension to higher dimensional subspaces is straightforward and is conducted in my publication [90]; the interested reader is referred to this publication for details. The basic idea is to apply a linear regression to the tangents observed over multiple time instants and to predict the current tangent based on this regression. Compared to the adaptive filter based approach proposed below, the regression has the advantage that the filter convergence phase of the adaptive filter is omitted, despite a slightly reduced prediction performance as shown in Section 4.2.3.

Adaptive filter based prediction of one-dimensional subspaces: The following predictor is derived for a prediction on the Grassmann manifold $\mathcal{G}\left(N_i, 1\right)$. An auto-regressive model of order

N_p is considered to describe the temporal evolution of the tangent random process. The prediction is based on quantized CSI. The tangent describing the geodesic between the quantized channel subspaces $\hat{\mathbf{h}}_{u,i}[n,k-1]$ and $\hat{\mathbf{h}}_{u,i}[n,k]$ at consecutive time instants is denoted

$$\mathbf{t}_{u,i}^{(q)}[n,k] = T\left(\hat{\mathbf{h}}_{u,i}[n,k-1], \hat{\mathbf{h}}_{u,i}[n,k]\right) \in \mathcal{T}\left(\hat{\mathbf{h}}_{u,i}[n,k-1]\right). \tag{4.71}$$

At time instant k, a prediction $\tilde{\mathbf{h}}_{u,i}^{(p)}[n,k]$ of the subspace $\tilde{\mathbf{h}}_{u,i}[n,k]$ is required by the quantizer to calculate the quantization codebook. This prediction is obtained indirectly by first predicting the tangent $\mathbf{t}_{u,i}[n,k] = T\left(\hat{\mathbf{h}}_{u,i}[n,k-1], \tilde{\mathbf{h}}_{u,i}[n,k]\right)$ using a linear predictor of order N_p

$$\mathbf{t}_{u,i}^{(p)}[n,k] = \sum_{j=1}^{N_p} \bar{\mathbf{t}}_{u,i}^{(q)}[n,k-j]\; a_{u,i}^{(j)}[n,k] = \mathbf{T}_{u,i}^{(q)}[n,k]\mathbf{a}_{u,i}[n,k], \tag{4.72}$$

$$\mathbf{T}_{u,i}^{(q)}[n,k] = \left[\bar{\mathbf{t}}_{u,i}^{(q)}[n,k-1], \ldots, \bar{\mathbf{t}}_{u,i}^{(q)}[n,k-N_p]\right] \in \mathcal{T}\left(\hat{\mathbf{h}}_{u,i}[n,k-1]\right), \tag{4.73}$$

$$\mathbf{a}_{u,i}[n,k] = \left[a_{u,i}^{(1)}[n,k], \ldots, a_{u,i}^{(N_p)}[n,k]\right]^{\mathrm{T}}. \tag{4.74}$$

A linear combination of tangent vectors as in (4.72) is only meaningful if the tangents are defined in the same tangent space. Hence, the tangents $\mathbf{t}_{u,i}^{(q)}[n,k-i], i \in \{1,\ldots,N_p\}$ cannot directly be employed in (4.72). Instead, it is necessary to transport the geometric information contained in $\mathbf{t}_{u,i}^{(q)}[n,k-i]$ from the respective tangent space $\mathcal{T}(\hat{\mathbf{h}}_{u,i}[n,k-i-1])$ to the current position $\mathcal{T}(\hat{\mathbf{h}}_{u,i}[n,k-1])$. This is enabled by means of parallel transport as defined in Equation (C.12). Thus, parallel transported versions $\bar{\mathbf{t}}_{u,i}^{(q)}[n,k-i]$ of the previously observed tangents $\mathbf{t}_{u,i}^{(q)}[n,k-i]$ are employed in (4.72). This is achieved by updating the matrix $\mathbf{T}_{u,i}^{(q)}[n,k]$ at each time instant k, as soon as $\hat{\mathbf{h}}_{u,i}[n,k]$ is observed, as follows:

1. Circularly shift the columns of $\mathbf{T}_{u,i}^{(q)}[n,k]$ by one to the right.
2. Calculate the tangent $\mathbf{t}_{u,i}^{(q)}[n,k]$ and replace the first column of $\mathbf{T}_{u,i}^{(q)}[n,k]$ with $\mathbf{t}_{u,i}^{(q)}[n,k]$.
3. Parallel transport the columns of $\mathbf{T}_{u,i}^{(q)}[n,k]$ along the geodesic defined by $\mathbf{t}_{u,i}^{(q)}[n,k]$ from $\mathcal{T}(\hat{\mathbf{h}}_{u,i}[n,k-1])$ to $\mathcal{T}(\hat{\mathbf{h}}_{u,i}[n,k])$, using (C.12).

$$\left[\mathbf{T}_{u,i}^{(q)}[n,k+1]\right]_{:,j} = \Pi\left(\hat{\mathbf{h}}_{u,i}[n,k-1], \hat{\mathbf{h}}_{u,i}[n,k], \left[\mathbf{T}_{u,i}^{(q)}[n,k]\right]_{:,j}\right). \tag{4.75}$$

The predictor coefficients $\mathbf{a}_{u,i}[n,k]$ are trained with a stochastic gradient algorithm, namely the normalized least mean squares algorithm [149]. Defining the tangent prediction error at time k as

$$\mathbf{e}_{u,i}^{(p)}[n,k] = \mathbf{t}_{u,i}^{(q)}[n,k] - \mathbf{t}_{u,i}^{(p)}[n,k] = \mathbf{t}_{u,i}^{(q)}[n,k] - \mathbf{T}_{u,i}^{(q)}[n,k]\mathbf{a}_{u,i}[n,k], \tag{4.76}$$

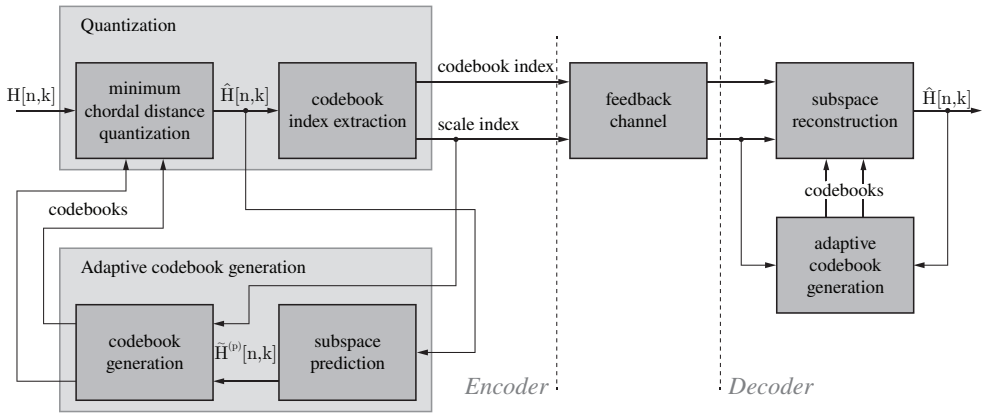


Figure 4.1.: Structure of the Grassmannian subspace quantizer visualizing the different components.

the filter coefficient update rule is obtained as

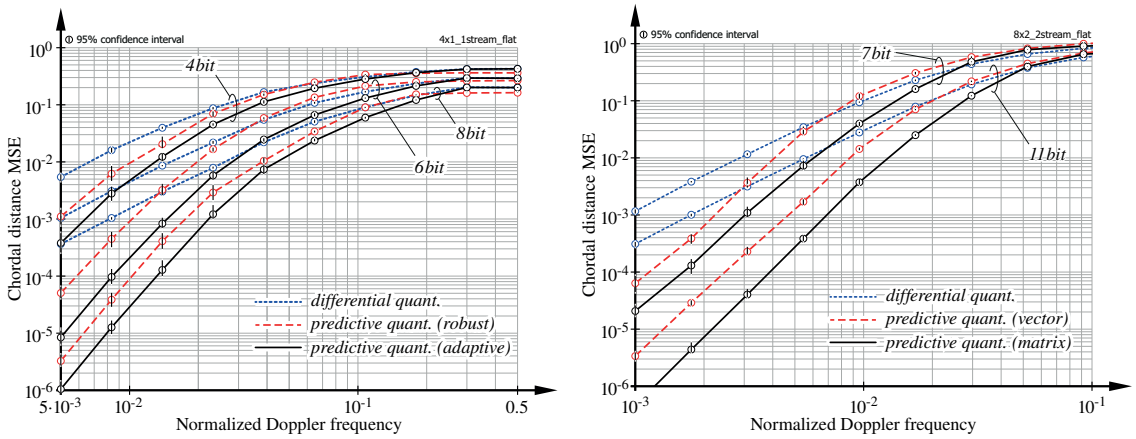
$$\mathbf{a}_{u,i}[n,k+1] = \mathbf{a}_{u,i}[n,k] + \mu \frac{\mathbf{T}_{u,i}^{(q)}[n,k]^{\mathrm{H}}}{\left\|\mathbf{T}_{u,i}^{(q)}[n,k]\right\|^2} \mathbf{e}_{u,i}^{(p)}[n,k]. \tag{4.77}$$

The step size μ, determining the trade-off between the filter convergence speed and the steady-state MSE, has to satisfy $0 < \mu < 2$ for convergence [149]. Notice that the error (4.76) is calculated with the tangent $\mathbf{t}_{u,i}^{(q)}[n,k]$ obtained from the quantized subspace $\hat{\mathbf{h}}_{u,i}[n,k]$ and not from the actual subspace $\tilde{\mathbf{h}}_{u,i}[n,k]$. Therefore, the accuracy of the prediction of $\tilde{\mathbf{h}}_{u,i}[n,k]$ is impacted by the quantization error. This approach is necessary to enable the prediction at the decoder, which has only access to quantized CSI.

In Figure 4.1, a schematic of the concatenation of the proposed quantization, prediction and codebook generation methods is illustrated, visualizing the interplay between the different components. The quantized subspace is passed from the quantizer to the predictor at both ends of the feedback link to enable the prediction of the next subspace. Based on this prediction and on information about the previously employed codebook scaling, a new quantization codebook is calculated that spans a certain volume on the Grassmann manifold around the predicted subspace. Over a dedicated feedback channel the quantized CSI is passed from the encoder to the decoder in form of the codebook index and the scale index. This information is sufficient for the decoder to reconstruct the quantized channel subspace and to prepare the codebook for the next time instant.

4.2.3. Evaluation of the Quantization MSE

In this section, the performance of the proposed Grassmannian quantization algorithms is investigated by means of Monte-Carlo simulations. The chordal distance MSE as defined in Equation (4.26),



(a) Predictive quantization on $\mathcal{G}(4,1)$.

(b) Predictive quantization on $\mathcal{G}(8,2)$.

Figure 4.2.: MSE performance comparison of the proposed predictive Grassmannian quantizer using different prediction algorithms.

providing a natural distance measure for points on the Grassmannian, is employed as performance metric. The chordal distance also determines the achievable rate loss of BD based MU-MIMO with quantized CSIT compared to perfect CSIT; see Section 4.3 for details.

In the first simulation, predictive Grassmannian quantization on $\mathcal{G}(4,1)$ is considered using the quantization codebook sizes $Q \in \{8, 32, 128\}$ which translates to a CSI feedback overhead of $b \in \{3, 5, 7\} + 1$ bit, due to the 1 bit overhead caused by the codebook scale index. The performance of differential quantization [78], robust Grassmannian prediction [82] and the adaptive filter based prediction proposed in the previous section is compared. The temporal correlation is determined by Clarke's model as described in Section 2.4. The results are plotted in dependency of the normalized channel Doppler frequency ν_d defined in Equation (3.33). The appropriate length N_p of the prediction filter is determined by the channel correlation according to Clarke's model. The length is set as

$$N_p = \min\left(\left\lceil \frac{1.5}{2\pi\nu_d} \right\rceil, 20\right), \tag{4.78}$$

employing past values with a correlation of ≥ 0.5 for prediction, where the argument of the Bessel function determining the correlation in Clarke's model is approximately 1.5 [112]. The maximum filter length is restricted to 20 taps to achieve moderate complexity and a reasonable filter convergence speed. The channel gain matrix (2.11) is assumed as an identity matrix.

The results of the simulation are shown in Figure 4.2a. The performance of the quantizer is strongly dependent on the Doppler frequency, which determines the channel correlation according to (2.13). With increasing Doppler frequency the channel correlation is reduced, leading to a larger prediction error and thus to an increased variance of the error tangent $\mathbf{T}_{u,i}[n,k]$. It can be seen that the slope of the MSE achieved with the adaptive filter based predictor and the robust predictor is similar.

Differential quantization is significantly outperformed by predictive quantization if the channel correlation is sufficiently large, i.e., at low Doppler frequencies. Notice though that this observation depends on the considered channel model. In [78] it is observed that quantizers using higher order prediction do not achieve a gain over differential quantizers in case the temporal channel evolution is determined by an autoregressive model of order 1. This is evident, because in this case the current channel is obtained by adding i.i.d. noise to the previous channel realization; as the i.i.d. noise cannot be predicted, the best approach is to use the previous quantized observation as prediction. Comparing adaptive prediction and robust prediction, it can be seen that the adaptive filter is able to extract a prediction gain already at higher Doppler frequencies. With a similar or even slightly improved slope, this gain is maintained over the full range of considered Doppler frequencies. The gain is achieved because the adaptive filter adjusts to the temporal statistics of the tangents, in contrast to the tangent regression model employed by [82], which does not exploit any statistical information.

Similar observations are obtained if quantization of higher dimensional subspaces is investigated. In Figure 4.2b, the quantization performance on $\mathcal{G}(8,2)$ is shown. In this case, a CSI feedback overhead of 7 bit and 11 bit is considered. A comparison of differential quantization and predictive quantization is conducted. The predictive quantizers employ the robust prediction algorithm of [90]. Again it is observed that a larger MSE slope is achieved with predictive quantization compared to differential quantization, leading to a significantly reduced quantization error at low Doppler frequencies. The gain obtained by jointly quantizing all subspace dimensions (denoted as matrix quantization) is investigated by comparing to an individual quantization of the subspace dimensions (vector quantization). In case of matrix quantization, a matrix codebook of size $Q_m \in \{64, 1024\}$ is employed, while with vector quantization the codebook size has to be reduced to $Q_v \in \{6, 23\}$ to achieve the same feedback overhead of $\log_2(64) + 1 \approx 2(\log_2(6) + 1) \approx 7$ bit respectively 11 bit. Due to the reduced codebook size, individual quantization of the subspace dimensions is outperformed by joint quantization. With a growing subspace dimension the performance difference is also increased.

In the next simulation, the effect of different channel gains experienced in a DAS on the quantization performance is considered, by assuming a channel gain matrix of the form

$$\mathbf{C}_{u,i} = \operatorname{diag}(1, 1, 1, 1, \gamma_s, \gamma_s, \gamma_s, \gamma_s),$$

with $\gamma_s \in [0, 1]$. Effectively this causes a gradual switching from the quantization on $\mathcal{G}(4,1)$ to $\mathcal{G}(8,1)$. A codebook size of $Q = 16$ is assumed. The performance of memoryless quantization, using the correlated RVQ codebook proposed in Equation (4.27), is compared to the MSE of predictive quantization, employing adaptive filter based prediction. The obtained quantization MSE is shown in Figure 4.3. It can be seen how the quantizers adapt to the varying channel gains. With decreasing γ_s the quantization MSE is improved from the pure 8×1 performance at $\gamma_s = 1$ to the pure 4×1 curve at $\gamma_s = 0$. Also, the performance improvement of predictive quantization over memoryless quantization at low to moderate Doppler frequencies is demonstrated in Figure 4.3.

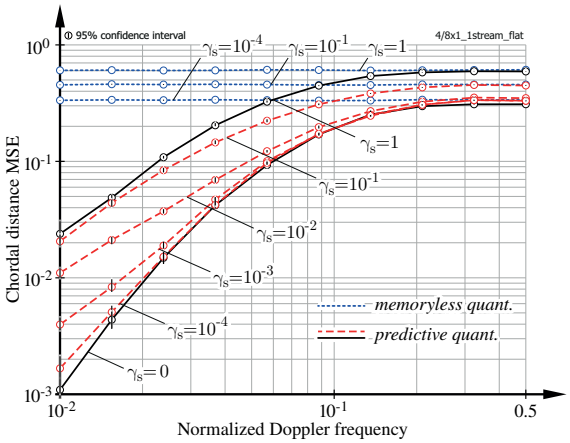


Figure 4.3.: Quantization MSE achieved with memoryless and predictive quantization in a DAS with varying channel gain differences.

4.3. Extension to Systems with Excess Antennas

The performance of ZF beamforming and BD precoding based MU-MIMO with quantized CSIT has been studied thoroughly in the literature, for the case that the number of data streams per user $\ell_{u,i}[k]$ is equal to the number of receive antennas $M_{u,i}$. It is shown in [45] that the CSI feedback overhead of ZF beamforming must be scaled linearly with the logarithmic SNR (the SNR in [dB]) to achieve the same multiplexing gain as in a system with perfect CSIT. A similar bit-scaling law is determined in [46] for BD precoding to multiple users, if the number of data streams per user is equal to the number of receive antennas. In [85], the results of [45] on ZF beamforming are extended to the case that the users are equipped with multiple receive antennas. An efficient antenna combining algorithm denoted as quantization based combining (QBC) is proposed, which exploits the excess antennas to minimize the CSI quantization error. With this strategy a significant reduction of the residual multi-user interference is achieved, implying a reduced slope of the feedback bit-scaling law.

In this section, the QBC algorithm of [85] is extended to multiple data-streams per user via BD precoding, for the case that the number of data streams per user is less than or equal to the number of receive antennas, i.e., $\ell_{u,i}[k] \leq M_{u,i}$. The performance of the obtained subspace quantization based combining (SQBC) algorithm is investigated analytically by deriving the statistics of the Gramian of the effective channel when including the antenna combiner. An upper bound on the rate loss of a BD system employing SQBC with quantized CSIT compared to a BD system with perfect CSIT is derived. It is shown that this bound generalizes the previous results [45, 46, 85]. The corresponding scaling law of the feedback overhead to achieve the same multiplexing gain as a system with perfect CSIT is calculated. Depending on the number of data streams per user, the slope of the bit-scaling law is significantly reduced compared to BD without antenna combining.

To set the basis for the proposed antenna combining algorithm, previous results on the rate loss of ZF

and BD precoding systems with quantized CSIT are summarized in Section 4.3.1. The SQBC strategy, derived and analyzed in Section 4.3.2, is compared to a conventional antenna combining method, namely maximum eigenmode transmission (MET), which considers maximization of the effective channel gain of a user. The performance of MET with quantized CSIT is evaluated in Section 4.3.3 by deriving an upper bound on the rate loss with respect to perfect CSIT. Although MET provides an advantage in case of perfect CSIT, a significant throughput improvement is obtained with SQBC in interference-limited scenarios if a reasonable CSI feedback overhead is considered. The RE index $[n, k]$ is partly omitted for intermediate steps and auxiliary variables to shorten notations.

4.3.1. Summary of Previous Results

Important results that assess the sensitivity of special cases of BD based MU-MIMO systems with respect to the CSI quantization error are summarized below under the assumption of i.i.d. Rayleigh fading channels, i.e., $[\mathbf{H}_{u,i}[n,k]]_{n,m} \sim \mathcal{N}_{\mathbb{C}}(0, \gamma_{u,i})$, with $\gamma_{u,i}$ denoting the macroscopic fading loss. A symmetric system is investigated where all users have the same number of antennas and receive an equal number of data streams $\ell_{u,i} = L,\ \forall u$. The schedule $\mathcal{S}_i[n,k]$ is supposed to be time independent, constantly serving the same set of $S_i = \frac{N_i}{L}$ users in parallel (supposed to be integer-valued). CSI quantization is achieved with an RVQ quantization codebook.

ZF beamforming with $L = M_{u,i} = 1$: The gap between the per-user transmission rate achieved with ZF beamforming based on perfect CSIT (see Equation (4.16)) and the rate of ZF beamforming with quantized CSIT (see Equation (4.25)) employing RVQ is upper-bounded by [45]

$$R_{\text{ZF}} - R_{\text{ZF-Quant}} \leq \log_2\left(1 + \frac{P_i\,\gamma_{u,i}}{\tilde{\sigma}_z^2} D\right),\ D = 2^{-\frac{b}{N_i - 1}}, \tag{4.79}$$

with D being the average distortion achieved with RVQ. To maintain a bounded rate gap, the number of feedback bits b must be scaled linearly with the logarithmic SNR, with a slope that is determined by the number of transmit antennas N_i.

BD precoding with $L = M_{u,i} \geq 1$: The performance of "pure" BD precoding without antenna combining is investigated in [46]. The number of streams L per user is here equal to the number of receive antennas $M_{u,i}$. The per-user rate gap between BD precoding with perfect CSIT and BD precoding with quantized CSIT is upper-bounded by

$$R_{\text{BD}} - R_{\text{BD-Quant}} \leq M_{u,i} \log_2\left(1 + \frac{P_i\,\gamma_{u,i}}{\tilde{\sigma}_z^2\, M_{u,i}} D\right),\ D = C_{\text{BD}}\, 2^{-\frac{b}{M_{u,i}(N_i - M_{u,i})}}. \tag{4.80}$$

As before it is observed that the number of feedback bits must be scaled linearly with the SNR in [dB] to maintain a bounded rate gap, with a slope that depends on both N_i and $M_{u,i}$. The constant C_{BD} is specified in [159]. Setting $L = M_{u,i} = 1$ this bound reduces to the result of [45].

ZF beamforming with $L = 1, M_{u,i} \geq 1$ and receive antenna combining: ZF beamforming with multiple receive antennas is considered in [85]. The QBC antenna combiner proposed in [85] results in an effective channel that can be quantized with minimal quantization error, given the quantization codebook $\mathcal{Q}_{u,i}$. In that way, the expected residual multi-user interference is minimized without requiring knowledge about the interference statistics. On the downside, the gain of the effective channel is reduced when applying QBC, causing a loss in the received power of the intended signal. Still, in interference limited scenarios this power-loss is outweighed by the reduction of the multi-user interference. The rate gap between ZF beamforming with $M_{u,i}^{(\mathrm{ZF})} = 1$ and perfect CSIT and ZF beamforming with $M_{u,i} > 1$ employing QBC and quantized CSIT is bounded by

$$R_{\mathrm{ZF}} - R_{\mathrm{QBC}}^{(M_{u,i})} \leq \log_2\left(1 + \frac{P_i\,\gamma_{u,i}}{\tilde{\sigma}_z^2}\frac{N_i - M_{u,i} + 1}{N_i} D\right) + \log_2(e) \sum_{\ell=N_i-M_{u,i}+1}^{N_i-1} \frac{1}{\ell}, \tag{4.81}$$

with D being proportional to $2^{-\frac{b}{N_i - M_{u,i}}}$ as specified in [85]. Hence, with $M_{u,i} > 1$, the distortion is reduced compared to Equation (4.79). The superscript $^{(M_{u,i})}$ is employed to highlight the important dependency of the performance of QBC on the number of available receive antennas. In contrast to the other bounds considered above, a constant residual rate loss (the second summand) is caused by the application of QBC, which does not depend on the quantization accuracy and thus cannot be reduced by increasing the feedback overhead. This loss is caused by the reduced channel gain of QBC compared to the single receive antenna system. With $M_{u,i} = 1$ the result of [45] is recovered.

4.3.2. Subspace Quantization based Combining

Considering the list of previous work on theoretical performance bounds for ZF and BD MU-MIMO with limited feedback presented in the previous section, it is noticed that the general case of BD precoding with $1 \leq \ell_{u,i}[k] \leq M_{u,i}$ and receive antenna combining is not evaluated. This general case is investigated below by first extending the QBC method to multi-stream transmission per user and then deriving an upper-bound on the rate loss compared to perfect CSIT, generalizing the results of [45, 46, 85] presented above.

SQBC algorithm: The proposed antenna combiner is designed such as to generate an effective channel that can be quantized with minimal subspace quantization error, given the quantization codebook $\mathcal{Q}_{u,i}[n,k]$ and the number of streams $\ell_{u,i}[k]$, that is,

$$\left\{\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k], \hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[n,k]\right\} = \underset{\mathbf{G},\mathbf{Q}_j}{\operatorname{argmin}}\, \mathrm{d_c^2}\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k], \mathbf{Q}_j\right) = \underset{\mathbf{G},\mathbf{Q}_j}{\operatorname{argmin}}\, \mathrm{d_c^2}\left(\mathbf{H}_{u,i}[n,k]\mathbf{G}, \mathbf{Q}_j\right),$$

$$\text{subject to: } \mathbf{G} \in \mathbb{C}^{M_{u,i} \times \ell_{u,i}[k]},\ \mathbf{G}^{\mathrm{H}}\mathbf{G} = \mathbf{I}_{\ell_{u,i}[k]},$$

$$\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k] = \left\{\mathbf{Q}_j \in \mathbb{C}^{N_i \times \ell_{u,i}[k]} | \mathbf{Q}_j^{\mathrm{H}}\mathbf{Q}_j = \mathbf{I}_{\ell_{u,i}[k]}, j \in \{1, \ldots, 2^b\}\right\}.$$

Considering any $\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]$, a decomposition into its range space and left null space components with respect to $\mathbf{H}_{u,i}[n,k]$ can be applied, using an orthonormal basis $\mathbf{B}_{u,i}$ for $\operatorname{span}\left(\mathbf{H}_{u,i}[n,k]\right)$

$$\mathbf{Q}_j = \mathbf{Q}_j^{(\mathrm{R})} + \mathbf{Q}_j^{(\mathrm{N})}, \tag{4.82}$$

$$\mathbf{Q}_j^{(\mathrm{R})} = \mathbf{H}_{u,i}[n,k]\left(\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]\right)^{-1}\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{Q}_j = \mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{Q}_j, \tag{4.83}$$

$$\mathbf{Q}_j^{(\mathrm{N})} = \left(\mathbf{I}_{N_i} - \mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\mathrm{H}}\right)\mathbf{Q}_j, \tag{4.84}$$

$$\operatorname{span}\left(\mathbf{B}_{u,i}\right) = \operatorname{span}\left(\mathbf{H}_{u,i}[n,k]\right),\ \mathbf{B}_{u,i} \in \mathbb{C}^{N_i \times M_{u,i}},\ \mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i} = \mathbf{I}_{M_{u,i}}. \tag{4.85}$$

As $\mathbf{Q}_j^{(\mathrm{R})}$ is in the range space of $\mathbf{H}_{u,i}[n,k]$, it is possible to find an antenna combiner $\mathbf{G}_{u,i}[n,k]$ such that the effective channel spans the same space as $\mathbf{Q}_j^{(\mathrm{R})}$, i.e., $\operatorname{span}\big(\mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}[n,k]\big) = \operatorname{span}\big(\mathbf{Q}_j^{(\mathrm{R})}\big)$. On the other hand, the subspace distance to the component $\mathbf{Q}_j^{(\mathrm{N})}$ in the orthogonal complement of $\operatorname{span}\big(\mathbf{H}_{u,i}[n,k]\big)$ cannot be reduced with antenna combining. Hence, with appropriate antenna combining, the subspace quantization error is only determined by $\mathbf{Q}_j^{(\mathrm{N})}$, and the minimal quantization error is obtained with that $\mathbf{Q}_j$ that is closest to the range space of $\mathbf{H}_{u,i}[n,k]$.

Theorem 4.1 (Subspace Quantization Based Combining)**.** *An $\ell_{u,i}[k] \leq M_{u,i}$ dimensional subspace of a channel matrix $\mathbf{H}_{u,i}[n,k] \in \mathbb{C}^{N_i \times M_{u,i}}$ is to be quantized with a given quantization codebook*

$$\mathcal{Q}_{u,i}[n,k] = \left\{\mathbf{Q}_j \in \mathbb{C}^{N_i \times \ell_{u,i}[k]} | \mathbf{Q}_j^{\mathrm{H}}\mathbf{Q}_j = \mathbf{I}_{\ell_{u,i}[k]}, j \in \{1,\ldots,2^b\}\right\}. \tag{4.86}$$

Applying an antenna combiner to generate an effective channel of dimension $N_i \times \ell_{u,i}[k]$, the minimal quantization error that can be obtained and the corresponding quantized channel subspace are

$$\mathrm{d}_{\mathrm{c,SQBC}}^2[n,k] = \min_{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]} \left\|\mathbf{Q}_j^{(\mathrm{N})}\right\|^2 = \min_{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]} \operatorname{tr}\left((\mathbf{Q}_j^{(\mathrm{N})})^{\mathrm{H}}\mathbf{Q}_j^{(\mathrm{N})}\right)$$

$$= \min_{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]} \ell_{u,i}[k] - \operatorname{tr}\left(\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{Q}_j\mathbf{Q}_j^{\mathrm{H}}\mathbf{B}_{u,i}\right) = \min_{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]} \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}[n,k], \mathbf{Q}_j\right), \tag{4.87}$$

$$\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[n,k] = \operatorname*{argmin}_{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]} \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}[n,k], \mathbf{Q}_j\right), \tag{4.88}$$

with $\operatorname{span}\left(\mathbf{B}_{u,i}\right) = \operatorname{span}\left(\mathbf{H}_{u,i}[n,k]\right),\ \mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i} = \mathbf{I}_{M_{u,i}}$. *The corresponding receive antenna combiner, generating the effective channel that achieves this error, is obtained from the conditions*

$$\operatorname{span}\left(\mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k]\right) \overset{!}{=} \operatorname{span}\left((\mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\mathrm{H}})\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[n,k]\right), \tag{4.89}$$

$$(\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k])^{\mathrm{H}}\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k] \overset{!}{=} \mathbf{I}_{\ell_{u,i}[k]}. \tag{4.90}$$

A formal proof of this theorem is provided in Appendix F.1. Notice that the quantization metric (4.88) is independent of the antenna combiner. Hence, it is not necessary for CSI quantization to calculate the combiner for each $\mathbf{Q}_j$, providing an advantage in terms of computational complexity.

A solution for the antenna combiner obtained from the conditions stated in Theorem 4.1 is unique only up to right-multiplication with any unitary $\ell_{u,i}[k] \times \ell_{u,i}[k]$ matrix. In the following a specific solution is derived that enables further investigations on the channel statistics; an equivalent alternative solution is provided in my publication [92]. The condition (4.89) can be written as

$$\mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k] = (\mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\mathrm{H}})\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[n,k]\tilde{\mathbf{K}}_{u,i}, \tag{4.91}$$

with an appropriate full-rank matrix $\tilde{\mathbf{K}}_{u,i} \in \mathbb{C}^{\ell_{u,i}[k]\times\ell_{u,i}[k]}$. The orthonormal basis $\mathbf{B}_{u,i}$ can be chosen such that the first $\ell_{u,i}[k]$ columns of $\mathbf{B}_{u,i}$ correspond to the range space component of $\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[n,k]$ with respect to $\mathbf{H}_{u,i}[n,k]$ and the remaining $M_{u,i}-\ell_{u,i}[k]$ columns are orthogonal to $\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[n,k]$

$$\mathbf{B}_{u,i}^{\mathrm{H}}\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[n,k] = \mathbf{W}_{u,i}\mathbf{R}_{u,i} = \begin{bmatrix} \mathbf{I}_{\ell_{u,i}} \\ \mathbf{0} \end{bmatrix}\mathbf{R}_{u,i}, \tag{4.92}$$

$$\mathbf{W}_{u,i} \in \mathbb{C}^{M_{u,i}\times\ell_{u,i}[k]},\ \mathbf{R}_{u,i} \in \mathbb{C}^{\ell_{u,i}[k]\times\ell_{u,i}[k]}.$$

The channel $\mathbf{H}_{u,i}[n,k]$ is decomposed in terms of $\mathbf{B}_{u,i}$, resulting in $\mathbf{H}_{u,i}[n,k] = \mathbf{B}_{u,i}\mathbf{D}_{u,i}$ with $\mathbf{D}_{u,i} = \mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}[n,k] \in \mathbb{C}^{M_{u,i}\times M_{u,i}}$. With these decompositions the solution for $\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k]$ is obtained as

$$\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k] = \mathbf{H}_{u,i}[n,k]^{\dagger}\mathbf{B}_{u,i}\mathbf{W}_{u,i}\underbrace{\mathbf{R}_{u,i}\tilde{\mathbf{K}}_{u,i}}_{\mathbf{K}_{u,i}} = \mathbf{D}_{u,i}^{-1}\mathbf{W}_{u,i}\mathbf{K}_{u,i}. \tag{4.93}$$

The undetermined matrix $\mathbf{K}_{u,i}$ of size $\ell_{u,i}[k] \times \ell_{u,i}[k]$ is obtained by invoking condition (4.90)

$$(\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k])^{\mathrm{H}}\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k] = \mathbf{K}_{u,i}^{\mathrm{H}}\left(\mathbf{W}_{u,i}^{\mathrm{H}}\left(\mathbf{D}_{u,i}\mathbf{D}_{u,i}^{\mathrm{H}}\right)^{-1}\mathbf{W}_{u,i}\right)\mathbf{K}_{u,i} \stackrel{!}{=} \mathbf{I}_{\ell_{u,i}[k]}, \tag{4.94}$$

$$\Rightarrow \mathbf{K}_{u,i} = \left(\mathbf{W}_{u,i}^{\mathrm{H}}\left(\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{B}_{u,i}\right)^{-1}\mathbf{W}_{u,i}\right)^{-1/2} \in \mathbb{C}^{\ell_{u,i}[k]\times\ell_{u,i}[k]}. \tag{4.95}$$

With this solution, the effective channel employing SQBC is given by

$$\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] = \mathbf{H}_{u,i}[n,k]\mathbf{G}_{u,i}^{(\mathrm{SQBC})}[n,k] = \mathbf{B}_{u,i}\mathbf{W}_{u,i}\mathbf{K}_{u,i} = \tilde{\mathbf{H}}_{u,i}[n,k]\mathbf{K}_{u,i}. \tag{4.96}$$

Channel statistics: To derive statements about the statistics of the effective channel obtained with SQBC, and to develop the throughput bound of the rate loss incurred with quantized CSIT compared to perfect CSIT, it is necessary to impose additional assumptions on the channel $\mathbf{H}_{u,i}[n,k]$ and the quantization codebook $\mathcal{Q}_{u,i}[n,k]$. Specifically, in the following two paragraphs the channel matrix is assumed as i.i.d. Rayleigh fading, i.e., $[\mathbf{H}_{u,i}[n,k]]_{n,m} \sim \mathcal{N}_{\mathbb{C}}(0,\gamma_{u,i})$, and RVQ is employed for channel subspace quantization. Also, the number of streams per user is assumed as time independent $\ell_{u,i}[k] = \ell_{u,i}$.

Lemma 4.1. *The subspaces spanned by the effective channels* $\mathbf{H}_{u,i}^{\text{eff}}[n,k], \forall u$ *are statistically independent and isotropically (uniformly) distributed on* $\mathcal{G}\left(N_i, \ell_{u,i}\right)$.

Proof. From (4.89) it is known that the subspace spanned by $\mathbf{H}_{u,i}^{\text{eff}}[n,k]$ is determined by the projection of the best quantization matrix $\hat{\mathbf{H}}_{u,i}[n,k]$ onto $\text{span}\left(\mathbf{H}_{u,i}[n,k]\right)$. Since the quantization matrices are isotropically distributed on $\mathcal{G}\left(N_i, \ell_{u,i}\right)$, their projections onto $\text{span}\left(\mathbf{H}_{u,i}[n,k]\right)$ are isotropically distributed within this subspace. This holds also true for the best quantization matrix, since it is chosen based solely on the Frobenius norm of the null space component. Furthermore, the subspace $\text{span}\left(\mathbf{H}_{u,i}[n,k]\right)$ itself is isotropically distributed on $\mathcal{G}\left(N_i, M_{u,i}\right)$, since we assume i.i.d. Rayleigh fading. Therefore, $\text{span}\left(\mathbf{H}_{u,i}^{\text{eff}}[n,k]\right)$ is isotropically distributed on $\mathcal{G}\left(N_i, \ell_{u,i}\right)$. Finally, the quantization codebooks and channels of different users are statistically independent, implying statistical independence of $\mathbf{H}_{1,i}^{\text{eff}}[n,k], \ldots, \mathbf{H}_{S_i,i}^{\text{eff}}[n,k]$. □

Lemma 4.2. *The Gramian of the effective channel* $(\mathbf{H}_{u,i}^{\text{eff}}[n,k])^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}[n,k]$ *is complex Wishart distributed of dimension* $\ell_{u,i}$*, with* $N_i - M_{u,i} + \ell_{u,i}$ *degrees of freedom and identity scale matrix*

$$(\mathbf{H}_{u,i}^{\text{eff}}[n,k])^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}[n,k] \sim \mathcal{W}_{\ell_{u,i}}^{\mathbb{C}}\left(N_i - M_{u,i} + \ell_{u,i}, \gamma_{u,i}\,\mathbf{I}_{\ell_{u,i}}\right). \tag{4.97}$$

The proof of this lemma is provided in Appendix F.2.

Throughput analysis: Utilizing the results of the previous paragraph, it is possible to derive an upper bound on the rate loss of BD precoding with quantized CSIT and excess receive antennas $M_{u,i} \geq \ell_{u,i}$, with respect to BD precoding with perfect CSIT but having no excess antennas at the receivers, i.e., $M_{u,i}^{(\text{BD})} = \ell_{u,i}$. In case of excess antennas, it is assumed that the users employ SQBC antenna combining to determine the channel subspace feedback according to the quantization metric in Theorem 4.1. The quantized CSIT is used by the base station to calculate the precoders from (4.4) and (4.22), respectively. The same symmetric scenario as in Section 4.3.1 is considered, i.e., all S_i users are equipped with the same number of $M_{u,i}$ receive antennas and are served over the same number of $\ell_{u,i} = L$ data streams. An RVQ quantization codebook is employed for quantization of the effective channel subspace.

According to (4.16), the achievable user rate of BD with perfect CSIT and no excess antennas is

$$R_{\text{BD}}^{(L)} = \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,\mathbf{H}_{u,i}[n,k]^{\text{H}}\tilde{\mathbf{F}}_{u,i}[n,k]\tilde{\mathbf{F}}_{u,i}[n,k]^{\text{H}}\mathbf{H}_{u,i}[n,k]\right),\ \rho = \frac{P_i}{\tilde{\sigma}_z^2\,S_i\,L}, \tag{4.98}$$

with $\mathbf{H}_{u,i}[n,k] = \mathbf{H}_{u,i}^{\text{eff}}[n,k]$ because $M_{u,i}^{(\text{BD})} = L$. The expected value is calculated with respect to the channel and the precoder. Similarly, with quantized CSIT, $L \leq M_{u,i}$ data streams and application

of SQBC antenna combining, the achievable rate is obtained as

$$R_{\text{SQBC}}^{(L,M_{u,i})} = \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\sum_{s=1}^{S_i}\mathbf{H}_{u,i}^{\text{eff}}[n,k]^{\text{H}}\tilde{\mathbf{F}}_{s,i}[n,k]\tilde{\mathbf{F}}_{s,i}[n,k]^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}[n,k]\right)$$

$$-\mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\sum_{s=1,s\neq u}^{S_i}\mathbf{H}_{u,i}^{\text{eff}}[n,k]^{\text{H}}\tilde{\mathbf{F}}_{s,i}[n,k]\tilde{\mathbf{F}}_{s,i}[n,k]^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}[n,k]\right),\ \rho = \frac{P_i}{\tilde{\sigma}_z^2\, S_i\, L}, \quad (4.99)$$

with $\mathbf{H}_{u,i}^{\text{eff}}[n,k]$ being determined by Equation (4.96). Here, the expected value is additionally calculated over quantization codebook realizations. Similar to the bounds proposed in [45, 46, 85], i.e., Equations (4.79) to (4.81), the throughput loss $R_{\text{BD}}^{(L)} - R_{\text{SQBC}}^{(L,M_{u,i})}$ can be upper bounded:

Theorem 4.2 (SQBC rate loss)**.** *Consider a broadcast system with N_i transmit antennas, $M_{u,i}$ receive antennas per user and transmit power P_i. The system serves S_i users with $L \leq M_{u,i}$ streams each, over i.i.d. Rayleigh fading channels with additive Gaussian receiver noise of variance $\tilde{\sigma}_z^2$. The per-user throughput loss of BD precoding employing SQBC with quantized CSIT compared to BD precoding with perfect CSIT, but having only $M_{u,i}^{(\text{BD})} = L$ receive antennas, is upper bounded by*

$$R_{\text{BD}}^{(L)} - R_{\text{SQBC}}^{(L,M_{u,i})} \leq L\log_2\left(1 + \rho\,\gamma_{u,i}\,\frac{N_i - M_{u,i} + L}{N_i - L}\,(S_i - 1)\,D\right)$$

$$+\log_2(e)\sum_{k=0}^{L-1}\sum_{\ell=N_i-M_{u,i}+L}^{N_i-1}\frac{1}{\ell-k},\ \ \rho = \frac{P_i}{\tilde{\sigma}_z^2\, S_i\, L}. \quad (4.100)$$

Here, the average quantization distortion in terms of subspace chordal distance achieved with RVQ is denoted D and the macroscopic pathloss is considered in $\gamma_{u,i}$.

