HABILITATIONSSCHRIFT

Mobile Wireless Communications for a Society in Motion

ausgeführt zum Zwecke der Erlangung der Lehrbefugnis für das Fach

Mobilkommunikation
(Mobile Communications)

eingereicht an der

Technischen Universität Wien

Fakultät für Elektrotechnik und Informationstechnik

von

Dipl. Ing. Dr. techn. Stefan Schwarz

September, 2018

Acknowledgements

This thesis collects a part of my research results obtained throughout the last eight years at the Institute of Telecommunications of TU Wien, in particular within my Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion. During this time I had the opportunity to collaborate with and provide advice to many talented researchers.

First of all, I would like to express my gratitude to Markus Rupp for his mentorship.

I would also like to thank my former and current fellow researchers: Martin Wrulich, Christian Mehlführer, Sebastian Caban, Qi Wang, Michal Simko, Josep Colom Ikuno, Philipp Svoboda, Robert Langwieser, Christoph Mecklenbräuker, Robert Heath, Martin Taranetz, Ronald Nissel, Martin Müller, Martin Lerch, Fjolla Ademaj, Ljiljana Marijanovic, Erich Zöchmann, Stefan Pratschner, Blanca Ramos Elbal, Bashar Tahir, Samira Homayouni, Lucas Nogueira Ribeiro.

I am grateful to my colleagues from our corporate partners: Birger Hätty, Gottfried Schnabl, Roland Gabriel, Gerd Saala, Waltraud Müllner, Martin Kogelbauer, Werner Weiler, Thomas Ergoth and Tal Philosof.

Finally, thanks to my family!

Abstract

This habilitation thesis is based on a collection of journal articles and conference publications dealing with single- and multi-point transmission optimization and analysis of mobile wireless communications, with main focus on time-variant scenarios. The motivating communication scenario underlying this thesis is determined by tomorrow's *Society in Motion*, in which large numbers of mobile users (humans and autonomous machines) with varying quality of service requirements (throughput, latency, reliability) will simultaneously demand high performance dependable wireless connectivity. Satisfying such requirements poses significant challenges on mobile wireless communications, in particular when users are moving with relatively high velocity through the network coverage area. In this thesis, I present my research within four closely connected research fields addressing these challenges to enable efficient and dependable wireless communications for the *Society in Motion*:

Link and System Level Modeling and Simulation: Link and system level simulations build the basis for accurate and realistic performance investigation of mobile communication systems. They are indispensable for rapid prototyping of novel signal processing methods and enable benchmarking against state-of-the-art standardized approaches. I present our link and system level simulation methodologies and abstraction techniques, which provide the means to analyze the transceiver designs developed in this thesis within realistic environments, in order to gauge their practical value.

Acquisition of Channel State Information at the Transmitter: Channel state information at the transmitter (CSIT) builds the fundamental basis for advanced multi-antenna and multi-point transceiver optimization. However, acquiring accurate CSIT with acceptable overhead is challenging, especially in high mobility scenarios and frequency division duplex (FDD) systems. I provide a general classification scheme of CSIT acquisition methods and present my contributions within the field of so-called limited feedback techniques. Specifically, I discuss my implicit and explicit channel state information (CSI) feedback algorithms, which exploit the algebraic topological structure imposed by specific classes of transmission schemes, as well as, the characteristics of the wireless propagation environment to achieve an efficient CSI quantization.

Multi-User Resource Allocation and Robust Transmission Optimization: Having CSIT available allows to optimize the multi-user resource allocation and the multi-antenna transmit processing to achieve satisfactory, efficient and reliable transmissions. However, CSIT imperfections due to estimation and quantization errors, as well as, channel aging, fundamentally impair the achievable performance. I provide a discussion on theoretical performance limitations imposed by imperfect CSIT concerning the outage probability and the achievable degrees of freedom of the wireless transmissions. I moreover present my transceiver

designs that account for CSIT imperfections to enhance the robustness and resilience of the communication system, in order to provide efficient and dependable wireless connectivity even with imperfect CSIT.

Coordinated Transmission and Interference Management: Spatial densification of wireless networks is the major approach to sustain the ever-increasing capacity demands of mobile networks. However, increasing mutual interference amongst transmission points restricts the achievable spectral efficiency and impairs the reliability of the wireless connectivity. These interference limitations can be mitigated through multi-point coordination and joint transmission. I provide an overview of multi-point interference management techniques for unicast and multicast transmissions. I discuss the importance of multicast transmissions for cellular-assisted vehicular communications, which is one of the central use-cases underlying the *Society in Motion*. I furthermore present my multi-point coordination techniques for the multi-antenna multicast interference channel, which support weighted sum-rate maximization through distributed transmission optimization.

The scientific results discussed within this thesis support reliable and efficient multi-antenna and multi-point wireless communications, while providing robustness with respect to CSI imperfections, thus building a corner stone for mobile communication systems that are capable of supporting the requirements imposed by the envisioned *Society in Motion*.

Contents

1. Overview of Contributions

1.1. Introduction and Motivation

Mobile wireless broadband connectivity has become a commodity. The demand for increased mobile network capacity is continuously growing, especially in cities and urban agglomerations [1]. This trend will continue for several decades to come, as tomorrow's society is shaped by increased urbanization. By 2050 the percentage of the world's population living in urban areas is predicted to increase from today's 55% to 68% [2]. Within such cities and their commuter belts high-rate wireless connectivity will not only be required by masses of (quasi-)static indoor users, but also by large numbers of people moving in public and private transportation networks [3]. Moreover, wireless connectivity is not restricted to people: Frictionless functioning of such a *Society in Motion* is supported by connected transportation vehicles (cars, trains, buses) and environmental sensors, exchanging monitoring and control information amongst each other and with the cloud, and relaying internet traffic of their users. Road-safety applications, for example, require reliable and low-latency exchange of safety-relevant information with vehicles to reduce the number of road accidents and fatalities [4]. As soon as wireless communications is employed to support safety critical applications, such as, driver assistance systems or even automated driving, dependability of the transmission link becomes a must, guaranteeing reliable communication under strict packet delay deadlines (timeliness) [5]. This leads to a new situation in future where a significant portion of the wireless users (people and machines) is permanently moving through the mobile network, and where best-effort entertainment services and applications with strict dependability requirements compete for the same resources.

Key challenges: Providing satisfactory and sustainable wireless connectivity for the envisioned *Society in Motion* gives rise to the following key challenges:

- Improving the data throughput of individual users, even at high mobility.
- Increasing the mobile network capacity and the supported connection density, not only for static scenarios but also for mobile situations.

- Enhancing the dependability of wireless connectivity to support applications that require beyond best-effort quality of service (QoS).

I provide an in-depth treatment of the challenges imposed by the outlined *Society in Motion* on mobile communications in general and in the specific context of cellular-assisted vehicular communications, as well as, a discussion of potential solution approaches in my accompanying magazine publications:

[Article 1] S. Schwarz and M. Rupp, "Society in motion: Challenges for LTE and beyond mobile communications," *IEEE Communications Magazine, Feature Topic: LTE Evolution*, vol. 54, no. 5, pp. 76–83, 2016

[Article 2] S. Schwarz, T. Philosof, and M. Rupp, "Signal processing challenges in cellular assisted vehicular communications," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 47–59, March 2017

1.2. Outline and Overview

The focus of my research work presented in thesis is on signal processing methods and transceiver optimization techniques to enable a sustainable support of the requirements imposed on mobile communication systems by the envisioned *Society in Motion*.