The proof of this theorem is provided in Appendix F.3. As can be seen from (4.87), the relevant distortion D for SQBC is the distortion achieved when quantizing subspaces from $\mathcal{G}(N_i, M_{u,i})$ using a quantization codebook with entries from $\mathcal{G}(N_i, L)$. This quantization problem is considered in [159]. It is shown in [159], that the average distortion with random isotropically distributed quantization codebooks, i.e., RVQ, is obtained as

$$D = \mathbb{E}\left(\text{d}_{\text{c,SQBC}}^2\right) \approx C_{\text{SQBC}}\, 2^{-\frac{b}{L(N_i-M_{u,i})}}, \quad (4.101)$$

with C_{SQBC} being a function of L, N_i and $M_{u,i}$ as specified in [159, Eqs. (8) and (11)][3], and 2^b being the size of the codebook. By setting appropriate values for L, N_i and $M_{u,i}$ and considering the case $S_i = \frac{N_i}{L}$ it can be verified that (4.79), (4.80) and (4.81) are contained in (4.100).

[3] For random quantization the upper bound of [159, Eq. (11)] is relevant; the $\mathcal{O}(1)$ term is omitted.

Discussion of the bound: The upper bound on the rate loss in (4.100) is composed of two summands capturing distinct effects of the SQBC antenna combiner. The second summand, which is abbreviated by Δa, is independent of the quantization accuracy D. From the argumentation in the proof in Appendix F.3, it follows that Δa gives a tight upper bound on the rate loss $R_{\text{BD}}^{(L)} - R_{\text{SQBC}}^{(L,M_{u,i})}$ for the case of perfect CSIT ($b \to \infty, D \to 0$) and $\rho\,\gamma_{u,i} \to \infty$, i.e., at high SINR. This high SINR rate loss is caused by the reduction of the effective channel gain due to the antenna combiner. Specifically, for BD with $M_{u,i}^{(\text{BD})} = L$, the term $\mathbf{H}_{u,i}^{\text{eff}}[n,k]^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}[n,k]$ is distributed according to $\mathcal{W}_L^{\mathbb{C}}\left(N_i, \gamma_{u,i}\,\mathbf{I}_L\right)$, while it is shown in Lemma 4.2 that it is distributed as $\mathcal{W}_L^{\mathbb{C}}\left(N_i - M_{u,i} + L, \gamma_{u,i}\,\mathbf{I}_L\right)$ for SQBC. This loss in DoF of the Wishart distribution is the cause for the constant rate offset Δa. Hence, SQBC is disadvantageous if the CSI quantization accuracy is already very high. In this case, it is better to invest the available DoF provided by the $M_{u,i} \geq L$ antennas to obtain a large effective channel gain, instead of further reducing the quantization error.

The more important insights for limited feedback systems are captured in the first summand of (4.100) (abbreviated by Δb). If b is fixed, this term grows to infinity as the SNR increases, meaning that the system becomes interference limited. If b is scaled with the SNR, however, the rate gap can be kept constant and a multiplexing gain of N_i can be achieved. Assuming $S_i = \frac{N_i}{L}$ the necessary feedback bit scaling is determined by setting Δb equal to a constant rate loss ΔR in [bits/s/Hz]

$$b = L\left(N_i - M_{u,i}\right)\left(\log_2\left(\frac{P_i}{\tilde{\sigma}_z^2}\,\frac{\gamma_{u,i}}{N_i}\right) + \log_2\left(\frac{N_i - M_{u,i} + L}{L}\,C_{\text{SQBC}}\right) - \log_2\left(2^{\frac{\Delta R}{L}} - 1\right)\right). \tag{4.102}$$

The important insight is the growth-rate of the number of feedback bits with the SNR β_{dB} in [dB]

$$\frac{db}{d\beta_{\text{dB}}} = L\left(N_i - M_{u,i}\right)\frac{\log_2\left(10\right)}{10} \approx \frac{L\left(N_i - M_{u,i}\right)}{3}, \quad \beta_{\text{dB}} = 10\log_{10}\left(\frac{P_i}{\tilde{\sigma}_z^2}\right).$$

It can be seen that the slope of the feedback overhead with SNR grows linearly with the number of streams L, and reduces with the number of receive antennas $M_{u,i}$. Hence, having more receive antennas with a fixed L the CSI feedback overhead can be decreased. Still, this does not give the full picture, because varying L and/or $M_{u,i}$ also impacts the absolute achievable throughput.

To investigate this influence, the high SINR sum rate difference between two SQBC systems with the same N_i and L, but different $M_{u,i} \in \{M_1, M_2\}$ is determined. This rate difference is obtained from (4.100) by considering the bound on $R_{\text{BD}}^{(L)} - R_{\text{SQBC}}^{(L,M_2)} - \left(R_{\text{BD}}^{(L)} - R_{\text{SQBC}}^{(L,M_1)}\right)$ at high SINR

$$\Delta R(M_1, M_2) = \frac{N_i}{L}\left(R_{\text{SQBC}}^{(L,M_1)} - R_{\text{SQBC}}^{(L,M_2)}\right) = \frac{N_i}{L}\log_2\left(e\right)\sum_{k=0}^{L-1}\sum_{\ell=N_i-M_2+L}^{N_i-M_1+L-1}\frac{1}{\ell-k}, \tag{4.103}$$

which results in a positive rate loss if $M_2 > M_1$. Thus, in contrast to most conventional antenna combining strategies, if SQBC is employed with perfect CSIT, a rate reduction is incurred at high SINR with growing number of receive antennas $M_{u,i}$. Similarly, it can be shown that the sum rate of SQBC is improved if the number of streams per user is increased and the number of users is correspondingly decreased; see [93] for details.

4.3.3. Maximum Eigenmode Transmission

In this section, the performance of an alternative interference-unaware receive antenna combining algorithm, namely MET, is investigated. With MET the effective channel generated by a user is composed of the L maximum eigenmodes of the channel matrix. Hence, with perfect CSIT the L-dimensional dominant subspace of each users' channel is kept free of interference, providing a potentially large channel gain. On the other hand, the CSI quantization error achieved with MET is significantly larger than with SQBC, and thus the residual multi-user interference has a much stronger impact on the performance of MET. The main reason for considering MET as an alternative strategy to SQBC is that it enables an instructive investigation on the trade off between investing the provided DoF of having $M_{u,i} \geq L$ excess antennas to maximizing the signal power by means of MET, in contrast to minimizing the CSI quantization error (and hence the expected interference power) utilizing SQBC. Also, the complexity and CSI feedback requirements (subspace information) of MET are very similar to those of SQBC. In [172], MET is combined with a coordinated eigenmode selection by the base station, such that users with close to orthogonal channels are served in parallel. In the following investigation, scheduling is not explicitly considered.

MET algorithm: The goal of MET is to generate an L-dimensional effective channel that maximizes the achievable transmission rate of a user in the absence of multi-user interference, by applying a semi-unitary antenna combiner. This is achieved, if the transmission to the user takes place over the L-dimensional dominant subspace of the channel matrix $\mathbf{H}_{u,i}[n,k]$. Consider an SVD of the channel matrix $\mathbf{H}_{u,i}[n,k]$ in compact form

$$\mathbf{H}_{u,i}[n,k] = \mathbf{U}_{u,i}[n,k]\mathbf{\Sigma}_{u,i}[n,k]\mathbf{V}_{u,i}[n,k]^{\mathrm{H}} \tag{4.104}$$

$$\mathbf{U}_{u,i}[n,k] \in \mathbb{C}^{N_i \times M_{u,i}},\ \mathbf{\Sigma}_{u,i}[n,k] \in \mathbb{C}^{M_{u,i} \times M_{u,i}}, \mathbf{V}_{u,i}[n,k] \in \mathbb{C}^{M_{u,i} \times M_{u,i}}.$$

Notice that $\mathbf{U}_{u,i}[n,k]$, $\mathbf{\Sigma}_{u,i}[n,k]$ and $\mathbf{V}_{u,i}[n,k]$ are statistically independent, and $\mathbf{U}_{u,i}[n,k]$ and $\mathbf{V}_{u,i}[n,k]$ are isotropic for i.i.d. Rayleigh fading. The channel subspace to be quantized is chosen as the first L columns of $\mathbf{U}_{u,i}[n,k]$, corresponding to the L largest singular values of $\mathbf{H}_{u,i}[n,k]$. The quantized subspace is obtained as

$$\hat{\mathbf{H}}_{u,i}^{(\mathrm{MET})}[n,k] = \underset{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[n,k]}{\operatorname{argmin}} \mathrm{d}_{\mathrm{c}}^2 \left([\mathbf{U}_{u,i}[n,k]]_{:,1:L}, \mathbf{Q}_j \right). \tag{4.105}$$

As with SQBC, the quantized channel subspace can be obtained without having to calculate the antenna combiner for each $\mathbf{Q}_j$. The codebook index of the quantized channel subspace is fed back to the base station by the users, and the BD precoder is calculated from quantized CSIT. With the provided subspace feedback power loading over the eigenmodes is not reasonable, because no information about the magnitude of the singular values is available at the base station. The

corresponding MET antenna combiner and the effective channel are given by

$$\mathbf{G}_{u,i}^{(\mathrm{MET})}[n,k] = [\mathbf{V}_{u,i}[n,k]]_{:,1:L}\,, \tag{4.106}$$

$$\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k] = [\mathbf{U}_{u,i}[n,k]]_{:,1:L}\,[\mathbf{\Sigma}_{u,i}[n,k]]_{1:L,1:L} = \mathbf{U}_{u,i}^{(L)}[n,k]\mathbf{\Sigma}_{u,i}^{(L)}[n,k] = \mathbf{H}_{u,i}^{(L)}[n,k]. \tag{4.107}$$

Throughput analysis: The performance of BD precoding with MET antenna combining and quantized CSIT can be evaluated in a similar way as the performance of SQBC under the same assumptions (symmetric scenario, i.i.d. Rayleigh fading, RVQ), by deriving an upper bound on the rate loss with respect to perfect CSIT. With the effective channel from Equation (4.107), the achievable per-user rate with perfect CSIT according to Equation (4.16) is

$$R_{\mathrm{MET}} = \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,\mathbf{H}_{u,i}^{(L)}[n,k]^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}[n,k]\tilde{\mathbf{F}}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}^{(L)}[n,k]\right). \tag{4.108}$$

Similarly to (4.99), with quantized CSIT the achievable per-user rate is obtained as

$$R_{\text{MET-Quant}} = \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\sum_{s=1}^{S_i}\mathbf{H}_{u,i}^{(L)}[n,k]^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}[n,k]\tilde{\mathbf{F}}_{s,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}^{(L)}[n,k]\right)$$

$$-\mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\sum_{s=1,s\neq u}^{S_i}\mathbf{H}_{u,i}^{(L)}[n,k]^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}[n,k]\tilde{\mathbf{F}}_{s,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}^{(L)}[n,k]\right). \tag{4.109}$$

Theorem 4.3 (MET rate loss)**.** *Consider a multi-user broadcast system with N_i transmit antennas, $M_{u,i}$ receive antennas per user and transmit power P_i. The system serves S_i users with $L \leq M_{u,i}$ spatial streams each, over i.i.d. Rayleigh fading channels with additive Gaussian receiver noise of variance $\tilde{\sigma}_z^2$. The per-user throughput loss of BD precoding employing MET antenna combining with quantized CSIT compared to perfect CSIT is upper bounded by*

$$R_{\mathrm{MET}} - R_{\text{MET-Quant}} \leq \sum_{\ell=1}^{L}\log_2\left(1 + \rho\,\bar{\sigma}_{\ell,u,i}^2\,\frac{S_i - 1}{N_i - L}\,D\right),\quad \rho = \frac{P_i}{\tilde{\sigma}_z^2\,S_i\,L}. \tag{4.110}$$

Here, the average quantization distortion in terms of subspace chordal distance achieved with RVQ is denoted D. The set $\left\{\bar{\sigma}_{1,u,i}^2\ldots,\bar{\sigma}_{L,u,i}^2\right\}$ is composed of the expected values of the L largest squared singular values of the channel matrix $\mathbf{H}_{u,i}[n,k]$. Hence, $\bar{\sigma}_{\ell,u,i}^2$ is equal to the expected value of the ℓ-th largest eigenvalue of $\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k] \sim \mathcal{W}_{M_{u,i}}^{\mathbb{C}}\left(N_i, \gamma_{u,i}\,\mathbf{I}_{M_{u,i}}\right)$, with $\gamma_{u,i}$ denoting the macroscopic pathloss.

The proof of this theorem is provided in Appendix F.4. Notice that the macroscopic pathloss $\gamma_{u,i}$ does not appear explicitly in (4.110) but is captured in the squared singular values. Closed form expressions for the expected eigenvalues of $\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{H}_{u,i}[n,k]$ are known [173], but the

expressions are involved and do not provide further analytical insights. In contrast to the distortion D in (4.101) achieved with SQBC, the average distortion D with MET is determined by quantizing an isotropically distributed subspace from $\mathcal{G}(N_i, L)$ with a random codebook of isotropically distributed entries from $\mathcal{G}(N_i, L)$ [159]

$$D \approx C_{\text{MET}}\, 2^{-\frac{b}{L(N_i-L)}}. \tag{4.111}$$

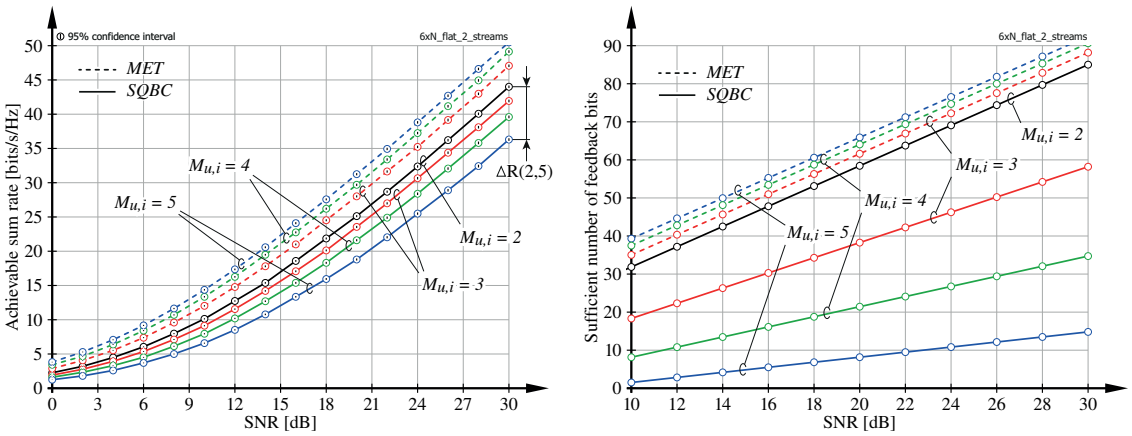
It is hard to derive a closed form solution for b in dependency of the rate loss as in (4.102), but the required number of bits can be evaluated numerically. The important fact to note is that with MET the exponent of the distortion scales inversely proportional to $L(N_i - L)$ in contrast to $L(N_i - M_{u,i})$, which is achieved with SQBC. Hence, with SQBC having $M_{u,i} > L$ leads to a reduction of the average quantization error, which is not the case with MET.

4.3.4. Achievable Rate Comparison of SQBC and MET

Whether SQBC or MET is considered as the preferred receive antenna combining strategy in combination with BD precoding is determined by the available CSI feedback resources. This is investigated and explained in more detail in this section. For that purpose, the achievable transmission rates of the two strategies in the limiting case of perfect CSIT, i.e., $b \to \infty$ is evaluated. The rates are calculated from the corresponding achievable rate equations (4.98), (4.99) by estimating the expected value by means of Monte-Carlo simulations. The achievable rate is calculated for a single OFDM subcarrier, assuming frequency flat Rayleigh fading with $[\mathbf{H}_{u,i}[n,k]]_{l,m} \sim \mathcal{N}_{\mathbb{C}}(0,1)$. The result is shown in Figure 4.4a for the case $N_i = 6$, $L = 2$ and $M_{u,i} \in \{2,3,4,5\}$. As scheduling is not considered in this investigation, it is supposed that the set of scheduled users $\mathcal{S}_i$ is composed of all $U_i = 3$ users in the cell (hence also $S_i = 3$).

It is observed in Figure 4.4a that the achievable sum rate of SQBC decreases with increasing number of receive antennas, while the throughput of MET improves. This is in conformance with our theoretical investigation of Section 4.3.2. More specifically, the throughput loss of SQBC at high SNR is given by the value $\Delta R(M_1, M_2)$ calculated in (4.103), e.g., $\Delta R(2,5) = 8.08$ bits/s/Hz according to (4.103). Hence, with perfect CSIT, SQBC is not a reasonable choice. This behavior can be explained by considering the channel statistics provided in Lemma 4.2. According to this lemma, the DoF of the Wishart distribution defining the statistics of $\mathbf{H}_{u,i}^{\text{eff}}[n,k]^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}[n,k]$ are reduced with a growing number of receive antennas $M_{u,i}$. Correspondingly, the eigenvalues of this Wishart matrix are decreased on average, causing a reduction of the average effective channel gain. On the other hand, with MET the channel gain is determined by the L maximum eigenvalues of a matrix with Wishart distribution $\mathcal{W}_{M_{u,i}}^{\mathbb{C}}\left(N_i, \mathbf{I}_{M_{u,i}}\right)$ according to Theorem 4.3. It is well known that these eigenvalues increase on average with growing dimension $M_{u,i}$.

The advantage of SQBC is visualized in Figure 4.4b. In this figure the sufficient number of CSI feedback bits to achieve a per-user rate loss of 1 bits/s/Hz compared to perfect CSIT is shown. With



(a) Achievable sum rate with perfect CSIT.

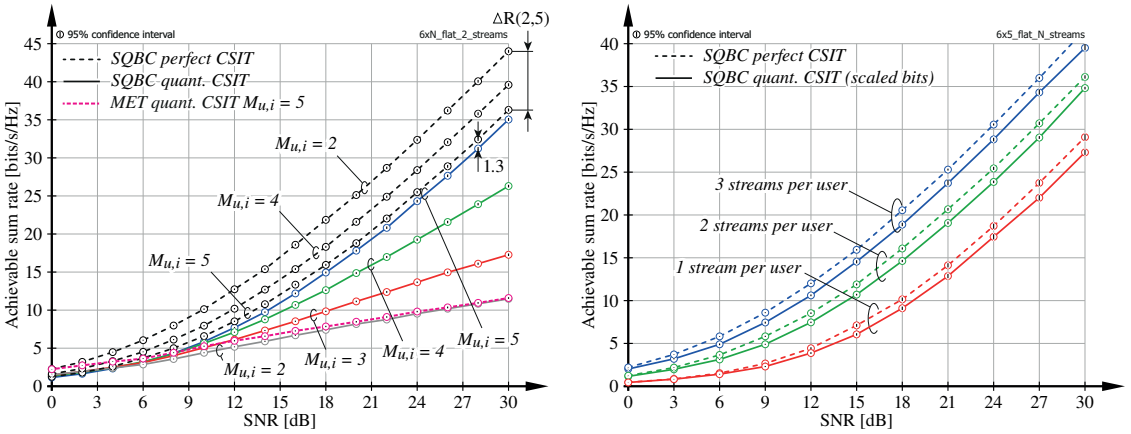
(b) Feedback bit-scaling for a per-user rate loss of 1 bits/s/Hz.

Figure 4.4.: Evaluation of the achievable rate performance and CSI feedback overhead requirement of BD precoding with SQBC and MET antenna combining. Three users are served over six transmit antennas via two streams per user.

MET, the slope of the feedback overhead versus the SNR for a given number of data streams L per user is independent of the number of receive antennas $M_{u,i}$ and the required feedback overhead is significantly larger than in case of SQBC. With SQBC, the feedback overhead can be substantially reduced by increasing the number of receive antennas $M_{u,i}$, for the cost of a moderate SNR offset in the achievable throughput. This reduction is due to the decreasing exponent in the average distortion (4.101) with increasing $M_{u,i}$, obtained from the degree of freedom to select the best L-dimensional subspace within an $M_{u,i}$-dimensional space during CSI quantization.

In Figure 4.5, the validity of the bit-scaling law derived in (4.102) is investigated. In the results shown in Figure 4.5a the base station is equipped with $N_i = 6$ transmit antennas. $S_i = 3$ users are served in parallel over $L = 2$ streams each. The CSI feedback overhead is scaled such as to achieve a sum rate loss of 1.5 bits/s/Hz for the case of SQBC with $M_{u,i} = 5$. At an SNR of $\beta_{\text{dB}} = 0\,\text{dB}$, a feedback overhead of $b = 0$ bits is required to maintain the intended rate loss, which is increased to $b = 17$ bits at $\beta_{\text{dB}} = 30\,\text{dB}$. The other configurations considered in Figure 4.5a are simulated with the same number of feedback bits. It is observed that the actually achieved rate loss is equal to 1.3 bits/s/Hz, which is close to the desired value. With the same feedback overhead a significant reduction in throughput is incurred when $M_{u,i}$ is decreased. This reduction is caused by the increased residual multi-user interference due to the CSI quantization error. If MET is employed with $M_{u,i} = 5$ and the same feedback overhead, a much worse performance compared to SQBC is observed if $\beta_{\text{dB}} > 10\,\text{dB}$. At low SNR, however, SQBC is outperformed by MET, because the transmission rate is limited by the noise rather than the multi-user interference. In this case, the interference reduction capabilities of SQBC are outweighed by the channel gain improvement of MET.

In Figure 4.5b, the performance of SQBC with $N_i = 6$, $M_{u,i} = 5$ and $L \in \{1, 2, 3\}$ is shown. The



(a) $L = 2$, $S_i = U_i = 3$ and $M_{u,i} \in \{2, \ldots, 5\}$.

(b) $M_{u,i} = 5$, $L \in \{1, 2, 3\}$ and $S_i = U_i = \frac{N_i}{L}$.

Figure 4.5.: Achievable sum rate of BD-based MU-MIMO systems employing SQBC with quantized and perfect CSIT and $N_i = 6$.

feedback overhead is scaled such as to achieve a sum rate loss of 2 bits/s/Hz with respect to perfect CSIT. This is obtained with an overhead of $b \in [0, 8]$ bits/user for $L = 1$, $b \in [0, 16.1]$ bits/user for $L = 2$ and $b \in [0, 18.3]$ bits/user for $L = 3$. With decreasing number of streams L per user a sum rate reduction is observed. This is because the interference between a larger number of users must be canceled by the BD precoder when the number of streams per user is reduced and the total number of streams is kept constant, leading to an SNR loss of the effective user channel [71].

Further simulation results investigating the tightness of the proposed bounds on the rate loss are provided in [93]. A trade-off between SQBC and MET can be achieved by restricting the SQBC algorithm to finding the best subspace within only a subset of the eigenmodes of the channel. This is explained in Section 4.4 when extending the method to frequency selective channels.

4.3.5. Adjustment of the Grassmannian CSI Feedback

In Section 4.2, memoryless and predictive Grassmannian quantization codebooks are proposed under the assumption that the number of data streams per user $\ell_{u,i}[k]$ is equal to the number of receive antennas $M_{u,i}$, rendering the application of an antenna combiner obsolete. In this case, the quantization metric is given by the chordal distance between the $M_{u,i}$-dimensional subspace defined by the channel matrix $\mathbf{H}_{u,i}[n, k]$ and the $M_{u,i}$-dimensional subspaces spanned by the elements $\mathbf{Q}_j$ of the codebook $\mathcal{Q}_{u,i}[n, k]$; see Equation (4.19).

With the antenna combiners of the previous section the quantization metric is modified, which can be considered in correlated RVQ to improve the quantizer efficiency. Specifically, with SQBC the chordal distance between the $M_{u,i}$-dimensional channel subspace and each of the $\ell_{u,i}[k]$-dimensional

subspaces defined by the elements of the codebook is minimized according to Equation (4.88). The minimum is achieved with that element of the codebook, that has the smallest left null space component with respect to the channel matrix. As the null space component of the codebook elements is the decisive factor during quantization, there exists no preferred subspace orientation within the range space of the channel matrix. Hence, efficient memoryless quantization is enabled if the quantization codebook is matched to the distribution of any $\ell_{u,i}[k]$-dimensional subspace of the channel matrix. With correlated RVQ, the quantization codebook can be generated according to

$$\mathcal{Q}_{u,i}^{(\mathrm{SQBC})}[n,k] = \Big\{ \mathbf{Q}_j^{(\mathrm{SQBC})} = \mathbf{Q}_j^{(\mathrm{corr})}\mathbf{U} | \mathbf{Q}_j^{(\mathrm{corr})} \in \mathcal{Q}_{u,i}^{(\mathrm{corr})}[n,k], \\ \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}} = \bar{\mathbf{H}} \in \mathbb{C}^{M_{u,i}\times \ell_{u,i}[k]},\ [\bar{\mathbf{H}}]_{m,n} \sim \mathcal{N}_{\mathbb{C}}(0,1) \Big\}. \tag{4.112}$$

Here, the codebook defined in (4.27) is employed to obtain semi-unitary matrices $\mathbf{Q}_j^{(\mathrm{corr})}$ of size $N_i \times M_{u,i}$ whose span is matched in distribution to the subspace $\mathrm{span}\left(\mathbf{H}_{u,i}[n,k]\right)$. The multiplication with the isotropic semi-unitary matrix $\mathbf{U} \in \mathbb{C}^{M_{u,i}\times \ell_{u,i}[k]}$ then generates an orthonormal basis $\mathbf{Q}_j^{(\mathrm{SQBC})}$ that spans a uniformly distributed subspace within $\mathrm{span}(\mathbf{Q}_j^{(\mathrm{corr})})$.

With MET, the quantization metric (4.105) is determined by the chordal distance between the subspace spanned by the $\ell_{u,i}[k]$ maximum eigenmodes of the channel matrix and the elements of the codebook. Efficient memoryless quantization is hence enabled by matching the quantization codebook to the distribution of these eigenmodes. This can, e.g., be achieved by modifying the codebook construction in Equation (4.27) to select only the first $\ell_{u,i}[k]$ columns of the matrix of left singular vectors

$$\mathcal{Q}_{u,i}^{(\mathrm{MET})}[n,k] = \left\{ \mathbf{Q}_j^{(\mathrm{MET})} = \left[\mathbf{Q}_j^{(\mathrm{corr})}\right]_{:,1:\ell_{u,i}[k]} \middle| \mathbf{Q}_j^{(\mathrm{corr})} \in \mathcal{Q}_{u,i}^{(\mathrm{corr})}[n,k] \right\}. \tag{4.113}$$

In case of predictive quantization, deriving the tangent statistics when MET or SQBC is applied was not successful. The tangent codebook is therefore generated assuming the channel subspace to be uniformly distributed on $\mathcal{G}\left(N_i, \ell_{u,i}[k]\right)$. This is not a critical issue, because the predictive quantizer automatically adjusts to the statistics of the subspace spanned by the effective channel using the adaptive codebook construction described in Section 4.2.2.

4.4. Extension to Frequency-Selective Systems

The CSI feedback algorithms and channel subspace selection methods proposed in Sections 4.2 and 4.3 are derived for frequency flat channels and are hence applicable to OFDM on a per-subcarrier basis. Providing CSI feedback for each RE, however, implies a large feedback overhead, which cannot be sustained in practical systems. The CSI feedback overhead can be reduced by employing the same approaches as in SU-MIMO, i.e., CSI feedback clustering and interpolation. CSI feedback

interpolation for ZF and BD based MU-MIMO is considered, e.g., in [86, 89, 174, 175]. Linear subspace interpolation on the Grassmannian is possible by sampling the geodesic between neighboring CSI pilots. Geodesic interpolation is reviewed in Section 4.4.1 for completeness. Higher order spline interpolation on the unit-sphere has been proposed in [176] in the context of graphics and animations. These algorithms are applicable to CSI interpolation for ZF beamforming, but suffer from high computational complexity. An issue of interpolation based methods is the high density of CSI pilots required to achieve sufficiently accurate results; see Section 4.4.3. If this cannot be sustained due to limitations on the feedback overhead, better performance is possible with suitable feedback clustering. To this end, a clustering approach is proposed in Section 4.4.1, in which a single representative subspace is determined for each resource block (RB) such that the average chordal distance is minimized. This method can be viewed as a natural extension of the SQBC idea to multiple REs, providing a significant performance improvement especially in systems with excess antennas. With the SQBC clustering approach, it is possible to exploit the time-frequency channel correlation within each RB separately. The residual interference in-between RBs, on the other hand, is utilized by the predictive quantizer proposed in Section 4.2.2. Combining the two methods, efficient CSI quantization is achieved, as demonstrated in Section 4.4.3.

In Section 4.4.2, channel quality feedback for transmission rate adaptation and multi-user scheduling in the space, time and frequency domain is considered. A combination of the semi-orthogonal user selection (SUS) algorithm [58] with proportional fair scheduling [56] is employed to determine the multi-user resource allocation. The scheduling is based on a proposed estimate of the achievable user data rate with BD precoding. The performance of the proposed techniques is evaluated by means of simulations in Section 4.4.3. Notice that the notation introduced for feedback clustering in Section 3.2.1 is reused in this section.

4.4.1. Grassmannian Interpolation and Clustering

Geodesic interpolation: Interpolation on the Grassmannian can be achieved by exploiting the differential geometry associated with the manifold. The concept of a straight line in the Euclidean space is generalized to curved spaces and manifolds with the geodesic, which is introduced in Appendix C.3. Linear interpolation on the manifold is possible by equidistantly sampling the geodesic between neighboring CSI pilots.

When CSI interpolation is employed, it is assumed that the OFDM time-frequency resource grid is partitioned into RBs as visualized in Figure 3.3. With the notation introduced in Section 3.2.1, the index pair $[\eta, \kappa]$ is employed to indicate the RBs. The channel matrix observed on the RE in the center of the RB is quantized and fed back by the users, employing the Grassmannian quantizers

of Section 4.2. The corresponding subcarrier index and symbol-time index are denoted as

$$n_\eta = (\eta - 1)N_{\text{clust}}^{(f)} + \left\lfloor \frac{N_{\text{clust}}^{(f)}}{2} \right\rfloor, \quad k_\kappa = (\kappa - 1)N_{\text{clust}}^{(t)} + \left\lceil \frac{N_{\text{clust}}^{(t)}}{2} \right\rceil. \tag{4.114}$$

The quantized subspace on RB $[\eta, \kappa]$ is thus obtained as

$$\hat{\mathbf{H}}_{u,i}[\eta, \kappa] = \underset{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[\eta,\kappa]}{\operatorname{argmin}} \mathrm{d}_{\mathrm{c}}^2 \left(\tilde{\mathbf{H}}_{u,i}[n_\eta, k_\kappa], \mathbf{Q}_j \right). \tag{4.115}$$

Because CSI feedback is provided only once per RB, the quantization codebook $\mathcal{Q}_{u,i}[\eta, \kappa]$ is adapted on a per-RB basis. To simplify the exposition, block fading is considered in this section; hence, the channel is assumed as temporally constant within each RB. Then 1D interpolation in the frequency domain is sufficient. A trivial extension to time-frequency selective channels is possible by consecutive 1D interpolation in time and frequency (or vice versa); more sophisticated multi-dimensional manifold interpolators are proposed in [177].

Considering two neighboring quantized channel subspaces $\hat{\mathbf{H}}_{u,i}[\eta, \kappa]$ and $\hat{\mathbf{H}}_{u,i}[\eta + 1, \kappa]$ the tangent defining the geodesic is obtained according to Equation (C.9)

$$\mathbf{T}_{u,i}[\eta, \kappa] = T \left(\hat{\mathbf{H}}_{u,i}[\eta, \kappa], \hat{\mathbf{H}}_{u,i}[\eta + 1, \kappa] \right) \in \mathcal{T} \left(\hat{\mathbf{H}}_{u,i}[\eta, \kappa] \right). \tag{4.116}$$

The linearly interpolated subspace between $\operatorname{span}(\hat{\mathbf{H}}_{u,i}[\eta, \kappa])$ and $\operatorname{span}(\hat{\mathbf{H}}_{u,i}[\eta + 1, \kappa])$ at distance $\Delta f = N_{\text{clust}}^{(f)} p,\ p \in [0, 1]$ from $\operatorname{span}(\hat{\mathbf{H}}_{u,i}[\eta, \kappa])$ can be calculated from the geodesic $\Gamma(\cdot)$ defined in Equation (C.11)

$$\hat{\mathbf{H}}_{u,i}[n_\eta + \Delta f, k_\kappa] = \Gamma \left(\hat{\mathbf{H}}_{u,i}[\eta, \kappa], \mathbf{T}_{u,i}[\eta, \kappa], p \right), \tag{4.117}$$

with $N_{\text{clust}}^{(f)}$ denoting the number of subcarriers in-between $\hat{\mathbf{H}}_{u,i}[\eta, \kappa]$ and $\hat{\mathbf{H}}_{u,i}[\eta + 1, \kappa]$. At the boundaries of the system bandwidth CSI extrapolation is necessary. Considering, e.g., the upper boundary, CSI extrapolation is achieved by extending the tangent between $\hat{\mathbf{H}}_{u,i}[N_{\text{RB}} - 1, \kappa]$ and $\hat{\mathbf{H}}_{u,i}[N_{\text{RB}}, \kappa]$ beyond the subcarrier $n_{N_{\text{RB}}}$, using Equation (4.117) with $p > 1$.

SQBC clustering: As an alternative to CSI interpolation, feedback clustering is considered, avoiding the need for an interpolator at the base station. The clustering approach is suitable when the distance $N_{\text{clust}}^{(f)}$ between CSI pilots is large compared to the channel coherence bandwidth, entailing unsatisfactory performance of linear interpolation because of significant channel variations within the RBs. For such situations it is proposed that the users determine the best $\ell_{u,i}[\kappa]$-dimensional subspace representation for each RB, assuming zeroth-order interpolation. The best subspace representation

$\tilde{\mathbf{H}}_{u,i}[\eta,\kappa]$ of RB $[\eta,\kappa]$ is defined by minimizing the average chordal distance over the RB

$$\begin{aligned}\tilde{\mathbf{H}}_{u,i}[\eta,\kappa] &= \underset{\tilde{\mathbf{H}}}{\operatorname{argmin}} \frac{1}{N_{\mathrm{RE}}} \sum_{\rho=1}^{N_{\mathrm{RE}}} \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{U}_{u,i}[n,k], \tilde{\mathbf{H}}\right) \\ &= \underset{\tilde{\mathbf{H}}}{\operatorname{argmin}} \frac{1}{N_{\mathrm{RE}}} \sum_{\rho=1}^{N_{\mathrm{RE}}} \ell_{u,i}[\kappa] - \operatorname{tr}\left(\tilde{\mathbf{H}}^{\mathrm{H}}(\mathbf{U}_{u,i}[n,k]\mathbf{U}_{u,i}[n,k]^{\mathrm{H}})\tilde{\mathbf{H}}\right), \end{aligned} \tag{4.118}$$

$$\text{subject to: } \tilde{\mathbf{H}} \in \mathbb{C}^{N_i \times \ell_{u,i}[\kappa]},\ \ \tilde{\mathbf{H}}^{\mathrm{H}}\tilde{\mathbf{H}} = \mathbf{I}_{\ell_{u,i}[\kappa]},$$

with $\mathbf{U}_{u,i}[n,k]$ being obtained from an SVD of the channel matrix as in (4.104), and the RE index $[n,k]$ being implicitly determined by $[\eta,\kappa]$ and ρ as specified in (3.13). Notice the similarity of this optimization problem to the SQBC optimization in Equation (4.88). In fact, (4.118) can be combined with SQBC to not only finding the best subspace representation in terms of zeroth-order interpolation, but minimizing both, the interpolation and the quantization error

$$\hat{\mathbf{H}}_{u,i}[\eta,\kappa] = \underset{\mathbf{Q}_j}{\operatorname{argmin}} \frac{1}{N_{\mathrm{RE}}} \sum_{\rho=1}^{N_{\mathrm{RE}}} \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{U}_{u,i}[n,k], \mathbf{Q}_j\right), \tag{4.119}$$

$$\text{subject to: } \mathbf{Q}_j \in \mathcal{Q}_{u,i}[\eta,\kappa] = \left\{\mathbf{Q}_j \in \mathbb{C}^{N_i \times \ell_{u,i}[\kappa]} \middle| \mathbf{Q}_j^{\mathrm{H}}\mathbf{Q}_j = \mathbf{I}_{\ell_{u,i}[\kappa]}\right\}.$$

The solution to problem (4.118) is obtained according to

$$\tilde{\mathbf{H}}_{u,i}[\eta,\kappa] = \left[\bar{\mathbf{U}}_{u,i}[\eta,\kappa]\right]_{:,1:\ell_{u,i}[\kappa]} = \bar{\mathbf{U}}_{u,i}^{(\ell_{u,i}[\kappa])}[\eta,\kappa], \tag{4.120}$$

$$\bar{\mathbf{U}}_{u,i}[\eta,\kappa]\bar{\mathbf{\Lambda}}_{u,i}[\eta,\kappa]\bar{\mathbf{U}}_{u,i}[\eta,\kappa]^{\mathrm{H}} = \bar{\mathbf{R}}_{u,i}[\eta,\kappa], \tag{4.121}$$

$$\bar{\mathbf{R}}_{u,i}[\eta,\kappa] = \frac{1}{N_{\mathrm{RE}}} \sum_{\rho=1}^{N_{\mathrm{RE}}} \mathbf{U}_{u,i}[n,k]\mathbf{U}_{u,i}[n,k]^{\mathrm{H}}, \tag{4.122}$$

with (4.121) denoting an eigendecomposition of $\bar{\mathbf{R}}_{u,i}[\eta,\kappa]$. Matrix $\bar{\mathbf{R}}_{u,i}[\eta,\kappa]$ can be interpreted as a *subspace correlation matrix*. Notice that the eigenvalues in $\bar{\mathbf{\Lambda}}_{u,i}[\eta,\kappa]$ are assumed in decreasing order. Similarly, the solution of (4.119) is determined by the quantization metric

$$\begin{aligned}\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[\eta,\kappa] &= \underset{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[\eta,\kappa]}{\operatorname{argmin}} \ell_{u,i}[\kappa] - \operatorname{tr}\left(\bar{\mathbf{\Lambda}}_{u,i}[\eta,\kappa](\bar{\mathbf{U}}_{u,i}[\eta,\kappa]^{\mathrm{H}}\mathbf{Q}_j\mathbf{Q}_j^{\mathrm{H}}\bar{\mathbf{U}}_{u,i}[\eta,\kappa])\right), \\ &\triangleq \underset{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[\eta,\kappa]}{\operatorname{argmin}} \mathrm{d}_{\mathrm{c,w}}^2\left(\bar{\mathbf{U}}_{u,i}[\eta,\kappa], \mathbf{Q}_j, \bar{\mathbf{\Lambda}}_{u,i}[\eta,\kappa]\right), \end{aligned} \tag{4.123}$$

$$\mathrm{d}_{\mathrm{c,SQBC}}^2[\eta,\kappa] = \min_{\mathbf{Q}_j \in \mathcal{Q}_{u,i}[\eta,\kappa]} \mathrm{d}_{\mathrm{c,w}}^2\left(\bar{\mathbf{U}}_{u,i}[\eta,\kappa], \mathbf{Q}_j, \bar{\mathbf{\Lambda}}_{u,i}[\eta,\kappa]\right). \tag{4.124}$$

Here, $\mathrm{d}_{\mathrm{c,w}}^2\left(\bar{\mathbf{U}}_{u,i}[\eta,\kappa],\mathbf{Q}_j,\bar{\mathbf{\Lambda}}_{u,i}[\eta,\kappa]\right)$ can be viewed as a *weighted chordal distance* with weighting matrix $\bar{\mathbf{\Lambda}}_{u,i}[\eta,\kappa]$. The importance of the individual eigenmodes of the subspace correlation matrix in the quantization metric (4.123) is specified by the diagonal weighting matrix.

By applying the SQBC antenna combiner of Section 4.3.2 to generate the effective channel on RE $[n,k]$ based on the quantized subspace $\hat{\mathbf{H}}_{u,i}^{(\mathrm{SQBC})}[\eta,\kappa]$, the quantization error (4.124) is obtained.