1.2.1. Main Scientific Approaches

I pursue the following three main approaches in this thesis:

- Accurate and efficient acquisition of channel state information (CSI) at the transmitter (CSIT) to enable the application of advanced multi-user transmission concepts; see Chapter 3.
- Robust multi-user transceiver design to support reliable and spectrally efficient wireless transmissions, even with imperfect CSIT; see Chapter 4.
- Multi-point coordination and transceiver optimization to exploit macroscopic diversity and to mitigate interference, thereby enhancing the reliability, capacity and area spectral efficiency (ASE) of mobile networks; see Chapter 5.

CSIT acquisition and robust transceiver design: Spatial multi-user multiplexing in multi-antenna systems is a central technique for enhancing the spectral efficiency of wireless communications. The achievable performance of advanced multi-user transmission concepts is determined by the quality of the CSIT. In wireless communications with moving users, obtaining accurate CSIT is difficult, as only a short time interval (the channel coherence time) is available for the CSI acquisition. In such situations, it is therefore imperative to enhance the resilience of the CSI acquisition with respect to channel variations, as well as, the robustness of transceiver designs with respect to CSI uncertainty, to enable reliable and spectrally efficient transmissions. These issues are the core of Chapters 3 and 4.

Multi-point coordination and interference mitigation: Over the last decades, densifying network structures was the main and most successful approach for boosting mobile network capacities by spatially reusing the available spectral resources. This approach, however, leads to increasing inter-cell interference in mobile networks, limiting the achievable spectral efficiency of users and impairing the reliability of the wireless connectivity. Multi-point transceiver coordination and optimization provides a way to mitigate interference or to even utilize multiple transmission/reception points to enhance the macroscopic diversity and thus the reliability and the efficiency of the system. Such techniques are particularly important in densely populated interference-limited scenarios, as is the case in most urban environments, and they are the focus of Chapter 5.

Link and system level simulation: Prior to presenting these multi-user and multi-point transceiver optimization methods in Chapters 3 to 5, I discuss in Chapter 2 accurate link and system level modeling and simulation approaches for mobile wireless communication systems. In wireless communications research, computer simulations are an important tool for validation and substantiation of analytic results. Link and system level simulations are the industry standard for rapid prototyping of novel technologies. Such simulations allow for the implementation of standard-compliant environments, which represent the baseline performance of state-of-the-art technologies and thereby provide a reference for benchmarking of novel techniques.

1.2.2. Summary of Contributions

Contributions of Chapter 2: Link and System Level Modeling and Simulation

In Chapter 2, I provide an overview of link and system level simulation approaches for mobile communications. The scope is thereby on an accurate and efficient modeling and abstraction of cellular communication systems, with special emphasis on high-mobility

scenarios. A critical issue of realizing complex simulation environments in general is a proper verification of the functionality and a rigorous validation of the correctness of simulation outcomes. I discuss corresponding means for verification and validation in Chapter 2. In my research work, the proposed modeling and abstraction approaches are specifically employed to simulate Third Generation Partnership Project (3GPP) standard-compliant fourth generation (4G) and fifth generation (5G) mobile communication systems. My corresponding scientific publications, providing further details on efficient and accurate link and system level modeling of mobile communication systems, are included at the end of Chapter 2 of this thesis.

Contributions of Chapter 3: Acquisition of Channel State Information at the Transmitter

In Chapter 3, I discuss techniques for efficient CSI acquisition at the transmitter-side, putting the scope chiefly on so-called limited feedback methods. Such methods utilize a dedicated limited capacity feedback link between the receiver and the transmitter to convey the quantized CSI. A key challenge in limited feedback systems is to achieve an adequate CSIT quality with a minimal feedback overhead, requiring an efficient CSI quantization. I present in this chapter a general classification scheme of existing limited feedback approaches, outline their applicability for certain classes of transmission strategies, discuss their advantages and drawbacks, and summarize my own contributions specifically in the field of predictive quantization on manifolds. Furthermore, I highlight the relevance of finite-scatterer channel models to accurately represent propagation conditions in full-dimension MIMO (FD-MIMO) systems. A major contribution of this chapter is to show that finite-scatterer channels permit an efficient CSI quantization by applying a sparse channel decomposition in the angular domain. My corresponding scientific publications, which provide further details on the described methods, are included at the end of Chapter 3 of this thesis.

Contributions of Chapter 4: Multi-User Resource Allocation and Robust Transmission Optimization

In Chapter 4, I proceed to discuss transceiver optimization and resource allocation in multi-user scenarios under the assumption of imperfect (quantized) CSIT. The focus here is on robust multi-antenna beamforming and precoding techniques that can achieve substantial spatial multi-user multiple-input multiple-output (MIMO) gains even with imperfect CSIT, thereby enhancing the spectral efficiency of the wireless communication system. In multi-user scenarios, a smart allocation of the time, frequency and spatial resources is a critical issue, which can decide on a successful accomplishment of QoS requirements.

In Chapter 4, I discuss an efficient framework for scheduling and resource allocation in MIMO bit-interleaved coded modulation (BICM) orthogonal frequency division multiple access (OFDMA) systems, such as, 4G/5G mobile communications, and I address outage-based scheduling and admission control with imperfect CSIT. Details on the presented contributions are accessible in my corresponding scientific publications included at the end of Chapter 4 of this thesis.

Contributions of Chapter 5: Coordinated Transmission and Interference Management

In Chapter 5, I put the scope on multi-point transceiver optimization. I discuss MIMO multi-cell coordination techniques for unicast and multicast transmission scenarios. My main focus thereby is on multicasting, which is considered as an enabling technology for efficient support of cellular-assisted vehicular communications. This is because in vehicular communication scenarios common information, such as, road traffic status information, frequently needs to be shared with many vehicles in a certain geographic area, which is most efficiently realized through multicasting. In Chapter 5, I discuss the underlying information-theoretic model of multi-cell multicasting, namely the multi-antenna multicast interference channel (MCIC), and I present a distributed optimization method for coordinated multicast transmission of multiple base stations. This method enables significant latency reductions and efficiency improvements for data packet distribution in cellular-assisted vehicular communications. My corresponding scientific publications are included at the end of Chapter 5 of this thesis.

Overview of Major Contributions

To summarize, the following major contributions constitute this thesis:

- Accurate link and system level modeling and abstraction techniques for mobile wireless communications, to enable a realistic performance investigation and benchmarking of novel technologies, with focus on dynamic time-variant scenarios.
- Limited-feedback based CSIT acquisition methods that provide resilience against temporal channel variations.
- Advanced multi-user MIMO transceivers that exhibit robustness with respect to uncertainty of the CSI available at the transmitter.
- Distributed optimization techniques for unicast and multicast transmission in multi-cell wireless communication systems, with application in cellular-assisted vehicular communications.

Society in Motion: Challenges for LTE and Beyond Mobile Communications

Stefan Schwarz and Markus Rupp

The authors highlight challenges that must be addressed by future mobile communications to enable efficient support of large numbers of highly mobile users in networks that are crowded with quasi-static (nomadic) users. They survey existing solutions, and put special emphasis on open issues and competing priorities.

Abstract

Tomorrow's society will see hosts of people constantly on the move, commuting between home and work, meeting up with family and friends or visiting shopping centers and leisure facilities. While on their way, these people use mobile devices to connect to the Internet, utilizing journey time for work-related tasks, entertainment, or socializing in online communities. Current fourth generation cellular networks are, however, not designed to efficiently serve large numbers of comparatively high-mobility users, often causing insufficient service quality while on the move. Machine-type communication will cause further aggravation, with wirelessly connected sensors that constantly monitor/record our environment, and vehicles that autonomously exchange traffic- and safety-relevant information. In this article, we highlight challenges that must be addressed by future mobile communications to enable efficient support of large numbers of highly mobile users in networks that are crowded with quasi-static (nomadic) users. We survey existing solutions and put special emphasis on open issues and competing priorities.

Introduction

We envision tomorrow's society as a Society in Motion. Hosts of people are constantly on the move, commuting between home and work, attending leisure events, meeting friends and family, and visiting shopping promenades and malls. Most such activities are concentrated in urban agglomerations, since urbanization is predicted to incorporate more than 80 percent of the developed world's population by 2025 [1]. However, at the same time, commuter traffic from rural areas is rising, since most workplaces are in and around cities. The European Environment Agency reported in 2013 that typical commuting times within larger cities are not unlikely to exceed one hour per trip; commuters from rural areas have to endure even longer trip durations. To make the most of this time spent in public and private transportation, people utilize mobile devices, such as smartphones and tablets, for entertainment (watching video clips, reading news and e-books, listening to music and audio books), shopping in online stores, preparing work, scheduling appointments, socializing on web platforms, and so on. All such services require Internet access, in some cases just for a few bytes while in other cases for entire data streaming, and thus depend on wireless connectivity.

Wireless connectivity while on the move is not restricted to people though; machine-type communication will add a significant portion of mobile data traffic if not even the majority . Frictionless functioning of the *Society in Motion* is supported by connected sensors and transportation vehicles (cars, trains, buses), autonomously exchanging monitoring and control information among themselves and with the cloud, and relaying Internet traffic of their users. Fleets of commuter trains, buses, and individual cars, for example, are nowadays equipped with wireless communication devices to enable monitoring of the vehicle's internal state, facilitating finding weak points early on before causing severe damage. Connected in-vehicle entertainment systems that support online video streaming and Internet access are increasingly recognized as important revenue drivers by car manufacturers. Additionally, road safety applications require reliable and low-latency exchange of safety-relevant information to enable realization of global road fatality reduction goals; for example, the European Union targets 50 percent reduction of road fatalities by 2020 compared to 2010 (more than 31,000 deaths on Europe's roads in 2010). As soon as wireless communications is employed to support safety-critical applications, such as driver assistance systems or even automated driving, dependability of the transmission link becomes a must, guaranteeing reliable low-latency communication under strict packet delay deadlines (timeliness) [2]. Even though dependability cannot be guaranteed by current mobile communications, vehicular communications strategies of companies and manufacturers still promote such technology, since it is cheaply available off the shelf and because mobile networks are almost ubiquitously accessible. In response, the Third Generation Partnership Project (3GPP) recently initiated a new study item within Release 14 of Universal Mobile Telecommunications System (UMTS) Long Term Evolution (LTE) on vehicular communications: *Study item on LTE support for V2X services* (*V2XLTE*); the goal is to

Stefan Schwarz is with the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion; Markus Rupp is with Technische Universität Wien.

0163-6804/16/$25.00 © 2016 IEEE

develop a set of LTE specifications for vehicular environments (LTE-V). Notice that vehicle to X (V2X) communication encompasses connectivity between vehicles — vehicle to vehicle (V2V) — to roadside infrastructure — vehicle to infrastructure (V2I) — and to people in proximity of connected cars — vehicle to pedestrian (V2P). Such connectivity can be achieved either directly, using device-to-device (D2D) transmission, or indirectly by employing base stations as transmission hubs. These developments will lead to a new situation in the future, where a significant portion of wireless users (people and machines) are permanently moving through the network, and where best effort entertainment services as well as applications with strict dependability requirements compete for the same resources. Future mobile communications will thus basically face two distinct groups of users:

- Masses of quasi-static/nomadic best effort users that (mostly indoors) require virtually bandwidth-unlimited wireless connectivity
- Large numbers of mobile (moving) users with varying quality of service (QoS) requirements demanding seamless mobility

3GPP LTE is designed to provide basic point-to-point connectivity even at very high mobility of up to 500 km/h. However, the promised gigabit-per-second data throughput with latency not exceeding 10 ms can only be sustained for very few static users; large numbers of highly mobile users lead to inefficiencies on the physical layer (PHY) as well as the signaling plane of LTE [3]. Even more, many current fifth generation (5G) proposals focus purely on the first group of quasi-static/nomadic users, hardly accounting for the impact of novel technological concepts, such as massive multiple-input multiple-output (MIMO), millimeter-wave (mmWave) transmission, and ultra high network densification, on high-mobility users. This is, on one hand, justified by the ever increasing demand for higher "static" network capacity; on the other hand, however, when carelessly deployed, such technologies can even worsen the performance of this second group of mobile users. The purpose of this article, hence, is to reveal shortcomings of the current LTE standard in the context of a *Society in Motion* and to provide potential solutions for the evolution of LTE as well as for 5G mobile communications. We address techniques that are readily applicable within LTE by enhancing network nodes and/or user equipment, as well as methods that require introduction of a new standard. Future mobile networks will likely comprise a variety of possibilities for access to the network as well as for direct communication between subscribers, as sketched in Fig. 1; this article attempts to shed light on the impact of such technologies on mobile users.

Physical Layer Challenges

Acquisition of Channel State Information

Accurate channel state information (CSI) at the receiver (CSIR) and the transmitter (CSIT) is a critical prerequisite for the exploitation of the spatial degrees of freedom provided by MIMO systems. Especially at high mobility, obtaining the necessary CSI accuracy can be challenging, as detailed below.

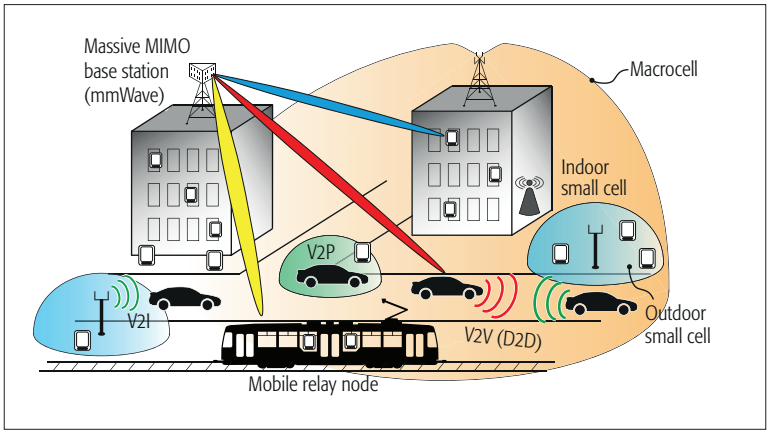


Figure 1. Illustration of a future mobile network in an urban area comprising various access technologies to efficiently support static and mobile users.