Dimensionality adaptation: As observed in Section 4.3.4, SQBC is subject to the problem that all $M_{u,i}$ modes of the channel matrix $\mathbf{H}_{u,i}[n,k]$ are equally treated in the quantization metric (4.88), potentially causing a weak channel gain of the effective channel matrix. This is also the case with SQBC clustering and its corresponding quantization metric (4.123). When the transmission is noise limited, rather than interference limited, the effective channel gain is of greater importance than the CSI quantization error. A trade-off between MET and SQBC can be achieved by considering only the subset of d maximum eigenmodes of the channel during quantization, with $\ell_{u,i}[\kappa] \leq d \leq M_{u,i}$. Then, the weighted chordal distance quantization metric (4.123) is replaced with

$$\hat{\mathbf{H}}_{u,i}^{(d)}[\eta,\kappa] = \underset{\mathbf{Q}_j\in\mathcal{Q}_{u,i}[\eta,\kappa]}{\operatorname{argmin}} \mathrm{d}_{\mathrm{c,w}}^2\left(\bar{\mathbf{U}}_{u,i}^{(d)}[\eta,\kappa],\mathbf{Q}_j,\bar{\mathbf{\Lambda}}_{u,i}^{(d)}[\eta,\kappa]\right), \tag{4.125}$$

$$\bar{\mathbf{U}}_{u,i}^{(d)}[\eta,\kappa]\bar{\mathbf{\Lambda}}_{u,i}^{(d)}[\eta,\kappa]\bar{\mathbf{U}}_{u,i}^{(d)}[\eta,\kappa]^{\mathrm{H}} = \bar{\mathbf{R}}_{u,i}^{(d)}[\eta,\kappa], \tag{4.126}$$

$$\bar{\mathbf{R}}_{u,i}^{(d)}[\eta,\kappa] = \frac{1}{N_{\mathrm{RE}}}\sum_{\rho=1}^{N_{\mathrm{RE}}}\mathbf{U}_{u,i}^{(d)}[n,k]\mathbf{U}_{u,i}^{(d)}[n,k]^{\mathrm{H}}, \tag{4.127}$$

$$\mathbf{U}_{u,i}^{(d)}[n,k] = [\mathbf{U}_{u,i}[n,k]]_{:,1:d}\,. \tag{4.128}$$

With this metric, the best $\ell_{u,i}[\kappa]$-dimensional subspace representation in the quantization codebook $\mathcal{Q}_{u,i}[\eta,\kappa]$ is determined, with respect to the average chordal distance to the d maximum eigenmodes of the channel matrices within the RB $[\eta,\kappa]$.

The natural question that arises when considering the quantization metric (4.125) is how to select the dimension d. To answer this question, an estimate of the pre-equalization achievable data rate with BD precoding and quantized CSIT is employed. The corresponding pre-equalization input-output relationship is given in Equation (2.4). According to [178], the instantaneous mutual information between the channel input and output, determining the achievable transmission rate, is

$$R_{u,i}^{(d)}[n,k] = \log_2\det\left(\mathbf{I}_{M_{u,i}} + \mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbf{S}_{u,i}[n,k]\mathbf{H}_{u,i}[n,k]\left(\tilde{\sigma}_z^2\,\mathbf{I}_{M_{u,i}} + \mathbf{R}_{u,i}^{(d)}[n,k]\right)^{-1}\right), \tag{4.129}$$

with $\mathbf{S}_{u,i}[n,k]$ denoting the covariance matrix of the channel input and $\mathbf{R}_{u,i}^{(d)}[n,k]$ being the interference covariance matrix. During the calculation of the CSI feedback, the precoders are unknown to

the users. To determine an estimate of the mutual information, the precoders are hence considered as random and are taken into account in the covariance matrices

$$\mathbf{S}_{u,i}[n,k] = \mathbb{E}\left(\mathbf{F}_{u,i}[n,k]\mathbf{x}_{u,i}[n,k](\mathbf{F}_{u,i}[n,k]\mathbf{x}_{u,i}[n,k])^{\mathrm{H}}\right) = \mathbb{E}\left(\mathbf{F}_{u,i}[n,k]\mathbf{F}_{u,i}[n,k]^{\mathrm{H}}\right), \quad (4.130)$$

$$\mathbf{R}_{u,i}^{(d)}[n,k] = \sum_{s=1,s\neq u}^{S_i[n,k]} \mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\mathbb{E}\left(\mathbf{F}_{s,i}[n,k]\mathbf{F}_{s,i}[n,k]^{\mathrm{H}}\right)\mathbf{H}_{u,i}[n,k], \quad (4.131)$$

with the expectation being taken with respect to the transmit signals and the precoders and considering the statistical independence of transmit signals corresponding to different streams and users. The precoder $\mathbf{F}_{u,i}[n,k]$ is determined by the channels of the other users according to Equation (4.4). As these channels are unknown to user u, $\mathbf{F}_{u,i}[n,k]$ is assumed as isotropically distributed, implying

$$\begin{aligned}\mathbf{S}_{u,i}[n,k] &= P_{u,i}[n,k]\,\mathbb{E}\left(\tilde{\mathbf{F}}_{u,i}[n,k]\tilde{\mathbf{F}}_{u,i}[n,k]^{\mathrm{H}}\right)\\ &= P_{u,i}[n,k]\,\frac{\ell_{u,i}[\kappa]}{N_i}\,\mathbf{I}_{\ell_{u,i}[\kappa]} = \frac{P_i}{N_i\,S_i[n,k]}\,\mathbf{I}_{\ell_{u,i}[\kappa]},\end{aligned} \quad (4.132)$$

with $\tilde{\mathbf{F}}_{u,i}[n,k]$ as defined in Equation (4.15). Due to the BD construction, the precoders $\mathbf{F}_{s,i}[n,k]$ of the other users are restricted to the left null space of $\hat{\mathbf{H}}_{u,i}^{(d)}[\eta,\kappa]$. Considering this knowledge in the calculation of $\mathbf{R}_{u,i}^{(d)}[n,k]$, as in Appendix H.2, and assuming that $S_i[n,k] = \frac{N_i}{\ell_{u,i}[\kappa]}$ users are served in parallel each over $\ell_{u,i}[\kappa]$ streams, the interference covariance matrix is obtained as (see also (H.11))

$$\mathbf{R}_{u,i}^{(d)}[n,k] = \frac{P_i}{N_i}\mathbf{H}_{u,i}[n,k]^{\mathrm{H}}\left(\mathbf{I}_{M_{u,i}} - \hat{\mathbf{H}}_{u,i}^{(d)}[\eta,\kappa]\hat{\mathbf{H}}_{u,i}^{(d)}[\eta,\kappa]^{\mathrm{H}}\right)\mathbf{H}_{u,i}[n,k]. \quad (4.133)$$

The *preferred dimensionality* $d_{u,i}[\eta,\kappa]$ is selected by maximizing the sum rate over the RB $[\eta,\kappa]$

$$d_{u,i}[\eta,\kappa] = \underset{\ell_{u,i}[\kappa]\leq d\leq M_{u,i}}{\operatorname{argmax}} \sum_{\rho=1}^{N_{\mathrm{RE}}} R_{u,i}^{(d)}[n,k], \quad (4.134)$$

where the RE index $[n,k]$ is implicitly determined by ρ and $[\eta,\kappa]$, according to Equation (3.13). Notice that knowledge of $d_{u,i}[\eta,\kappa]$ is not required at the base station; hence the dimensionality adaptation does not imply an additional feedback overhead.

The sum rate in (4.134) could potentially also be considered as quantization metric for the channel subspace. This is not followed up in this dissertation for two reasons. Firstly, calculating (4.134) for all quantization matrices is computationally expensive. The second reason is only relevant for predictive quantization. The performance of the prediction algorithm in the proposed quantizer of Section 4.2.2 is strongly impacted by the chordal distance quantization error, because the prediction is based on quantized CSI. Not choosing the quantized subspace according to the minimal chordal distance can entail a significant degradation of the prediction accuracy.

4.4.2. Channel Quality Feedback and Multi-User Scheduling

CQI feedback for limited feedback BD: To determine the multi-user resource allocation in the time, frequency and spatial domain, it is necessary to calculate the achievable data rate of a given schedule at the base station. In the spatial domain this rate is dependent on the number of users that is served in parallel, because the available transmit power is equally distributed among users and spatial streams according to Equation (4.15). Therefore, it must be possible to update the CQI feedback at the base station to account for the number of users served in parallel. For this reason, the pre-equalization mutual information estimated in the previous paragraph in Equation (4.129) is not employed as channel quality feedback, because the power allocation is hidden within the $\log_2 \det(\cdot)$ and cannot be updated subsequently.

The achievable data rate of a realistic communication system is determined by the post-equalization SINR, as argued in Section 3.2.2. It is therefore proposed to employ an estimate of the post-equalization SINR, achieved with limited feedback based BD precoding, as CQI feedback. An accurate SINR estimate cannot be obtained during feedback calculation, because the applied precoders are unknown at that time, in contrast to SU-MIMO with codebook based precoding. To circumvent this problem, a lower bound on the expected value of the SINR is utilized instead, similar to the proposal in [179] for ZF beamforming. The derivation of this lower bound is provided in Appendix G. According to Equation (G.19), the *lower bound on the expected SINR* of stream ν is expressed as

$$\tilde{\beta}_{\nu,u,i}[n,k] = \frac{c_S[n,k]\,\sigma^{\text{eff}}_{\nu,u,i}[n,k]^2}{\tilde{\sigma}_z^2 + c_I[n,k]\,\sigma^{\text{eff}}_{\nu,u,i}[n,k]^2}, \tag{4.135}$$

$$\boldsymbol{\Sigma}^{\text{eff}}_{u,i}[n,k] = \operatorname{diag}\left(\sigma^{\text{eff}}_{1,u,i}[n,k], \ldots, \sigma^{\text{eff}}_{\ell_{u,i},u,i}[n,k]\right),$$

$$c_S[n,k] = \frac{P_i}{S_i[n,k](N_i - \bar{\ell}_i[n,k])}\left(1 - \frac{\mathrm{d}_{\mathrm{c}}^2(\hat{\mathbf{H}}_{u,i}[n,k], \mathbf{B}_{u,i}[n,k])}{\ell_{u,i}[k]}\right)\left(1 - \frac{\mathrm{d}_{\mathrm{c}}^2(\tilde{\mathbf{H}}_{u,i}[n,k], \hat{\mathbf{H}}_{u,i}[n,k])}{\ell_{u,i}[k]}\right),$$

$$c_I[n,k] = \frac{P_i}{N_i\,\ell_{u,i}[k]}\,\mathrm{d}_{\mathrm{c}}^2(\tilde{\mathbf{H}}_{u,i}[n,k], \hat{\mathbf{H}}_{u,i}[n,k]).$$

with $\boldsymbol{\Sigma}^{\text{eff}}_{u,i}[n,k]$ denoting the matrix of singular values of the effective user channel $\mathbf{H}^{\text{eff}}_{u,i}[n,k]$, $\tilde{\mathbf{H}}_{u,i}[n,k]$ as defined in (4.6), $\hat{\mathbf{H}}_{u,i}[n,k]$ being the quantized channel subspace, $\mathbf{B}_{u,i}[n,k]$ denoting an orthonormal basis for the orthogonal complement of the other served users' effective channels (see Equation (G.5)) and $\bar{\ell}_i[n,k] = \sum_{s \in \mathcal{S}_i[n,k], s \neq u} \ell_{s,i}[k]$.[4] The residual multi-user interference due to imperfect CSI feedback is determined by the chordal distance quantization error in $c_I[n,k]$, justifying the application of the chordal distance as CSI quantization metric.

[4] Notice that the orthonormal basis $\mathbf{U}^{\text{eff}}_{u,i}[n,k]$ in Equations (G.11) and (G.17) has been replaced with the equivalent orthonormal basis $\tilde{\mathbf{H}}_{u,i}[n,k]$ in Equation (4.135) for the calculation of the chordal distance.

Considering the constant $c_S[n,k]$, the first two factors cannot be determined by the user u during feedback calculation, because neither the number of served users $S_i[n,k]$ nor the number of interfering streams $\bar{\ell}_i[n,k]$ is known, and also the other users' quantized effective channels are unknown. The last term, however, depending on the quantization error, can be calculated by the user. Also, the constant $c_I[n,k]$ and the singular values in $\mathbf{\Sigma}^{\text{eff}}_{u,i}[n,k]$ are available at the user, because they depend on local CSI only. It is therefore proposed to employ the following value as *per-stream CQI feedback*

$$\text{CQI}_{\nu,u,i}[n,k] = \frac{\sigma^{\text{eff}}_{\nu,u,i}[n,k]^2 \left(1 - \frac{\text{d}_\text{c}^2\left(\tilde{\mathbf{H}}_{u,i}[n,k], \hat{\mathbf{H}}_{u,i}[n,k]\right)}{\ell_{u,i}[k]}\right)}{\tilde{\sigma}_z^2 + \frac{P_i}{N_i\,\ell_{u,i}[k]}\,\text{d}_\text{c}^2(\tilde{\mathbf{H}}_{u,i}[n,k], \hat{\mathbf{H}}_{u,i}[n,k])\,\sigma^{\text{eff}}_{\nu,u,i}[n,k]^2}. \tag{4.136}$$

With this CQI feedback, it is possible to obtain an estimate of the achievable user rate $R_{u,i}[n,k]$ according to Equation (G.18) for a given schedule $\mathcal{S}_i[n,k]$, because the remaining two factors of $c_S[n,k]$ can be calculated by the base station, enabling the calculation of the SINR

$$\tilde{\beta}_{\nu,u,i}[n,k] = \frac{P_i}{S_i[n,k](N_i - \bar{\ell}_i[n,k])}\left(1 - \frac{\text{d}_\text{c}^2(\hat{\mathbf{H}}_{u,i}[n,k], \mathbf{B}_{u,i}[n,k])}{\ell_{u,i}[k]}\right)\text{CQI}_{\nu,u,i}[n,k], \tag{4.137}$$

$$R_{u,i}[n,k] \approx \sum_{\nu=1}^{\ell_{u,i}[k]} \log_2\left(1 + \tilde{\beta}_{\nu,u,i}[n,k]\right). \tag{4.138}$$

In case of feedback clustering or interpolation, CQI feedback is provided only once per RB. In the proposed SU-MIMO feedback algorithms of Section 3.2.2, mutual information effective SINR mapping (MIESM) is applied to obtain a single SNR that represent the average channel quality of an RB. This approach is not applicable here, because the CQI in (4.136) represents only a scaled version of the SINR (4.135), and MIESM is not a linear function; hence, without the correct scaling the application of MIESM is pointless. Instead, it is proposed to calculate an average CQI for RB $[\eta,\kappa]$, by linearly averaging the chordal distance quantization error and the squared singular values

$$\text{CQI}_{\nu,u,i}[\eta,\kappa] = \frac{\bar{\sigma}^{\text{eff}}_{\nu,u,i}[\eta,\kappa]^2\left(1 - \frac{\bar{\text{d}}_\text{c}^2[\eta,\kappa]}{\ell_{u,i}[\kappa]}\right)}{\tilde{\sigma}_z^2 + \frac{P_i}{N_i\,\ell_{u,i}[\kappa]}\,\bar{\text{d}}_\text{c}^2[\eta,\kappa]\,\bar{\sigma}^{\text{eff}}_{\nu,u,i}[\eta,\kappa]^2}, \tag{4.139}$$

$$\bar{\sigma}^{\text{eff}}_{\nu,u,i}[\eta,\kappa]^2 = \frac{1}{N_\text{RE}}\sum_{\rho=1}^{N_\text{RE}} \sigma^{\text{eff}}_{\nu,u,i}[n,k]^2, \;\; \bar{\text{d}}_\text{c}^2[\eta,\kappa] = \frac{1}{N_\text{RE}}\sum_{\rho=1}^{N_\text{RE}} \text{d}_\text{c}^2(\tilde{\mathbf{H}}_{u,i}[n,k], \hat{\mathbf{H}}_{u,i}[n,k]),$$

with $[n,k]$ being determined by $[\eta,\kappa]$ and ρ as specified in Equation (3.13).

The performance of the proposed CQI feedback for multi-user scheduling is evaluated in Section 4.4.3, demonstrating reasonably close to optimal results with quantized CSIT. Notice though that the CQI is not sufficiently accurate for transmission rate adaptation. As the CQI is based on a lower bound on the expected value of the SINR, the instantaneous SINR achieved with a given schedule can

be considerably above or below this value. It is therefore proposed to employ (4.136) only as an initial CQI for multi-user scheduling. As soon as the schedule is fixed, the users are able to estimate the instantaneous SINR defined in (2.7). Using this instantaneous SINR in combination with the CQI feedback algorithm of Section 3.2.2, accurate transmission rate adaptation can be achieved. Naturally this approach is only useful if the schedule is kept fixed for several TTIs, implying a loss of temporal multi-user diversity. However, accounting for the downlink signaling overhead involved in changing the multi-user resource allocation, the performance loss is negligible in the considered low to moderate mobility scenarios.

Multi-user scheduling: The sum throughput achieved with ZF beamforming and BD precoding is very much dependent on the selected set of users that is served in parallel [58]. If the channels of the selected users are orthogonal, the signal energy of the user can be steered into the user's channel subspace without causing any interference. This effect is also observable in the term $c_S[n,k]$ of the proposed SINR lower bound in Equation (4.135): when the channels of the selected users are orthogonal, the chordal distance $\mathrm{d}_{\mathrm{c}}^2\big(\hat{\mathbf{H}}_{u,i}[n,k], \mathbf{B}_{u,i}[n,k]\big)$ is equal to zero, implying no reduction of the channel gain by the precoder. In principal, the schedule can hence be obtained from an exhaustive search employing the achievable rate (4.138) as the scheduling metric. To avoid the complexity of an exhaustive search, greedy scheduling based on the SUS algorithm is instead considered in this dissertation. To achieve some level of fairness among the users in the cell and to avoid user starvation, proportional fairness is utilized as scheduling metric [56]. Scheduling is applied on an RB basis, due to the availability of RB-specific CQI feedback. The scheduling procedure is summarized below for RB $[\eta,\kappa]$; the same approach is applied on all RBs. To simplify the presentation, it is assumed that $\ell_{s,i}[\eta,\kappa] = L, \forall s \in \mathcal{U}_i$ and that $\frac{N_i}{L}$ is integer-valued. Otherwise, an explicit validation of the feasibility conditions of BD precoding must be performed by the scheduler.

0. Initialize the weighted sum rate $R_i[\eta,\kappa] = 0$, the set of scheduled users $\mathcal{S}_i[\eta,\kappa] = \{\}$, the set of potential users $\mathcal{P}_i[\eta,\kappa] = \mathcal{U}_i$ and the number of scheduled users $S_i[\eta,\kappa] = 0$.
   Repeat the following steps until the maximum number of $S_i[\eta,\kappa] = \frac{N_i}{L}$ users is served:
1. Calculate an orthonormal basis $\mathbf{B}_i[\eta,\kappa]$ for the space spanned by the channels of the served users in $\mathcal{S}_i[\eta,\kappa]$.
2. Find the semi-orthogonal user set $\tilde{\mathcal{S}}_i[\eta,\kappa]$, by determining all users $s \in \mathcal{P}_i[\eta,\kappa]$ for which the SUS condition is fulfilled:
$$\mathrm{tr}\left(\hat{\mathbf{H}}_{s,i}[\eta,\kappa]^{\mathrm{H}}\mathbf{B}_i[\eta,\kappa]\mathbf{B}_i[\eta,\kappa]^{\mathrm{H}}\hat{\mathbf{H}}_{s,i}[\eta,\kappa]\right) \leq \alpha_{\mathrm{SUS}}\, L. \tag{4.140}$$
3. If the SUS user set is empty, stop the algorithm and serve the users in $\mathcal{S}_i[\eta,\kappa]$.
4. For all users $s \in \tilde{\mathcal{S}}_i[\eta,\kappa]$, calculate the estimated achievable rate $R_{s,i}[\eta,\kappa]$ in (4.138) (replacing $\tilde{\beta}_{\nu,s,i}[n,k]$ with $\tilde{\beta}_{\nu,s,i}[\eta,\kappa]$ as obtained from $\mathrm{CQI}_{\nu,s,i}[\eta,\kappa]$) under the assumption that user s is served in parallel with the users in $\mathcal{S}_i[\eta,\kappa]$.

5. Determine the user $s \in \tilde{\mathcal{S}}_i[\eta, \kappa]$ that achieves the largest weighted rate

$$\hat{s} = \underset{s \in \tilde{\mathcal{S}}_i[\eta,\kappa]}{\operatorname{argmax}} \frac{R_{s,i}[\eta, \kappa]}{T_{s,i}}, \tag{4.141}$$

   with $T_{s,i}$ denoting the average throughput of user s achieved over the past.
6. Update the estimated achievable rates $R_{s,i}[\eta, \kappa]$ of the users $s \in \mathcal{S}_i[\eta, \kappa]$, assuming that $\hat{s}$ is served in addition to the users in $\mathcal{S}_i[\eta, \kappa]$.
7. Calculate the weighted sum rate of the schedule $\{\mathcal{S}_i[\eta, \kappa], \hat{s}\}$

$$R = \sum_{s \in \{\mathcal{S}_i[\eta,\kappa],\hat{s}\}} \frac{R_{s,i}[\eta, \kappa]}{T_{s,i}}. \tag{4.142}$$

8. If $R \geq R_i[\eta, \kappa]$, add user $\hat{s}$ to $\mathcal{S}_i[\eta, \kappa]$ and update $S_i[\eta, \kappa]$, set $R_i[\eta, \kappa] = R$ and remove $\hat{s}$ from $\mathcal{P}_i[\eta, \kappa]$. Otherwise, stop the algorithm and serve the users in $\mathcal{S}_i[\eta, \kappa]$

In step 2 of the scheduling algorithm, a pre-selection of users is performed, based on their subspace distance to the already served users. Only if a user is close to orthogonal to the previously served users, he is considered as a potential additional user. The exact meaning of "close to orthogonal" is determined by the choice of the SUS parameter α_{SUS}; see [58] for details on the selection of this parameter. In the presented simulations, the parameter $\alpha_{\text{SUS}} = 0.35$ turned out as a good choice.

4.4.3. Performance Investigation

In this section, the performance of CSI feedback using subspace clustering and interpolation for BD precoding based MU-MIMO transmission over a frequency-selective OFDM broadcast channel is investigated. Monte-Carlo simulations of the quantization MSE are conducted and the corresponding throughput achieved in the downlink of a single LTE compliant cell is evaluated. The frequency-selectivity of the wireless channel is characterized by the coherence bandwidth defined in Equation (2.12) or equivalently by the root mean square (RMS) delay spread τ_{RMS} of the channel multipath power delay profile. When the sensitivity of the proposed methods with respect to the frequency-selectivity of the channel is investigated, the simulation results are presented in dependence of the normalized sampling bandwidth

$$B_s = \Delta f_s \, N_{\text{clust}}^{(f)} \, \tau_{\text{RMS}}, \tag{4.143}$$

with Δf_s [Hz] being the OFDM subcarrier spacing ($\Delta f_s = 15\,\text{kHz}$ in case of LTE) and $N_{\text{clust}}^{(f)}$ denoting the size of a feedback cluster in the frequency domain in multiples of subcarriers. The advantage of SQBC clustering in terms of the chordal distance quantization MSE is demonstrated, but also its downside, i.e., the implied channel gain reduction, is highlighted, which can cause a significant throughput degradation, especially in case of strongly correlated receive antennas. In order

to enable efficient CSI feedback operation, channel prediction, subspace clustering and Grassmannian quantization are combined at the user. The MSE contribution of the individual components is investigated to reveal the corresponding limitations and operating regimes of the proposed feedback algorithms. Finally, the performance achieved with the proposed SUS based multi-user scheduler, utilizing the derived CQI feedback, is evaluated by comparing to the throughput attained with the optimal schedule, as determined from an exhaustive search.

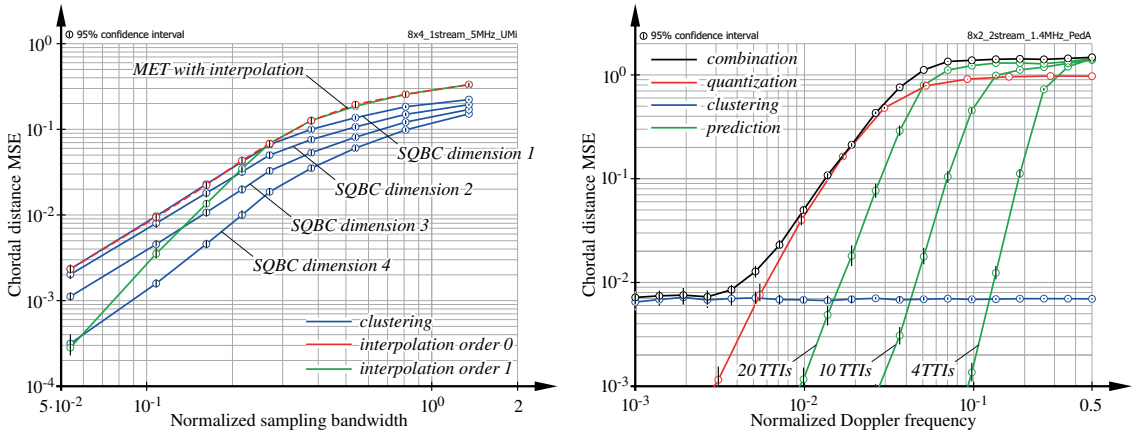
Chordal distance MSE evaluation: In the first simulation, the chordal distance MSE obtained with MET subspace selection, according to Section 4.3.3, in combination with zeroth- and first-order interpolation, as detailed in Section 4.4.1, is compared to SQBC clustering. The channel matrix is of size $N_i \times M_{u,i} = 8 \times 4$, and an $\ell_{u,i} = 1$-dimensional subspace is selected as CSI feedback. Unquantized feedback of this representative subspace is considered; hence the error is caused only by the subspace clustering and the interpolation, respectively. In case of MET, the maximum eigenmode of the channel matrix experienced in the center of each RB is used as feedback information, while with SQBC clustering the subspace feedback is determined by solving the optimization problem (4.118). The MSE is estimated by means of Monte-Carlo simulations as

$$D \approx \frac{1}{N_{\text{tot}}} \frac{1}{K} \sum_{n=1}^{N_{\text{tot}}} \sum_{k=1}^{K} \mathrm{d}_{\mathrm{c}}^2 \left(\hat{\mathbf{H}}_{u,i}[n,k], \mathbf{H}_{u,i}[n,k] \right), \tag{4.144}$$

where the interpolated subspace $\hat{\mathbf{H}}_{u,i}[n,k] \in \mathbb{C}^{N_i \times \ell_{u,i}}$ is obtained from Equation (4.117) in case of first-order interpolation, while $\hat{\mathbf{H}}_{u,i}[n,k] = \tilde{\mathbf{H}}_{u,i}[\eta,\kappa]$ is employed with zeroth-order interpolation and SQBC clustering ($[\eta,\kappa]$ is defined in Equation (3.14)).

The effect of dimensionality adaptation, as defined in Equation (4.125), on the MSE performance of SQBC clustering is investigated. A system bandwidth of 5 MHz is assumed and the power delay profile of the channel is determined by the SCME *urban micro* channel model [180], having a coherence bandwidth of $B_C = 680\,\text{kHz}$.

The results of the simulation are shown in Figure 4.6a. Considering MET subspace selection with CSI interpolation, it is observed that zeroth-order interpolation is outperformed by first-order (linear) geodesic interpolation, as soon as the distance between the CSI pilots is sufficiently small. With SQBC clustering, a substantial MSE reduction is achieved by increasing the dimensionality d of the optimization problem (4.125), which determines the subspace selection. Notice that the dimensionality d has no impact on the dimension $\ell_{u,i}$ of the subspace feedback; it is rather the dimension of the *search space* in the optimization problem (4.125) that is specified with d. Comparing CSI interpolation and clustering, a significant MSE improvement is observed with clustering when a small density of CSI pilots, i.e., a large sampling bandwidth B_s, is applied. This MSE improvement, however, is obtained at the cost of a channel gain reduction of the effective channel, causing a throughput loss at low SNR, as demonstrated below.



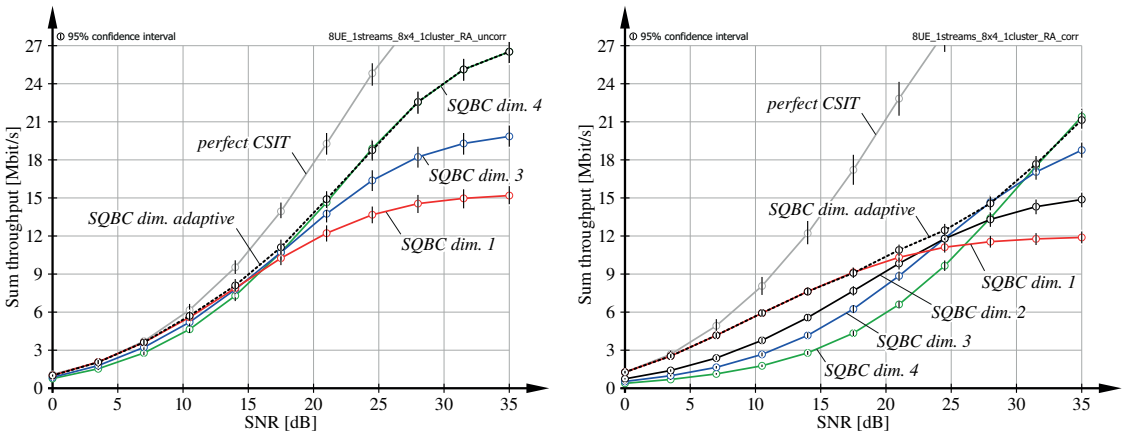
(a) Comparison of feedback clustering and interpolation.

(b) Contribution of CSI prediction, clustering and quantization.

Figure 4.6.: Chordal distance MSE investigation of the proposed CSI feedback methods for frequency-selective channels.

In the next simulation, predictive CSI quantization and Grassmannian subspace clustering are combined to enable efficient limited feedback operation. Additionally, the effect of a delay in the feedback path is investigated. To compensate for this delay, FIR based channel prediction, as described in Section 3.3.2, is employed by the user. Notice that this channel predictor is only required at the user and not at the base station and it can therefore be based on perfect CSI. Channel prediction for delay compensation is not related to the operation of the predictive quantizer; the quantizer is working independently on top of the channel predictor output. Feedback of an $\ell_{u,i} = 2$-dimensional subspace of an $N_i \times M_{u,i} = 8 \times 2$ dimensional channel matrix is considered. The *PedA* channel model [108], with $B_C = 4.4\,\text{MHz}$, is employed and a single feedback cluster is applied to represent the $1.4\,\text{MHz}$ system bandwidth, resulting in a feedback overhead of $7\,\text{bit}$ per TTI ($7\,\text{kbit/s}$ assuming the LTE subframe duration of $1\,\text{ms}$).

The results of the investigation are shown in Figure 4.6b. The individual MSE contributions of the channel prediction, the subspace clustering and the quantization, as well as the performance of the concatenated system are separately plotted. At low normalized Doppler frequencies (see (3.33)), i.e., when the channel variation over time is slow, the overall MSE is dominated by the subspace representation error, due to feedback clustering. In this case, the adaptive codebook construction of the predictive quantizer is able to track the channel variation very well, causing a negligible quantization error. A reduction of the observed error floor is only possible by reducing the cluster size, implying an increased feedback overhead. At intermediate Doppler frequencies, the overall MSE is determined by the quantization error. This error can be reduced by enlarging the quantization codebook, for the cost of increasing the feedback rate. Only at high Doppler frequencies and with a large feedback delay, the error caused by the channel prediction comes into play. In the region of interest for predictive quantization ($\nu_d \approx 10^{-2}$), even a feedback delay of $20\,\text{TTIs}$ is irrelevant; hence, the feedback delay is not further considered in the remaining simulations of this thesis.



(a) No receive antenna correlation $\alpha_{\text{corr}} = 0$.

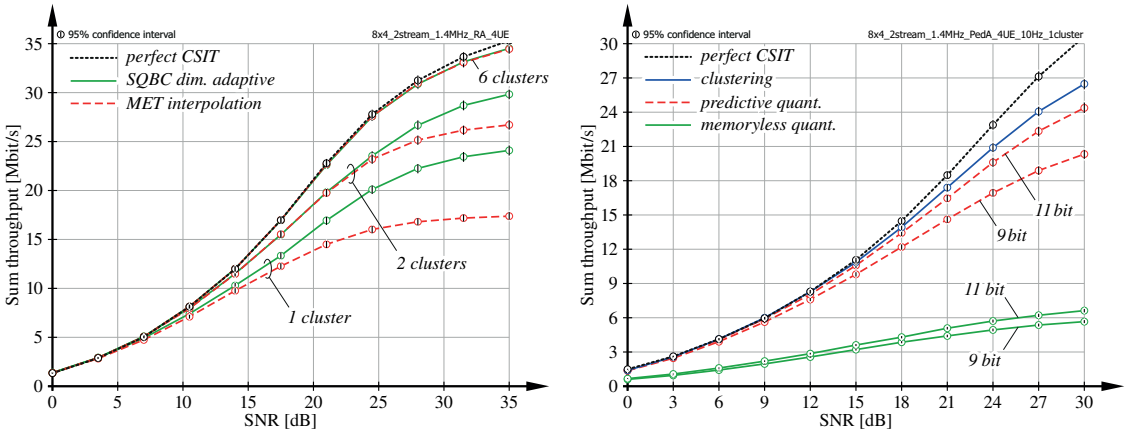
(b) Strong receive antenna correlation $\alpha_{\text{corr}} = 0.9$.

Figure 4.7.: Sum throughput achieved with ZF beamforming and SQBC feedback clustering in an $N_i \times M_{u,i} = 8 \times 4$ system, serving $S_i = 8$ users over $\ell_{u,i} = 1$ spatial stream each. The impact of dimensionality adaptation is investigated.

Throughput evaluation: Next, the throughput performance of BD based MU-MIMO is investigated, when the proposed limited feedback algorithms are applied to provide CSIT. In the first simulation, the transmission rate reduction with respect to perfect CSIT, due to the subspace representation error caused by SQBC clustering, is evaluated, demonstrating the effect of dimensionality adaptation on the achieved throughput in dependence of the equivalent average transmit SNR (3.32).

The base station is equipped with $N_i = 8$ transmit antennas; $S_i = 8$ users, having $M_{u,i} = 4$ receive antennas, are served in parallel, each over a single spatial stream $\ell_{u,i} = 1$. The power delay profile of the channel is specified by the *rural area* model [109], which has a coherence bandwidth of $B_C = 2\,\text{MHz}$. The carrier bandwidth of $1.4\,\text{MHz}$ is represented with a single CSI feedback cluster, using unquantized CSI feedback. Correlated receive antennas are assumed according to Equation (2.17) with $\alpha_{\text{corr}} \in \{0, 0.9\}$. The interference-averaged MMSE equalizer proposed in Appendix H.2 is applied by the users to detect their data streams. The throughput simulations are conducted with the standard compliant Vienna LTE link level simulator [59].

The obtained simulations results are shown in Figure 4.7. At low SNR, the best performance is achieved with a dimensionality of $d = 1$. With $d = 1$, only the maximum eigenmodes of the channels experienced over the subcarriers of an RB are taken into account for the calculation of the subspace representation according to Equation (4.125). The calculated subspace is kept free of interference by the BD precoder; hence, the interference-free subspace has a potentially large channel gain. Due to clustering, however, the single subspace representation is imperfect for a given RE, implying residual multi-user interference after precoding. Therefore, an interference-limitation is observed at high SNR. The strength of the residual multi-user interference is impacted by the dimensionality d of the feedback clustering. By increasing d the *search space* for the subspace representation in (4.125)



(a) MET feedback with interpolation versus SQBC clustering.

(b) Predictive versus memoryless quantization.

Figure 4.8.: Sum throughput achieved with BD precoding and the proposed limited feedback clustering and quantization algorithms in an $N_i \times M_{u,i} = 8 \times 4$ system, serving $S_i = 4$ users over $\ell_{u,i} = 2$ spatial streams each.

is extended, enabling a reduction of the chordal distance error, as shown in Figure 4.6a, which implies reduced residual multi-user interference according to (4.135). The optimal dimensionality is dependent on the SNR. The proposed dimensionality adaptation (4.134) is able to identify the optimal dimensionality, as demonstrated in Figure 4.7. Comparing Figures 4.7a and 4.7b, it is observed that the throughput difference between different dimensionalities is increased with growing antenna correlation. This behavior is due to the increased singular value spread of the channel matrix with larger α_{corr}. Notice the similarity of the performance observed with dimensionality adaptation and with transmission rank adaptation, as shown in Figure 3.5a. In the current simulations, however, the transmission rank is not changed; the total number of data streams is always equal to eight.

In Figure 4.8a, a similar scenario is evaluated where four users are served over $\ell_{u,i} = 2$ streams each. The throughput of SQBC clustering with dimensionality adaptation is compared to MET subspace selection with linear geodesic interpolation. The number of feedback clusters is varied from $N_{\text{RB}} = 1$ to $N_{\text{RB}} = 6$ and uncorrelated receive antennas, $\alpha_{\text{corr}} = 0$, are assumed. It is observed that MET is outperformed by SQBC at high SNR if the feedback cluster size is large, because the residual multi-user interference achieved with SQBC is smaller due to the improved subspace representation. SQBC clustering is never surpassed by MET feedback, not even at low SNR when the throughput is determined by the effective channel gain, due to the application of dimensionality adaptation, which trades-off the effective channel gain for the residual multi-user interference depending on the SNR.

Finally, in Figure 4.8b, the performance of the combination of feedback clustering and quantization is evaluated. The $1.4\,\text{MHz}$ channel generated with the *PedA* model [108], assuming strongly correlated receive antennas, $\alpha_{\text{corr}} = 0.9$, is quantized with a single feedback cluster. Nine and eleven bit of feedback per TTI are considered, corresponding to a feedback rate of $9\,\text{kbit/s}$ and $11\,\text{kbit/s}$,

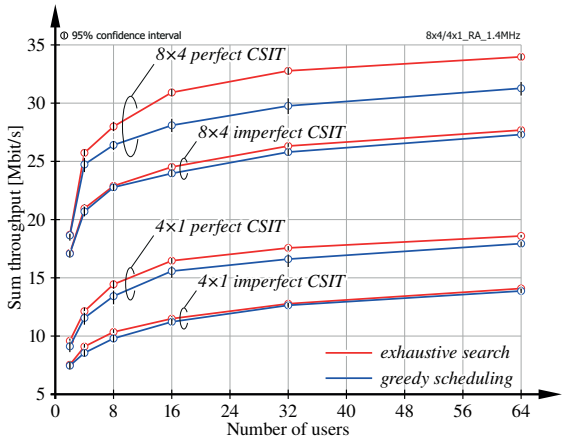


Figure 4.9.: Performance of the proposed SUS scheduling algorithm with perfect and imperfect CSIT compared to an exhaustive search. Two configurations are considered: $N_i \times M_{u,i} = 4 \times 1$ with $\ell_{u,i} = 1$ and $N_i \times M_{u,i} = 8 \times 4$ with $\ell_{u,i} = 2$.

respectively. A maximum normalized Doppler frequency of $\nu_d = 0.01$ is assumed, representing, at a center frequency of 2 GHz, a walking user with a speed of approximately 5 km/h.

With these parameters, a negligible throughput degradation is caused by the subspace clustering, compared to perfect CSIT, up to an SNR of approximately 15 dB. Above this value, a throughput loss is observed, due to the residual multi-user interference induced by the frequency selectivity of the channel. When the SQBC subspace selection is conveyed to the base station by means of predictive quantization, close to optimal performance is achieved with 11 bit of feedback per TTI over a large SNR range. Naturally, a throughput deterioration is incurred when the quantization codebook size is reduced. Considering memoryless quantization, however, the obtained CSIT accuracy is insufficient to ensure reliable transmission to four users in parallel, each being served over $\ell_{u,i} = 2$ streams. In this case, better performance is possible by employing a multi-user scheduler to select only a subset of the users for transmission, implying reduced multi-user interference.

Multi-user scheduling: In this section, the efficiency of the SUS based greedy multi-user scheduler, proposed in Section 4.4.2, is investigated by comparing the achieved sum throughput to the optimal schedule, as obtained from an exhaustive search. Two different antenna configurations are considered, i.e., $N_i \times M_{u,i} = 4 \times 1$ and $N_i \times M_{u,i} = 8 \times 4$. In the 4×1 system, a single stream is transmitted per user, $\ell_{u,i} = 1$, while with 8×4 antennas $\ell_{u,i} = 2$ streams per user are employed. At most four users can thus be spatially multiplexed in both cases. The *rural area* channel model [109] is used and the antenna correlation parameter α_{corr} is set equal to zero. In case of imperfect CSIT, unquantized feedback of a single feedback cluster is considered for the total system bandwidth of 1.4 MHz using SQBC clustering. All users are simulated with the same SNR of 20 dB.

The simulation results are presented in Figure 4.9, in terms of the sum throughput achieved in

the downlink of the cell versus the number of served users. It can be seen that a similar multi-user diversity is attained with the exhaustive search scheduler and the proposed greedy scheduling algorithm, albeit an approximately constant rate loss that is dependent on the settings of the system. Especially in the realistic situation of imperfect CSIT, the throughput obtained with the greedy algorithm is close to the rate achieved with the optimal schedule.