Channel Estimation and Pilot Designs: LTE employs pilot-symbol-based channel estimation with different pilot patterns in up- and downlink directions as illustrated in Fig. 2a. The pilot pattern for downlink homogeneously fills a portion of the time and frequency grid, thus equally supporting channels with varying delay and Doppler spread. In the uplink direction pilots fill the full bandwidth at fixed times, to conserve the favorable peak-to-average-power ratio (PAPR) of single carrier frequency-division multiplexing (SC-FDM) transmission; the density of pilots along the time axis, however, is reduced, making them less suitable for high-velocity scenarios.

As shown in Fig. 2b, LTE's downlink pilot pattern with $N_t \in \{1, 2\}$ transmit antennas and applying least squares channel estimation supports velocities up to 325 km/h at 2 GHz center frequency, corresponding to a Doppler frequency of $f_d \approx 600$ Hz, with minimal performance degradation. With $N_t = 4$ transmit antennas the pilot density along the time axis is reduced, explaining the performance drop already at $f_d = 300$ Hz. Above these Doppler frequencies throughput deteriorates due to increasing channel estimation errors and inter-carrier interference (ICI).

Several 5G proposals promote highly flexible multi-carrier waveform designs that adapt subcarrier spacing, cyclic prefix, prototype pulse, and so on, to support varying channel characteristics and traffic demands. A first step toward such flexibility can already be taken without requiring a novel PHY, by adapting pilot symbols according to channel characteristics. In [4], the authors propose to adapt the pilot distance in time and frequency (D_t and D_f in Fig. 2a) as well as the pilot power according to channel delay/Doppler spread and signal-to-noise ratio (SNR), to maximize the achievable throughput of the system. The performance of this scheme is shown in Fig. 2b, demonstrating substantial performance gains, especially for larger antenna configurations. Notice that performance can be improved even at low Doppler frequency, since the pilot density can be reduced compared to the LTE pattern, and thus, the reference symbol overhead can be decreased. Large parts of these gains can be conserved by employing a small number of different pilot patterns optimized for different regimes of

channel characteristics, enabling pilot adaptation with minimal feedback overhead from the receivers. Since channel delay and Doppler spread do not vary quickly with the carrier frequency, feedback information can even be avoided by adapting downlink pilot patterns according to uplink measurements, provided the frequency-division duplex (FDD) distance is not too large. Such a scheme can in principle readily be incorporated into a future release of LTE by enhancing user-specific reference symbols, without impairing backward compatibility.

Inter-Carrier Interference: At high velocity, multi-carrier transmissions suffer from increasing ICI due to the Doppler spread introduced by the channel, deteriorating the quality of channel estimation and symbol detection. Such effects are especially pronounced in orthogonal frequency-division multiplexing (OFDM) due to poor spectral confinement of the applied rectangular transmit pulse. One approach to alleviate performance degradation is ICI mitigation, for example, employing iterative approaches [5] where, in each decoding round, the ICI contributions are estimated in parallel to channel estimation and successively cancelled. Recent developments in [6] achieve almost perfect ICI-free OFDM transmission by iteratively estimating the doubly dispersive wireless channel and equalizing its distortions. The authors demonstrate that already three iterations are sufficient for velocities up to 400 km/h at 2 GHz center frequency ($f_d \approx 740$ Hz), even without incorporating forward-error-correction coding into the iterations; hence, the incurred processing delay is practically negligible. Since such methods only require enhancement of the user equipment, they can be implemented without any modification of the LTE standard. Alternatively, novel multi-carrier waveforms can be incorporated into a new standard that inherently provide robustness with respect to channel Doppler spread, as discussed further below.

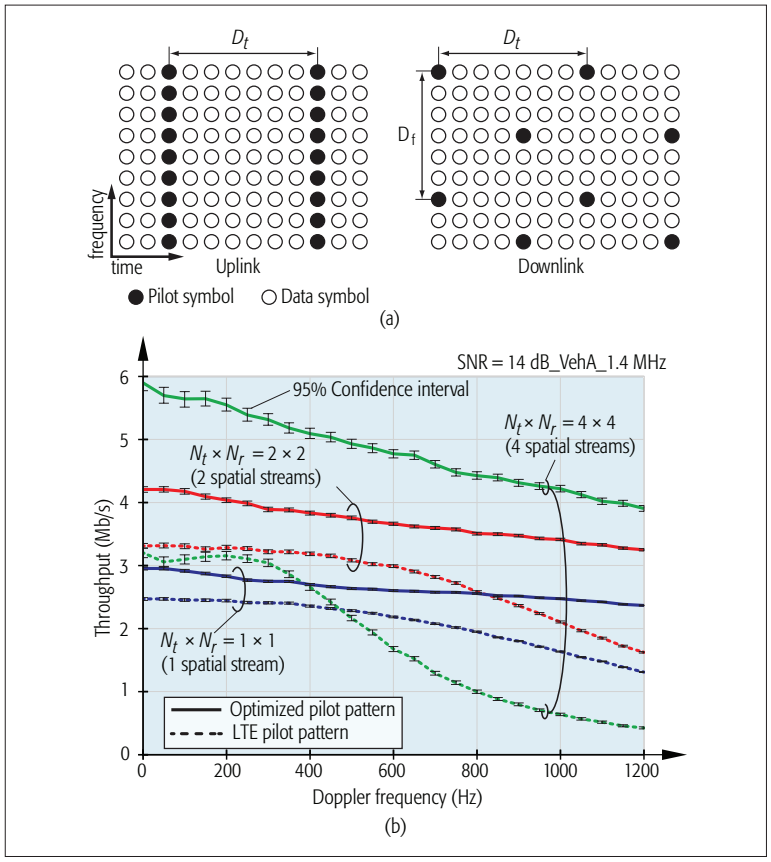


Figure 2. Optimization of pilot power and distance in dependence of the delay and Doppler spread introduced by the channel: a) pilot patterns as employed in LTE up- and downlink; b) comparison of the downlink throughput at 2 GHz center frequency dependent on the maximum channel Doppler frequency achieved with the LTE pilot pattern and with optimized pilot designs according to [4].

CSI Feedback in FDD Systems: Over the last years, significant research effort was dedicated to the exploitation of the spatial degrees of freedom provided by multiple antenna systems, whether it be in the context of antenna arrays controlled by single transmitters (single- and multi-user MIMO) or for coordination of multiple spatially distributed transmitters within coordinated multipoint (CoMP) transmission concepts.