Additional simulation results, involving distributed transmit antennas, multi-cell networks causing out-of-cell interference and users that experience different SNRs (determined by a macroscopic fading model), are provided in Chapter 5, when comparing several cellular networking architectures.

4.5. Summary

In this chapter, limited feedback algorithms for MU-MIMO transmission in wireless communications are proposed and evaluated. The focus is put on linear transceiver architectures, employing BD precoding at the transmitter to orthogonalize the transmissions to several users, and semi-unitary antenna combining at the receivers to separate the interference-free subspace, occupied by the intended signal, from the interference-contaminated subspace. To achieve orthogonal transmissions, the base station must be informed about the subspaces spanned by the users' channel matrices with very high precision. Efficient feedback of the required CSI from the receiver to the transmitter is enabled with the proposed predictive quantization algorithm that exploits the structure of the subspace information by operating on the underlying Grassmannian manifold. Close to optimal performance is demonstrated in low to moderate mobility scenarios, in terms of the subspace quantization error. When applied for CSIT acquisition in wireless communications, virtually perfect performance is observed with a reasonable feedback overhead at practically relevant SNRs ($\leq 30\,\mathrm{dB}$).

When users are equipped with excess antennas, that is, when the number of user antennas is larger than the number of data streams per user, it is proposed to perform BD precoding over appropriate subspaces of the users' channel matrices, which are selected by the users based on selfish arguments. Although better performance is possible with a joint optimization of the precoders and the antenna combiners at the base station, the proposed approach can help reducing the required feedback information. Specifically, instead of requiring knowledge of the full channel matrix of all users at the base station, information about the selected subspace is sufficient, which can be conveyed efficiently using Grassmannian quantization. With channel subspace feedback, the obtained user throughput is strongly dependent on the selection of the subspace within the space spanned by the channel matrix. Depending on the SNR of the user, the achieved data rate is determined either by the channel gain experienced over the selected subspace, or by the error that is incurred during CSI quantization, causing residual multi-user interference. To this end, two extremes of the subspace selection are proposed that either put all focus on maximizing the channel gain of the subspace or on minimizing its quantization error. The corresponding antenna combiners generating these subspaces

are derived. The performance of both strategies is investigated analytically by deriving upper bounds on the amount of feedback information necessary to achieve a given rate loss with respect to perfect CSIT. A trade-off between the two strategies is proposed using an estimate of the achievable user rate given a specific subspace selection. In that way, efficient CSI feedback and data transmission is enabled over the full range of considered SNRs.

The CSI feedback algorithms and subspace selection strategies are extended to frequency-selective OFDM by means of feedback clustering and interpolation. Especially with a large cluster size, corresponding to a low density of CSI pilots, it is shown that interpolation is significantly outperformed by the proposed clustering approach, minimizing the CSI quantization error obtained over a cluster of OFDM subcarriers. To enable efficient multi-user scheduling, channel quality information is required in addition to the channel subspace feedback at the transmitter, to facilitate estimation of the achievable throughput. For this purpose, a lower bound on the expected per-stream post-equalization SINR is derived, which quantifies the channel quality of a user without requiring a-priori knowledge of the multi-user schedule and the applied precoders. The efficiency of the resource allocation obtained from this SINR lower bound is determined by means of Monte-Carlo simulations, demonstrating the same multi-user diversity as attained with an optimal schedule based on perfect CSIT.

Considering the proposed methods self-critically, the following assumptions may be worth rethinking: The out-of-cell interference is treated as additional Gaussian noise by the receivers. In case the interference is determined by a few dominant sources, this assumption may be violated and better performance can be achieved by incorporating the actual interference distribution in the derivation of the algorithms. The approximate out-of-cell interference model provided in Appendix I may be a valid starting point for such investigations.

When treating predictive quantization in Section 4.2.2, it is assumed that the innovation noise (the non-deterministic part of the channel evolution) and the prediction error in the Euclidean space are both Gaussian distributed. This assumption can be justified in case of Rayleigh fading, when the channel matrix is itself Gaussian distributed [169, 181], but may be violated otherwise. The most significant part of the performance improvement compared to memoryless quantization, however, is not so much obtained by matching the quantization codebook to the error distribution, but rather by fitting the variance of the prediction error, i.e., the volume of the Grassmann manifold that is covered by the quantization codebook. This is achieved with the proposed variance tracking algorithm, which can be used to track the variance of any distribution.

The theoretical performance investigation of SQBC and MET antenna combining in Section 4.3 is based on the assumption that the channel matrix is i.i.d. Gaussian distributed. Hence, the correlation of the wireless channel in the spatial, temporal and frequency domain is not taken into account in this analysis. Extending the investigation to correlated channels is hard and was not yet successful, mostly because theoretical results on the performance of RVQ with correlated channels are not available, which I consider as a worthwhile problem for future research.

Chapter 5.

Application Scenarios

> It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong.
>
> *(Richard P. Feynman)*

In this chapter, the proposed single-user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO) feedback algorithms and transceiver architectures of Chapters 3 and 4 are applied to obtain a realistic performance comparison of three different concepts of cellular networking architectures, which are currently in the focus of many research papers. In the first part of this simulation based study, the area spectral efficiency (ASE) achieved in a centralized antenna system (CAS) is contrasted to the efficiency attained with distributed antennas. The impact of several parameters, such as the number of users per cell, the temporal and spatial channel correlation as well as the feedback overhead, on the performance of SU- and MU-MIMO is evaluated. A similar investigation is then conducted with small cells as well, considering only MU-MIMO transmission. To obtain a fair comparison, the total number of transmit antennas as well as the total transmit power are the same in all scenarios. The simulations are carried out with an extended version of the Vienna LTE link level simulator [59], supporting a pathloss model to determine the SINR experienced by the users. For complexity reasons, only a single cell of the cellular network, as illustrated in Figure 5.1, is explicitly simulated. Inter-cell interference between the three sectors of the cell, including the remote radio units (RRUs) and small cells within the nominal cell area, is accurately simulated. Out-of-cell interference from other cells surrounding the cell of interest is generated with the interference model proposed in Appendix I. To this end, two tiers of interfering macro base stations, which are identical copies of the cell of interest in terms of radio equipment and geometry, are taken into account.

This chapter is organized as follows: In Section 5.1, the considered networking architectures are introduced and the simulation assumptions are specified. Furthermore, the holistic performance metric applied to quantify the efficiency of the networks, i.e., the area spectral efficiency [182], is defined. In Section 5.2, the comparison between the CAS and the distributed antenna system (DAS) is provided, employing SU- and MU-MIMO transmission. The ASE obtained with MU-MIMO in small cells is presented in Section 5.3 and concluding remarks are provided in Section 5.4.

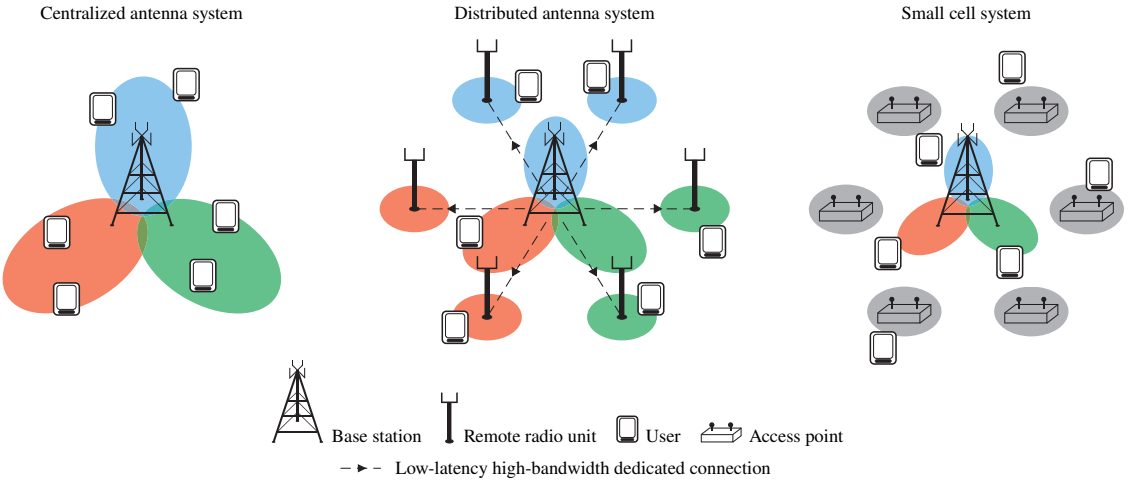


Figure 5.1.: Investigated cellular networking architectures.

5.1. Cellular Networking Architectures

In the simulations presented in this chapter, the three cellular networking architectures illustrated in Figure 5.1 are considered, assuming a 2D propagation model. The networking architectures are characterized as follows:

- The centralized antenna system, shown in the left part of Figure 5.1, is employed to represent the classical hexagonal tessellation of the network area using macro base stations only. In the simulations, sectorization of the cell area into 120° sectors is considered. The 2D antenna gain pattern of the applied model is defined in [183], having a 3 dB beam width of 65° and a maximum attenuation of 20 dB. In the main pointing direction of the sectorized antenna, a gain of 15 dB over the isotropic radiator is achieved. Each sector has a total transmit power of 43 dBm and is equipped with eight transmit antennas.
- In the distributed antenna system, the macro base station is augmented with two RRUs per sector, which are connected to the macro base station via low-latency high-bandwidth dedicated links, facilitating coherent transmission from all antenna arrays within the sector. The RRUs are regularly placed on a ring of radius $2/3\, r_c$, with $r_c = 500$ m denoting the cell radius, as shown in the center part of Figure 5.1. The RRUs are equipped with two omni-directional antennas each. To conserve the total number of transmit antennas, the macro base station has only four antennas per sector available. The distribution of the total transmit power of 43 dBm among the antenna arrays within each sector is determined by the applied precoder.
- In the small cell system, the RRUs of the DAS are replaced with autonomous micro base stations (access points), as shown in the right part of Figure 5.1. The micro base stations are equipped with two omni-directional antennas each and have a transmit power of 37 dBm. The

Table 5.1.: Simulation parameters applied for the cellular networking architecture comparison.

Parameter	Value
System bandwidth	$1.4\,\text{MHz}$
Carrier frequency	$f_c = 2\,\text{GHz}$
Noise power spectral density	$-174\,\text{dBm/Hz}$
Channel model	pedestrian A [108]
Number of receive antennas	$M_{u,i} \in \{2, 4\}$
Number of transmit antennas	$N_i = 8$
Number of users	$U_i \in [2, 32]$
Number of feedback clusters	$N_{\text{RB}} = 1$
Number of feedback bits	$b \in [1, 10]$
Spatial correlation parameter	$\alpha_{\text{corr}} \in [0, 0.95]$
Maximum channel Doppler frequency	$f_d \in [10, 500]\,\text{Hz}$
Multi-user scheduling	proportional fair (SUS)
MIMO receiver	MMSE (interference averaged)

transmit power of the sectors of the macro base station is reduced to $40\,\text{dBm}$ and the number of transmit antennas is decreased to four. Thus, the total transmit power per sector is equal to $43\,\text{dBm}$ and the total number of transmit antennas is equal to eight.

Shadow fading is not considered in the presented simulations. This simplifies the simulations, because the region of interest is clearly delimited by the geometry of the cell. Due to the regular placement of the RRUs and the small cells, the hexagonal tessellation of the CAS is also valid with the other architectures. Users are uniformly placed within the nominal cell area of the macro base station. To simplify the calculation of the ASE, the hexagonal cell is approximated with a circle of radius

$$r_n = \frac{r_c + r_c/\cos(30^\circ)}{2} \approx 538\,\text{m}, \tag{5.1}$$

i.e., the arithmetic mean between the incircle and the circumcircle of the hexagon is employed as the nominal cell radius. In the small cell system, the users within a sector are associated either with the macro base station or with one of the micro base stations, depending on the received signal strength. As shadow fading is not simulated, the received signal strength and the channel gain matrix (2.11) are determined by the user position dependent pathloss and the antenna gain. The assumed pathloss model is specified in [183, Section 4.5.2]; see also [95]. Load balancing between the base stations is not considered. This is not a significant restriction here, because mostly small user populations are investigated. A review of load balancing methods is provided in [184]. Receive antenna correlation according to Equation (2.17), with varying correlation parameter α_{corr}, is assumed. Further simulation parameters are summarized in Table 5.1.

With these assumptions, the sum ASE can be estimated from Monte-Carlo simulations as in [32]

$$\text{ASE} \approx \frac{1}{r_n^2 \pi} \sum_{k=1}^{K} \bar{T}_u(r_k) \left(U \, \frac{r_k^2 - r_{k-1}^2}{r_n^2} \right), \tag{5.2}$$

with $r_n = 538\,\text{m}$ as defined in Equation (5.1), $\bar{T}_u(r_k)$ denoting the average user throughput at a distance to the macro base station between r_k and r_{k-1} and U being the total number of users in all three sectors. Notice that the last term in brackets is equal to the expected number of users within the ring between r_{k-1} and r_k, because the probability to find a user within a certain area of the cell is proportional to the size of that area, due to the assumed uniform user distribution. The sum is calculated over all such rings within the cell area from $r_0 = 0$ to $r_K = r_n$. The value of $\bar{T}_u(r_k)$ is estimated from 240 random positions per user.[1] For each position realization, the throughput is averaged over 150 LTE subframes. The ASE is calculated assuming a radius increment of $\Delta r = r_k - r_{k-1} = 2\,\text{m}$.

The simulations are conducted with perfect and quantized channel state information (CSI) at the transmitter (CSIT). The following transmit modes and CSI feedback configurations are considered:

- **SU-MIMO** according to Chapter 3:
  - **Perfect CSIT**: *singular value decomposition (SVD)* based precoding, as detailed in Section 3.1.2, is employed.
  - **Quantized CSIT**: codebook based precoding is applied, using LTE's *closed loop spatial multiplexing (CLSM)* codebooks [8]. The CSI feedback is determined as described in Section 3.2.2. The MMSE equalizer defined in Equation (3.12) is utilized to separate the spatial streams of the user. If RRUs are considered, CLSM is combined with *antenna selection* as described in Section 3.2.4.
- **MU-MIMO** according to Chapter 4, employing block diagonalization (BD) precoding:
  - **Perfect CSIT**: maximum eigenmode transmission (MET) is employed to determine the unquantized subspace feedback for each resource element (RE); see Section 4.3.3.
  - **Quantized CSIT**: subspace quantization based combining (SQBC) feedback clustering, according to Section 4.4.1, is applied with a cluster size of $N_{\text{RB}} = 1$ to determine a single subspace representation for the full system bandwidth of $1.4\,\text{MHz}$. Dimensionality adaptation, as proposed in Section 4.4.1, is considered to maximize the achievable throughput of the multi-user transmission.
    - *Memoryless quantization*: random vector quantization (RVQ), according to Section 4.2.1, is employed to quantize the subspace determined by SQBC clustering. With RRUs, correlated RVQ is applied as well, to improve the quantization accuracy.
    - *Predictive quantization*: the predictive quantizer proposed in Section 4.2.2 is used to quantize the subspace determined by SQBC clustering.

[1] E.g., with eight users per sector a total of $240 \cdot 8 \cdot 3 = 5760$ user positions is simulated.

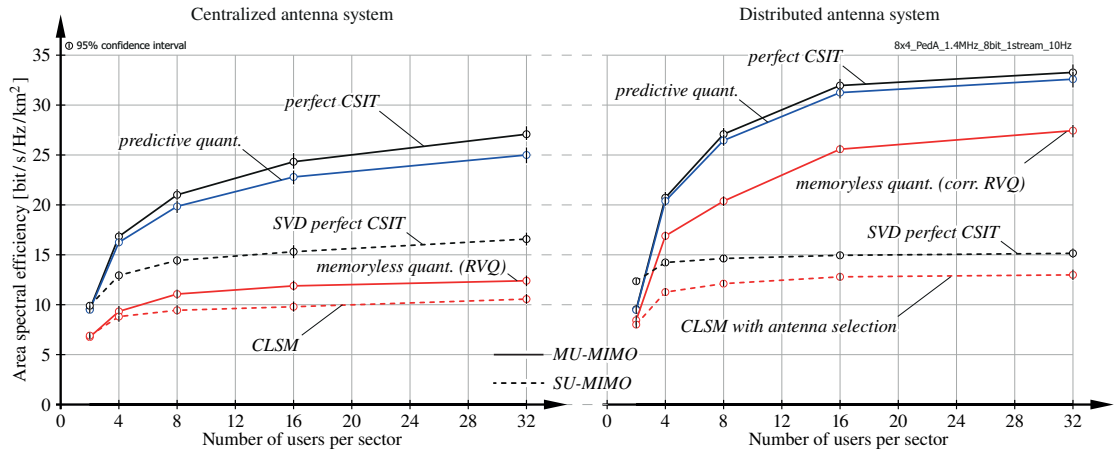


Figure 5.2.: Area spectral efficiency achieved in the CAS and the DAS in dependence of the number of users per sector. An $N_i \times M_{u,i} = 8 \times 4$ antenna configuration is considered, transmitting a single stream per user, $\ell_{u,i} = 1$, to at most eight users in parallel in case of MU-MIMO, and up to $\ell_{u,i} = 4$ streams with SU-MIMO.

In case of MU-MIMO transmission, either $\ell_{u,i} = 1$ or $\ell_{u,i} = 2$ spatial streams per user are transmitted. Notice that per-user transmission rank adaptation is not considered for MU-MIMO, i.e., the number of data streams per user is fixed. Only the number of users served in parallel is adapted, employing the scheduler of Section 4.4.2. With SU-MIMO, on the other hand, the per-user transmission rank is adapted. The interference-averaged MMSE equalizer, derived in Appendix H.2, is applied by the users, to separate their intended data streams from the residual multi-user interference, and to cancel interference in-between the intended streams.

5.2. Centralized versus Distributed Antenna Systems

In this section, a systematic comparison between the CAS and the DAS illustrated in Figure 5.1 is provided. The impact of important system parameters on the performance of SU- and MU-MIMO transmission in both networking architectures is evaluated. In all simulations presented in this section the users are equipped with $M_{u,i} = 4$ receive antennas.

The improvement of the ASE with the number of users per sector is investigated in the first simulation, shown in Figure 5.2. A feedback overhead of 8 bit per subframe is considered and a normalized Doppler frequency of $\nu_d = 0.01$ as well as uncorrelated receive antennas $\alpha_{\text{corr}} = 0$ are assumed. Comparing SU-MIMO (dashed) to MU-MIMO (solid), it is observed that single-user spatial multiplexing is outperformed by multi-user spatial multiplexing, as soon as the number of users is large enough and the CSIT accuracy is sufficiently good to achieve a higher multiplexing gain with MU-MIMO, irrespective of the networking architecture. In the CAS, the MU-MIMO performance attained with memoryless quantization, employing the RVQ codebook described in Section 4.2.1, is

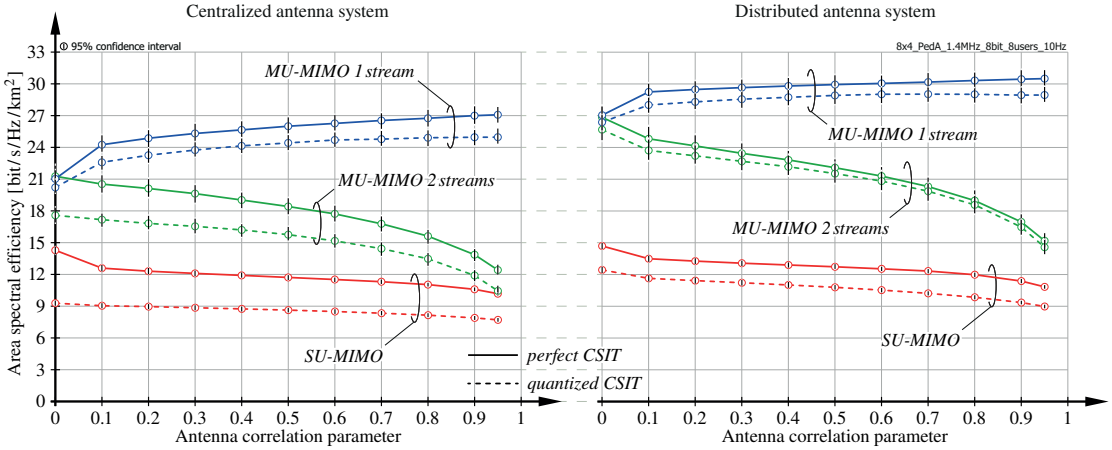


Figure 5.3.: Area spectral efficiency achieved in the CAS and the DAS in dependence of the receive antenna correlation. An $N_i \times M_{u,i} = 8 \times 4$ antenna configuration is considered, serving $U_i = 8$ user over $\ell_{u,i} \in \{1, 2\}$ streams per user in case of MU-MIMO, and over up to $\ell_{u,i} = 4$ streams in case of SU-MIMO.

significantly below the efficiency obtained with perfect CSIT as well as with predictive quantization, using the quantizer of Section 4.2.2. In the DAS, this gap is reduced if the pathloss differences experienced with respect to the RRUs are utilized during CSI quantization, by means of the correlated RVQ codebook proposed in Section 4.2.1. A similar effect, albeit not as significant, is noticed with SU-MIMO. Here, the performance of CLSM is improved in the DAS if antenna subset selection, as proposed in Section 3.2.4, is employed. Comparing the ASE of the CAS and the DAS, a substantial improvement of the MU-MIMO performance in the DAS is observed, while the SU-MIMO efficiency is very similar in both systems. Considering MU-MIMO with memoryless quantization, an improvement of approximately $100\,\%$ is achieved with the DAS over the CAS, if the number of users U_i is larger than four. In the following simulations $U_i = 8$ is assumed.

Next, in Figure 5.3, the impact of the antenna correlation parameter α_{corr} on the SU- and MU-MIMO throughput is investigated. In case of MU-MIMO, zero forcing (ZF) beamforming with $\ell_{u,i} = 1$ as well as BD precoding with $\ell_{u,i} = 2$ are considered, assuming perfect CSIT and predictive quantization. A strong performance degradation of BD precoding with growing antenna correlation is observed in both, the CAS and the DAS. This is in line with theoretical investigations, e.g., [185], and is caused by the increased imbalance between the singular values of the channel matrix with stronger antenna correlation [153]. The same effect leads to an improvement of the ASE achieved with ZF beamforming, because the magnitude of the maximum singular value grows with the antenna correlation. With uncorrelated antennas, $\alpha_{\text{corr}} = 0$, very similar efficiency is attained with ZF beamforming and BD precoding. Theoretically, ZF should be outperformed by BD at this point. This is not observed here, because the multiplexed data streams of a user are independently detected using linear MMSE equalization, instead of joint maximum likelihood detection. To optimize the performance of the MU-MIMO transmission, the per-user transmission rank $\ell_{u,i}$ should hence be

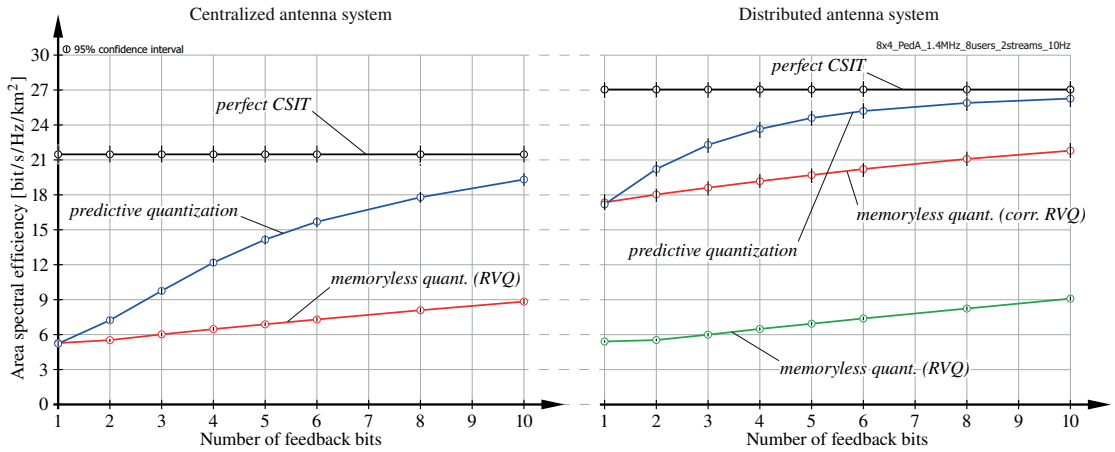


Figure 5.4.: Area spectral efficiency achieved in the CAS and the DAS in dependence of the feedback overhead. An $N_i \times M_{u,i} = 8 \times 4$ antenna configuration is considered, serving $U_i = 8$ user over $\ell_{u,i} = 2$ streams per user employing BD based MU-MIMO.

adapted to the channel conditions, which is not considered in this thesis. The impact of the antenna correlation on the SU-MIMO performance is not significant, because mostly low rank transmission is employed in this case, due to the low SINR experienced by the users.

In the remaining simulations, the focus is put on MU-MIMO, due to the similar performance of SU-MIMO in the CAS and the DAS. Also, $\alpha_{\text{corr}} = 0$ is optimistically assumed from now on. The improvement of the ASE achieved with BD precoding ($\ell_{u,i} = 2$) and increasing feedback overhead is evaluated in Figure 5.4. In the CAS, even 10 bit of feedback per subframe is by far insufficient to achieve close-to-optimal performance with memoryless quantization. This observation applies to the DAS as well, if the gain differences between the RRUs are not exploited during quantization (RVQ). The important observation from Figure 5.4 is that the gap between perfect CSIT and memoryless quantization is substantially reduced in the DAS, if knowledge of the channel gain matrix (2.11) is exploited during quantization (corr. RVQ), to improve the quantization accuracy of those RRUs that are received strongly.

Finally, the impact of the temporal channel correlation on the ASE obtained with predictive quantization is investigated for the case of ZF beamforming in Figure 5.5. The temporal correlation is quantified with the normalized channel Doppler frequency (3.33). As shown in Figure 5.5, perfect performance is achieved by the predictive quantizer at low Doppler frequencies. With reducing temporal channel correlation, i.e., with growing Doppler frequency, the efficiency is deteriorated to that of memoryless quantization. The actual Doppler frequency at which this transition occurs is determined by the size of the quantization codebook (5 bit of feedback per transmission time interval (TTI) are considered here). Very similar behavior is observed in the CAS and in the DAS. In the DAS, however, the performance of predictive quantization is only reduced to that of correlated RVQ at high Doppler frequencies. Hence, even if the channel is temporally uncorrelated, the spatial

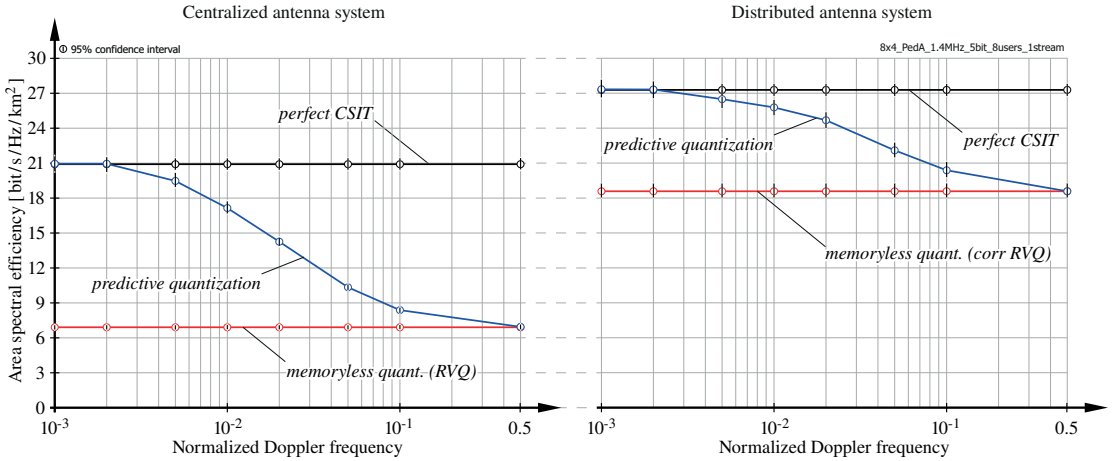


Figure 5.5.: Area spectral efficiency achieved in the CAS and the DAS in dependence of the channel Doppler frequency. An $N_i \times M_{u,i} = 8 \times 4$ antenna configuration is considered, serving $U_i = 8$ user over $\ell_{u,i} = 1$ stream per user employing ZF based MU-MIMO.

correlation is still exploited by the predictive quantizer, because the quantizer automatically adjusts to the statistics of the channel using the adaptive codebook construction described in Section 4.2.2.

5.3. Distributed Antenna Systems versus Small Cells

Small cells are considered as one of the most promising tools to cope with the exponentially increasing demand on network capacity, by placing autonomous access points (micro, pico or femto base stations) at user hot spot locations, to offload traffic from the macro cellular network. This approach is advantageous if the small cells are well isolated from the macro network (e.g., through walls of a building), such that the access points do not cause excessive interference to the macro base stations. If such isolation is not in place, coordination of the macro and micro layers of the cellular network may be required to handle the out-of-cell interference. Compared to DASs, small cells can be deployed cheaper and more flexible, because the dedicated low-latency high-bandwidth connection to the macro network is not required. The access points, however, are autonomous base station and are thus composed of more complex and expensive signal processing hardware. Considering these points, it can be difficult for network operators to make a sensible decision on hardware investments.

In this section, a performance comparison between the DAS and the small cell system shown in Figure 5.1 is conducted. It is assumed that there is no isolation, e.g., in the form of a wall loss, between the small cells, respectively the RRUs, and the macro network. The considered access points can be classified as micro base stations [186], due to their relatively high transmission power. There are two sides to this investigation: firstly, it can help making decisions on investments into new radio equipment. Secondly, and maybe even more importantly, the investigation can be viewed

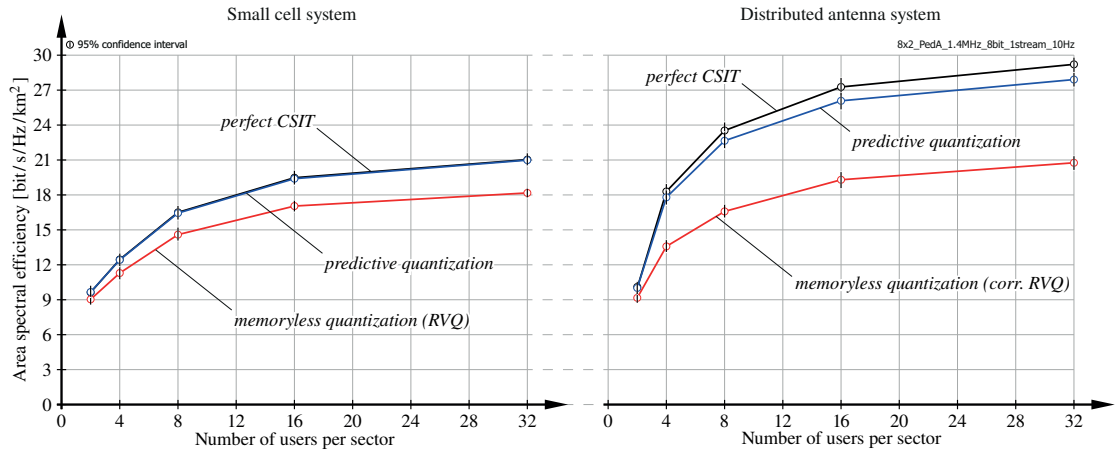


Figure 5.6.: Area spectral efficiency achieved in the DAS and the small cell system in dependence of the number of users per sector. An $N_i \times M_{u,i} = 8 \times 2$ antenna configuration is considered, transmitting $\ell_{u,i} = 1$ stream per user employing ZF based MU-MIMO.

as a comparison between non-cooperative transmission and joint transmission coordinated multi-point (CoMP) between the macro and micro layers of an already existing network, hence, evaluating the potential of coordinated transmission in the downlink of cellular networks. Notice though that the power constraints in the small cell system and the DAS are not the same. In the DAS, a total power constraint of $43\,\text{dBm}$ is considered, irrespective of how this power divides onto the individual antennas; in the small cell system, however, individual power constraints are invoked per base station to ensure the same total power consumption (see Section 5.1 for details).

The focus in this section is put on MU-MIMO, because the gain achieved with distributed antennas and SU-MIMO is relatively small, as shown in the previous section. ZF beamforming based MU-MIMO is considered with users that are equipped with $M_{u,i} = 2$ receive antennas. In the small cell system, feedback is provided by the users only to that micro or macro base station they are attached to. Based on this CSIT, the precoders are calculated individually for each access point and each sector of the macro base station. A normalized Doppler frequency of $\nu_d = 0.01$ is considered.

The ASE in dependence of the number of users per sector is shown in Figure 5.6, assuming a feedback overhead of $8\,\text{bit}$ per subframe, i.e., $8\,\text{kbit/s}$. It can be observed that the difference between perfect CSIT and memoryless quantization in the small cell system is even smaller than in the DAS. This is because the channel matrix is of smaller dimensions.[2] Also, the ASEs attained with predictive quantization and perfect CSIT are overlapping. Comparing the two networking architectures with perfect CSIT, a gain of approximately $40 - 45\,\%$ is achieved with the DAS, enabling coherent data transmission from multiple non-colocated antenna arrays. With memoryless quantization, this gain is reduced to $15 - 25\,\%$, depending on the number of served users. Therefore, the ASE of the cellular

[2]The small cell channel matrix is either 4×2 or 2×2, depending on whether the user is attached to a macro or micro base station. The DAS channel, on the other hand, is of size 8×2.

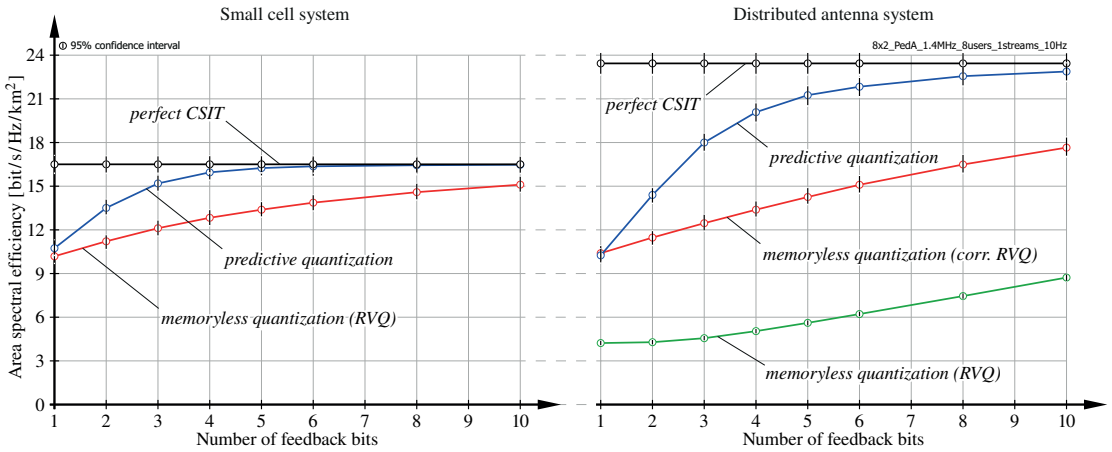


Figure 5.7.: Area spectral efficiency achieved in the DAS and the small cell system in dependence of the feedback overhead. An $N_i \times M_{u,i} = 8 \times 2$ antenna configuration is considered, serving $U_i = 8$ user over $\ell_{u,i} = 1$ stream per user employing ZF based MU-MIMO.

network can be improved by implementing joint transmission CoMP. Whether the attained gain pays off is determined by the expenditure involved in facilitating coherent data transmission from multiple base stations. To provide a more comprehensive answer, the improvement of the cell edge throughput should be evaluated as well, which is considered as the main application scenario of CoMP.

In the second simulation, the impact of the feedback overhead on the ASE is determined, to evaluate if the gains observed above are robust with respect to the accuracy of the CSIT. The simulation results are shown in Figure 5.7, assuming eight users per sector. In the small cell system, almost perfect performance is achieved with memoryless quantization and 10 bit of feedback per TTI; the same performance is obtained with predictive quantization using only 3 bit of feedback in the considered low mobility scenario. In the DAS, the difference between memoryless quantization and perfect CSIT is larger, thus requiring a larger feedback overhead to get equally close-to-optimal. The important observation from Figure 5.7 is that the efficiency achieved in the DAS is always above that of the small cell system, provided the pathloss differences of the channel gain matrix are exploited during quantization, i.e., with correlated RVQ and predictive quantization.

5.4. Summary

In this chapter, the ASE obtained with the transceivers and limited feedback algorithms proposed in Chapters 3 and 4 is compared, assuming three different cellular networking architectures. In the presented simulations, the worst performance is observed in the classical CAS, consisting of macro base stations only. The efficiency of the network can be improved by augmenting the macro base stations with either RRUs, to generate a DAS, or with autonomous small cells. If the small

cells are not well isolated from the macro network, excessive interference between the layers of the network is the consequence. In this case, it is possible to improve the network efficiency by enabling coordination among several transmission points. This is investigated in this chapter by comparing the small cell system to the DAS, demonstrating ASE gains in the order of $20 - 40\,\%$ with distributed antennas. In conclusion, the best performance is attained with the DAS, facilitating coherent data transmission from several non-colocated antenna arrays. This gain, however, comes at the significant cost of providing low-latency high-bandwidth dedicated connections between the RRUs and the macro base station to enable data and CSI exchange.

Although the performance of practical communication systems can be represented more realistically with sophisticated link and system level simulations compared to achievable rate calculations, often important variables and factors cannot be represented accurately by the physical layer models and by the applied abstractions. Important restrictions of the employed simulation environment, which should be kept in mind when evaluating the results, are noticed below.

The presented investigations are conducted with a hybrid of link and system level simulations. With this approach, the physical layer details of the cell of interest, consisting of three macro cellular sectors plus additional radio equipment such as RRUs and access points, are fully covered by the employed simulator to facilitate application of the proposed transceivers and limited feedback algorithms. Out-of-cell interference from other cells of the network, however, is only modeled, utilizing the interference model described in Appendix I. This approach, albeit greatly reducing computational complexity compared to full link level simulations, has the disadvantage that the mutual coupling in the scheduling decisions of the cell of interest and the other cells is not considered, i.e., the other cells are assumed as fully loaded all the time, causing maximal interference.

Another important influence of the wireless channel, which has not been considered in the presented simulations, is macroscopic fading due to shadowing, caused by obstacles such as buildings. Shadowing has been neglected to limit the size of the network that has to be simulated by means of link level simulations, because without shadow fading the area served by each cell is purely determined by the geometry of the network. RRUs and small cells are actually considered as important tools to mitigate the impact of shadowing, by providing additional macro-diversity. When shadow fading is taken into account, also the placement of the additional radio equipment should be optimized such as to maximize, e.g., the capacity or the coverage of the network. Such effects, however, are not in the scope of the presented simulations, which rather focus on the advantage provided by RRUs and small cells in terms of CSI requirements and network efficiency.

A possible point of criticism is the regular structure of the investigated networks. Especially when considering the increasing heterogeneity of cellular networks, the relevance of such a hexagonal network tessellation is decreasing; cf. [187]. In general, the applied out-of-cell interference model can be used with any network geometry. As with shadowing, the regular structure is considered to obtain a simple geometric boundary between the cells. It is not the intention of my simulations to

represent the performance of a typical cell of a real wireless network, but rather to demonstrate the advantage of MU-MIMO transmission in DASs and small cell systems compared to CASs.

Finally, a few remarks to the applied physical layer models: the correlation of the wireless channel in space and frequency is determined by power delay profile based channel models in connection with a spatial correlation matrix in Kronecker form (see Section 2.4). These models are parametrized by a few variables and facilitate interpretation of the obtained simulations results. More realistic models have been published that account for the antenna geometry at the transmitter and the receiver, e.g., [188]. With these models, however, the correlation parameters are implicitly determined by the antenna geometry and the considered surrounding environment, complicating the evaluation of the simulation outcome. Also, the applied models are only two-dimensional; hence, three-dimensional signal propagation, enabling 3D MIMO and elevation beamforming, is not covered.

Chapter 6.

Conclusions

> There is a single light of science, and to brighten it anywhere is to brighten it everywhere.
>
> *(Isaac Asimov)*

Life without ubiquitous possibilities to connect to the Internet is hard to imagine nowadays. Wireless access is provided even in the most remote places by cellular networks. Consumer acceptance of novel mobile devices, such as tablet computers and smartphones, and associated multimedia and Internet driven applications is high, causing an immense growth in mobile data traffic and forcing network providers, standardization bodies as well as researcher to react and to devise feasible solutions, in order to improve the quality of service of wireless communication systems.

The basic cellular networking technology for the next ten years at least is specified by 3GPP's LTE and is currently being globally deployed by mobile operators. Accordingly, viable short- and mid-term technological improvements have to be compatible with the standardized design underlying LTE. Often such restrictions are disregarded by researchers for the sake of generality and mathematical tractability, hampering the consideration of state-of-the-art research in the standardization process.