Simple codebook-based precoding schemes, restricting spatial pre-processing of the transmit signal to limited sets of precoders, are successfully established for single-user MIMO transmission in LTE and provide valuable performance gains with reasonable complexity and signaling overhead. The required CSIT in FDD systems can be obtained with minimal feedback from the receivers, whereas time-division duplex (TDD) systems might even rely on channel reciprocity. However, such schemes fall short of delivering noticeable capacity improvements in multi-user MIMO and CoMP, due to low spatial resolution provided by the limited codebook size. Furthermore, already such a seemingly simple task as transmission rate adaptation through limited feedback can utterly fail at high velocity, in case the feedback information is not delivered in time, as illustrated in Fig. 3. This figure demonstrates the impact of feedback delay on LTE's channel quality indicator (CQI) dependent on the user mobility given in terms of the maximum Doppler shift of the signal. Without channel prediction and at 1 ms feedback delay, the throughput degrades already at 50 km/h ($f_d \approx 93$ Hz) by more than 20 percent due to deteriorating block error ratio (BLER). If the receiver is able to perform complex channel prediction, employing the recursive least squares (RLS) algorithm, feedback delay can be partly compensated. If this is not feasible, due to complexity restrictions, rate adaptation should be based on channel statistics instead of instantaneous CSI to keep the BLER acceptably low; but such an approach prohibits exploitation of channel and multi-user diversity. The minimal feedback delay achievable in FDD communications is practically limited by the length of the transmission time interval (TTI) (1 ms in LTE). Hence, future mobile communication systems may have to provide the option for reduced frame length to improve the efficiency of high-mobility users.

To accommodate more complex multi-user MIMO and CoMP schemes, LTE provides basic support of non-codebook-based precoding relying on explicit CSIT to optimally design pre-

coders. Such methods, however, only perform well with accurate CSIT, otherwise suffering from residual interference. In FDD, this implies enhancing the feedback algorithms supported by LTE. Corresponding differential and predictive feedback schemes that operate on manifolds to minimize the required overhead are available; see, for example, [7, references therein], where the performance of multi-user MIMO with predictive limited feedback is analyzed. These existing methods, however, achieve gains over memoryless CSI quantization only at low to moderate user mobility, efficiently supporting static users and pedestrians. Nevertheless, we do see potential for differential/predictive methods even at high velocity: imagine a highway situation with many cars traversing the same stretch of road. At a given position on the road, the preferred transmission rate and beamforming direction of consecutive vehicles will not vary much. Thus, if each vehicle provides only little differential update information on the preferred rate and beamformer, gradually the transmitter will obtain an accurate picture about optimal transmit processing along the highway. Differential/predictive CSI feedback enhancements in principle can be incorporated within a future release of LTE; but it does require standardization efforts to implement a common predictor structure, similar to modern video codecs.

Impact on Massive MIMO: Massive MIMO promises order of magnitude spectral efficiency gains by employing hundreds of antennas at the base stations to spatially multiplex tens of users. A major requirement to achieve such gains is the availability of accurate CSI to avoid multi-user interference, which is especially pronounced in massive MIMO due to the high spatial resolution achievable with large antenna arrays. The common consensus is that channel reciprocity in TDD systems will intrinsically provide the required CSIT; however, pilot contamination puts deterministic limits on the signal-to interference-plus-noise ratio (SINR) and the achievable rate. Several methods to mitigate such effects have been proposed, frequently employing not only pilot signals but also data symbols to improve channel estimates. Pilot contamination becomes even more pronounced in high-mobility situations, since the channel coherence time is short, thus restricting the length of pilot signals. This, in turn, limits the number of available orthogonal pilot sequences, reducing the pilot reuse distance between interfering base stations. Furthermore, channel aging in TDD-based massive MIMO systems, implying outdated CSI during transmission, causes enormous performance losses at high mobility [8]. In the long run, full duplex transmission might alleviate such issues, provided up- and downlink signal processing chains can be calibrated with sufficient accuracy.

Most existing LTE deployments, however, employ FDD. Since such systems have to rely on explicit CSI feedback to obtain channel estimates at the transmitter, the situation is even more challenging. Current three-dimensional (3D) beamforming and full dimension MIMO (FD-MIMO) developments discussed within 3GPP foresee a

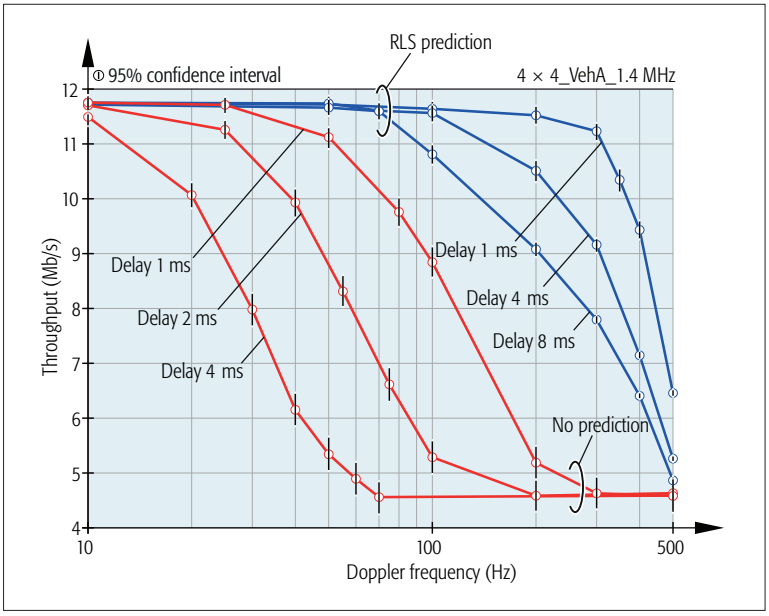


Figure 3. Impact of CSI feedback delay on throughput performance employing LTE's closed-loop spatial multiplexing at 2 GHz center frequency.

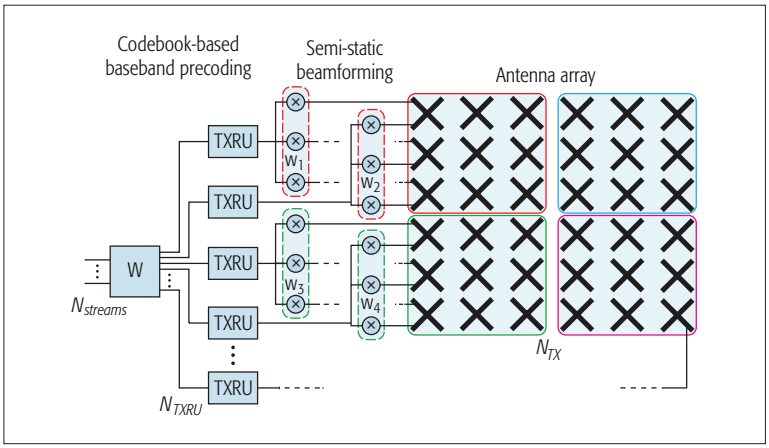


Figure 4. Transceiver unit virtualization as considered for Rel. 13 of LTE.

combination of codebook-based precoding and semi-static beamforming to extend the standard to large-scale antenna arrays; this approach is denoted transceiver unit (TXRU) virtualization, and is illustrated in Fig. 4. The method basically partitions the large-scale antenna array into several sub-arrays and applies semi-static beamforming weights along the vertical and horizontal directions of these sub-arrays to generate distinct radiation patterns. Based on CSI feedback from the users, reporting preferred baseband beamformers/precoders, the transmitter opportunistically selects a set of users that can be served in parallel with minimal interference. The achievable multiplexing gain is limited by the number of available TXRUs. At high user mobility, the spatial resolution of the baseband beamformer/precoder codebook will have to be reduced to ensure that the (delayed) CSI feedback is still valid during transmission; this, in turn, limits the number of non-overlapping beams that can be generated and thus the multiplexing capabilities.

Wireless communications in the mmWave band is of interest for 5G mobile networks, since large amounts of untapped spectrum are available in this regime, promising multi-gigibit-per-second transmission and substantially increased cell capacities. However, transmission in the mmWave band comes with its own challenges.

Alternatively, especially along predetermined paths of motion, such as, streets, highways and railroads, predictive beam-steering approaches, which predict beamforming directions based on users' trajectories, can enable higher resolution. Furthermore, location information (e.g., provided by GPS) can be helpful to achieve more accurate beam-steering [9].

Multi-Carrier Modulation with Adaptive Transmission Parameters

The advantages of non-orthogonal pulse-shapes for multi-carrier transmission over doubly dispersive channels compared to OFDM were already recognized by Kozek and Molisch in the 1990s [10]. Cyclic prefix (CP)-OFDM, as employed by most state-of-the-art wireless communication systems, can be designed robustly with respect to inter-symbol interference (ISI), caused by the delay spread of the wireless channel, by appending a CP of sufficient length. This, however, implies reduced spectral efficiency since only part of the symbol carries useful information. Furthermore, the rectangular pulse-shape employed by OFDM is prone to ICI caused by the frequency dispersion of the channel, which is especially problematic at high user mobility. Optimal pulse-shaping has thus gained attention in recent years, with the goal of finding efficient and robust successor waveforms for 5G wireless communications. Several modulation formats and multi-carrier waveforms currently compete for the succession of OFDM/SC-FDM as employed by LTE [11]. Arguments put forward for such novel waveform designs include, among others, spectral efficiency gains, enabled by omission of cyclic-prefix and by reduction of out-of-band emission; latency reduction, as required for real-time monitoring/control in the tactile Internet; and reduction of time/frequency synchronization sensitivity, supporting energy efficient machine-type-communications.

Introducing novel multi-carrier waveforms in future 5G mobile communication is an opportunity to enhance the performance of high-mobility users. Below we highlight some important design aspects and considerations for parameter-adaptive multi-carrier modulation that are imposed by the envisioned *Society in Motion*.

- Indoor and outdoor users with varying mobility will observe different channel dispersion characteristics (delay and Doppler spread). Hence, filter-bank parameters, such as the sub-carrier spacing and the applied prototype pulse, should be adaptable to the time-frequency dispersion of the channel. This allows matching waveform parameters to channel conditions to achieve a favorable trade-off between residual ICI/ISI and spectral efficiency.
- The TTI length should be adjustable to support different types of traffic with varying QoS demands. For the highest efficiency and robustness with respect to microscopic fading, channel coding over long blocks of data is required. However, for applications such as the tactile Internet, ultra low-latency transmission is essential to achieve the desired instantaneous remote response. Vehicular communications adds an extra dimension between these two, requiring reliable data transmission within certain hard deadlines (timeliness).
- The parameters addressed in the previous two bullets should be modifiable for individual users or groups of users with similar channel conditions, to enhance multi-user support with diverse channel and traffic characteristics. The most challenging is to accommodate a broad range of requirements, enabling efficient support of static indoor users with almost flat channels as well as high-mobility outdoor users with highly variable channels in time and frequency. In practice, certain discrete sets of compatible parameters will have to be applied to cover several ranges of dispersion characteristics with minimal extra signaling overhead for parameter adaptation.

Millimeter-Wave Transmission

Wireless communications in the mmWave band is of interest for 5G mobile networks, since large amounts of untapped spectrum are available in this regime, promising multi-gigabit-per-second transmission and substantially increased cell capacities. However, transmission in the mmWave band comes with its own challenges, such as hardware complexity constraints impacting beamforming and precoding algorithms; the requirement for highly directive beams to compensate for increased path loss; and significant probability of signal outages due to shadowing and reduced multipath propagation. Especially the latter two issues require careful investigation in the context of high-mobility and dependable communications, since accurate beamforming is challenging in highly time-variant scenarios, and signal outages cannot be tolerated in safety-relevant applications. To reduce signal outage probability, macro-diversity has to be enhanced through multi-connectivity approaches, as described below.

Wireless Channel Models

In many scenarios of interest for highly mobile users, for example, on highways and along railroads, the wireless channel behaves markedly different as compared to prevailing wireless channel models, which are utilized for evaluation of signal processing techniques within 3GPP and other standardization bodies. In vehicular environments, propagation is characterized by shadowing through other vehicles, high Doppler shifts with usually only few dominant scatterers (other vehicles, road signs, and other structures along the road) and inherent nonstationarity of the channel statistics [12]. Channel models for such scenarios exist and are utilized mostly within the community of IEEE 802.11p researchers and engineers. However, to the best of our knowledge, there are no such channel models available to date that cover novel techniques such as mmWave transmission, 3D beamforming, and FD-MIMO. Furthermore, it must be mentioned that the geometry-based stochastic 3GPP 3D channel model, presented in 3GPP Technical Report TR 36.873, is not well suited for investigation of high-mobility scenarios. Imagine a user who moves along a street; large-scale channel fading parameters, such as the azimuth and elevation spreads of the signal's arrival and departure angles, are generated in this 3D

channel model based on the network geometry and the user's position. These parameters are thus correlated between consecutive user positions. Based on the large-scale fading parameters, small-scale fading parameters, such as the actual arrival and departure angles, are randomly generated. As long as the position of the user stays constant, the obtained small-scale channel realizations vary smoothly over time and frequency. However, as soon as the position of the user changes, new random realizations of small-scale fading parameters are generated, leading to unrealistic non-smoothness of the channel realizations over time. Thus, this model in its current form is not applicable for the investigation, for example, of CSI feedback delay and predictive beamforming techniques, which can be a critical factor for the evaluation of future LTE-V. Such investigations are only meaningful if the channel varies smoothly over time, calling for revision of the 3GPP 3D channel model.

System-Level Challenges

In this section, we discuss system-level challenges imposed by high-mobility users. We do not treat the important use case of high-speed trains here; the interested reader is referred to the corresponding recent Feature Topic on Future Railway Communications in *IEEE Communications Magazine* [13].

LTE-Based Vehicular Communications

The goal of 3GPP's V2XLTE study item is to provide LTE support of vehicular communications. Such V2X specifications allow generating revenue via high-bandwidth infotainment applications for in-car users and proximity services, as well as through support of traffic telematics and intelligent transport systems (ITS), complementing or even replacing IEEE 802.11p-based dedicated short-range communications (DSRC). Currently three technologies are considered as central for the realization of LTE-based vehicular communications:

- Dual connectivity to support high user mobility in dense heterogeneous networks (HetNets)
- LTE-based broadcast services (public warning system, PWS, and enhanced multimedia broadcast/multicast service, eMBMS) for efficient distribution of messages among vehicles
- Proximity service (ProSe) including D2D communication to realize connectivity between vehicles as well as between connected cars and handheld terminals (pedestrians)

In cooperative ITS (C-ITS), vehicles communicate with each other and with roadside infrastructure to exchange information about vehicles' status, their locations, and the road environment (active road safety). For this purpose the European Telecommunications Standard Institute (ETSI) defined two basic types of messages: cooperative awareness messages (CAMs) and decentralized environmental notification messages (DENMs) . DENMs are event-driven and thus generated only sporadically, for example, in case of a traffic accident to warn upcoming vehicles; their distribution in certain notification areas is most efficiently handled through PWS, which allows geographical information to be considered. CAMs are periodically generated messages utilized to exchange vehicle status information with cars in the vicinity; the equivalent message in the U.S. Society of Automotive Engineers (SAE) J2735 standard is denoted basic safety message (BSM). CAMs/BSMs can be exchanged either directly via DSRC or indirectly utilizing the cellular network; both approaches have been shown to require substantial amounts of bandwidth to guarantee timely packet delivery (see, e.g., reports of the EU FP7 METIS project), more than is currently foreseen for road safety-critical communication in the assigned 5.9 GHz band.