The main contribution of this dissertation is the development of limited feedback algorithms and transceiver architectures to enable the implementation of advanced single-user (SU-) and multi-user MIMO (MU-MIMO) techniques in LTE compliant cellular networks. Even though the concepts underlying the proposed methods are derived having generality of the obtained solutions and analytical tractability in mind, their practical relevance and applicability is considered of equal value. Below, a summary of the most significant contributions and findings of this thesis is provided.

6.1. Summary of Contributions

The proposed SU-MIMO limited feedback algorithms are designed to optimize the downlink throughput achieved with the currently prevalent implementation of single-user spatial multiplexing in standardized communication technologies, such as LTE, WiMAX and WiFi. In such systems, a

combination of codebook based precoding and transmission rank adaptation is employed to trade-off the beamforming capabilities of the transmitter for the spatial multiplexing gain achieved with MIMO, depending on the SNR operating regime. In combination with adaptive modulation and coding, a flexible transceiver is obtained that supports a wide range of SNRs and is able to exploit the diversity provided by the wireless channel.

In Chapter 3 of this thesis, SU-MIMO transmission in compliance with LTE's closed-loop spatial-multiplexing transmission mode is considered. Dynamic adjustment of the transmission parameters to the instantaneous channel quality is enabled by the proposed limited feedback algorithms. Although instantaneous feedback is prone to a delay in the feedback path, it is demonstrated that reasonable lags can be compensated in low to moderate mobility scenarios, using channel prediction at the receivers. Without delay compensation, the proposed instantaneous channel state information (CSI) feedback is restricted to quasi-stationary situations.

LTE's single-user transmission capabilities, employing the proposed feedback algorithms, are evaluated by means of detailed link level simulations of the LTE physical layer (PHY). A comparison of the downlink throughput of LTE and the MIMO channel capacity is conducted, revealing a throughput loss of $60 - 70\,\%$, even without consideration of impairments caused by the signal-processing (e.g., channel estimation, synchronization) and the hardware (e.g., I/Q imbalance, amplifier nonlinearities). The reasons of the performance degradation are investigated by deriving upper bounds on the achievable throughput, accounting for the constraints imposed by the standard. The throughput loss cannot be attributed to a single dominant source, but rather divides evenly among signaling overhead and architectural restrictions. Hence, significant improvements of the LTE PHY are possible only with a comprehensive revision of the standard. With the proposed feedback algorithms a throughput within $1\,\%$ of the optimal performance of LTE, as obtained from exhaustive search, is achieved.

The second major part of this thesis is devoted to spatial multiplexing of several users. MU-MIMO has several important benefits over SU-MIMO, which make it attractive for practical implementations. Most importantly, the potential spatial multiplexing gain in the downlink of the cellular network is not confined by the number of receive antennas per user, as in SU-MIMO. Therefore, the hardware complexity of the communication system can be offloaded to the base stations of the access network, facilitating cheap and small user equipments. MU-MIMO operation is not yet standardized in LTE, although the foundation has been set with the incorporation of non-codebook based precoding in Rel'10 of the standard specifications. In Chapter 4, a viable transceiver architecture for limited feedback based MU-MIMO in LTE is proposed, encompassing multi-user scheduling and precoding at the base station, as well as antenna combining and CSI feedback calculation at the users.

At the transmitter, block diagonalization (BD) precoding is employed to orthogonalize the transmissions to multiple users. The subspace spanned by the user's channel matrix is identified as the feedback information required at the base station for precoder calculation. Accurate CSI at the transmitter (CSIT) is critical to minimize the residual multi-user interference observed by the

users. Efficient quantization of the subspace is enabled with the proposed predictive Grassmannian quantizer, exploiting the temporal correlation of the wireless channel and the invariances of the subspace information to achieve high fidelity quantization in low to moderate mobility scenarios with a reasonable feedback overhead.

When additional degrees of freedom (DoF) are provided by the users via excess receive antennas, the residual multi-user interference (or the feedback overhead) can further be reduced by applying the proposed subspace quantization based combining (SQBC) subspace selection and antenna combining strategy. It is shown that a linear reduction of the slope of the feedback bit scaling law, governing the feedback overhead to obtain a constant loss with respect to perfect CSIT, with the number of excess antennas is achieved by application of SQBC antenna combining. Subspace selection based on SQBC is also an effective approach for extending the feedback methods to OFDM by means of subcarrier clustering. Combining feedback clustering with predictive quantization, close to optimal throughput performance is demonstrated with a reasonable feedback overhead over the practically relevant SNR range from $0\,\mathrm{dB}$ to $30\,\mathrm{dB}$.

The proposed methods are applied in Chapter 5 to evaluate the performance of LTE, utilizing three different concepts of cellular networking architectures. In this simulation based study, the highest area spectral efficiency (ASE) is achieved with the distributed antenna system (DAS), enabling coherent data transmission from all transmit antennas within the cell, provided sufficiently accurate CSIT is available. Both, the centralized antenna system (CAS) as well as the small cell system, are significantly outperformed by the DAS when predictive quantization is considered. Only when the CSIT in the DAS is not sufficiently accurate to enable efficient MU-MIMO operation, the best performance is achieved with the small cell system. Correspondingly, with accurate CSIT it is possible to improve the operation of small cells within the macro network by applying joint transmission coordinated multi-point (CoMP), effectively transforming the small cell network into a DAS.

6.2. Open Issues and Outlook

Despite the effort put in the work on this dissertation, there are still a few issues left that require some further investigations and research work to enable practical implementations of the proposed methods. One issue, considering the practical applicability of the derived limited feedback algorithms, is computational complexity. Complexity has not been taken into account during the development of the proposed techniques; the focus was instead put on the achieved performance. Complex operations, such as matrix inversions and singular value decompositions (SVDs), are extensively employed in the methods, despite their feasibility in practical realizations. Hence, potentials for savings in terms of algorithmic complexity have to be investigated to enable efficient implementations.

Another issue that is not sufficiently researched yet is the determination of the required feedback overhead to achieve sufficiently accurate CSIT in case of feedback clustering, predictive quantization

and correlated random vector quantization (RVQ). Only for RVQ of independent and identically distributed (i.i.d.) Rayleigh fading channels, it was possible to derive the feedback bit scaling law versus the SNR, because bounds on the average chordal distance quantization distortion are available in literature. Deriving similar bounds for spatially correlated channels was not yet successful. Similarly, it is unknown how large the theoretical potential of predictive Grassmannian quantization of temporally correlated subspaces is. Hence, the question how close-to-optimal the proposed quantizer performs cannot be answered yet. Information theoretic results on the rate-distortion trade-off of this Grassmannian quantization problem would be valuable.

The impact of impairments in the feedback channel has to be further investigated. Specifically, feedback errors are problematic in case of predictive quantization, because the prediction operation causes error propagation, similar to the effects observed in video codecs. Such errors can be resolved with a simple cyclic-redundancy-check in combination with a selective-repeat automatic repeat-request protocol or with a synchronized reset of the quantizers at the encoder and the decoder. Alternatively, a concept similar to the I-frame in video coding [189] can be employed to periodically provide a valid reference. In either case, the additional feedback and signaling overhead must be taken into account.

Finally, the BD precoding strategy itself has to be scrutinized. Albeit the fact that block diagonalization is DoF optimal for the MIMO broadcast channel, a performance loss is incurred at low SNR, when the multi-user interference is dominated by the noise. Precoding approaches that are less susceptible to the receiver noise have been proposed, e.g., regularized block diagonalization [190]. It is unclear if such methods can be adjusted to operate with subspace feedback or if full feedback of the channel matrix is required. Alternatively, an approach similar to shape-gain vector quantization [76] is conceivable, in which subspace quantization is combined with an independent quantization of the singular values of the channel matrix, to additionally provide information about the channel gain.

6.3. Conclusion

Despite the fact that already Shannon in 1956 showed that feedback does not increase the capacity of memoryless point-to-point channels with infinite code block length [191], feedback has still proved to be one of the most successful tools in wireless communications to enable robust and efficient data transmission. With feedback it is possible to adapt the transmission parameters to the varying channel conditions, thereby exploiting the diversity of the wireless channel and improving the achievable data rate with finite code block length [192]. Even more important, the capacity of multi-point channels is substantially increased with feedback, see, e.g., [193], [194]. I am therefore confident that this dissertation provides a valuable contribution towards advancing limited feedback based single- and multi-user MIMO communications, to improve the efficiency and capacity of cellular networks.

Appendix A.

List of Abbreviations

3GPP	Third Generation Partnership Project
AMC	adaptive modulation and coding
ASE	area spectral efficiency
AWGN	additive white Gaussian noise
BD	block diagonalization
BER	bit-error ratio
BICM	bit-interleaved coded modulation
BICM LR	bit-interleaved coded modulation (BICM) with linear receiver
BLER	block-error ratio
CAS	centralized antenna system
CDMA	code division multiple access
CLMI	closed loop mutual information
CLMI LR	CLMI with linear receiver
CLSM	closed loop spatial multiplexing
CoMP	coordinated multi-point
CQI	channel quality indicator
CSI	channel state information
CSIT	CSI at the transmitter
DAS	distributed antenna system
DoF	degrees of freedom
DPC	dirty paper coding
EDGE	Enhanced Data Rates for GSM Evolution
EESM	exponential effective SINR mapping
ESM	effective SINR mapping
ETSI	European Telecommunications Standard Institute
FDD	frequency division duplex
FEC	forward error correction code
FIR	finite impulse response

GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HARQ	hybrid automatic repeat request
HetNets	heterogeneous networks
HSPA	High Speed Packet Access
IEEE	Institute of Electrical and Electronics Engineers
i.i.d.	independent and identically distributed
IMT-A	International Mobile Telecommunications-Advanced
ITU	International Telecommunications Union
LDPC	low-density parity-check code
LTE	Long Term Evolution
LTE-A	LTE advanced
MCS	modulation and coding scheme
MET	maximum eigenmode transmission
MIMO	multiple-input multiple-output
MIESM	mutual information effective SINR mapping
MMSE	minimum mean-squared error
MSE	mean-squared error
MU-MIMO	multi-user MIMO
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
PHY	physical layer
PMI	precoding matrix indicator
PSK	phase shift keying
QAM	quadratur amplitude modulation
QBC	quantization based combining
RB	resource block
RE	resource element
RI	rank indicator
RLS	recursive least squares
RMS	root mean square
RRU	remote radio unit
RVQ	random vector quantization
SINR	signal to interference and noise ratio
SISO	single-input single-output
SNR	signal to noise ratio
SQBC	subspace quantization based combining
SU-MIMO	single-user MIMO
SUS	semi-orthogonal user selection

SVD	singular value decomposition
TTI	transmission time interval
UMTS	Universal Mobile Telecommunications System
WCDMA	wideband code division multiple access
ZF	zero forcing

Appendix B.

Notation

The following notation is used throughout this dissertation:

Table B.1.: Definition of the employed mathematical notation.

Symbol	Meaning
$a, A \in \mathbb{C}$ $(\mathbb{R})$	complex-valued (real-valued) scalars
$\mathbf{a} \in \mathbb{C}^{n\times 1}$ $(\mathbb{R}^{n\times 1})$	length n complex-valued (real-valued) column vector
$\mathbf{A} \in \mathbb{C}^{n\times m}$ $(\mathbb{R}^{n\times m})$	complex-valued (real-valued) matrix with n rows and m columns
$\mathcal{A}$	set of numbers or other elements
$\mathbf{I}_n$	$n \times n$ dimensional identity matrix
$\mathbf{0}_{n\times m}$	$n \times m$ dimensional matrix with all entries equal to zero
$\lvert a \rvert$	magnitude of a scalar
$\lVert \mathbf{a} \rVert$	l_2 norm of a vector
$\lVert \mathbf{A} \rVert$	Frobenius norm of a matrix
$\lvert \mathcal{A} \rvert$	size of a set
$\mathbf{A}^{\mathrm{T}}$	transpose of a matrix
$\mathbf{A}^{\mathrm{H}}$	conjugate-transpose of a matrix
$\mathbf{A}^{1/2}$	square root of a matrix
$\mathbf{A}^{-1}$	inverse of a matrix
$\mathbf{A}^{\dagger}$	Moore-Penrose pseudo inverse of a matrix
$\mathbb{E}\left(a\right)$	expected value of a
$\mathrm{var}\left(a\right)$	variance of a
$\mathrm{tr}\left(\mathbf{A}\right)$	trace of a matrix
$\det\left(\mathbf{A}\right)$	determinant of a matrix
$\mathrm{diag}\left(\mathbf{a}\right)$	diagonal matrix whose main diagonal entries equal $\mathbf{a}$
$\log_{\mathrm{b}}\left(a\right)$	base b logarithm of a
$\mathrm{vec}\left(\mathbf{A}\right)$	column vector consisting of the stacked columns of $\mathbf{A}$

Notation (continued from previous page)

Symbol	Meaning
$\operatorname{span}(\mathbf{A})$	space spanned by the columns of $\mathbf{A}$
$\operatorname{null}(\mathbf{A})$	kernel (null space) of $\mathbf{A}$
$\mathrm{d}_{\mathrm{c}}^2(\mathbf{Q}_1, \mathbf{Q}_2)$	squared chordal distance between $\operatorname{span}(\mathbf{Q}_1)$, $\operatorname{span}(\mathbf{Q}_2)$
$\mathrm{d}_{\mathrm{c,w}}^2(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{\Lambda})$	squared weighted chordal distance between $\operatorname{span}(\mathbf{Q}_1)$, $\operatorname{span}(\mathbf{Q}_2)$ with diagonal weighting matrix $\mathbf{\Lambda}$
$[\mathbf{A}]_{m:n,i:j}$	selects the submatrix of $\mathbf{A}$ consisting of rows m to n and columns i to j
$[\mathbf{A}]_{:,i}$	selects the i-th column of $\mathbf{A}$
$[\mathbf{A}]_{m,:}$	selects the m-th row of $\mathbf{A}$
$\beta(a, b)$	beta distribution with parameters a and b
$\Gamma(k, \theta)$	gamma distribution with shape k and scale θ
$\mathcal{N}_{\mathbb{C}}(\mathbf{m}, \mathbf{R})$	circularly symmetric complex Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{R}$
$\mathcal{W}_n^{\mathbb{C}}(m, \mathbf{S})$	central complex Wishart distribution of dimension n with m degrees of freedom and scale matrix $\mathbf{S}$
$\mathcal{W}_n^{-\mathbb{C}}(m, \mathbf{S})$	inverse central complex Wishart distribution of dimension n with m degrees of freedom and scale matrix $\mathbf{S}$
$\mathcal{G}(n, m)$	Grassmann manifold of m-dimensional subspaces in the n-dimensional complex Euclidean space ($m \leq n$)
$\mathcal{U}_n$	group of $n \times n$ unitary matrices
$\mathcal{T}(\mathbf{A})$	tangent space of the matrix $\mathbf{A}$

Appendix C.

Grassmann Manifold Basics

In the limited feedback algorithms proposed in this dissertation symmetry and invariance properties of the cost functions and constraints underlying the considered optimization problems are exploited to minimize the required feedback information exchange between the user and the base station, and to improve the efficiency of the derived algorithms and methods. Specifically, the considered precoding strategies are based on channel subspace information, facilitating transmission of the data over a preferred subspace of the channel matrix in SU-MIMO (see Section 3.1.2), or enabling multi-user interference-nulling over certain subspaces of the channel matrices in MU-MIMO (see Section 4.1). Subspace information can efficiently be represented and conveyed using the Grassmann manifold.

The Grassmann manifold has been successfully employed in several areas of MIMO wireless communications, including capacity evaluations of non coherent multiple-antenna channels [70], codebook design for limited feedback single- and multi-user MIMO as well as interference alignment [40, 81, 195], space-time code design [196, 197] and many others. Therefore, in this chapter a short introduction to the Grassmann manifold is provided, covering its matrix representation and a few necessary notions of geometry utilized in this thesis. For a more detailed discussion of the Grassmannian and other related manifolds the interested reader is referred to [198, 199].

C.1. Definition of the Grassmann Manifold

A manifold in general is defined as a topological space that resembles Euclidean space in a neighborhood around each point, i.e., there exists a neighborhood around each point that is homeomorphic to an open subset of the Euclidean space; see, e.g., [200] for a rigorous formal definition. Familiar examples are circles, spheres, parabolas and so on. The Grassmann manifold, also known as the Grassmannian, $\mathcal{G}(m, n, \mathbb{K})$ with $n \leq m$ is the set of all n-dimensional subspaces in the m-dimensional vector space $\mathbb{K}^m$, for example with $\mathbb{K} = \mathbb{C}$. The Grassmannian can also be defined as a quotient manifold. Consider the *non compact Stiefel manifold* of full-rank $m \times n$ matrices defined as

$$\mathcal{S}(m, n, \mathbb{K}) := \left\{ \mathbf{S} \in \mathbb{K}^{m \times n} : \det\left(\mathbf{S}^{\mathrm{H}}\mathbf{S}\right) \neq 0 \right\}, \tag{C.1}$$

and the equivalence relation between points $\mathbf{S}$ and $\mathbf{T}$ on $\mathcal{S}(m, n, \mathbb{K})$

$$\mathbf{S} \equiv \mathbf{T} \Leftrightarrow \mathbf{S} = \mathbf{T}\mathbf{M}, \;\; \mathbf{M} \in \mathcal{GL}(n, \mathbb{K}), \tag{C.2}$$

$$\mathcal{GL}(n, \mathbb{K}) := \left\{ \mathbf{N} \in \mathbb{K}^{n \times n} : \det(\mathbf{N}) \neq 0 \right\}, \tag{C.3}$$

with $\mathcal{GL}(n, \mathbb{K})$ denoting the general linear group of degree n. Elements of the Grassmann manifold are equivalence classes of matrices $\mathbf{S} \in \mathcal{S}(m, n, \mathbb{K})$ with respect to the equivalence relation $\equiv$. Thus, the *Grassmann manifold* is identical to the quotient space [199]

$$\mathcal{G}(m, n, \mathbb{K}) \triangleq \mathcal{S}(m, n, \mathbb{K}) / \mathcal{GL}(n, \mathbb{K}). \tag{C.4}$$

In this dissertation, the vector space $\mathbb{K}$ underlying the considered Grassmannian is the Euclidean space of complex numbers; for simplicity it is written as $\mathcal{G}(m, n) = \mathcal{G}(m, n, \mathbb{C})$.

A point $H \in \mathcal{G}(m, n)$ on the Grassmann manifold can be represented by any matrix $\mathbf{H} \in \mathbb{C}^{m \times n}$ whose columns span the subspace defined by H, i.e., $H = \operatorname{span}(\mathbf{H})$. To unify this representation, orthonormal bases (semi-unitary matrices) are employed throughput this thesis to identify points on the Grassmannian

$$\tilde{\mathbf{H}} \in \mathcal{G}(m, n) \Leftrightarrow \tilde{\mathbf{H}}^{\mathrm{H}} \tilde{\mathbf{H}} = \mathbf{I}_n. \tag{C.5}$$

Notice that $\tilde{\mathbf{H}}$ represents an entire equivalence class in the sense of Equation (C.2). This matrix representation was advocated in [198] and is commonly employed in the corresponding scientific literature.

C.2. Distance Measures on the Grassmannian

When designing quantization codebooks for the Grassmann manifold it is necessary to determine the distance between points on the manifold. Several distance measures have been defined in the literature. In limited feedback MIMO wireless communications the most important distances between two subspaces $\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2 \in \mathcal{G}(m, n)$ are [40]:[1]

- The *chordal distance*

$$\mathrm{d}_{\mathrm{c}}^2\left(\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2\right) = n - \operatorname{tr}\left(\tilde{\mathbf{H}}_1^{\mathrm{H}} \tilde{\mathbf{H}}_2 \tilde{\mathbf{H}}_2^{\mathrm{H}} \tilde{\mathbf{H}}_1\right). \tag{C.6}$$

- The *projection two-norm*

$$\mathrm{d}_{\mathrm{p}}^2\left(\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2\right) = \left\| \tilde{\mathbf{H}}_1 \tilde{\mathbf{H}}_1^{\mathrm{H}} - \tilde{\mathbf{H}}_2 \tilde{\mathbf{H}}_2^{\mathrm{H}} \right\|_{\mathrm{F}}^2. \tag{C.7}$$

[1] Notice that $\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$ must be orthonormal bases, as in (C.5).

- The *Fubini-Study distance*

$$\mathrm{d}_{\mathrm{f}}^2\left(\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2\right) = \left(\arccos\left|\det \tilde{\mathbf{H}}_1^{\mathrm{H}} \tilde{\mathbf{H}}_2\right|\right)^2. \tag{C.8}$$

These distances have been related to different optimization criteria for the design of SU-MIMO precoder codebooks for codebook based precoding in [40]. In more recent work investigating limited feedback based MU-MIMO and MIMO interference alignment [45, 46, 53], the importance of the chordal distance as CSI quantization metric is recognized, because it determines the residual multi-user interference experienced by the users. Hence, the MU-MIMO feedback algorithms proposed in Chapter 4 are based on the chordal distance.

C.3. Geometry on the Grassmannian

The predictive CSI quantizer developed in Section 4.2 is based on geometric concepts associated with the differentiable (smooth) Grassmann manifold, which are specified and described below.

Two points $\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2 \in \mathcal{G}(m, n)$ on the Grassmannian can be related using the concept of a *tangent* $\mathbf{T}_1$ [198, 199, 201]

$$\mathbf{T}_1 = T\left(\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2\right) = \mathbf{U}\mathbf{\Phi}\mathbf{V}^{\mathrm{H}}, \tag{C.9}$$

$$\mathbf{U}\tan(\mathbf{\Phi})\,\mathbf{V}^{\mathrm{H}} = \mathbf{\Theta},\ \mathbf{U} \in \mathbb{C}^{m\times n},\ \mathbf{\Phi} \in \mathbb{C}^{n\times n},\ \mathbf{V} \in \mathbb{C}^{n\times n},$$

$$\mathbf{\Theta} = \left(\mathbf{I}_m - \tilde{\mathbf{H}}_1\tilde{\mathbf{H}}_1^{\mathrm{H}}\right)\tilde{\mathbf{H}}_2\left(\tilde{\mathbf{H}}_1^{\mathrm{H}}\tilde{\mathbf{H}}_2\right)^{-1} \in \mathbb{C}^{m\times n}.$$

Here, the tangent $\mathbf{T}_1 \in \mathcal{T}(\tilde{\mathbf{H}}_1)$ is determined by the matrices $\mathbf{U}$, $\mathbf{\Phi}$ and $\mathbf{V}$, which are obtained from the compact-form SVD of the auxiliary matrix $\mathbf{\Theta}$. Notice that the diagonal matrix $\tan(\mathbf{\Phi})$ is composed of the singular values of $\mathbf{\Theta}$. The matrix $\mathbf{\Phi}$ is obtained by calculating the arctangent of the diagonal elements of $\tan(\mathbf{\Phi})$. The Euclidean *tangent space* associated with the point $\tilde{\mathbf{H}}_1 \in \mathcal{G}(m, n)$ is defined as [198]

$$\mathcal{T}\left(\tilde{\mathbf{H}}_1\right) := \left\{\mathbf{T} \in \mathbb{C}^{m\times n} \middle| \tilde{\mathbf{H}}_1^{\mathrm{H}}\mathbf{T} = \mathbf{0}_{n\times n}\right\}. \tag{C.10}$$

The shortest path between the two points $\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2 \in \mathcal{G}(m, n)$ that lies entirely on $\mathcal{G}(m, n)$, also known as the geodesic, is specified by the tangent $\mathbf{T}_1$ of Equation (C.9). Geodesics on manifolds are a generalization of the concept of straight lines in the Euclidean space. A straight line is defined as a curve with zero acceleration. This can be generalized to curves on surfaces or manifolds by requiring that the acceleration lies only in the normal space of the manifold [198]. Intuitively, this means that the curve only changes direction to stay on the manifold and not to move sideways along

the manifold. The *geodesic* defined by the tangent $\mathbf{T}_1$ in (C.9) parametrized by $p \in [0, 1]$ is obtained as [198]

$$\mathbf{\Gamma}_1(p) = \Gamma\left(\tilde{\mathbf{H}}_1, \mathbf{T}_1, p\right) = \tilde{\mathbf{H}}_1 \mathbf{V} \cos\left(\mathbf{\Phi}\, p\right) \mathbf{V}^{\mathrm{H}} + \mathbf{U} \sin\left(\mathbf{\Phi}\, p\right) \mathbf{V}^{\mathrm{H}}. \tag{C.11}$$

Notice that $\mathbf{\Gamma}_1(0) = \tilde{\mathbf{H}}_1$ and that $\mathbf{\Gamma}_1(1) \equiv \tilde{\mathbf{H}}_2$ in the sense of Equation (C.2). It can also be verified that the derivative of the geodesic at $p = 0$ is equal to the tangent: $\frac{d}{dp}\mathbf{\Gamma}_1(p)\Big|_{p=0} = \mathbf{T}_1$. In the proposed predictive CSI quantizer of Section 4.2 the prediction of points on the Grassmannian is achieved by translating the problem to the tangent space, thereby avoiding the explicit manifold constraint and facilitating reuse of well-known prediction algorithms from the Euclidean geometry.

For that purpose, it is necessary to transport the geometric information contained in a tangent $\mathbf{T} \in \mathcal{T}\left(\tilde{\mathbf{H}}_1\right)$ from one point $\tilde{\mathbf{H}}_1 \in \mathcal{G}\left(m, n\right)$ on the manifold to another point $\tilde{\mathbf{H}}_2 \in \mathcal{G}\left(m, n\right)$, which can be achieved by parallel transport. Parallel transport is a means to move the base of a tangent along a smooth curve on the manifold, such that a valid tangent is obtained at each point along the curve, preserving the inner product with respect to the curve. *Parallel transport* of $\mathbf{T} \in \mathcal{T}\left(\tilde{\mathbf{H}}_1\right)$ from $\tilde{\mathbf{H}}_1$ to $\tilde{\mathbf{H}}_2$ along the geodesic defined in (C.11) is achieved with [198]

$$\mathbf{\Pi}(p) = \Pi\left(\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2, \mathbf{T}, p\right) = -\tilde{\mathbf{H}}_1 \mathbf{V} \sin\left(\mathbf{\Phi} p\right) \mathbf{U}^{\mathrm{H}} \mathbf{T} + \mathbf{U} \cos\left(\mathbf{\Phi} p\right) \mathbf{U}^{\mathrm{H}} \mathbf{T} + \left(\mathbf{I}_m - \mathbf{U}\mathbf{U}^{\mathrm{H}}\right) \mathbf{T}. \tag{C.12}$$

Here, $\mathbf{U}$, $\mathbf{\Phi}$ and $\mathbf{V}$ are obtained from the tangent $\mathbf{T}_1$ of Equation (C.9). With this approach, a valid tangent $\mathbf{\Pi}(1) \in \mathcal{T}\left(\tilde{\mathbf{H}}_2\right)$ is calculated that can be utilized to smoothly extend the geodesic between $\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$ beyond $\tilde{\mathbf{H}}_2$. This approach has been exploited in [81] to achieve a simple one step ahead prediction. In this thesis this idea is extended to an arbitrary prediction order, by interpreting the tangents describing the evolution of the channel subspace over time as a stochastic process that is predicted by means of linear filtering; see Section 4.2.

Appendix D.

Performance Bounds for MIMO OFDM

In this chapter, several increasingly tight and restrictive performance bounds for single user single cell MIMO OFDM wireless communication systems are developed. Since many years the channel capacity of multi-antenna Gaussian channels is well known [38], but still current wireless communication technology is far from achieving throughputs close to capacity. Although modern turbo codes and low-density parity-check codes (LDPCs) are performing close to ideal Shannon codes [202], there are several other sources of performance loss that all together accumulate to a large fraction of the theoretically possible throughput. The most dominant factors of performance degradation are systematically identified below. Starting from Shannon's channel capacity of multi-antenna Gaussian channels, tighter upper bounds on the achievable throughput are derived gradually, incorporating design constraints encountered in practical implementations, such as the system overhead, the restriction to codebook based precoding and the application of BICM instead of Gaussian signaling. These bounds are derived for a system architecture according to Figure 2.2. Closed form expressions for the bounds are not available, but they can be evaluated by Monte Carlo simulations similar to the water-filling solution of the Shannon capacity. Such simulations are performed in Section 3.3 of this thesis, where the derived throughput bounds are applied to investigate the performance of LTE and to analyze the potential sources of the significant throughput loss observed with respect to the Shannon capacity. The presented throughput bounds are a subset of the bounds proposed in my corresponding publication [63], accounting for the most dominant factors of performance degradation. The additive noise is assumed to be i.i.d. Gaussian distributed with variance $\tilde{\sigma}_z^2$. The user and cell indices are omitted for brevity, as only a single cell single user scenario is considered.

Channel capacity: The amount of information that any communication system can transmit reliably over a given channel is upper bounded by the well known Shannon channel capacity [203]. In MIMO OFDM, channel capacity can be achieved by SVD based precoding and reception on each OFDM subcarrier n, along with water-filling power allocation among spatial modes and subcarriers [106]. Power allocation over time is not considered, i.e., an instantaneous power constraint

per OFDM symbol is assumed. According to Equation (2.5), the post-equalization MIMO input-output relationship of the system is

$$\mathbf{y}[n,k] = (\mathbf{H}[n,k]\mathbf{G}[n,k])^{\mathrm{H}}\mathbf{F}[n,k]\mathbf{x}[n,k] + \mathbf{G}[n,k]^{\mathrm{H}}\tilde{\mathbf{z}}[n,k]. \tag{D.1}$$

The SVD of the channel matrix $\mathbf{H}[n,k]$ at subcarrier n and OFDM symbol k is written as

$$\mathbf{H}[n,k] = \mathbf{U}[n,k]\mathbf{\Sigma}[n,k]\mathbf{V}[n,k]^{\mathrm{H}}, \tag{D.2}$$

$$\mathbf{\Sigma}[n,k] = \mathrm{diag}\left(\sigma_1[n,k], \ldots, \sigma_{\ell_{\max}}[n,k]\right),$$

where the number of spatial modes is bounded as $\ell_{\max} \leq \min(N, M)$ and is determined by the rank of the channel matrix (it is assumed that $\ell_{\max}$ does not change over subcarriers). Provided the precoder and the receive filter are set as

$$\mathbf{F}[n,k] = \mathbf{U}[n,k]\mathbf{P}[n,k]^{1/2}, \tag{D.3}$$

$$\mathbf{G}[n,k] = \mathbf{V}[n,k], \tag{D.4}$$

with $\mathbf{P}[n,k] = \mathrm{diag}\left(p_1[n,k], \ldots, p_{\ell_{\max}}[n,k]\right)$ being a diagonal power loading matrix, the MIMO channel is decomposed into a set of $\ell_{\max}$ non-interfering AWGN SISO channels

$$[\mathbf{y}[n,k]]_\nu = \sqrt{p_\nu[n,k]}\sigma_\nu[n,k]\,[\mathbf{x}[n,k]]_\nu + \tilde{z}[n,k], \quad \nu \in \{1, \ldots, \ell_{\max}\}, \tag{D.5}$$

with $\tilde{z}[n,k] \sim \mathcal{N}_{\mathbb{C}}\left(0, \tilde{\sigma}_z^2\right)$. The optimal power distribution over subcarriers and spatial streams is obtained from the water-filling power allocation policy [106], solving the optimization problem

$$C[k] = \max_{p_\nu[n,k]} \sum_{\nu=1}^{\ell_{\max}} \sum_{n=1}^{N_{\mathrm{tot}}} \log_2\left(1 + \frac{p_\nu[n,k](\sigma_\nu[n,k])^2}{\tilde{\sigma}_z^2}\right) \tag{D.6}$$

$$\text{subject to: } \sum_{\nu=1}^{\ell_{\max}} \sum_{n=1}^{N_{\mathrm{tot}}} p_\nu[n,k] = N_{\mathrm{tot}}P, \tag{D.7}$$

with P being the average transmit power per subcarrier and N_{tot} denoting the total number of OFDM subcarriers.

Achievable capacity: The channel capacity does not consider the most obvious performance loss encountered in practical systems, due to overhead such as the transmission of the OFDM cyclic prefix to avoid inter symbol interference [106], the insertion of reference symbols for timing- and frequency-synchronization [204] as well as channel estimation [39], and the inclusion of guard bands to ensure negligible interference to other frequency bands. Knowing the specifications of the considered technology, these constraints can easily be accounted for by multiplying the channel capacity with an efficiency factor assessing the loss of available bandwidth for data transmission caused by the technology [59].

Closed loop mutual information: With codebook based precoding, the precoders that can be applied at the base station are restricted to a predefined set (codebook) $\mathcal{Q}_\ell^{(N)}$, e.g., as defined in (3.8). Also, the variability of the precoding in the time-frequency domain is often confined by clustering, in order to reduce the CSI feedback overhead (see Section 3.2). With clustering, the same precoder is used for a set of $N_{\text{clust}}^{(f)}$ consecutive subcarriers n and a set of $N_{\text{clust}}^{(t)}$ consecutive time instants k, dividing the time-frequency resource grid into a set of N_{RB} resource blocks (RBs) per time slot κ, as detailed in Section 3.2. The same transmission rank $\ell[\kappa]$ is thereby applied over all frequency domain clusters $\eta \in \{1, \ldots, N_{\text{RB}}\}$ per time slot. In that case, an optimum power allocation cannot be found by means of water-filling, because the applied precoder does in general not diagonalize the channel. Hence, uniform power allocation is commonly applied in practice and is assumed in the following for simplicity. The bound can easily be accommodated for other forms of power allocation, e.g., to consider pathloss compensation.

Taking these restrictions into account, the achievable transmission rate on resource element (RE) $[n, k]$, assuming the application of precoder $\mathbf{Q}[\eta, \kappa] \in \mathcal{Q}_{\ell[\kappa]}^{(N)}$, is determined by the mutual information [142]

$$I[n,k] = \log_2 \det \left(\mathbf{I}_M + \frac{P}{\tilde{\sigma}_z^2 N} \, \mathbf{H}[n,k]^{\text{H}} \mathbf{Q}[\eta,\kappa] \mathbf{Q}[\eta,\kappa]^{\text{H}} \mathbf{H}[n,k] \right), \tag{D.8}$$

which can be obtained by applying a maximum likelihood MIMO detector at the receiver.[1] The precoder $\mathbf{Q}[\eta, \kappa]$ is here indexed by the RB index $[\eta, \kappa]$ instead of the RE index $[n, k]$ to take the clustering into account. The relationship between these index pairs is defined in Equation (3.14). For simplicity the same notation is employed in both cases; the correct interpretation follows from the context. Similarly, the transmission rank $\ell[\kappa]$ is indexed by the slot index κ instead of the OFDM symbol index k.

The *closed loop mutual information (CLMI)* bound is defined as the maximum sum mutual information over clusters η, with respect to the precoders and with respect to the transmission rank $\ell[\kappa] \in \{1, \ldots, \ell_{\text{max}}\}$

$$I^{(\text{CL})}[\kappa] = \max_{\ell[\kappa]} \sum_{\eta=1}^{N_{\text{RB}}} \max_{\mathbf{Q}[\eta,\kappa] \, \in \mathcal{Q}_{\ell[\kappa]}^{(N)}} \sum_{\rho=1}^{N_{\text{RE}}} \log_2 \det \left(\mathbf{I}_M + \frac{P}{\tilde{\sigma}_z^2 N} \, \mathbf{H}[n,k]^{\text{H}} \mathbf{Q}[\eta,\kappa] \mathbf{Q}[\eta,\kappa]^{\text{H}} \mathbf{H}[n,k] \right). \tag{D.9}$$

Here, the RB specific RE index ρ, defined in Section 3.2.1, is employed to simplify notations. The corresponding absolute RE index $[n, k]$ is implicitly determined by Equation (3.13). Comparing the channel capacity and the CLMI bound, the following inequality holds

$$I^{(\text{CL})}[\kappa] \leq C[\kappa] = \sum_{k=(\kappa-1)N_{\text{clust}}^{(t)}+1}^{\kappa N_{\text{clust}}^{(t)}} C[k], \tag{D.10}$$

[1] Notice the application of the semi-unitary precoder $\mathbf{Q}$ here, instead of its scaled version $\mathbf{F}$ as in (3.9).

with $C[k]$ from Equation (D.6), and the sum going over all symbol indices k that correspond to slot index κ; cf. Section 3.2.1.

Closed loop mutual information with linear receiver: In practical implementations, maximum likelihood detection at the users is often too complex. Instead, linear receive filters are frequently employed to separate the spatial data streams, followed by independent decoding of the individual streams. If the precoder $\mathbf{Q}[n,k]$ and the receive filter $\mathbf{G}[n,k]$ are applied on RE $[n,k]$, the per-stream post-equalization SINR is obtained from Equation (2.7). In the considered single cell single user case, the post-equalization SINR of stream ν with equal power allocation among streams is

$$\beta_\nu[n,k] = \frac{\frac{P}{\ell[k]} \left| \mathbf{g}_\nu[n,k]^{\mathrm{H}} \mathbf{H}[n,k]^{\mathrm{H}} \mathbf{q}_\nu[n,k] \right|^2}{\frac{P}{\ell[k]} \sum_{\mu=1,\mu\neq\nu}^{\ell[k]} \left| \mathbf{g}_\nu[n,k]^{\mathrm{H}} \mathbf{H}[n,k]^{\mathrm{H}} \mathbf{q}_\mu[n,k] \right|^2 + \tilde{\sigma}_z^2 \left\| \mathbf{g}_\nu[n,k] \right\|^2}, \tag{D.11}$$

with $\mathbf{g}_\nu[n,k] = [\mathbf{G}[n,k]]_{:,\nu}$ and $\mathbf{q}_\nu[n,k] = [\mathbf{Q}[n,k]]_{:,\nu}$. Assuming Gaussian signaling, the achievable transmission rate on RE $[n,k]$ with linear receiver is hence given by

$$R[n,k] = \sum_{\nu=1}^{\ell[k]} \log_2 \left(1 + \beta_\nu[n,k]\right). \tag{D.12}$$

Taking into account the precoder clustering and the constant transmission rank per time slot as before, the *CLMI with linear receiver (CLMI LR)* throughput bound is obtained by maximizing the sum rate over clusters with respect to the precoders and the transmission rank

$$R^{(\mathrm{CL,LR})}[\kappa] = \max_{\ell[\kappa]} \sum_{\nu=1}^{N_{\mathrm{RB}}} \max_{\mathbf{Q}[\nu,\kappa] \in \mathcal{Q}_{\ell[\kappa]}^{(N)}} \sum_{\rho=1}^{N_{\mathrm{RE}}} \sum_{\nu=1}^{\ell[\kappa]} \log_2 \left(1 + \beta_\nu[n,k]\right) \leq I^{(\mathrm{CL})}[\kappa], \tag{D.13}$$

with $[n,k]$ from Equation (3.13).

Bit-interleaved coded modulation with linear receiver: To achieve the channel capacity and mutual information bounds derived above, it is required to apply Gaussian signaling for data transmission [132]. Hence, the information theoretic correct approach of modulation and coding is to apply signal-space coding, which was first considered in Ungerböck's work on coded modulation [205]. Still, for complexity reasons, separate modulation and coding is preferred in practice. Since the discovery of strong binary forward error correction codes (FECs) such as turbo codes and LDPCs, performance close to Shannon capacity is nevertheless possible provided the coding and modulation stages are joined over a powerful bit-interleaver, forming a BICM architecture [123]. The capacity of BICM is derived in [122] and is compared in Figure 3.2 to Shannon capacity assuming a SISO AWGN channel, illustrating the good performance of BICM over a wide SNR range.

In the following, a BICM architecture is assumed that supports a set $\mathcal{A}$ of modulation alphabets. Denoting the SISO AWGN BICM capacity achieved with modulation alphabet $A \in \mathcal{A}$ in dependence of the SNR β as $B_A(\beta)$ (cf., Figure 3.2), the BICM system capacity is defined as

$$B(\beta) = \max_{A \in \mathcal{A}} B_A(\beta). \tag{D.14}$$

The throughput bound of ***BICM with linear receiver (BICM LR)*** is obtained by replacing the Shannon capacity $\log_2(1+\beta)$ in Equation (D.13) with the BICM system capacity $B(\beta)$

$$B^{(\mathrm{CL,LR})}[\kappa] = \max_{\ell[\kappa]} \sum_{\eta=1}^{N_{\mathrm{RB}}} \max_{\mathbf{Q}[\eta,\kappa] \in \mathcal{Q}_{\ell[\kappa]}^{(N)}} \sum_{\rho=1}^{N_{\mathrm{RE}}} \sum_{\nu=1}^{\ell[\kappa]} B(\beta_\nu[n,k]) \leq R^{(\mathrm{CL,LR})}[\kappa]. \tag{D.15}$$

As before, codebook based precoding with equal power allocation and precoder clustering with a constant transmission rank over clusters is assumed.