To improve the performance of CAM/BSM delivery employing LTE, eMBMS should be extended to support multiple-antenna transmission. Several beamforming techniques for multiple-input single-output (MISO) multicast transmission have been proposed over the last decade. The latest developments are published in [14], where the so-called MISO multicast interference channel is considered, containing several multicast transmitters that interfere with each other. The authors of [14] propose leakage-based multicast (LBM) beamforming, a method that maximizes the minimum SINR of the intended multicast users while controlling the amount of interference leaked to users of other base stations. With an iterative optimization of interference leakage parameters among base stations, the achievable rate region of the involved base stations can be expanded as shown in Fig. 5;

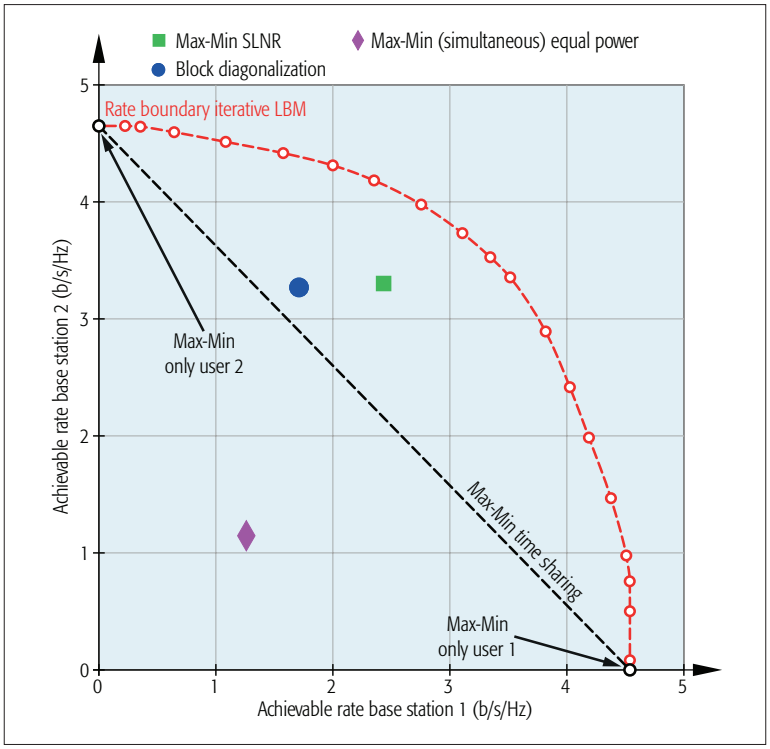


Figure 5. Achievable rate tuples of different transmission schemes for the multiple-input single-output multicast interference channel; two equally strong base stations, each equipped with eight antennas, serve six users each [14].

Despite the commonly agreed 5G targets of enhancing network capacity and reducing transmission latency, we thus view provisioning of dependable wireless connectivity for a Society in Motion as one of the major challenges and enablers of future mobile communications.

LBM outperforms existing alternatives such as block diagonalization and signal-to-leakage-plus-noise ratio (SLNR) precoding. However, the method is not well suited for high-mobility scenarios, since it applies iterative optimization and exhibits comparatively slow convergence speed; thus, further improvements are required.

Small Cell Enhancements

Small cells are currently mostly deployed indoors to improve capacity at user hotspot locations. More recently, though, outdoor rollouts have garnered industry interest to complement existing macrocell infrastructure; for example, Swisscom is currently testing in-house developed underground microcells for cable conduits. These microcells are connected to existing fixed network conduits that are buried in cable shafts below streets and public places in urban areas. Such developments enable large network capacities, by minimizing the distance to users and enabling ultra-dense deployments; however, they face difficulties in providing user mobility, since small cell sizes imply frequent handovers, increasing the signaling load of the network and degrading dependability of the wireless connection due to handover failures. To tackle such issues, the concept of macro-assisted small cells, or phantom cells, has been proposed for LTE, splitting the control plane and user plane of the network among macro and small cells to improve handover performance. This concept was standardized in Release 12 of LTE as dual connectivity. Further enhancement to multi-connectivity, maintaining multiple parallel connections to several macro/small cells, is expected in future releases of LTE and 5G; this promises increased data throughput over multiple parallel data streams, improved reliability due to additional macro-diversity, and enhanced robustness with respect to mobility, since hard handovers can be avoided. However, it comes at the cost of requiring sophisticated coordination of multiple transmission points to maintain efficiency of the network, implying substantial backhaul signaling overhead. In this context of self-organizing network functionalities, reinforcement learning techniques have gained interest, since they provide significant improvements in network capacity and mobility support with minimal information exchange among base stations(e.g., see [15]). Mobility management and network optimization in future mobile networks will additionally be complicated by vehicle mounted mobile relay nodes and vehicular small cells, which promise to offer premium wireless connectivity to their passengers even at highest velocities; vehicular small cells are currently being pushed into the market by all major car manufacturers.

Distributed Antennas

Distributed antenna systems (DASs) were originally proposed for indoor coverage improvement of cellular networks, but they also proved instrumental for reducing outage probability and increasing network capacity in outdoor environments. In contrast to small cells, which act as autonomous network entities, DASs are composed of several remote radio heads (RRHs) that are controlled by a single base station, commonly over radio-over-fiber links or dedicated mmWave connections. DASs are well suited to support highest mobility whenever users follow predetermined paths; hence, they find application in wireless communication systems for high-speed railways and highway scenarios. Here, they can reduce the amount of handovers between cells by virtually moving cells along with a user, radiating the associated signal from the nearest RRH. Further throughput gains are possible by coordinating transmissions from two or more neighboring RRHs to enhance signal quality [16]. Future mobile networks will likely blur the boundaries between small cells and DASs, since backhaul connections are becoming increasingly powerful, and CoMP schemes facilitate removing cell edges. Especially in the context of mmWave transmission, we see large potential in DASs for improving dependability of the wireless connection: via on-demand mmWave backhaul links spatially distributed RRHs can dynamically form DASs in order to provide macro-diversity for the otherwise outage-susceptible user connection.

Conclusion

In a future society, where constantly large numbers of wirelessly connected people and machines are on the move, providing satisfactory QoS, even at the highest velocities, will demand careful design and validation of prospective mobile networking technologies. Specifically, if the scope of mobile communications is to be extended beyond best effort connections and entertainment services, to support even critical information exchange, such as for road safety applications, dependability (timeliness and reliability) of data transmission is imperative. Despite the commonly agreed 5G targets of enhancing network capacity and reducing transmission latency, we thus view provisioning of dependable wireless connectivity for a *Society in Motion* as one of the major challenges and enablers of future mobile communications.