To summarize, the following practical constraints are incrementally incorporated in the proposed throughput bounds:

- The throughput loss caused by the system overhead is considered in the achievable capacity.
- The restrictions imposed by codebook based precoding and feedback clustering are taken into account by the CLMI bound.
- The performance degradation due to the application of a linear receiver instead of a maximum likelihood MIMO detector is considered in the CLMI LR bound.
- The loss caused by the BICM architecture is accounted for in the BICM LR bound.

Notice, these restrictions are added on top of each other, i.e., the BICM LR bound accounts for all of the constraints.

Appendix E.

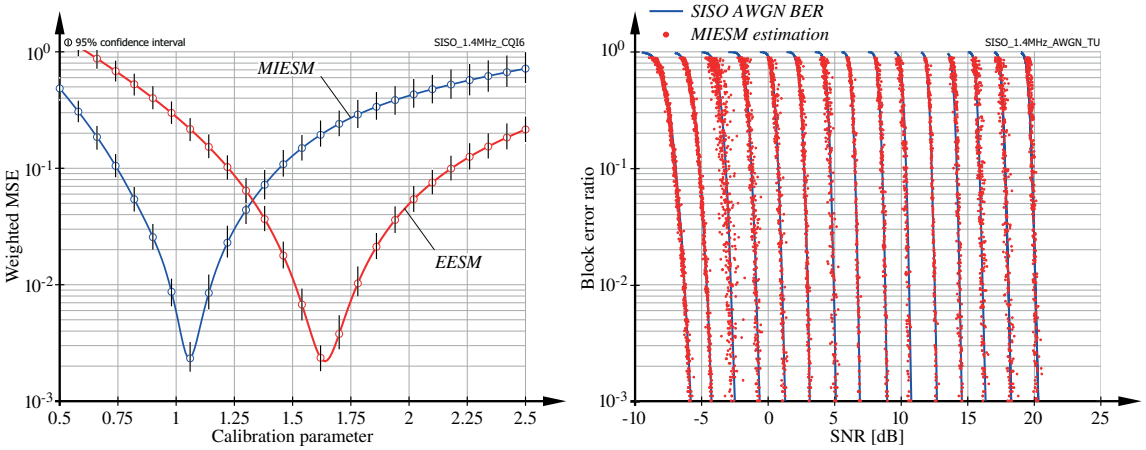
Overview and Calibration of ESM

The limited feedback algorithms proposed in this dissertation for the selection of the channel quality indicator (CQI) rely on effective SINR mapping (ESM) to estimate the block-error ratio (BLER) and the corresponding achievable data rate of a user over a fading channel; see, e.g., Section 3.2.2. In literature, ESM is mostly employed for PHY layer abstraction in system level evaluations to simplify the simulation of large cellular networks [206, 207]. In this chapter, an overview of the ESM methods considered in this thesis is given and the calibration issue related with such techniques is discussed. The optimal calibration parameters for mutual information effective SINR mapping (MIESM) and exponential effective SINR mapping (EESM) are determined for the modulation and coding schemes (MCSs) utilized by LTE. Simulation results illustrating the accuracy of the obtained BLER prediction are provided.

In a wireless channel, the instantaneous SINR of RE $[n, k]$, defined in Equation (2.7), is in general subject to significant variations over frequency n and time k, due to multipath interference caused by reflections/refractions in the environment and due to movement of the transmitter, receiver and/or obstacles. While in a static AWGN channel the link performance of a coded digital communications system can be characterized by the BLER versus the SNR, this is not possible in a fading environment as the SINR does not remain constant over a block of coded symbols. Determining the average BLER over a fading channel from the linear average SINR is not effective, because the diversity order of the channel is not taken into account with this approach. Higher order moments of the SINR can be considered to estimate the diversity order, leading to multi-dimensional sets of parameters that must be accurately calibrated. Alternatively, to reduce the number of parameters involved, ESM has been proposed in [141, 208] to map the varying SINRs of different REs to a single effective SNR corresponding to an AWGN channel that achieves the same BLER as the fading channel.

Assuming transmission over N_{RE} REs with corresponding SINRs $\{\beta[1], \ldots, \beta[N_{\text{RE}}]\}$, the MCS $m \in \mathcal{M}$ dependent mapping employed by *ESM* to determine the effective SNR is

$$\beta^{(m)} = \phi_m f_m^{-1} \left(\frac{1}{N_{\text{RE}}} \sum_{\rho=1}^{N_{\text{RE}}} f_m \left(\frac{\beta[\rho]}{\phi_m} \right) \right). \tag{E.1}$$



(a) Weighted MSE of MIESM and EESM.

(b) MIESM SNR averaging accuracy.

Figure E.1.: Calibration of effective SNR mapping and comparison of the MIESM abstraction for a 1.4 MHz typical urban channel to the average BLERs of the 15 LTE defined MCSs achieved over a 1.4 MHz AWGN channel.

Here, the averaging function $f_m(\cdot)$ is determined by the considered ESM method; its inverse is denoted $f_m^{-1}(\cdot)$. EESM is based on the Chernoff upper bound for the error probability of binary phase shift keying over an AWGN channel [209]. In that case, $f_m(\cdot)$ turns out to be the exponential function: $f_m(\cdot) = e^{(\cdot)}$. In case of MIESM, the modulation alphabet dependent BICM capacity, shown in Figure 3.2, is used as averaging function [141]. The scalar parameters ϕ_m are required for calibration purposes to adapt ESM to the performance of the different MCSs.

The calibration of the parameters ϕ_m is based on extensive link level simulations as detailed in [210] applying the weighted error function defined in [211]. The basic idea is to simulate the average BLERs of the considered communication technology for many different channel realizations, using different fading channel models. Then, the error between the calculated effective SNR $\beta^{(m)}$ for each channel realization and the SNR of an AWGN channel achieving the same BLER is minimized by tuning the parameter ϕ_m. The outcome of such a simulation is shown in Figure E.1a for MCS 6 of LTE (4 QAM with a coding rate of 0.59 [11]). Irrespective of the MCS, the optimal calibration parameters of MIESM are always found to be close to one in my simulations (between 0.85 and 1.2) while those of EESM vary significantly with the MCS (from 0.8 to 35), suggesting that MIESM provides a more natural representation of the performance of LTE than EESM. Hence, MIESM is employed throughout this dissertation. The abstraction performance of MIESM obtained for the 15 different MCSs of LTE is shown in Figure E.1b. In this figure, the AWGN BLERs of the 15 MCSs are represented by the solid lines, while the dots are pairs of simulated BLERs and calculated effective SNRs as obtained from MIESM for a typical urban channel model [109]. As these pairs lie mostly close to the AWGN BLERs, accurate BLER estimation can be achieved by mapping the effective SNR calculated with MIESM to the corresponding BLER via AWGN BLER look-up tables.

Appendix F.

Derivations and Proofs of SQBC

In this appendix, the detailed proofs and derivations of the theorems and lemmas related to the SQBC method of Section 4.3.2 are provided. The RE index $[n, k]$ is omitted for brevity.

F.1. Derivation of the SQBC Condition

The goal of the SQBC antenna combiner is to minimize the chordal distance quantization error between the obtained effective channel $\mathbf{H}_{u,i}^{\text{eff}} = \mathbf{H}_{u,i}\mathbf{G}_{u,i} \in \mathbb{C}^{N_i \times \ell_{u,i}}$ and the subspace spanned by the orthonormal basis $\mathbf{Q}_j \in \mathcal{Q}_{u,i}$. The effective channel is therefore considered as a starting point for the derivation of the corresponding condition that must be fulfilled by the antenna combiner $\mathbf{G}_{u,i}$

$$\begin{aligned} \mathbf{H}_{u,i}^{\text{eff}} = \mathbf{H}_{u,i}\mathbf{G}_{u,i} = \mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}} = \mathbf{B}_{u,i}\mathbf{W}_G\mathbf{R}_G, \\ \mathbf{W}_G \in \mathbb{C}^{M_{u,i} \times \ell_{u,i}},\ \mathbf{W}_G^{\text{H}}\mathbf{W}_G = \mathbf{I}_{\ell_{u,i}}. \end{aligned} \tag{F.1}$$

Here, an orthonormal basis $\mathbf{B}_{u,i} \in \mathbb{C}^{N_i \times M_{u,i}}$ of $\operatorname{span}(\mathbf{H}_{u,i})$ is employed to express the projection of $\mathbf{H}_{u,i}^{\text{eff}}$ onto $\operatorname{span}(\mathbf{H}_{u,i})$ in terms of a QR decomposition $\mathbf{B}_{u,i}^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}} = \mathbf{W}_G\mathbf{R}_G$. Hence, an orthonormal basis for $\operatorname{span}(\mathbf{H}_{u,i}^{\text{eff}})$ is given by $(\mathbf{B}_{u,i}\mathbf{W}_G)$. The chordal distance quantization error, according to (4.20), between $\operatorname{span}(\mathbf{H}_{u,i}^{\text{eff}})$ and any $\mathbf{Q}_j \in \mathcal{Q}_{u,i}$ can therefore be written as[1]

$$\begin{aligned} \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}^{\text{eff}}, \mathbf{Q}_j\right) &= \ell_{u,i} - \operatorname{tr}\left((\mathbf{B}_{u,i}\mathbf{W}_G)^{\text{H}}\mathbf{Q}_j\mathbf{Q}_j^{\text{H}}\mathbf{B}_{u,i}\mathbf{W}_G\right) \\ &= \ell_{u,i} - \operatorname{tr}\left(\mathbf{W}_G^{\text{H}}\mathbf{B}_{u,i}^{\text{H}}\mathbf{Q}_j^{(\text{R})}(\mathbf{Q}_j^{(\text{R})})^{\text{H}}\mathbf{B}_{u,i}\mathbf{W}_G\right), \\ \mathbf{Q}_j^{(\text{R})} &= \mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\text{H}}\mathbf{Q}_j. \end{aligned} \tag{F.2}$$

[1] Notice that the chordal distance is calculated using orthonormal bases specifying the subspaces.

Here, the range space component of $\mathbf{Q}_j$ with respect to $\mathrm{span}(\mathbf{H}_{u,i})$ is introduced. This is possible because $\mathbf{B}_{u,i}$ is defined as an orthonormal basis; hence $\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i} = \mathbf{I}_{M_{u,i}}$. Next, a QR decomposition is applied to $\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{Q}_j$

$$\mathbf{Q}_j^{(\mathrm{R})} = \mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{Q}_j = \mathbf{B}_{u,i}\mathbf{W}_Q\mathbf{R}_Q, \tag{F.3}$$

$$\mathbf{W}_Q \in \mathbb{C}^{M_{u,i}\times\ell_{u,i}},\ \mathbf{W}_Q^{\mathrm{H}}\mathbf{W}_Q = \mathbf{I}_{\ell_{u,i}}.$$

With this, the quantization error is obtained as

$$\mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}^{\mathrm{eff}}, \mathbf{Q}_j\right) = \ell_{u,i} - \mathrm{tr}\left(\mathbf{W}_G^{\mathrm{H}}\mathbf{W}_Q\mathbf{R}_Q\mathbf{R}_Q^{\mathrm{H}}\mathbf{W}_Q^{\mathrm{H}}\mathbf{W}_G\right). \tag{F.4}$$

Since both, $\mathbf{W}_G$ and $\mathbf{W}_Q$, are orthonormal bases of $\ell_{u,i}$-dimensional subspaces of the $M_{u,i}$-dimensional Euclidean space, it holds that

$$\mathrm{d}_{\mathrm{c}}^2\left(\mathbf{H}_{u,i}^{\mathrm{eff}}, \mathbf{Q}_j\right) = \ell_{u,i} - \mathrm{tr}\left(\mathbf{W}_G^{\mathrm{H}}\mathbf{W}_Q\mathbf{R}_Q\mathbf{R}_Q^{\mathrm{H}}\mathbf{W}_Q^{\mathrm{H}}\mathbf{W}_G\right) \geq \ell_{u,i} - \mathrm{tr}\left(\mathbf{R}_Q\mathbf{R}_Q^{\mathrm{H}}\right), \tag{F.5}$$

with equality if and only if

$$\mathbf{W}_Q^{\mathrm{H}}\mathbf{W}_G(\mathbf{W}_Q^{\mathrm{H}}\mathbf{W}_G)^{\mathrm{H}} = \mathbf{I}_{\ell_{u,i}}. \tag{F.6}$$

The lower bound is therefore achieved if and only if $\mathbf{W}_G$ and $\mathbf{W}_Q$ define the same subspace. Consequently, the minimal quantization error is obtained when

$$\mathrm{span}\left(\mathbf{B}_{u,i}\mathbf{W}_G\right) = \mathrm{span}\left(\mathbf{B}_{u,i}\mathbf{W}_Q\right) \Leftrightarrow \mathrm{span}\left(\mathbf{H}_{u,i}\mathbf{G}_{u,i}\right) = \mathrm{span}\left(\mathbf{Q}_j^{(\mathrm{R})}\right). \tag{F.7}$$

The corresponding minimal quantization error is

$$\begin{aligned}\mathrm{d}_{\mathrm{c,min}}^2 &= \ell_{u,i} - \mathrm{tr}\left(\mathbf{R}_Q\mathbf{R}_Q^{\mathrm{H}}\right) = \ell_{u,i} - \mathrm{tr}\left(\mathbf{R}_Q^{\mathrm{H}}\mathbf{R}_Q\right)\\ &= \mathrm{tr}\left(\mathbf{I}_{\ell_{u,i}} - (\mathbf{Q}_j^{(\mathrm{R})})^{\mathrm{H}}\mathbf{Q}_j^{(\mathrm{R})}\right) = \mathrm{tr}\left((\mathbf{Q}_j^{(\mathrm{N})})^{\mathrm{H}}\mathbf{Q}_j^{(\mathrm{N})}\right).\end{aligned} \tag{F.8}$$

It follows that the best quantization matrix $\mathbf{Q}_j \in \mathcal{Q}_{u,i}$, in terms of the chordal distance quantization error, is the one that has the smallest left null space component $\mathbf{Q}_j^{(\mathrm{N})}$ with respect to $\mathrm{span}(\mathbf{H}_{u,i})$.

F.2. Proof of Lemma 4.2

The following proof is provided under the essential assumptions that the channel matrix $\mathbf{H}_{u,i}$ is i.i.d. Rayleigh fading, i.e., $[\mathbf{H}_{u,i}]_{m,n} \sim \mathcal{N}_{\mathbb{C}}\left(0, \gamma_{u,i}\right), \forall m, n$, and that the quantization codebook $\mathcal{Q}_{u,i}$ consists of statistically independent semi-unitary matrices $\mathbf{Q}_i \in \mathcal{Q}_{u,i}$ that are uniformly distributed on $\mathcal{G}\left(N_i, \ell_{u,i}\right)$.

To proof lemma 2 the effective channel generated by the SQBC antenna combiner according to Equation (4.96) is considered

$$\mathbf{H}_{u,i}^{\text{eff}} = \mathbf{B}_{u,i}\mathbf{W}_{u,i}\mathbf{K}_{u,i}, \tag{F.9}$$

$$\mathbf{B}_{u,i} \in \mathbb{C}^{N_i \times M_{u,i}},\ \mathbf{W}_{u,i} \in \mathbb{C}^{M_{u,i} \times \ell_{u,i}}\ \mathbf{K}_{u,i} \in \mathbb{C}^{\ell_{u,i} \times \ell_{u,i}},$$

where both, $\mathbf{B}_{u,i}$ and $\mathbf{W}_{u,i}$, are semi-unitary matrices (orthonormal bases). Using the solution of $\mathbf{K}_{u,i}$ from Equation (4.95) the channel Gramian $(\mathbf{H}_{u,i}^{\text{eff}})^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}$ is obtained as

$$(\mathbf{H}_{u,i}^{\text{eff}})^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}} = \mathbf{K}_{u,i}^{\text{H}}\mathbf{K}_{u,i} = \left(\mathbf{W}_{u,i}^{\text{H}}\left(\mathbf{B}_{u,i}^{\text{H}}\mathbf{H}_{u,i}\mathbf{H}_{u,i}^{\text{H}}\mathbf{B}_{u,i}\right)^{-1}\mathbf{W}_{u,i}\right)^{-1}. \tag{F.10}$$

To proceed with the proof, an SVD in compact form is applied to the channel matrix $\mathbf{H}_{u,i}$

$$\mathbf{H}_{u,i} = \mathbf{U}_{u,i}\mathbf{\Sigma}_{u,i}\mathbf{V}_{u,i}^{\text{H}}, \tag{F.11}$$

$$\mathbf{U}_{u,i} \in \mathbb{C}^{N_i \times M_{u,i}},\ \mathbf{\Sigma}_{u,i} \in \mathbb{C}^{M_{u,i} \times M_{u,i}}, \mathbf{V}_{u,i} \in \mathbb{C}^{M_{u,i} \times M_{u,i}},$$

Remember that $\mathbf{B}_{u,i}$ is defined as an orthonormal basis of $\text{span}(\mathbf{H}_{u,i})$. Throughout this proof, the basis $\mathbf{B}_{u,i}$ specified by Equation (4.92) is employed. Since both, $\mathbf{B}_{u,i}$ and $\mathbf{U}_{u,i}$, are orthonormal bases for the same subspace, it is possible to express $\mathbf{U}_{u,i}$ in terms of $\mathbf{B}_{u,i}$ with an appropriate unitary rotation matrix $\mathbf{\Theta}_{u,i} \in \mathbb{C}^{M_{u,i} \times M_{u,i}}$

$$\mathbf{U}_{u,i} = \mathbf{B}_{u,i}\mathbf{\Theta}_{u,i}. \tag{F.12}$$

With this equality and with (F.11) the center term of Equation (F.10) can be written as

$$\mathbf{B}_{u,i}^{\text{H}}\mathbf{H}_{u,i}\mathbf{H}_{u,i}^{\text{H}}\mathbf{B}_{u,i} = \mathbf{\Theta}_{u,i}\mathbf{\Sigma}_{u,i}^2\mathbf{\Theta}_{u,i}^{\text{H}}. \tag{F.13}$$

Notice that the diagonal matrix $\mathbf{\Sigma}_{u,i}$ contains the non-zero singular values of an $N_i \times M_{u,i}$ dimensional matrix consisting of i.i.d. Gaussian elements of variance $\gamma_{u,i}$.

Since the quantization codebook consists of isotropically distributed matrices on $\mathcal{G}(N_i, \ell_{u,i})$ and because the best quantization matrix $\hat{\mathbf{H}}_{u,i}$ is solely determined by its component in the orthogonal complement of $\text{span}(\mathbf{H}_{u,i})$ according to Theorem 4.1, it follows that the range space component of $\hat{\mathbf{H}}_{u,i}$ with respect to $\text{span}(\mathbf{H}_{u,i})$ is isotropically distributed within $\text{span}(\mathbf{H}_{u,i})$. Hence, also the basis $\mathbf{B}_{u,i}$, which is determined by this range space component according to (4.92), is isotropically distributed in the subspace $\text{span}(\mathbf{H}_{u,i})$. It follows that $\mathbf{\Theta}_{u,i} \in \mathbb{C}^{M_{u,i} \times M_{u,i}}$ is uniformly distributed on the unitary group $\mathcal{U}_{M_{u,i}}$. Also, $\mathbf{\Theta}_{u,i}$ is independent of $\mathbf{\Sigma}_{u,i}$, since $\mathbf{U}_{u,i}$ is independent of $\mathbf{\Sigma}_{u,i}$ for i.i.d. Rayleigh fading. Hence, the statistical properties of $\mathbf{\Theta}_{u,i}$ and $\mathbf{\Sigma}_{u,i}$ correspond to the properties of the right singular vector matrix and the singular value matrix of an i.i.d. Gaussian matrix $\mathbf{X} \in \mathbb{C}^{N_i \times M_{u,i}}$ with elements of variance $\gamma_{u,i}$. Therefore the product

$$\mathbf{\Theta}_{u,i}\mathbf{\Sigma}_{u,i}^2\mathbf{\Theta}_{u,i}^{\text{H}} = \mathbf{X}^{\text{H}}\mathbf{X} \tag{F.14}$$

is distributed according to a central complex Wishart distribution $\mathcal{W}_{M_{u,i}}^{\mathbb{C}}\left(N_i, \gamma_{u,i}\mathbf{I}_{M_{u,i}}\right)$ [212] of dimension $M_{u,i}$ and with N_i DoF. Consequently, the inverse $\left(\boldsymbol{\Theta}_{u,i}\boldsymbol{\Sigma}_{u,i}^2\boldsymbol{\Theta}_{u,i}^{\mathrm{H}}\right)^{-1}$ follows an inverse-Wishart distribution with parameters $\mathcal{W}_{M_{u,i}}^{-\mathbb{C}}\left(N_i, \gamma_{u,i}^{-1}\mathbf{I}_{M_{u,i}}\right)$. The product

$$\mathbf{W}_{u,i}^{\mathrm{H}}\left(\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}\mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i}\right)^{-1}\mathbf{W}_{u,i}$$

$$= \left[\mathbf{I}_{\ell_{u,i}}, \mathbf{0}\right]\left(\boldsymbol{\Theta}_{u,i}\boldsymbol{\Sigma}_{u,i}^2\boldsymbol{\Theta}_{u,i}^{\mathrm{H}}\right)^{-1}\begin{bmatrix}\mathbf{I}_{\ell_{u,i}}\\ \mathbf{0}\end{bmatrix} = \left[\left(\boldsymbol{\Theta}_{u,i}\boldsymbol{\Sigma}_{u,i}^2\boldsymbol{\Theta}_{u,i}^{\mathrm{H}}\right)^{-1}\right]_{1:\ell_{u,i},1:\ell_{u,i}}, \tag{F.15}$$

with $\mathbf{W}_{u,i}$ from Equation (4.92), consists of the first $\ell_{u,i}$ rows and columns of the inverse-Wishart matrix. It is known that the marginal distribution of this $\ell_{u,i} \times \ell_{u,i}$ upper-left block is itself inverse-Wishart distributed, with reduced DoF compared to the full matrix [213], i.e.,

$$\left[\left(\boldsymbol{\Theta}_{u,i}\boldsymbol{\Sigma}_{u,i}^2\boldsymbol{\Theta}_{u,i}^{\mathrm{H}}\right)^{-1}\right]_{1:\ell_{u,i},1:\ell_{u,i}} \sim \mathcal{W}_{\ell_{u,i}}^{-\mathbb{C}}\left(N_i - M_{u,i} + \ell_{u,i}, \gamma_{u,i}^{-1}\,\mathbf{I}_{\ell_{u,i}}\right). \tag{F.16}$$

Hence, the matrix $(\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}$ is distributed as

$$(\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}} = \left(\left[\left(\boldsymbol{\Theta}_{u,i}\boldsymbol{\Sigma}_{u,i}^2\boldsymbol{\Theta}_{u,i}^{\mathrm{H}}\right)^{-1}\right]_{1:\ell_{u,i},1:\ell_{u,i}}\right)^{-1} \sim \mathcal{W}_{\ell_{u,i}}^{\mathbb{C}}\left(N_i - M_{u,i} + \ell_{u,i}, \gamma_{u,i}\,\mathbf{I}_{\ell_{u,i}}\right). \tag{F.17}$$

Although a specific basis $\mathbf{B}_{u,i}$ is employed throughout this proof such that the well-known marginalization result can be applied in Equation (F.16), the result holds for arbitrary bases, which can easily be seen by applying a change of coordinates.

F.3. Proof of Theorem 4.2

In this section, the proof of Theorem 4.2, stating an upper bound on the per-user rate loss of BD using SQBC with quantized CSIT and $M_{u,i} \geq L$ receive antennas, compared to BD with perfect CSIT and $M_{u,i}^{(BD)} = L$, is provided. This proof relies on the validity of Lemma 4.2; hence, the same assumptions as in Appendix F.2 apply here as well. Also, to simplify notations, the number of data streams per user is assumed as equal for all served users $\ell_{u,i} = L$, $\forall u \in \mathcal{S}_i$.

The achievable rate of SQBC with quantized CSIT, as defined in (4.99), is used as a starting point of this proof. The first summand of the achievable rate can be lower bounded, by applying Minkowski's determinant theorem [214] to remove the positive definite interference terms from this summand

$$\mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\sum_{s\in\mathcal{S}_i}(\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}\right) \geq \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,(\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}\right), \tag{F.18}$$

which is tight in case of perfect CSIT, due to the construction of the BD precoders: $(\mathbf{H}_{u,i}^{\text{eff}})^{\text{H}}\mathbf{F}_{s,i} = \mathbf{0}$ with perfect CSIT.

Next, the effective channel is decomposed by applying an SVD

$$\mathbf{H}_{u,i}^{\text{eff}} = \mathbf{X}_{u,i}\mathbf{\Lambda}_{u,i}\mathbf{Y}_{u,i}^{\text{H}} = \tilde{\mathbf{H}}_{u,i}\mathbf{\Phi}_{u,i}\mathbf{\Lambda}_{u,i}\mathbf{Y}_{u,i}^{\text{H}}, \tag{F.19}$$

$$\mathbf{X}_{u,i} \in \mathbb{C}^{N_i \times L},\ \mathbf{\Lambda}_{u,i} \in \mathbb{C}^{L\times L}, \mathbf{Y}_{u,i} \in \mathbb{C}^{L\times L},$$

with $\mathbf{\Phi}_{u,i}$ uniformly distributed on $\mathcal{U}_L$, because $\tilde{\mathbf{H}}_{u,i}$ is a random isotropically distributed basis of $\text{span}(\mathbf{H}_{u,i}^{\text{eff}})$. Using the notation from Equation (4.23), $\tilde{\mathbf{H}}_{u,i}$ can be written in terms of the quantized channel subspace as

$$\tilde{\mathbf{H}}_{u,i} = \left(\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\text{H}}\right)\tilde{\mathbf{H}}_{u,i} + \hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\text{H}}\tilde{\mathbf{H}}_{u,i}, \tag{F.20}$$

$$\hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\text{H}}\tilde{\mathbf{H}}_{u,i} = \mathbf{S}_{u,i}\mathbf{R}_{u,i}. \tag{F.21}$$

Here, $\mathbf{S}_{u,i} \in \mathbb{C}^{N_i\times L}$ is an orthonormal basis for an isotropically distributed L-dimensional plane in the $(N_i - L)$-dimensional orthogonal complement of $\hat{\mathbf{H}}_{u,i}$, and $\mathbf{R}_{u,i} \in \mathbb{C}^{L\times L}$ satisfies $\text{tr}\left(\mathbf{R}_{u,i}^{\text{H}}\mathbf{R}_{u,i}\right) = \text{d}_{\text{c}}^2\left(\mathbf{H}_{u,i}^{\text{eff}}, \hat{\mathbf{H}}_{u,i}\right)$. This follows from [46, Lemma 1], which is valid here as well, because $\mathbf{H}_{u,i}^{\text{eff}}$ is isotropic according to Lemma 4.1, and can be proved by applying a QR decomposition to $(\hat{\mathbf{H}}_{u,i}^{\perp})^{\text{H}}\tilde{\mathbf{H}}_{u,i}$. Plugging Equations (F.19) to (F.21) into the second summand of Equation (4.99), and remembering that $\hat{\mathbf{H}}_{u,i}^{\text{H}}\mathbf{F}_{s,i} = 0, \forall s \neq u$ due to BD, the following expression is obtained

$$\mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\sum_{s\in\mathcal{S}_i, s\neq u}(\mathbf{H}_{u,i}^{\text{eff}})^{\text{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}\right)$$

$$= \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\sum_{s\in\mathcal{S}_i, s\neq u}\mathbf{R}_{u,i}^{\text{H}}\mathbf{S}_{u,i}^{\text{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\text{H}}\mathbf{S}_{u,i}\mathbf{R}_{u,i}\,\mathbf{\Phi}_{u,i}\mathbf{\Lambda}_{u,i}^2\mathbf{\Phi}_{u,i}^{\text{H}}\right), \tag{F.22}$$

where the order of the matrices has been rearranged in accordance with Sylvester's determinant theorem [214]. Equation (F.22) is abbreviated as Δb.

With Equation (F.18) and Equation (F.22), the rate loss between SQBC with quantized CSIT and $M_{u,i} = M_{u,i}^{(\text{SQBC})} \geq L$ and BD with perfect CSIT and $M_{u,i}^{(\text{BD})} = L$ is upper bounded as

$$R_{\text{BD}}^{(L)} - R_{\text{SQBC}}^{(L,M_{u,i})}$$

$$\leq \underbrace{\mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,\mathbf{H}_{u,i}^{\text{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\text{H}}\mathbf{H}_{u,i}\right) - \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,(\mathbf{H}_{u,i}^{\text{eff}})^{\text{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\text{H}}\mathbf{H}_{u,i}^{\text{eff}}\right)}_{\Delta a} + \Delta b. \tag{F.23}$$

The proceed with the proof, the term Δa is considered in more detail. Using the decompositions from Equations (F.11), (F.12) and (F.19)) and applying Sylvester's determinant theorem, Δa can be rewritten as

$$\Delta a = \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,\mathbf{B}_{u,i}^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i}\,\boldsymbol{\Theta}_{u,i}\boldsymbol{\Sigma}_{u,i}^2\boldsymbol{\Theta}_{u,i}^{\mathrm{H}}\right)$$
$$-\mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,\tilde{\mathbf{H}}_{u,i}^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\tilde{\mathbf{H}}_{u,i}\,\boldsymbol{\Phi}_{u,i}\boldsymbol{\Lambda}_{u,i}^2\boldsymbol{\Phi}_{u,i}^{\mathrm{H}}\right). \tag{F.24}$$

Notice that the precoders $\tilde{\mathbf{F}}_{u,i}$ in both summands are in general not the same (the first summand corresponds to the precoders obtained with BD and perfect CSIT; the second summand is obtained with quantized CSIT and SQBC). Still, in terms of distribution they are equivalent, because $\mathbf{F}_{u,i}$ is determined solely by the subspaces spanned by the channels of the other users, which are isotropically distributed on $\mathcal{G}(N_i, L)$ for both summands.

The matrices $\mathbf{U}_{u,i} = \mathbf{B}_{u,i}\boldsymbol{\Theta}_{u,i}$ and $\boldsymbol{\Sigma}_{u,i}$ are statistically independent because $\mathbf{H}_{u,i}$ is i.i.d. Gaussian distributed and thus distribution-invariant with respect to left-multiplication by an arbitrary unitary matrix. Also, $\mathbf{B}_{u,i}$ and $\boldsymbol{\Theta}_{u,i}$ are independent, because $\mathbf{B}_{u,i}$ is a randomly chosen basis for $\mathrm{span}(\mathbf{H}_{u,i})$. Similar arguments can be applied to show that $\tilde{\mathbf{H}}_{u,i}$, $\boldsymbol{\Lambda}_{u,i}$ and $\boldsymbol{\Phi}_{u,i}$ are independent as well.

Furthermore, $\mathbf{B}_{u,i}$ and $\tilde{\mathbf{H}}_{u,i}$ are both isotropic on $\mathcal{G}(N_i, L)$, because $\mathbf{H}_{u,i}$ (i.i.d. Gaussian) as well as $\mathbf{H}_{u,i}^{\mathrm{eff}}$ (see Lemma 4.1) are isotropically distributed. Hence, the product $\mathbf{C}_{u,i}^{\mathrm{H}}\mathbf{C}_{u,i} = \mathbf{B}_{u,i}^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i}$ is equivalent in distribution to $\tilde{\mathbf{H}}_{u,i}^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\tilde{\mathbf{H}}_{u,i}$, and in the calculation of the expectation one can be substituted for the other.

Since $\boldsymbol{\Sigma}_{u,i}^2$ and $\boldsymbol{\Lambda}_{u,i}^2$ contain the eigenvalues of L-dimensional Wishart matrices with a scale matrix that is a scaled identity and N_i respectively $N_i - M_{u,i} + L$ DoF, and since $\boldsymbol{\Theta}_{u,i}$ as well as $\boldsymbol{\Phi}_{u,i}$ are both isotropic unitary matrices, it can be concluded that

$$\mathbf{T}_{u,i} = \boldsymbol{\Theta}_{u,i}\boldsymbol{\Sigma}_{u,i}^2\boldsymbol{\Theta}_{u,i}^{\mathrm{H}} \sim \mathcal{W}_L^{\mathbb{C}}\left(N_i, \gamma_{u,i}\,\mathbf{I}_L\right), \tag{F.25}$$

$$\bar{\mathbf{T}}_{u,i} = \boldsymbol{\Phi}_{u,i}\boldsymbol{\Lambda}_{u,i}^2\boldsymbol{\Phi}_{u,i}^{\mathrm{H}} \sim \mathcal{W}_L^{\mathbb{C}}\left(N_i - M_{u,i} + L, \gamma_{u,i}\,\mathbf{I}_L\right). \tag{F.26}$$

With these results the term Δa can be developed as

$$\Delta a = \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,\mathbf{C}_{u,i}\mathbf{T}_{u,i}\mathbf{C}_{u,i}^{\mathrm{H}}\right) - \mathbb{E}\log_2\det\left(\mathbf{I}_L + \rho\,\mathbf{C}_{u,i}\bar{\mathbf{T}}_{u,i}\mathbf{C}_{u,i}^{\mathrm{H}}\right) = \mathbb{E}\,\log_2\frac{\det\mathbf{T}_{u,i}}{\det\bar{\mathbf{T}}_{u,i}}$$
$$+\Big(\mathbb{E}\log_2\det\left((\mathbf{C}_{u,i}\mathbf{T}_{u,i}\mathbf{C}_{u,i}^{\mathrm{H}})^{-1} + \rho\,\mathbf{I}_L\right) - \mathbb{E}\log_2\det\left((\mathbf{C}_{u,i}\bar{\mathbf{T}}_{u,i}\mathbf{C}_{u,i}^{\mathrm{H}})^{-1} + \rho\,\mathbf{I}_L\right)\Big), \tag{F.27}$$

where $\mathbf{C}_{u,i}$ can be removed from the first summand, due to the multiplicativity of the determinant. Considering the expected value with respect to $\mathbf{T}_{u,i}$ and $\bar{\mathbf{T}}_{u,i}$, it can be seen that the difference in brackets is negative, since $(\mathbf{C}_{u,i}\mathbf{T}_{u,i}\mathbf{C}_{u,i}^{\mathrm{H}})^{-1}$ and $(\mathbf{C}_{u,i}\bar{\mathbf{T}}_{u,i}\mathbf{C}_{u,i}^{\mathrm{H}})^{-1}$ are both inverse Wishart

distributed with the same scale matrix $\mathbf{C}_{u,i}\mathbf{C}_{u,i}^{\mathrm{H}}\gamma_{u,i}$, but the first term has smaller eigenvalues on average then the second term, because it has more DoF. Hence, Δa can be upper bounded as

$$\Delta a \leq \mathbb{E}\log_2\det\left(\mathbf{T}_{u,i}\right) - \mathbb{E}\log_2\det\left(\bar{\mathbf{T}}_{u,i}\right) = \log_2\left(e\right)\sum_{k=0}^{L-1}\sum_{\ell=N_i-M_{u,i}+L}^{N_i-1}\frac{1}{\ell-k}. \tag{F.28}$$

Notice that this bound is tight for $\rho\,\gamma_{u,i}\to\infty$, since then $\mathbf{I}_L$ is negligible compared to $\rho\,\mathbf{C}_u\mathbf{T}_u\mathbf{C}_u^{\mathrm{H}}$. The closed form solution is obtained from basic results about Wishart matrices, e.g., [212].

To find an upper bound on Δb, Jensen's inequality is applied

$$\Delta b \leq \log_2\det\left(\mathbf{I}_L + \rho\sum_{s\in\mathcal{S}_i, s\neq u}\mathbb{E}\left(\mathbf{R}_{u,i}^{\mathrm{H}}\mathbf{S}_{u,i}^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\mathrm{H}}\mathbf{S}_{u,i}\mathbf{R}_{u,i}\right)\mathbb{E}\left(\bar{\mathbf{T}}_{u,i}\right)\right). \tag{F.29}$$

According to [46, Lemma 1], $\mathbf{R}_{u,i}$ and $\mathbf{S}_{u,i}$ are statistically independent. Furthermore, $\mathbf{S}_{u,i}$ and $\tilde{\mathbf{F}}_{s,i}, \forall s\neq u$ are isotropically distributed in the (N_i-L)-dimensional left null space of $\hat{\mathbf{H}}_{u,i}$, due to the BD construction. Hence, it is possible to express them in the same orthonormal basis $\hat{\mathbf{H}}_{u,i}^{\perp}$, i.e., $\mathbf{S}_{u,i} = \hat{\mathbf{H}}_{u,i}^{\perp}\bar{\mathbf{S}}_{u,i}$, $\tilde{\mathbf{F}}_{s,i} = \hat{\mathbf{H}}_{u,i}^{\perp}\bar{\mathbf{F}}_{s,i}$, and thus $\mathbf{S}_{u,i}^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i} = \bar{\mathbf{S}}_{u,i}^{\mathrm{H}}\bar{\mathbf{F}}_{s,i}$. The matrices $\bar{\mathbf{S}}_{u,i}$, $\bar{\mathbf{F}}_{s,i}\in\mathbb{C}^{N_i-L\times L}$ are semi-unitary, statistically independent and isotropically distributed on $\mathcal{G}\left(N_i-L, L\right)$.

With these results the first expectation in Equation (F.29) can be evaluated as follows

$$\begin{aligned}\mathbb{E}\left(\mathbf{R}_{u,i}^{\mathrm{H}}\mathbf{S}_{u,i}^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\mathrm{H}}\mathbf{S}_{u,i}\mathbf{R}_{u,i}\right) &= \mathbb{E}\left(\mathbf{R}_{u,i}^{\mathrm{H}}\bar{\mathbf{S}}_{u,i}^{\mathrm{H}}\mathbb{E}\left(\bar{\mathbf{F}}_{s,i}\bar{\mathbf{F}}_{s,i}^{\mathrm{H}}\right)\bar{\mathbf{S}}_{u,i}\mathbf{R}_{u,i}\right)\\ = \frac{L}{N_i-L}\mathbb{E}\left(\mathbf{R}_{u,i}^{\mathrm{H}}\bar{\mathbf{S}}_{u,i}^{\mathrm{H}}\bar{\mathbf{S}}_{u,i}\mathbf{R}_{u,i}\right) &= \frac{L}{N_i-L}\mathbb{E}\left(\mathbf{R}_{u,i}^{\mathrm{H}}\mathbf{R}_{u,i}\right) = \frac{L}{N_i-L}\,\frac{1}{L}\,D\,\mathbf{I}_L,\end{aligned} \tag{F.30}$$

where the first equality follows from the independence of $\bar{\mathbf{F}}_{s,i}$ from the other matrices, the second equality is due to the isotropy of $\bar{\mathbf{F}}_{s,i}$ on $\mathcal{G}\left(N_i-L, L\right)$, i.e., $\mathbb{E}\left(\bar{\mathbf{F}}_{s,i}\bar{\mathbf{F}}_{s,i}^{\mathrm{H}}\right) = \frac{L}{N_i-L}\,\mathbf{I}_{N_i-L}$, and the third equality is obtained from the semi-unitarity of $\bar{\mathbf{S}}_{u,i}$. The last equality is reasoned in [46, Appendix B], with D denoting the average chordal distance quantization distortion, where it is argued that the distortion on average distributes equally over all L dimensions of the quantized subspace.

Finally, the expected value of the Wishart matrix $\bar{\mathbf{T}}_{u,i}$ can be evaluated as $\mathbb{E}\left(\bar{\mathbf{T}}_{u,i}\right) = \gamma_{u,i}\left(N_i - M_{u,i} + L\right)\mathbf{I}_L$. Thus, the upper bound for Δb is obtained as

$$\Delta b \leq L\log_2\left(1 + \rho\,\gamma_{u,i}\,\frac{N_i - M_{u,i} + L}{N_i - L}\left(S_i - 1\right)D\right). \tag{F.31}$$

The factor $S_i - 1$ accounts for the summation over all other users in Equation (F.29). Plugging (F.28) and (F.31) into (F.23) completes the proof.

F.4. Proof of Theorem 4.3

The following proof is very similar in nature to the derivation in Appendix F.3. For a better understanding of the steps involved in this proof, it is recommended to first work through the details of Appendix F.3. Also, the assumptions of the proof are the same as in the previous section.

Similar to the previous section, Minkowski's determinant theorem is employed to find a lower bound on the first summand of the achievable rate of maximum eigenmode transmission (MET) based BD with quantized CSIT, as defined in Equation (4.109)

$$\mathbb{E}\log_2\det\left(\mathbf{I}_L+\rho\sum_{s\in\mathcal{S}_i}(\mathbf{H}_{u,i}^{(L)})^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{(L)}\right)\geq\mathbb{E}\log_2\det\left(\mathbf{I}_L+\rho\,(\mathbf{H}_{u,i}^{(L)})^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{(L)}\right). \tag{F.32}$$

Plugging this result into Equation (4.109) and subtracting from Equation (4.108), the rate gap between MET based BD with perfect and quantized CSIT can be upper bounded as

$$R_{\text{MET}}-R_{\text{MET-Quant}}\leq\log_2\det\left(\mathbf{I}_L+\rho\sum_{s\in\mathcal{S}_i,s\neq u}(\mathbf{H}_{u,i}^{(L)})^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{(L)}\right). \tag{F.33}$$

Applying Jensen's inequality and exploiting the independence of $\mathbf{U}_u^{(L)}$ and $\mathbf{\Sigma}_u^{(L)}$, defined in Equation (4.107), which is satisfied for an i.i.d. Gaussian channel matrix, the rate loss can further be bounded as

$$R_{\text{MET}}-R_{\text{MET-Quant}}\leq\log_2\det\left(\mathbf{I}_L+\rho\sum_{s\in\mathcal{S}_i,s\neq u}\mathbb{E}\left((\mathbf{U}_{u,i}^{(L)})^{\mathrm{H}}\tilde{\mathbf{F}}_{s,i}\tilde{\mathbf{F}}_{s,i}^{\mathrm{H}}\mathbf{U}_{u,i}^{(L)}\right)\mathbb{E}\left(\mathbf{\Sigma}_{u,i}^{(L)^2}\right)\right). \tag{F.34}$$

The same argumentation as in Appendix F.3, Equations (F.21) and (F.30), can be applied to evaluate the first expectation

$$R_{\text{MET}}-R_{\text{MET-Quant}}\leq\log_2\det\left(\mathbf{I}_L+\rho\,\frac{S_i-1}{N_i-L}\,D\,\mathbb{E}\left(\mathbf{\Sigma}_{u,i}^{(L)^2}\right)\right). \tag{F.35}$$

The matrix $\mathbf{\Sigma}_{u,i}^{(L)^2}$ is diagonal and contains the L largest squared singular values $\sigma_{1,u,i}^2,\ldots,\sigma_{L,u,i}^2$ of $\mathbf{H}_{u,i}$. The determinant is hence obtained as the product of the diagonal elements

$$R_{\text{MET}}-R_{\text{MET-Quant}}\leq\sum_{\ell=1}^{L}\log_2\left(1+\rho\,\frac{S_i-1}{N_i-L}\,D\,\mathbb{E}\left(\sigma_{\ell,u,i}^2\right)\right), \tag{F.36}$$

completing the proof.

Appendix G.

SINR Lower Bound for Block Diagonalization

In this appendix, the derivation of the SINR lower bound for limited feedback based BD is provided, which is considered as CQI feedback in Section 4.4.2. The RE index $[n, k]$ is omitted for brevity.

Signal term: The post-equalization input-output relationship of BD MU-MIMO with limited feedback, given in Equation (4.24), is considered as starting point in the derivation. With (4.24) and the precoder normalization in (4.15), the covariance matrix of the intended signal of user u is

$$\begin{aligned}\mathbf{S}_{u,i} &= \mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\mathbf{F}_{u,i}\mathbf{x}_{u,i}\mathbf{x}_{u,i}^{\mathrm{H}}\mathbf{F}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}\right) = P_{u,i}\,\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}\right) \\ &= P_{u,i}\,\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff,r}} + \mathbf{H}_{u,i}^{\mathrm{eff,n}})^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}(\mathbf{H}_{u,i}^{\mathrm{eff,r}} + \mathbf{H}_{u,i}^{\mathrm{eff,n}})\right), \end{aligned} \tag{G.1}$$

where the channel decomposition of Equation (4.23) has been employed. At the time the CQI feedback is calculated, the precoder $\tilde{\mathbf{F}}_{u,i}$ is unknown. Because the precoder is determined solely by the channels of the other served users according to (4.4), which are unknown to user u, the precoder is assumed as isotropically distributed. With this assumption, the expected value of the mixed terms between $\mathbf{H}_{u,i}^{\mathrm{eff,r}}$ and $\mathbf{H}_{u,i}^{\mathrm{eff,n}}$ with respect to the precoder are equal to zero, e.g.,

$$(\mathbf{H}_{u,i}^{\mathrm{eff,r}})^{\mathrm{H}}\mathbb{E}\left(\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\right)\mathbf{H}_{u,i}^{\mathrm{eff,n}} = \frac{\ell_{u,i}}{N_i}(\mathbf{H}_{u,i}^{\mathrm{eff,r}})^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff,n}} = \mathbf{0}, \tag{G.2}$$

$$\Rightarrow \mathbf{S}_{u,i} = P_{u,i}\,\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff,r}})^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff,r}}\right) + P_{u,i}\,\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff,n}})^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff,n}}\right). \tag{G.3}$$

Neglecting the signal power received over $\mathbf{H}_{u,i}^{\mathrm{eff,n}}$, the covariance matrix is lower bounded as

$$\mathbf{S}_{u,i} \succeq \tilde{\mathbf{S}}_{u,i} = P_{u,i}\,\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff,r}})^{\mathrm{H}}\tilde{\mathbf{F}}_{u,i}\tilde{\mathbf{F}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff,r}}\right), \tag{G.4}$$

in the sense that $\mathbf{S}_{u,i} - \tilde{\mathbf{S}}_{u,i}$ is positive semidefinite. Notice that this bound is tight in case of unquantized feedback or if the scheduled users have orthogonal channels, because then $\tilde{\mathbf{F}}_{u,i} = \hat{\mathbf{H}}_{u,i}$. Also, the achievable rate obtained with $\mathbf{S}_{u,i}$ is lower bounded by the rate calculated with $\tilde{\mathbf{S}}_{u,i}$, due to Minkowski's determinant theorem on the sum of positive semidefinite matrices [214].

According to the BD construction in (4.4) and (4.22), the precoder of user u is found in the left null space of the other users' quantized channels. Considering an orthonormal basis $\mathbf{B}_{u,i}$ for this space,

$$\hat{\overline{\mathbf{H}}}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i}=\mathbf{0},\ \ \mathbf{B}_{u,i}\in\mathbb{C}^{N_i\times N_i-\bar{\ell}_i},\ \ (\mathbf{B}_{u,i})^{\mathrm{H}}\mathbf{B}_{u,i}=\mathbf{I}_{N_i-\bar{\ell}_i}, \tag{G.5}$$

with $\hat{\overline{\mathbf{H}}}_{u,i}$ from (4.22) and $\bar{\ell}_i=\sum_{s\in\mathcal{S}_i,s\neq u}\ell_{s,i}$, the precoder can be expressed as

$$\tilde{\mathbf{F}}_{u,i}=\mathbf{B}_{u,i}\mathbf{Q}_{u,i},\ \ \mathbf{Q}_{u,i}\in\mathbb{C}^{N_i-\bar{\ell}_i\times\ell_{u,i}},\ \ \mathbf{Q}_{u,i}^{\mathrm{H}}\mathbf{Q}_{u,i}=\mathbf{I}_{\ell_{u,i}}. \tag{G.6}$$

With the assumption that $\tilde{\mathbf{F}}_{u,i}$ is isotropically distributed, the same holds true for $\mathbf{Q}_{u,i}$. Also, $\mathbf{Q}_{u,i}$ is statistically independent of $\mathbf{B}_{u,i}$. Therefore, the expected value of $\tilde{\mathbf{S}}_{u,i}$ with respect to $\mathbf{Q}_{u,i}$ is

$$\begin{aligned}\tilde{\mathbf{S}}_{u,i}&=P_{u,i}\,\frac{\ell_{u,i}}{N_i-\bar{\ell}_i}\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff,r}})^{\mathrm{H}}\mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff,r}}\right)\\&=P_{u,i}\,\frac{\ell_{u,i}}{N_i-\bar{\ell}_i}\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i}\mathbf{B}_{u,i}^{\mathrm{H}}\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}\right),\end{aligned} \tag{G.7}$$

where $\mathbf{H}_{u,i}^{\mathrm{eff,r}}$ has been substituted from (4.23). Next, $\mathbf{B}_{u,i}$ is decomposed with respect to the quantized subspace $\hat{\mathbf{H}}_{u,i}\in\mathbb{C}^{N_i\times\ell_{u,i}}$

$$\begin{aligned}\mathbf{B}_{u,i}&=\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i}+\left(\mathbf{I}_{N_i}-\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\right)\mathbf{B}_{u,i}\\&=\hat{\mathbf{H}}_{u,i}\mathbf{U}_{u,i}\cos\left(\mathbf{\Phi}_{u,i}\right)\mathbf{V}_{u,i}^{\mathrm{H}}+\left(\mathbf{I}_{N_i}-\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\right)\mathbf{B}_{u,i},\end{aligned} \tag{G.8}$$

where the second equality follows from a compact SVD of $\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\mathbf{B}_{u,i}$. The CSI feedback of the other users and the schedule $\mathcal{S}_i$ are unknown to user u during feedback calculation; hence, $\mathbf{B}_{u,i}$ cannot be determined by the user and is therefore assumed as uniformly distributed on $\mathcal{G}\left(N_i,N_i-\bar{\ell}_i\right)$. Correspondingly, $\mathbf{U}_{u,i}$, $\cos\left(\mathbf{\Phi}_{u,i}\right)$ and $\mathbf{V}_{u,i}$ are statistically independent, and the $\ell_{u,i}\times\ell_{u,i}$ dimensional matrix $\mathbf{U}_{u,i}$ is isotropically distributed unitary [170, Theorem1]. Diagonal matrix $\mathbf{\Phi}_{u,i}$ is composed of the principal angles between $\mathrm{span}(\hat{\mathbf{H}}_{u,i})$ and $\mathrm{span}(\mathbf{B}_{u,i})$; hence, $\mathrm{tr}\big(\cos\left(\mathbf{\Phi}_{u,i}\right)^2\big)=\ell_{u,i}-\mathrm{d}_{\mathrm{c}}^2(\hat{\mathbf{H}}_{u,i},\mathbf{B}_{u,i})$. Plugging (G.8) into (G.7), $\tilde{\mathbf{S}}_{u,i}$ is obtained as

$$\begin{aligned}\tilde{\mathbf{S}}_{u,i}&=P_{u,i}\,\frac{\ell_{u,i}}{N_i-\bar{\ell}_i}\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\hat{\mathbf{H}}_{u,i}\left(\mathbf{U}_{u,i}\cos\left(\mathbf{\Phi}_{u,i}\right)^2\mathbf{U}_{u,i}^{\mathrm{H}}\right)\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}\right)\\&=P_{u,i}\,\frac{\ell_{u,i}}{N_i-\bar{\ell}_i}\mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{\mathrm{eff}}\right)\left(1-\frac{\mathrm{d}_{\mathrm{c}}^2\left(\hat{\mathbf{H}}_{u,i},\mathbf{B}_{u,i}\right)}{\ell_{u,i}}\right),\end{aligned} \tag{G.9}$$

where the second equality is due to the isotropy of $\mathbf{U}_{u,i}$. Finally, an SVD is applied to the effective channel $\mathbf{H}_{u,i}^{\mathrm{eff}}=\mathbf{U}_{u,i}^{\mathrm{eff}}\mathbf{\Sigma}_{u,i}^{\mathrm{eff}}(\mathbf{V}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}$, and $\hat{\mathbf{H}}_{u,i}$ is decomposed with respect to $\mathbf{U}_{u,i}^{\mathrm{eff}}\in\mathbb{C}^{N_i\times\ell_{u,i}}$

$$\begin{aligned}\hat{\mathbf{H}}_{u,i}&=\mathbf{U}_{u,i}^{\mathrm{eff}}(\mathbf{U}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\hat{\mathbf{H}}_{u,i}+\left(\mathbf{I}_{N_i}-\mathbf{U}_{u,i}^{\mathrm{eff}}(\mathbf{U}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\right)\hat{\mathbf{H}}_{u,i}\\&=\mathbf{U}_{u,i}^{\mathrm{eff}}\mathbf{Y}_{u,i}\cos\left(\mathbf{\Theta}_{u,i}\right)\mathbf{W}_{u,i}^{\mathrm{H}}+\left(\mathbf{I}_{N_i}-\mathbf{U}_{u,i}^{\mathrm{eff}}(\mathbf{U}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\right)\hat{\mathbf{H}}_{u,i},\end{aligned} \tag{G.10}$$

with the second equality being obtained from a compact SVD of $(\mathbf{U}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}\hat{\mathbf{H}}_{u,i}$, similar to (G.8). For a fixed quantized subspace $\hat{\mathbf{H}}_{u,i}$, this expression cannot further be simplified. However, considering the randomness of the quantization codebook, it is reasonable to argue that for a fixed subspace quantization error $\mathrm{d_c^2}(\mathbf{U}_{u,i}^{\mathrm{eff}}, \hat{\mathbf{H}}_{u,i}) = \ell_{u,i} - \mathrm{tr}\big(\cos(\boldsymbol{\Theta}_{u,i})^2\big)$ the range space component of $\hat{\mathbf{H}}_{u,i}$ within $\mathrm{span}(\mathbf{U}_{u,i}^{\mathrm{eff}})$ is isotropically distributed, implying that $\mathbf{Y}_{u,i}$ is an isotropic unitary matrix. Using (G.10) in (G.9) and exploiting the isotropy of $\mathbf{Y}_{u,i}$, the covariance matrix is

$$\tilde{\mathbf{S}}_{u,i} = \underbrace{P_{u,i}\,\frac{\ell_{u,i}}{N_i - \bar{\ell}_i}\left(1 - \frac{\mathrm{d_c^2}(\hat{\mathbf{H}}_{u,i}, \mathbf{B}_{u,i})}{\ell_{u,i}}\right)\left(1 - \frac{\mathrm{d_c^2}(\mathbf{U}_{u,i}^{\mathrm{eff}}, \hat{\mathbf{H}}_{u,i})}{\ell_{u,i}}\right)}_{c_S} \mathbf{V}_{u,i}^{\mathrm{eff}}\left(\boldsymbol{\Sigma}_{u,i}^{\mathrm{eff}}\right)^2 (\mathbf{V}_{u,i}^{\mathrm{eff}})^{\mathrm{H}}. \tag{G.11}$$

In this equation, the overlap between the quantized user subspace $\hat{\mathbf{H}}_{u,i}$ and the other users' channels is measured by the first term in brackets. If $\hat{\mathbf{H}}_{u,i}$ is in the null space of the other users, i.e., $\mathrm{d_c^2}(\hat{\mathbf{H}}_{u,i}, \mathbf{B}_{u,i}) = 0$, the transmit signal of user u can be steered completely into the subspace $\mathrm{span}(\hat{\mathbf{H}}_{u,i})$ without causing any interference to the other users. Hence, the channel gain of the user is not reduced by the precoder. The term involving $\mathrm{d_c^2}(\mathbf{U}_{u,i}^{\mathrm{eff}}, \hat{\mathbf{H}}_{u,i})$ is just a consequence of the lower bound (G.4).

Interference term: Next, the covariance matrix of the interference is determined. To obtain a single CQI per spatial stream, it is necessary to consider additional assumptions on the interference. As the user u does not have any knowledge about the number of streams of the other users, it is assumed that all users are served over $\ell_{s,i} = \ell_{u,i}$ streams, implying that the transmit power $P_{s,i} = P_{u,i}$ is the same for all users. Also, the worst case number of users $S_i = \frac{N_i}{\ell_{u,i}}$, in terms of residual multi-user interference, is considered. Without these assumptions separate feedback of the signal power and the interference power would be required.

Due to the statistical independence of the transmit signals of different users, the inference covariance matrix is obtained as the sum of the individual covariance matrices caused by the transmissions to the interfering users. From the input-output relationship (4.24), the sum interference is obtained as

$$\mathbf{R}_{u,i} = P_{u,i} \sum_{s \in \mathcal{S}_i, s \neq u} \mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff,n}})^{\mathrm{H}} \tilde{\mathbf{F}}_{s,i} \tilde{\mathbf{F}}_{s,i}^{\mathrm{H}} \mathbf{H}_{u,i}^{\mathrm{eff,n}}\right) = P_{u,i} \sum_{s \in \mathcal{S}_i, s \neq u} \mathbf{R}_{u,i}^{(s)}. \tag{G.12}$$

Notice that due to the BD construction the interference is not effective over $\mathbf{H}_{u,i}^{\mathrm{eff,r}}$. With an orthonormal basis $\hat{\mathbf{H}}_{u,i}^{\perp} \in \mathbb{C}^{N_i \times N_i - \ell_{u,i}}$ of the orthogonal complement of $\hat{\mathbf{H}}_{u,i}$, a single interference term can be written as

$$\mathbf{R}_{u,i}^{(s)} = \mathbb{E}\left((\mathbf{H}_{u,i}^{\mathrm{eff}})^{\mathrm{H}} \hat{\mathbf{H}}_{u,i}^{\perp} (\hat{\mathbf{H}}_{u,i}^{\perp})^{\mathrm{H}} \left(\hat{\mathbf{H}}_{u,i}^{\perp} \mathbf{Q}_{s,i} \mathbf{Q}_{s,i}^{\mathrm{H}} (\hat{\mathbf{H}}_{u,i}^{\perp})^{\mathrm{H}}\right) \hat{\mathbf{H}}_{u,i}^{\perp} (\hat{\mathbf{H}}_{u,i}^{\perp})^{\mathrm{H}} \mathbf{H}_{u,i}^{\mathrm{eff}}\right), \tag{G.13}$$

where $\mathbf{H}_{u,i}^{\mathrm{eff,n}}$ has been substituted from (4.23) and the BD construction has been exploited to express $\tilde{\mathbf{F}}_{s,i}$ in terms of the basis $\hat{\mathbf{H}}_{u,i}^{\perp}$ using the semi-unitary matrix $\mathbf{Q}_{s,i} \in \mathbb{C}^{N_i - \ell_{u,i} \times \ell_{s,i}}$. Due to the

unknown precoders, $\mathbf{Q}_{s,i}$ is assumed as uniformly distributed on $\mathcal{G}\left(N_i - \ell_{u,i}, \ell_{s,i}\right)$. With an SVD of $\mathbf{H}_{u,i}^{\text{eff}}$, as applied in (G.10), the covariance matrix is equal to

$$\mathbf{R}_{u,i}^{(s)} = P_{u,i}\,\frac{\ell_{s,i}}{N_i - \ell_{u,i}}\mathbf{V}_{u,i}^{\text{eff}}\mathbf{\Sigma}_{u,i}^{\text{eff}}\,\mathbb{E}\left((\mathbf{U}_{u,i}^{\text{eff}})^{\text{H}}\hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\text{H}}\mathbf{U}_{u,i}^{\text{eff}}\right)\mathbf{\Sigma}_{u,i}^{\text{eff}}(\mathbf{V}_{u,i}^{\text{eff}})^{\text{H}}. \tag{G.14}$$

To determine the expected value in the center, $\hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\text{H}}$ is replaced with $(\mathbf{I}_{N_i} - \hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\text{H}})$

$$\mathbb{E}\left((\mathbf{U}_{u,i}^{\text{eff}})^{\text{H}}\hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\text{H}}\mathbf{U}_{u,i}^{\text{eff}}\right) = \mathbf{I}_{N_i} - \mathbb{E}\left((\mathbf{U}_{u,i}^{\text{eff}})^{\text{H}}\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\text{H}}\mathbf{U}_{u,i}^{\text{eff}}\right). \tag{G.15}$$

Applying the decomposition and argumentation from (G.10) and below, the expected value is

$$\mathbb{E}\left((\mathbf{U}_{u,i}^{\text{eff}})^{\text{H}}\hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\text{H}}\mathbf{U}_{u,i}^{\text{eff}}\right) = \mathbf{I}_{N_i} - \mathbf{I}_{N_i}\left(1 - \frac{\mathrm{d_c^2}(\mathbf{U}_{u,i}^{\text{eff}}, \hat{\mathbf{H}}_{u,i})}{\ell_{u,i}}\right) = \frac{\mathrm{d_c^2}(\mathbf{U}_{u,i}^{\text{eff}}, \hat{\mathbf{H}}_{u,i})}{\ell_{u,i}}\mathbf{I}_{N_i}. \tag{G.16}$$

With the assumptions $\ell_{s,i} = \ell_{u,i}$, $P_{s,i} = P_{u,i}$ and $S_i = \frac{N_i}{\ell_{u,i}}$, the sum interference is obtained as

$$\mathbf{R}_{u,i} = \underbrace{P_{u,i}\frac{\ell_{u,i}}{N_i - \ell_{u,i}}\left(\frac{N_i}{\ell_{u,i}} - 1\right)\frac{\mathrm{d_c^2}(\mathbf{U}_{u,i}^{\text{eff}}, \hat{\mathbf{H}}_{u,i})}{\ell_{u,i}}}_{c_I}\mathbf{V}_{u,i}^{\text{eff}}\left(\mathbf{\Sigma}_{u,i}^{\text{eff}}\right)^2(\mathbf{V}_{u,i}^{\text{eff}})^{\text{H}}. \tag{G.17}$$

SINR lower bound: Assuming that a semi-unitary antenna combiner, as proposed in Section 4.3, is applied by the user, the post-equalization noise is i.i.d. Gaussian of variance $\tilde{\sigma}_z^2$. Using the covariance estimates derived above, the achievable user rate is lower bounded as [178]

$$R_{u,i} = \log_2\det\left(\mathbf{I}_{\ell_{u,i}} + \mathbf{S}_{u,i}\left(\tilde{\sigma}_z^2\mathbf{I}_{\ell_{u,i}} + \mathbf{R}_{u,i}\right)^{-1}\right) \geq \log_2\det\left(\mathbf{I}_{\ell_{u,i}} + \tilde{\mathbf{S}}_{u,i}\left(\tilde{\sigma}_z^2\mathbf{I}_{\ell_{u,i}} + \mathbf{R}_{u,i}\right)^{-1}\right). \tag{G.18}$$

Notice that both $\tilde{\mathbf{S}}_{u,i}$ and $\mathbf{R}_{u,i}$ are multiplied with $\mathbf{V}_{u,i}^{\text{eff}}$ and $(\mathbf{V}_{u,i}^{\text{eff}})^{\text{H}}$ from the left and right, respectively. These unitary matrices hence cancel out in the achievable rate lower bound. Based on this observation, the per stream post-equalization SINR lower bound of user u is defined as

$$\mathbb{E}\left(\beta_{\nu,u,i}\right) \geq \tilde{\beta}_{\nu,u,i} = \frac{c_S\left(\sigma_{\nu,u,i}^{\text{eff}}\right)^2}{\tilde{\sigma}_z^2 + c_I\left(\sigma_{\nu,u,i}^{\text{eff}}\right)^2}, \quad \mathbf{\Sigma}_{u,i}^{\text{eff}} = \text{diag}\left(\sigma_{1,u,i}^{\text{eff}}, \ldots, \sigma_{\ell_{u,i},u,i}^{\text{eff}}\right). \tag{G.19}$$

Thus, with scalar feedback of these SINRs, the lower bound on the achievable user rate can be calculated by the base station, facilitating efficient multi-user scheduling.

Appendix H.

MMSE Equalizers for BD based MU-MIMO

The channel subspace selection methods and the corresponding antenna combiners proposed in Section 4.3 for BD precoding with excess antennas at the users are based on knowledge of the channel matrix $\mathbf{H}_{u,i}[n,k]$ only, requiring neither any information about the precoders nor about the residual multi-user interference. This is an appropriate approach for the selection of the preferred subspace $\mathrm{span}\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]\right)$ that is kept free of interference by the BD precoder (see Section 4.1), because the residual multi-user interference is unknown ahead of time an can therefore not be considered during feedback calculation. However, as soon as the transmission takes place, the users may be able to determine instantaneous or statistical information about the residual multi-user interference. Then, instead of simply applying the SQBC or MET antenna combiner to filter-out the subspace $\mathrm{span}\left(\mathbf{H}_{u,i}^{\mathrm{eff}}[n,k]\right)$, more sophisticated equalizers can be applied to reduce the impact of the multi-user interference. Two MIMO minimum mean-squared error (MMSE) equalizers, which are published in [92], are derived below exploiting in one case full knowledge of the interfering precoders and in the other case knowledge of the BD precoder construction to estimate the average interference. To simplify notations the RE index $[n,k]$ is omitted.

H.1. Interference-Aware MMSE Equalizer

As a benchmark combiner, an interference-aware MMSE receiver [11, 215] for BD precoding with imperfect CSIT is proposed, which is based on the assumption of full knowledge of the interfering precoders at the user. For that purpose, the channel matrix is decomposed in terms of the quantized channel subspace $\hat{\mathbf{H}}_{u,i}$ as

$$\mathbf{H}_{u,i} = \left(\hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}} + (\mathbf{I}_{M_{u,i}} - \hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}})\right)\mathbf{H}_{u,i} = \mathbf{H}_{u,i}^{(r)} + \mathbf{H}_{u,i}^{(n)}, \tag{H.1}$$

with $\mathbf{H}_{u,i}^{(r)}$ and $\mathbf{H}_{u,i}^{(n)}$ denoting the component of the channel in the range-space and the left null space of $\hat{\mathbf{H}}_{u,i}$, respectively. With (H.1) the input-output relationship for BD with imperfect CSIT can be

written as

$$\mathbf{r}_{u,i} = \mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{F}_{u,i}\mathbf{x}_{u,i} + (\mathbf{H}_{u,i}^{(n)})^{\mathrm{H}} \sum_{s=1,s\neq u}^{S_i} \mathbf{F}_{s,i}\mathbf{x}_{s,i} + \tilde{\mathbf{z}}_{u,i}, \tag{H.2}$$

where the range-space component in the multi-user interference term vanishes due to BD. The interference-aware MMSE receiver is obtained from the optimization problem

$$\mathbf{G}_{u,i}^{(\text{MMSE-IA})} = \underset{\mathbf{G}_{u,i}\in\mathbb{C}^{M_{u,i}\times\ell_{u,i}}}{\operatorname{argmin}} \mathbb{E}_{\mathbf{x}_{u,i},\mathbf{x}_{s,i},\tilde{\mathbf{z}}_{u,i}} \left(\left\| \mathbf{x}_{u,i} - \mathbf{G}_{u,i}^{\mathrm{H}}\mathbf{r}_{u,i} \right\| \right), \tag{H.3}$$

with the expectation being taken with respect to the transmit signals and the additive noise. The solution is obtained as

$$\begin{aligned}\mathbf{G}_{u,i}^{(\text{MMSE-IA})} = \Bigg(& \tilde{\sigma}_z^2\mathbf{I}_{M_{u,i}} + \mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{F}_{u,i}\mathbf{F}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i} \\ & + \sum_{s=1,s\neq u}^{S_i} (\mathbf{H}_{u,i}^{(n)})^{\mathrm{H}}\mathbf{F}_{s,i}\mathbf{F}_{s,i}^{\mathrm{H}}\mathbf{H}_{u,i}^{(n)} \Bigg)^{-1} \mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{F}_{u,i},\end{aligned} \tag{H.4}$$

where the transmit signal normalization of Equation (2.3) is exploited.

This interference-aware receiver cannot be implemented with the signaling structure of LTE, because the users are not able to estimate the interfering precoders. An extension of the user-specific reference symbol structure of LTE is hence required to facilitate advanced receivers.

H.2. Interference-Averaged MMSE Equalizer

If the interfering precoders are assumed as unknown, an alternative MMSE optimization problem can be considered by averaging the mean squared error over the unknown interferers as well. Then it is not necessary for the users to estimate the interfering precoders, reducing the complexity of the receiver and the downlink signaling overhead considerably. The MMSE receiver for averaged interference is obtained from the optimization problem

$$\mathbf{G}_{u,i}^{(\text{MMSE})} = \underset{\mathbf{G}_{u,i}\in\mathbb{C}^{M_{u,i}\times\ell_{u,i}}}{\operatorname{argmin}} \mathbb{E}_{\mathbf{F}_{s,i},\mathbf{x}_{u,i},\mathbf{x}_{s,i},\tilde{\mathbf{z}}_{u,i}} \left(\left\| \mathbf{x}_{u,i} - \mathbf{G}_{u,i}^{\mathrm{H}}\mathbf{y}_{u,i} \right\| \right). \tag{H.5}$$

To solve this optimization problem it is necessary to calculate the interference covariance matrix

$$\begin{aligned}\mathbf{R}_{u,i} &= \mathbb{E}_{\mathbf{F}_{s,i},\mathbf{x}_{s,i}} \left((\mathbf{H}_{u,i}^{\mathrm{H}} \sum_{s=1,s\neq u}^{S_i} \mathbf{F}_{s,i}\mathbf{x}_{s,i})(\mathbf{H}_{u,i}^{\mathrm{H}} \sum_{s=1,s\neq u}^{S_i} \mathbf{F}_{s,i}\mathbf{x}_{s,i})^{\mathrm{H}} \right) \\ &= \sum_{s=1,s\neq u}^{S_i} \mathbb{E}_{\mathbf{F}_{s,i}} \left(\mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{F}_{s,i}\mathbf{F}_{s,i}^{\mathrm{H}}\mathbf{H}_{u,i} \right),\end{aligned} \tag{H.6}$$

which is obtained by considering the statistical independence of the transmit signals of different users. With (H.1) and considering the BD construction, the terms in the sum can be rewritten as

$$\mathbb{E}_{\mathbf{F}_{s,i}}\left(\mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{F}_{s,i}\mathbf{F}_{s,i}^{\mathrm{H}}\mathbf{H}_{u,i}\right) = (\mathbf{H}_{u,i}^{(n)})^{\mathrm{H}}\mathbb{E}\left(\mathbf{F}_{s,i}\mathbf{F}_{s,i}^{\mathrm{H}}\right)\mathbf{H}_{u,i}^{(n)}. \tag{H.7}$$

For a given quantized channel subspace $\hat{\mathbf{H}}_{u,i}$, the precoders $\mathbf{F}_{s,i}$ are chosen in the left null space of $\hat{\mathbf{H}}_{u,i}$ due to the BD construction (4.4). Without further knowledge of the precoders, it is assumed that each $\mathbf{F}_{s,i}$ is isotropically distributed in the orthogonal complement of $\mathrm{span}(\hat{\mathbf{H}}_{u,i})$, leading to the decomposition

$$\mathbf{F}_{s,i} = \hat{\mathbf{H}}_{u,i}^{\perp}\mathbf{C}_{s,i},\ \mathbf{C}_{s,i} \in \mathbb{C}^{(N_i-\ell_{u,i})\times\ell_{s,i}}, \tag{H.8}$$

$$\hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\mathrm{H}} = \mathbf{I} - \hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}. \tag{H.9}$$

Here, the orthogonal complement of $\mathrm{span}(\hat{\mathbf{H}}_{u,i})$ is spanned by the orthonormal basis $\hat{\mathbf{H}}_{u,i}^{\perp}$. The coordinate matrix $\mathbf{C}_{s,i}$ is composed of $\ell_{s,i}$ independent and isotropically distributed vectors $\mathbf{c}_{s,i}^{(j)}$ in the $(N_i - \ell_{u,i})$-dimensional left null space of $\hat{\mathbf{H}}_{u,i}$. With these assumptions the expected value in (H.7) can be evaluated as

$$\begin{aligned}\mathbb{E}\left(\mathbf{F}_{s,i}\mathbf{F}_{s,i}^{\mathrm{H}}\right) &= \hat{\mathbf{H}}_{u,i}^{\perp}\mathbb{E}\left(\mathbf{C}_{s,i}\mathbf{C}_{s,i}^{\mathrm{H}}\right)(\hat{\mathbf{H}}_{u,i}^{\perp})^{\mathrm{H}} \\ = \hat{\mathbf{H}}_{u,i}^{\perp}\sum_{j=1}^{\ell_{s,i}}\mathbb{E}\left(\mathbf{c}_{s,i}^{(j)}(\mathbf{c}_{s,i}^{(j)})^{\mathrm{H}}\right)(\hat{\mathbf{H}}_{u,i}^{\perp})^{\mathrm{H}} &= \frac{P_i}{\ell_{s,i}\,S_i}\,\frac{\ell_{s,i}}{N_i-\ell_{u,i}}\,\hat{\mathbf{H}}_{u,i}^{\perp}(\hat{\mathbf{H}}_{u,i}^{\perp})^{\mathrm{H}}.\end{aligned} \tag{H.10}$$

Here, the first fraction is due to the assumption of uniform power allocation over users and spatial streams, and the second term is obtained from the isotropy of the $\ell_{s,i}$ vectors $\mathbf{c}_{s,i}^{(j)}$ in the $(N_i - \ell_{u,i})$-dimensional orthogonal complement of $\mathrm{span}(\hat{\mathbf{H}}_{u,i})$. Substituting (H.7), (H.9) and (H.10) into (H.6) and exploiting the idempotence of the projection matrix (H.9), the interference covariance matrix is obtained as

$$\mathbf{R}_{u,i} = \frac{P_i}{S_i(N_i-\ell_{u,i})}(S_i-1)\mathbf{H}_{u,i}^{\mathrm{H}}\left(\mathbf{I}_{M_{u,i}} - \hat{\mathbf{H}}_{u,i}\hat{\mathbf{H}}_{u,i}^{\mathrm{H}}\right)\mathbf{H}_{u,i}. \tag{H.11}$$

With this result the MMSE receiver for averaged interference is

$$\mathbf{G}_{u,i}^{(\mathrm{MMSE})} = \left(\tilde{\sigma}_z^2\mathbf{I}_{M_{u,i}} + \mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{F}_{u,i}\mathbf{F}_{u,i}^{\mathrm{H}}\mathbf{H}_{u,i} + \mathbf{R}_{u,i}\right)^{-1}\mathbf{H}_{u,i}^{\mathrm{H}}\mathbf{F}_{u,i}, \tag{H.12}$$

which can be calculated from local CSI only.

H.3. Performance Comparison

In this section, the throughput performance of the proposed interference-aware and interference-averaged MMSE equalizers is compared to the interference-unaware SQBC and MET antenna

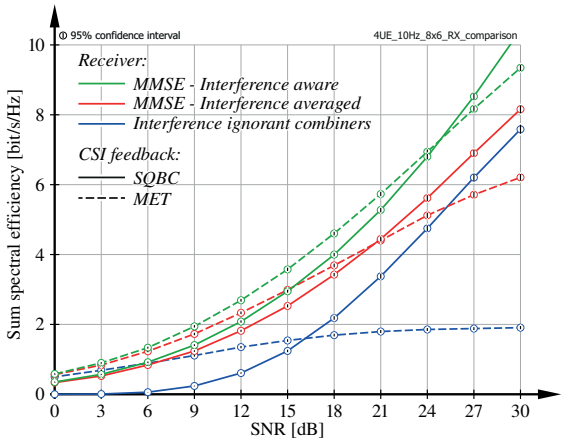


Figure H.1.: Comparison of different combinations of channel subspace feedback strategies and receive equalizers.

combiners followed by single-user MIMO MMSE equalization to separate the streams of a user. The interference-aware/averaged MMSE receivers are combined with either SQBC or MET channel subspace feedback.

A system with $N_i = 8$ transmit antennas, $M_{u,i} = 6$ receive antennas and $S_i = 4$ served users, receiving $\ell_{u,i} = 2$ streams each, is considered. Memoryless CSI quantization with a random Grassmannian codebook of size 10 bit is assumed. The channel matrices are generated according to Section 2.4, assuming a receive antenna correlation parameter of $\alpha_{\text{corr}} = 0.9$, temporally independent channel realizations and frequency flat fading.

The results are shown in Figure H.1. Solid lines correspond to results obtained with SQBC feedback and dashed lines correspond to MET feedback. Different colors are used to distinguish between different receiver strategies.[1] Irrespective of the applied equalizer, it can be observed that MET feedback outperforms SQBC feedback in the low SNR regime. In this range, the residual multi-user interference is dominated by the noise and a large channel gain is important to maximize the power of the desired signal. In the high SNR regime, however, the achievable rate is determined by the multi-user interference. As the interference experiences the same channel gain as the desired signal, the achievable rate is maximized by minimizing the interference, which is achieved with SQBC feedback. As expected, the best performance is obtained with the interference-aware MMSE equalizer. Especially in the high SNR regime the other proposals are significantly outperformed by the interference-aware receiver, justifying the increased signaling overhead required to determine the interfering precoders at the users. A large performance gain compared to the interference-unaware receivers can also be attained with the interference-averaged MMSE combiner, without posing a significant burden on interference estimation and downlink signaling, hence providing a reasonable trade-off that is compatible with the LTE standard.

[1] Notice that the combination of color and line-style specifies the applied feedback and receiver strategy in Figure H.1.

Appendix I.

Out-Of-Cell Interference Models

The simulation results presented in Chapter 5 are obtained with a hybrid of link and system level simulations. When simulating a DAS, it is important to consider multiple cells of the cellular network, because the out-of-cell interference landscape is significantly impacted by distributing RRUs over the cell area. For complexity reasons it was not possible to perform system level simulations of multi-cell DASs, because detailed modeling of the physical layer of the wireless channel is required for the proposed transceivers and feedback algorithms, entailing heavy requirements on memory and computational power. To circumvent this problem, a statistical interference model is employed to generate the interference caused by other cells of the network, reducing the necessary link level simulations to a single cell only. The out-of-cell interference gamma distribution model of [32] is employed and generalized to represent the interference from other cells.

In this appendix, the parameters of the out-of-cell interference gamma distribution model are derived, generalizing the results of [32] to multi-stream transmission per user. The derived model is applicable to SU-MIMO as well as ZF and BD based MU-MIMO. The pre-equalization out-of-cell interference is generated for each receive antenna individually, following the same distribution. Hence, in the following the out-of-cell interference is derived for an arbitrary receive antenna $r \in \{1, \ldots, M_{u,i}\}$. The calculations are obtained under the assumption that the small-scale fading is Rayleigh distributed, implying that the squared magnitude of each individual interfering channel is chi-squared distributed, which is a special case of the gamma distribution. Due to the different pathloss characteristics of the RRUs, the sum interference is in general not gamma distributed, but it can be reasonably well approximated with a gamma distribution following the approach of [32], as detailed below.

From Equation (2.4), the received signal on the r-th antenna is given by[1]

$$r_{u,i} = \mathbf{h}_{u,i}^{\mathrm{H}} \mathbf{F}_{u,i} \mathbf{x}_{u,i} + \mathbf{h}_{u,i}^{\mathrm{H}} \sum_{\substack{s \in \mathcal{S}_i \\ s \neq u}} \mathbf{F}_{s,i} \mathbf{x}_{s,i} + \sum_{j=0, j \neq i}^{I} \mathbf{h}_{u,i}^{(j)\mathrm{H}} \sum_{s \in \mathcal{S}_j} \mathbf{F}_{s,j} \mathbf{x}_{s,j} + \mathbf{z}_{u,i}, \tag{I.1}$$

$$r_{u,i} = [\mathbf{r}_{u,i}]_r\,, \;\; \mathbf{h}_{u,i} = \left[\mathbf{H}_{u,i}\right]_{:,r}, \;\; \mathbf{h}_{u,i}^{(j)} = \left[\mathbf{H}_{u,i}^{(j)}\right]_{:,r}$$

[1] The RE index is suppressed to shorten notations.

The out-of-cell interference model of [32] is based on a moment matching approach. The basic idea is to match the first- and second-order moments of the power of each out-of-cell interference term, i.e., each summand of the third term in Equation (I.1), to a gamma distribution. Then, employing [32, Lemma 7], the first- and second-order moments of the sum of these gamma random variables are calculated. These moments are matched to yet another gamma random variable by applying [32, Proposition 8], resulting in the approximate statistics of the out-of-cell interference.

Exploiting the statistical independence of the transmit symbols of different users (2.8), the instantaneous out-of-cell interference power is obtained as

$$I_{u,i} = \mathbb{E}_{\mathbf{x}_{s,j}} \left(\left\| \sum_{j=0, j\neq i}^{I} \mathbf{h}_{u,i}^{(j)\mathrm{H}} \sum_{s\in\mathcal{S}_j} \mathbf{F}_{s,j} \mathbf{x}_{s,j} \right\|^2 \right) = \sum_{j=0, j\neq i}^{I} \sum_{s\in\mathcal{S}_j} \left\| \mathbf{h}_{u,i}^{(j)\mathrm{H}} \mathbf{F}_{s,j} \right\|^2 . \tag{I.2}$$

A single interference term of Equation (I.2) can be written as

$$I_{u,i}^{(s,j)} = \left\| \mathbf{h}_{u,i}^{(j)\mathrm{H}} \mathbf{F}_{s,j} \right\|^2 = \left\| \mathbf{F}_{s,j}^{\mathrm{H}} \mathbf{h}_{u,i}^{(j)} \right\|^2 . \tag{I.3}$$

To further proceed with this term, the precoder $\mathbf{F}_{s,j}$ is assumed to be an isotropically distributed scaled semi-unitary matrix specifying a point on the Grassmann manifold of $\ell_{s,j}$-dimensional subspaces in the N_j dimensional Euclidean space $\mathcal{G}\left(N_j, \ell_{s,j}\right)$, with

$$\mathbf{F}_{s,j}^{\mathrm{H}} \mathbf{F}_{s,j} = \frac{P_j}{\ell_{s,j}\, S_j} \mathbf{I}_{\ell_{s,j}}, \tag{I.4}$$

where P_j is the transmit power of cell j and S_j is the number of served users in cell j. Utilizing results of [216] (after generalization to the complex-valued case), it can be shown that the term

$$\frac{\ell_{s,j}\, S_j}{P_j} \frac{\left\| \mathbf{F}_{s,j}^{\mathrm{H}} \mathbf{h}_{u,i}^{(j)} \right\|^2}{\left\| \mathbf{h}_{u,i}^{(j)} \right\|^2}$$

is beta distributed with parameters $\beta\left(\ell_{s,j}, N_j - \ell_{s,j}\right)$ for a fixed channel $\mathbf{h}_{u,i}^{(j)}$. Using the properties of a beta distribution, the conditional mean and variance of each interference term are obtained as

$$\begin{aligned} \mu_{|\mathbf{h}_{u,i}^{(j)}}\left(I_{u,i}^{(s,j)}\right) &:= \mathbb{E}\left(\left\| \mathbf{h}_{u,i}^{(j)\mathrm{H}} \mathbf{F}_{s,j} \right\|^2 \middle| \mathbf{h}_{u,i}^{(j)} \right) = \frac{P_j}{\ell_{s,j}\, S_j} \left\| \mathbf{h}_{u,i}^{(j)} \right\|^2 \mathbb{E}\left(\beta\left(\ell_{s,j}, N_j - \ell_{s,j}\right)\right) \\ &= \frac{P_j}{\ell_{s,j}\, S_j} \left\| \mathbf{h}_{u,i}^{(j)} \right\|^2 \frac{\ell_{s,j}}{N_j} = \frac{P_j}{S_j\, N_j} \left\| \mathbf{h}_{u,i}^{(j)} \right\|^2 , \end{aligned} \tag{I.5}$$

$$\begin{aligned} \sigma^2_{|\mathbf{h}_{u,i}^{(j)}}\left(I_{u,i}^{(s,j)}\right) &:= \mathrm{var}\left(\left\| \mathbf{h}_{u,i}^{(j)\mathrm{H}} \mathbf{F}_{s,j} \right\|^2 \middle| \mathbf{h}_{u,i}^{(j)} \right) \\ &= \left(\frac{P_j}{\ell_{s,j}\, S_j} \right)^2 \left\| \mathbf{h}_{u,i}^{(j)} \right\|^4 \frac{\ell_{s,j}(N_j - \ell_{s,j})}{N_j^2 (N_j + 1)} = \frac{N_j - \ell_{s,j}}{\ell_{s,j}(N_j+1)} \mu_{|\mathbf{h}_{u,i}^{(j)}}\left(I_{u,i}^{(s,j)}\right)^2 . \end{aligned} \tag{I.6}$$

Notice that the conditional mean is independent of the number of data streams $\ell_{s,j}$, due to the assumption of equal power allocation over streams. Removing the conditioning on $\mathbf{h}_{u,i}^{(j)}$ by employing the result in [32, Proposition 11], the unconditional moments are calculated as

$$\mu_{u,i}\left(I_{u,i}^{(s,j)}\right) = \frac{1}{N_j}\sum_{n=1}^{N_j}\frac{P_j}{S_j}\gamma_{u,i}^{(n,j)} = \frac{1}{N_j}\sum_{n=1}^{N_j}\sigma_{u,i}^{(n,j)^2}, \tag{I.7}$$

$$\sigma_{u,i}^2\left(I_{u,i}^{(s,j)}\right) = \frac{N_j-\ell_{s,j}}{N_j^2\,\ell_{s,j}(N_j+1)}\left(2\sum_{n=1}^{N_j}\sigma_{u,i}^{(n,j)^4} + \sum_{n=1}^{N_j}\sum_{l=1,l\neq n}^{N_j}\sigma_{u,i}^{(n,j)^2}\sigma_{u,i}^{(l,j)^2}\right). \tag{I.8}$$

with $\gamma_{u,i}^{(n,j)}$ being defined in Equation (2.11). This result is a generalization of the out-of-cell interference models presented in [96] for the special cases of SU-MIMO and ZF beamforming based MU-MIMO to BD with an arbitrary number of users per cell and an arbitrary number of streams per user. Utilizing [32, Lemma 7], the second-order gamma match of a single interference term is obtained as

$$I_{u,i}^{(s,j)} \sim \Gamma\left(k_{u,i}^{(s,j)}, \theta_{u,i}^{(s,j)}\right), \tag{I.9}$$

$$k_{u,i}^{(s,j)} = \frac{\mu_{u,i}\left(I_{u,i}^{(s,j)}\right)^2}{\sigma_{u,i}^2\left(I_{u,i}^{(s,j)}\right)},\quad \theta_{u,i}^{(s,j)} = \frac{\sigma_{u,i}^2\left(I_{u,i}^{(s,j)}\right)}{\mu_{u,i}\left(I_{u,i}^{(s,j)}\right)}, \tag{I.10}$$

with $\Gamma\left(k_{u,i}^{(s,j)}, \theta_{u,i}^{(s,j)}\right)$ denoting the gamma distribution with shape $k_{u,i}^{(s,j)}$ and scale $\theta_{u,i}^{(s,j)}$. Notice that the number of streams per user $\ell_{s,j}$ must be strictly less than the total number of transmit antennas N_j to obtain a useful match. With [32, Proposition 8] the gamma match of the sum interference defined in Equation (I.2) can be calculated as

$$I_{u,i} \sim \Gamma\left(k_{u,i}, \theta_{u,i}\right), \tag{I.11}$$

$$k_{u,i} = \frac{\left(\sum_{j=0,j\neq i}^{I}\sum_{s\in\mathcal{S}_j} k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)}\right)^2}{\sum_{j=0,j\neq i}^{I}\sum_{s\in\mathcal{S}_j} k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)^2}},\quad \theta_{u,i} = \frac{\sum_{j=0,j\neq i}^{I}\sum_{s\in\mathcal{S}_j} k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)^2}}{\sum_{j=0,j\neq i}^{I}\sum_{s\in\mathcal{S}_j} k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)}}. \tag{I.12}$$

In the presented simulations, the number of data streams per user $\ell_{s,j}$ is equal for all users s of a cell j, entailing that $I_{u,i}^{(s,j)}$ is independent of the user index s. Then the parameters of the gamma match of the sum interference simplify to

$$\tilde{k}_{u,i} = \frac{\left(\sum_{j=0,j\neq i}^{I} S_j\, k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)}\right)^2}{\sum_{j=0,j\neq i}^{I} S_j\, k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)^2}},\quad \tilde{\theta}_{u,i} = \frac{\sum_{j=0,j\neq i}^{I} S_j\, k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)^2}}{\sum_{j=0,j\neq i}^{I} S_j\, k_{u,i}^{(s,j)}\,\theta_{u,i}^{(s,j)}}, \tag{I.13}$$

with $k_{u,i}^{(s,j)}$ and $\theta_{u,i}^{(s,j)}$ taken from an arbitrary user s of cell j.

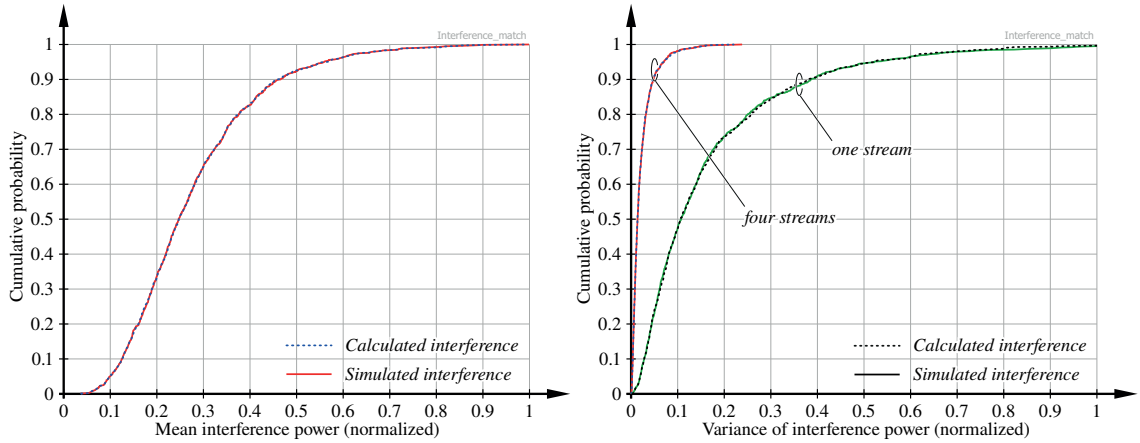


Figure I.1.: Cumulative probability of the mean and variance of the out-of-cell interference power of a single interferer receiving one or four spatially multiplexed data streams.

Compared to the results in [32], the novel part of this derivation are the conditional moments (I.5) and (I.6). To demonstrate the validity of these moments, empirical cumulative distribution functions of $\mu_{u,i}\left(I_{u,i}^{(s,j)}\right)$ and $\sigma_{u,i}^2\left(I_{u,i}^{(s,j)}\right)$ are compared in Figure I.1 to the calculated values according to (I.5) and (I.6). The Monte-Carlo simulations of the out-of-cell interference are conducted for a single non-sectorized interfering cell comprising one base station and six single antenna RRUs, transmitting either one or four parallel data streams. It can be seen that a perfect match of the body of the calculated and simulated cumulative distribution functions is obtained.[2] It is observed in Figure I.1 that the mean interference is independent of the number of transmitted data streams, while the variance decreases with a growing number of data streams. The accuracy of the remaining parts of the considered moment matching approach is investigated in detail in [32].

Notice that the tail probability of the conditional moments is not investigated here, because it is not of significant importance in this thesis. In the conducted performance investigations the focus is put on average throughputs and mean spectral efficiencies. These average values are only marginally impacted by the tail probabilities of the interference.

[2]The results are normalized to the maximum observed value.

Bibliography

[1] UMTS Forum, "Mobile traffic forecasts 2010 - 2020 report," UMTS Forum Report 44, Jan. 2011.

[2] Ericsson, "Ericsson mobility report," white paper, June 2013.

[3] Cisco Systems Inc., "Cisco visual networking index: forecast update, 2012-2017," white paper, Feb. 2013.

[4] Ericsson, "Traffic exploration tool," http://www.ericsson.com/TET, [Online; accessed 03-July-2013].

[5] T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE Performance: Evolution Towards 3G/UMTS*. John Wiley and Sons Ltd, 2003.

[6] 3GPP, "Technical specification group services and system aspects; release 1999 specifications 3G TS 21.101 version 1.0.0," http://www.3gpp.org/ftp/Specs/html-info/21101.htm, Oct. 1999.

[7] 3GPP, "Technical specification group radio access network; high speed downlink packet access (HSDPA); overall description; stage 2 (release 11)," http://www.3gpp.org/ftp/Specs/html-info/25308.htm, March 2013.

[8] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (release 8)," http://www.3gpp.org/ftp/Specs/html-info/36211.htm, Sept. 2009.

[9] International Telecommunication Union ITU, "Requirements related to technical performance for IMT-advanced radio interface(s), ITU-R M.2134," http://www.itu.int/pub/R-REP-M.2134-2008/en.

[10] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (release 10)," http://www.3gpp.org/ftp/Specs/html-info/36211.htm, Dec. 2010.

[11] S. Caban, C. Mehlführer, M. Rupp, and M. Wrulich, *Evaluation of HSDPA and LTE: From Testbed Measurements to System Level Performance*. UK: John Wiley & Sons, 2012.

[12] IEEE 802.16, "IEEE standard for air interface for broadband wireless access systems," June 2012.

[13] IEEE 802.11, "IEEE standard for information technology - specification requirements part 11: wireless LAN MAC and PHY specifications," March 2012.

[14] Rhode & Schwarz, "Carrier aggregation - (one) key enabler for LTE-advanced," white paper, Oct. 2012.

[15] Electronic communications committee within the European conference of postal and telecommunications administrations, "The European table of frequency allocations and applications in the frequency range 8.3 kHz to 3000 GHz (ECA table)," ERC report 25, Feb. 2013.

[16] Tolaga research, "The spectrum rubric for mobile broadband - A European perspective," Sept. 2010.

[17] H. Arslan and M. E. Scahin, "Cognitive UWB-OFDM: Pushing ultra-wideband beyond its limit via opportunistic spectrum usage," *Journal of Comm. and Networks*, vol. 8, no. 2, pp. 151–157, 2006.

[18] F. Peng, Y. Gao, and L. Cuthbert, "Reviews on cognitive access to TV white space," in *7th International ICST Conference on Communications and Networking in China*, Aug. 2012, pp. 727–732.

[19] K. Harrison, S. Mishra, and A. Sahai, "How much white-space capacity is there?" in *IEEE Symposium on New Frontiers in Dynamic Spectrum*, April 2010, pp. 1–10.

[20] S. Akoum, O. El Ayach, and R. Heath, Jr., "Coverage and capacity in mmWave cellular systems," in *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 688–692.

[21] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, 2013.

[22] A. Bleicher, "The 5G phone future," *IEEE Spectrum*, vol. 50, no. 7, July 2013.

[23] G. Stüber, *Principles of Mobile Communication*. Kluwer Academic Publishers, 2001.

[24] Y. Liang, A. Goldsmith, G. Foschini, R. Valenzuela, and D. Chizhik, "Evolution of base stations in cellular networks: Denser deployment versus coordination," in *IEEE International Conference on Communications*, June 2008, pp. 4128–4132.

[25] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks," *CoRR*, vol. abs/1109.2992, 2011.

[26] M. Taranetz and M. Rupp, "Performance of femtocell access point deployments in user hot-spot scenarios," in *Australasian Telecomm. Networks and Applications Conference*, Nov. 2012, pp. 1–5.

[27] A. Saleh, A. Rustako, and R. Roman, "Distributed antennas for indoor radio communications," *IEEE Transactions on Communications*, vol. 35, no. 12, Dec. 1987.

[28] M. Riaz, R. Nielsen, J. Pedersen, N. Prasad, and O. Madsen, "On radio over fiber for heterogeneous wireless networks," in *International Conference on Wireless and Optical Communications Networks*, April 2009.

[29] P. Chow, A. Karim, V. Fung, and C. Dietrich, "Performance advantages of distributed antennas in indoor wireless communication systems," in *IEEE 44th Vehicular Technology Conference*, vol. 3, June 1994.

[30] K. Kerpez and S. Ariyavisitakul, "A radio access system with distributed antennas," in *IEEE Global Telecommunications Conference*, vol. 3, Nov. 1994.

[31] W. Choi and J. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, Jan. 2007.

[32] R. Heath, Jr., T. Wu, Y. H. Kwon, and A. C. K. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, Oct. 2011.

[33] L. Garcia, K. Pedersen, and P. Mogensen, "On open versus closed LTE-advanced femtocells and dynamic interference coordination," in *IEEE Wireless Communications and Networking Conference*, 2010, pp. 1–6.

[34] H.-S. Jo, P. Xia, and J. Andrews, "Open, closed, and shared access femtocells in the downlink," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 363, 2012.

[35] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[36] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, 2002.

[37] A. Paulraj, D. Gore, R. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to Gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, 2004.

[38] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, Nov. 1999.

[39] M. Simko, C. Mehlfuhrer, M. Wrulich, and M. Rupp, "Doubly dispersive channel estimation with scalable complexity," in *International ITG Workshop on Smart Antennas*, Feb. 2010, pp. 251–256.

[40] D. Love and R. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967–2976, 2005.

[41] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2636–2647, 2003.

[42] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. on Information Theory*, vol. 51, no. 7, pp. 2491–2509, 2005.

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

[44] D. Gesbert, M. Kountouris, R. Heath, Jr., C. Chae, and T. Sälzer, "From single user to multiuser communications: Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, p. 36, Oct. 2007.

[45] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, p. 5, Nov. 2006.

[46] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE Journal on Sel. Areas in Communications*, vol. 26, no. 8, pp. 1473 –1482, Oct. 2008.

[47] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.

[48] X. Tao, X. Xu, and Q. Cui, "An overview of cooperative communications," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 65–71, 2012.

[49] S. Schwarz, J. Ikuno, M. Simko, M. Taranetz, Q. Wang, and M. Rupp, "Pushing the limits of LTE: A survey on research enhancing the standard," *IEEE Access*, vol. 1, pp. 51–62, 2013.

[50] E. Pateromichelakis, M. Shariat, A. Quddus, and R. Tafazolli, "On the evolution of multi-cell scheduling in 3GPP LTE / LTE-A," *IEEE Communications Surveys Tutorials*, vol. 15, no. 2, pp. 701–717, 2013.

[51] C.-B. Chae, I. Hwang, R. Heath, Jr., and V. Tarokh, "Interference aware-coordinated beamforming in a multi-cell system," *IEEE Trans. on Wireless Communications*, vol. 11, no. 10, pp. 3692–3703, 2012.

[52] M. Khan, M.-H. Lee, and M. Mustary, "Zero-forcing based multiuser beamforming for coordinated multi-point transmission," in *International Symposium on Communications and Information Technologies*, 2012, pp. 589–593.

[53] M. Rezaeekheirabadi and M. Guillaud, "Limited feedback for interference alignment in the k-user MIMO interference channel," in *Proc. Information Theory Workshop*, Lausanne, Switzerland, Sept. 2012, pp. 1–5.

[54] L. Qiang, Y. Yang, F. Shu, and W. Gang, "SLNR precoding based on QBC with limited feedback in downlink CoMP system," in *International Conference on Wireless Communications and Signal Processing*, 2010, pp. 1–5.

[55] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multi-point transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 148–155, 2012.

[56] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," DEC, Tech. Rep. TR-301, Sept. 1984.

[57] Z. Sun, C. Yin, and G. Yue, "Reduced-complexity proportional fair scheduling for OFDMA systems," in *IEEE International Conference on Communications, Circuits and Systems*, vol. 2, 2006, pp. 1221–1225.

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

[59] C. Mehlführer, J. C. Ikuno, M. Simko, S. Schwarz, and M. Rupp, "The Vienna LTE simulators — Enabling reproducibility in wireless communications research," *EURASIP Journal on Advances in Signal Processing; Special Issue on Reproducible Research*, vol. 2011, no. 1, p. 29, 2011.

[60] D. Love, R. Heath, Jr., V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, Oct. 2008.

[61] S. Schwarz, M. Wrulich, and M. Rupp, "Mutual information based calculation of the precoding matrix indicator for 3GPP UMTS/LTE," in *International ITG Workshop on Smart Antennas*, Bremen, Germany, Feb. 2010, pp. 52–58.

[62] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *6th Conf. on Wireless Advanced*, London, UK, June 2010, pp. 1–6.

[63] S. Schwarz, M. Simko, and M. Rupp, "On performance bounds for MIMO OFDM based wireless communication systems," in *IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, San Francisco, CA, June 2011, pp. 311–315.

[64] S. Schwarz and M. Rupp, "Throughput maximizing feedback for MIMO OFDM based wireless communication systems," in *IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, San Francisco, CA, June 2011, pp. 316–320.

[65] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439 – 441, May 1983.

[66] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Transcations on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[67] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936 –3964, Sept. 2006.

[68] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195 – 202, Jan. 2005.

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

[70] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[71] J. Lee and N. Jindal, "High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding," *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4787 –4792, Dec. 2007.

[72] J. S. Kim, H. Kim, C. S. Park, and K. B. Lee, "On the performance of multiuser MIMO systems in WCDMA/HSDPA: Beamforming, feedback and user diversity," *IEICE Transactions*, pp. 2161–2169, 2006.

[73] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," *IEEE Trans. on Wireless Communications*, vol. 11, no. 9, pp. 3037–3046, 2012.

[74] B. Nosrat-Makouei, J. Andrews, and R. Heath, Jr., "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2783 –2794, June 2011.

[75] B. Zhou, L. Jiang, S. Zhao, and C. He, "BER analysis of TDD downlink multiuser MIMO systems with imperfect channel state information," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 104, 2011.

[76] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*, ser. The Kluwer international series in engineering and computer science : communications and information theory. Kluwer Academic Publishers, 1992.

[77] T. Kim, D. Love, and B. Clerckx, "MIMO systems with limited rate differential feedback in slowly varying channels," *IEEE Transactions on Communications*, vol. 59, no. 4, pp. 1175 –1189, April 2011.

[78] O. El Ayach and R. Heath, Jr., "Grassmannian differential limited feedback for interference alignment," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6481 –6494, Dec. 2012.

[79] L. Liu and H. Jafarkhani, "Novel transmit beamforming schemes for time-selective fading multiantenna systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, Dec. 2006.

[80] D. Zhu, Y. Zhang, G. Wang, and M. Lei, "Grassmannian subspace prediction for precoded spatial multiplexing MIMO with delayed feedback," *IEEE Signal Processing Letters*, vol. 18, no. 10, pp. 555 –558, Oct. 2011.

[81] T. Inoue and R. Heath, Jr., "Grassmannian predictive coding for limited feedback multiuser MIMO systems," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, May 2011, pp. 3076 –3079.

[82] Y. Zhang and M. Lei, "Robust Grassmannian prediction for limited feedback multiuser MIMO systems," in *IEEE Wireless Communications and Networking Conference*, April 2012, pp. 863–867.

[83] R. Heath, Jr., T. Wu, and A. Soong, "Progressive refinement of beamforming vectors for high-resolution limited feedback," *EURASIP Journal on Advances in Signal Proc.*, vol. 2009, no. 1, pp. 743–747, 2009.

[84] A. Spanias, "Speech coding: a tutorial review," *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1541–1582, Oct. 1994.

[85] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3834 –3844, Oct. 2008.

[86] T. Inoue and R. Heath, Jr., "Grassmannian predictive frequency domain compression for limited feedback beamforming," in *Information Theory and Applications Workshop*, Feb. 2010, pp. 173–177.

[87] S. Schwarz and M. Rupp, "Adaptive channel direction quantization based on spherical prediction," in *IEEE International Conference on Communications*, Ottawa, Canada, June 2012, pp. 3757 – 3762.

[88] S. Schwarz and M. Rupp, "Adaptive channel direction quantization – enabling multi user MIMO gains in practice," in *IEEE International Conference on Communications*, Ottawa, Canada, June 2012, pp. 6947 – 6952.

[89] S. Schwarz and M. Rupp, "Adaptive channel direction quantization for frequency selective channels," in *20th European Signal Processing Conference*, Bucarest, Romania, Aug. 2012, pp. 2536–2540.

[90] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems," in *38th International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013.

[91] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on a Grassmann-manifold for limited feedback beamforming systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4450–4462, 2013.

[92] S. Schwarz and M. Rupp, "Antenna combiners for block-diagonalization based multi-user MIMO with limited feedback," in *IEEE International Conference on Communications, Workshop: Beyond LTE-A*, Budapest, Hungary, June 2013.

[93] S. Schwarz and M. Rupp, "Subspace quantization based combining for limited feedback block-diagonalization," *IEEE Trans. on Wireless Communications*, vol. PP, no. 99, pp. 1–12, Oct. 2013.

[94] S. Schwarz and M. Rupp, "Distributed downlink multi-user MIMO-OFDM with limited feedback," Oct. 2013, submitted to IEEE Transactions on Wireless Communications.

[95] S. Schwarz, R. Heath, Jr., and M. Rupp, "Multiuser MIMO in distributed antenna systems with limited feedback," in *IEEE 4th Int. Workshop on Heterogeneous and Small Cell Networks, GLOBECOM*, Anaheim, CA, Dec. 2012.

[96] S. Schwarz, R. Heath Jr., and M. Rupp, "Single-user MIMO versus multi-user MIMO in distributed antenna systems with limited feedback," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, p. 54, 2013.

[97] R. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell Systems Technical Journal*, vol. 45, pp. 1775–1796, Dec. 1966.

[98] S. Weinstein and P. Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE Trans. on Communication Technology*, vol. 19, no. 5, pp. 628–634, 1971.

[99] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Transactions on Communications*, vol. 42, no. 10, pp. 2908–2914, 1994.

[100] Y. Zhao and S.-G. Haggman, "Sensitivity to Doppler shift and carrier frequency errors in OFDM systems-the consequences and solutions," in *IEEE 46th Vehicular Technology Conference*, vol. 3, 1996, pp. 1564–1568.

[101] Q. Wang, S. Caban, C. Mehlführer, and M. Rupp, "Measurement based throughput evaluation of residual frequency offset compensation in WiMAX," in *Proc. 51th International Symposium ELMAR-2009*, Zadar, Croatia, Sept. 2009, pp. 233–236.

[102] S. Haykin, *An introduction to analog and digital communications*. Wiley, 1989.

[103] T. Chiueh and P. Tsai, *OFDM Baseband Receiver Design for Wireless Communications*. Wiley, 2008.

[104] Y. Cho, J. Kim, W. Yang, and C. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. Wiley, 2010.

[105] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 1990.

[106] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, ser. Wiley series in telecommunications. Cambridge, MA: Cambridge University Press, 2005.

[107] S. Geng and P. Vainikainen, "Frequency and bandwidth dependency of UWB propagation channels," in *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Helsinki, Finland, 2006, pp. 1–5.

[108] ITU, "Recommendation ITU-R M.1225: Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000," ITU, Tech. Rep., 1997.

[109] 3GPP, "Technical specification group radio access networks; Deployment aspects (Release 8)," http://www.3gpp.org/ftp/Specs/html-info/25943.htm, Dec. 2008.

[110] X. Chen, P.-S. Kildal, C. Orlenius, and J. Carlsson, "Channel sounding of loaded reverberation chamber for over-the-air testing of wireless devices: Coherence bandwidth versus average mode bandwidth and delay spread," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 678–681, 2009.

[111] T. Rappaport, *Wireless Communications: Principles and Practice*, ser. Prentice Hall communications engineering and emerging technologies series. Dorling Kindersley, 2009.

[112] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Systems Technical Journal*, vol. 47, pp. 957–1000, 1968.

[113] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.

[114] T. Zemen and C. Mecklenbrauker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3597–3607, Sept. 2005.

[115] C. Oestges, "Validity of the Kronecker model for MIMO correlated channels," in *IEEE 63rd Vehicular Technology Conference*, vol. 6, Melbourne, Australia, May 2006, pp. 2818–2822.

[116] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception," http://www.3gpp.org/ftp/Specs/html-info/36101.htm, Dec. 2010.

[117] A. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," Sept. 1994, U.S. patent 5,345,599.

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

[119] A. Dabbagh and D. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2190–2202, 2006.

[120] T. Pande, D. Love, and J. Krogmeier, "Reduced feedback MIMO-OFDM precoding and antenna selection," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2284–2293, 2007.

[121] J. Choi and R. Heath, Jr., "Interpolation based transmit beamforming for MIMO-OFDM with limited feedback," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4125–4135, 2005.

[122] G. Caire, G. Taricco, and E. Biglieri, "Capacity of bit-interleaved channels," *Electron. Lett.*, vol. 32, no. 12, June 1996.

[123] A. Guillen i Fabregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundation and Trends in Communications and Information Theory*, vol. 5, pp. 1–153, 2008.

[124] P. Viswanath, "Opportunistic communication: a system view," in *Space-Time Wireless Systems*. Cambridge University Press, 2006, pp. 426–442, Cambridge Books Online.

[125] E. Armanious, D. Falconer, and H. Yanikomeroglu, "Adaptive modulation, adaptive coding, and power control for fixed cellular broadband wireless systems: some new insights," in *IEEE Wireless Communications and Networking*, vol. 1, 2003, pp. 238–242.

[126] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); multiplexing and channel coding (release 10)," http://www.3gpp.org/ftp/Specs/html-info/36212.htm, Dec. 2010.

[127] L.-J. Chen, T. Sun, M. Y. Sanadidi, and M. Gerla, "Improving wireless link throughput via interleaved FEC," in *Proceedings of the Ninth International Symposium on Computers and Communications*, Washington, DC, USA, 2004, pp. 539–544.

[128] A. Chindapol and J. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 944–957, 2001.

[129] M. Wrulich, C. Mehlführer, and M. Rupp, "Managing the interference structure of MIMO HSDPA: A multi-user interference aware MMSE receiver with moderate complexity," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1472–1482, 2010.

[130] T. Yucek and H. Arslan, "MMSE noise power and SNR estimation for OFDM systems," in *IEEE Sarnoff Symposium*, 2006, pp. 1–4.

[131] J. Zheng and B. Rao, "Capacity analysis of MIMO systems using limited feedback transmit precoding schemes," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2886–2901, 2008.

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

[133] J. Roh and B. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2874–2886, 2006.

[134] D. Yang, L.-L. Yang, and L. Hanzo, "DFT-based beamforming weight-vector codebook design for spatially correlated channels in the unitary precoding aided multiuser downlink," in *IEEE International Conference on Communications*, Cape Town, South Africa, 2010, pp. 1–5.

[135] D. Ryan, I. Vaughan, L. Clarkson, I. Collings, D. Guo, and M. Honig, "QAM codebooks for low-complexity limited feedback MIMO beamforming," in *IEEE International Conference on Communications*, Glasgow, Scotland, 2007, pp. 4162–4167.

[136] T. Inoue and R. Heath, Jr., "Kerdock codes for limited feedback MIMO systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, 2008, pp. 3113–3116.

[137] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical layer procedures (release 10)," http://www.3gpp.org/ftp/Specs/html-info/36213.htm, Dec. 2010.

[138] S. Sesia, I. Toufik, and M. Baker, *LTE, the UMTS Long Term Evolution: From Theory to Practice*, ser. Wiley InterScience online books. Wiley, 2009.

[139] H. Zhang, S. Wei, G. Ananthaswamy, and D. Goeckel, "Adaptive signaling based on statistical characterizations of outdated feedback in wireless communications," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2337–2353, 2007.

[140] M. Wu, C. Shen, and Z. Qiu, "Feedback reduction based on clustering in MIMO-OFDM beamforming systems," in *5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, 2009, pp. 1–4.

[141] L. Wan, S. Tsai, and M. Almgren, "A fading-insensitive performance metric for a unified link quality model," in *IEEE Wireless Communications and Networking Conference*, vol. 4, 2006, pp. 2110–2114.

[142] Z. Wang and G. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Transactions on Information Theory*, vol. 50, no. 4, pp. 657–662, 2004.

[143] M. Jordan, A. Dimofte, X. Gong, and G. Ascheid, "Conversion from uplink to downlink spatio-temporal correlation with cubic splines," in *IEEE 69th Vehicular Technology Conference*, April 2009, pp. 1 –5.

[144] D. Lim, K. Choi, and H. Liu, "Optimum power allocation for distributed antenna systems with large-scale fading-only feedback," in *Sixth Int. Conf. on Information Techn.: New Generations*, April 2009.

[145] S. Schwarz, C. Mehlführer, and M. Rupp, "Low complexity approximate maximum throughput scheduling for LTE," in *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, Nov. 2010, pp. 1563–1569.

[146] T. Weber, A. Sklavos, and M. Meurer, "Imperfect channel-state information in MIMO transmission," *IEEE Transactions on Communications*, vol. 54, no. 3, pp. 543–552, 2006.

[147] M. Simko, S. Pendl, S. Schwarz, Q. Wang, J. Ikuno, and M. Rupp, "Optimal pilot symbol power allocation in LTE," in *IEEE 74th Vehicular Technology Conference*, San Francsico, CA, 2011, pp. 1–5.

[148] T. Cui, F. Lu, V. Sethuraman, A. Goteti, S. Rao, and P. Subrahmanya, "Throughput optimization in high speed downlink packet access (HSDPA)," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 474–483, 2011.

[149] S. Haykin, *Adaptive filter theory*, 4th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[150] S. Schwarz, C. Mehlführer, and M. Rupp, "Throughput maximizing multiuser scheduling with adjustable fairness," in *International Conference on Communications*, Kyoto, Japan, June 2011, pp. 1–5.

[151] D. Gesbert and M.-S. Alouini, "Selective multi-user diversity," in *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2003, pp. 162–165.

[152] S. Sanayei and A. Nosratinia, "Exploiting multiuser diversity with only 1-bit feedback," in *IEEE Wireless Communications and Networking Conference*, vol. 2, 2005, pp. 978–983.

[153] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, 2000.

[154] A. Molisch, "A generic model for MIMO wireless propagation channels in macro- and microcells," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 61–71, 2004.

[155] H. Bizaki and A. Falahati, "Tomlinson-harashima precoding with imperfect channel state information," *IET Communications*, vol. 2, no. 1, pp. 151–158, 2008.

[156] D. Ryan, I. Collings, I. V. L. Clarkson, and R. Heath, Jr., "Performance of vector perturbation multiuser MIMO systems with limited feedback," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2633–2644, 2009.

[157] A. Razi, D. Ryan, I. Collings, and J. Yuan, "Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 356–365, 2010.

[158] B. Mondal, S. Dutta, and R. Heath, Jr., "Quantization on the Grassmann manifold," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4208–4216, Aug. 2007.

[159] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on Grassmann manifolds and applications to MIMO communications," *IEEE Trans. on Information Theory*, vol. 54, no. 3, pp. 1108 –1123, March 2008.

[160] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. on Information Theory*, vol. 54, no. 8, pp. 3425 –3441, Aug. 2008.

[161] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. on Inform. Theory*, vol. 54, no. 8, pp. 3457 –3470, Aug. 2008.

[162] J. Park, B. Lee, and B. Shim, "A MMSE vector precoding with block diagonalization for multiuser MIMO downlink," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 569–577, 2012.

[163] I. S. Dhillon, R. Heath, Jr., T. Strohmer, and J. A. Tropp, "Constructing packings in Grassmannian manifolds via alternating projection," *ArXiv e-prints*, Sept. 2007.

[164] W. Santipach and M. Honig, "Capacity of a multiple-antenna fading channel with a quantized precoding matrix," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1218–1234, 2009.

[165] D. Love and R. Heath, Jr., "Limited feedback diversity techniques for correlated channels," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 718–722, March 2006.

[166] B. Clerckx, G. Kim, and S. Kim, "MU-MIMO with channel statistics-based codebooks in spatially correlated channels," in *IEEE Global Telecommunications Conference*, Nov. 2008, pp. 1 –5.

[167] E. Park, H. Kim, H. Park, and I. Lee, "Feedback bit allocation schemes for multi-user distributed antenna systems," *IEEE Communications Letters*, vol. 17, no. 1, pp. 99–102, 2013.

[168] D. McNamara, M. Beach, and P. Fletcher, "Experimental investigation of the temporal variation of MIMO channels," in *IEEE 54th Vehicular Technology Conference*, vol. 2, 2001, pp. 1063–1067.

[169] K. Baddour and N. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1650–1662, July 2005.

[170] B. Hassibi, "Random matrices, integrals and space-time systems," in *DIMACS Workshop on Algebraic Coding Theory and Information Theory*, Piscataway, NJ, Dec. 2003.

[171] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proc. - Radar, Sonar and Navigation*, vol. 147, no. 4, pp. 162–168, 2000.

[172] F. Boccardi and H. Huang, "A near-optimum technique using linear precoding for the MIMO broadcast channel," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 3, April 2007, pp. III–17–III–20.

[173] A. Zanella and M. Chiani, "The PDF of the lth largest eigenvalue of central Wishart matrices and its application to the performance analysis of MIMO systems," in *IEEE Global Telecommunications Conference*, New Orleans, LA, Dec. 2008, pp. 1 –6.

[174] T. Li, L. Yang, and Z. He, "Interpolation-based multiuser precoding for MIMO-OFDM system with limited feedback," in *International Conf. on Neural Networks and Signal Processing*, 2008, pp. 205–209.

[175] J. Chang, I. Lu, and Y. Li, "Adaptive codebook-based channel prediction and interpolation for multiuser multiple-input multiple-output orthogonal frequency division multiplexing systems," *IET Communications*, vol. 6, no. 3, pp. 281–288, 2012.

[176] S. R. Buss and J. P. Fillmore, "Spherical averages and applications to spherical splines and interpolation," *ACM Transactions on Graphics*, vol. 20, pp. 95–126, 2001.

[177] D. Amsallem, "Interpolation on manifolds of cfd-based fluid and finite element-based structural reduced-order models for on-line aeroelastic predictions," Dissertation, Standford University, 2010.

[178] R. Blum, "MIMO capacity with interference," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 793–801, 2003.

[179] M. Trivellato, F. Boccardi, and F. Tosate, "User selection schemes for MIMO broadcast channels with limited feedback," in *IEEE 65th Vehicular Technology Conference*, Dublin, Ireland, April 2007, pp. 2089–2093.

[180] D. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)," in *IEEE 61st Vehicular Technology Conference*, vol. 5, Stockholm, Sweden, 2005, pp. 3132–3136.

[181] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, ser. Springer Series in Statistics. Springer, 2009.

[182] M.-S. Alouini and A. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1047–1066, 1999.

[183] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); radio frequency (RF) system scenarios (release 10)," http://www.3gpp.org/ftp/Specs/html-info/36942.htm, Oct. 2010.

[184] A. Sharma, A. Roy, S. Ghosal, R. Chaki, and U. Bhattacharya, "Load balancing in cellular network: A review," in *Third International Conference on Computing Communication Networking Technologies*, 2012, pp. 1–5.

[185] E. Bjornson, M. Bengtsson, and B. Ottersten, "Receive combining vs. multistream multiplexing in multiuser MIMO systems," in *IEEE Swedish Comm. Technologies Workshop*, 2011, pp. 103–108.

[186] 3GPP, "Technical specification group radio access network; FDD base station (BS) classification (release 11)," http://www.3gpp.org/ftp/Specs/html-info/25951.htm, Sept. 2012.

[187] M. Haenggi, J. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1029–1046, 2009.

[188] L. Hentila, P. Kyosti, M. Kaske, M. Narandzic, and M. Alatossava, "MATLAB implementation of the WINNER phase II channel model ver1.1," http://www.ist-winner.org/phase_2_model.html, Dec. 2007.

[189] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[190] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 953–961, 2008.

[191] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[192] Y. Polyanskiy, H. Poor, and S. Verdu, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, 2011.

[193] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 100–102, 1975.

[194] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the gaussian broadcast channel with feedback (corresp.)," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 667–671, 1984.

[195] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for mimo interference channels," in *IEEE International Symposium on Information Theory Proceedings*, 2010, pp. 619–623.

[196] I. Kammoun and J. C. Belfiore, "A new family of grassmann space-time codes for non-coherent MIMO systems," *IEEE Communications Letters*, vol. 7, no. 11, pp. 528–530, 2003.

[197] Z. Utkovski, P.-C. Chen, and J. Lindner, "Some geometric methods for construction of space-time codes in grassmann manifolds," in *46th Annual Allerton Conference on Communication, Control, and Computing*, 2008, pp. 111–118.

[198] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[199] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008.

[200] J. Lee, *Introduction to Smooth Manifolds*, ser. Graduate Texts in Mathematics. Springer, 2003.

[201] K. Schober, P. Janis, and R. Wichman, "Geodesical codebook design for precoded MIMO systems," *IEEE Communications Letters*, vol. 13, no. 10, pp. 773–775, 2009.

[202] C. Schlegel and L. Perez, "On error bounds on turbo-codes," *IEEE Communication letters*, vol. 3, no. 7, pp. 205–207, July 1999.

[203] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, pp. 379–423, July 1948.

[204] Q. Wang, C. Mehlführer, and M. Rupp, "Carrier frequency synchronization in the downlink of 3GPP LTE," in *Proceeding of the 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Istanbul, Turkey, Sept. 2010.

[205] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, 1982.

[206] M. Wrulich and M. Rupp, "Computationally efficient MIMO HSDPA system-level modeling," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 1, p. 382501, 2009.

[207] J. Colom-Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in *IEEE 71st Vehicular Technology Conference*, Taipei, Taiwan, 2010, pp. 1–5.

[208] S. Tsai and A. Soong, "Effective-SNR mapping for modeling frame error rates in multiple-state channels," 3GPP2, Tech. Rep. 3GPP2-C30-20030429-010, April 2003.

[209] R. Sandanalakshmi, T. G. Palanivelu, and K. Manivannan, "Effective SNR mapping for link error prediction in OFDM based systems," in *International Conference on Information and Communication Technology in Electrical Sciences*, 2007, pp. 684–687.

[210] A. Cipriano, R. Visoz, and T. Salzer, "Calibration issues of PHY layer abstractions for wireless broadband systems," in *IEEE 68th Vehicular Technology Conference*, Calgary, Alberta, 2008, pp. 1–5.

[211] X. He, K. Niu, Z. He, and J. Lin, "Link layer abstraction in MIMO-OFDM system," in *International Workshop on Cross Layer Design*, 2007, pp. 41–44.

[212] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Communications and Information Theory*, vol. 1, no. 1, pp. 1 –182, June 2004.

[213] A. K. Gupta, D. K. Nagar, and A. M. Velez-Carvajal, "Unitary invariant and residual independent matrix distributions," *Computational & Applied Mathematics*, vol. 28, pp. 63 – 86, 2009.

[214] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)*, ser. Princeton reference. Princeton University Press, 2009.

[215] J. Andrews, W. Choi, and R. Heath, Jr., "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Communications*, vol. 14, no. 6, pp. 95 –104, Dec. 2007.

[216] J. Nielsen, "The distribution of volume reductions induced by isotropic random projections," *Advances in Applied Probability*, vol. 31, no. 4, pp. 985–994, Dec. 1999.