Acknowledgment

The financial support of the Austrian Federal Ministry of Science, Research and Economy, the National Foundation for Research, Technology and Development, and by TU Wien is gratefully acknowledged. This work has been co-financed by A1 Telekom Austria AG, Kathrein-Werke KG, and Nokia Solutions and Networks GmbH & Co. KG.

References

[1] UN, "World Urbanization Prospects [highlights], 2014 revision," 2014, esa.un.org/unpd/wup/.

[2] A. Avizienis *et al.*, "Basic Concepts and Taxonomy of Dependable and Secure Computing," *IEEE Trans. Dependable Secure Computing*, vol. 1, no. 1, Jan. 2004, pp. 11–33.

[3] D. Xenakis *et al.*, "Mobility Management for Femtocells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 1, Jan. 2014, pp. 64–91.

[4] M. Simko *et al.*, "Adaptive Pilot-Symbol Patterns for MIMO OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, Sept. 2013, pp. 4705–15.

[5] A. Molisch, M. Toeltsch, and S. Vermani, "Iterative Methods for Cancellation of Intercarrier Interference in OFDM systems," *IEEE Trans. Vehic. Tech.*, vol. 56, no. 4, July 2007, pp. 2158–67.

[6] R. Nissel and M. Rupp, "Doubly-Selective MMSE Channel Estimation and ICI Mitigation for OFDM Systems," *Proc. IEEE ICC '15*, London, U.K., June 2015.

[7] S. Schwarz and M. Rupp, "Predictive Quantization on the Stiefel Manifold," *IEEE Signal Process. Lett.*, vol. 22, no. 2, Feb. 2015, pp. 234–38.

[8] K. Truong and R. W. Heath Jr., "Effects of Channel Aging in Massive MIMO Systems," *J. Commun. and Networks*, vol. 15, no. 4, Aug 2013, pp. 338–51.
[9] V. Va, X. Zhang, and R. W. Heath, Jr., "Beam Switching for Millimeter Wave Communication to Support High Speed Trains," *Proc. 82nd IEEE VTC*, Boston, MA, Sept. 2015.
[10] W. Kozek and A. Molisch, "Nonorthogonal Pulseshapes for Multicarrier Communications in Doubly Dispersive Channels," *IEEE JSAC*, vol. 16, no. 8, Oct 1998, pp. 1579–89.
[11] P. Banelli *et al*., "Modulation Formats and Waveforms for 5G Networks: Who Will Be the Heir of OFDM?: An Overview of Alternative Modulation Schemes for Improved Spectral Efficiency," *IEEE Signal Processing Mag*., vol. 31, no. 6, Nov. 2014, pp. 80–93.
[12] C. Mecklenbräuker *et al*., "Vehicular Channel Characterization and Its Implications for Wireless System Design and Performance," *Proc. IEEE*, vol. 99, no. 7, July 2011, pp. 1189–1212.
[13] D. Matolak *et al*., "Future Railway Communications," *IEEE JSAC*, vol. 53, no. 10, Oct. 2015, pp. 60–123.
[14] S. Schwarz and M. Rupp, "Transmit Optimization for the MISO Multicast Interference Channel," *IEEE Trans. Commun*., vol. 63, no. 12, Dec. 2015, pp. 4936–49.
[15] M. Simsek, M. Bennis, and I. Guvenc, "Mobility Management in HetNets: A Learning-Based Perspective," *EURASIP J. Wireless Commun. and Networking*, vol. 2015, no. 1, 2015.
[16] M. Müller, M. Taranetz, and M. Rupp, "Providing Current and Future Cellular Services to High Speed Trains," *IEEE Commun. Mag*., vol. 10, 2015, pp. 96–101.

Biographies

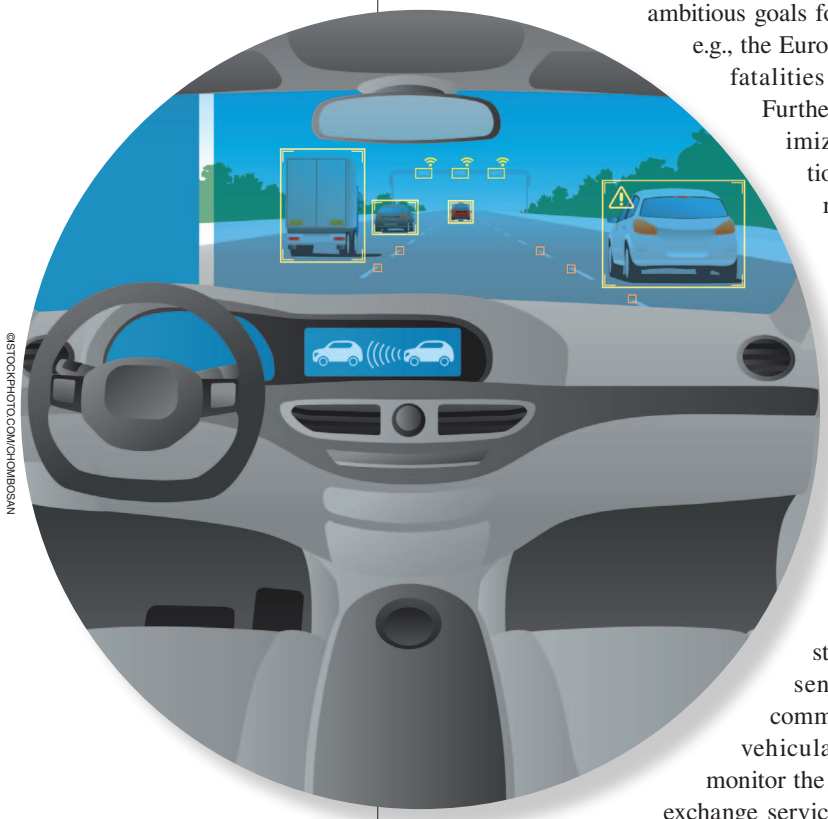
Stefan Schwarz (sschwarz@nt.tuwien.ac) received his Dr.techn. degree in telecommunications engineering in 2013 at Technische Universität (TU) Wien. In 2010 he received the honorary prize of the Austrian Minister of Science and Research and in 2014 he received the INiTS ICT award. From 2008 to 2014 he worked at the Institute of Telecommunications (ITC) of TU Wien. Since 2016 he has headed the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion.

Markus Rupp received his Dr.-Ing. degree in 1993 at Technische Universität Darmstadt. From 1993 until 1995, he had a postdoctoral position at the University of Santa Barbara, California. From October 1995 until August 2001 he was a member of technical staff in the Wireless Technology Research Department of Bell Labs, Crawford Hill, New Jersey. Since October 2001 he has been a full professor in Digital Signal Processing in Mobile Communications at TU Wien.

Stefan Schwarz, Tal Philosof, and Markus Rupp

Signal Processing Challenges in Cellular-Assisted Vehicular Communications

Efforts and developments within 3GPP LTE and beyond



©ISTOCKPHOTO.COM/CHOMBOSAN

Vehicular communications is an important enabler for enhancing the safety on roads by supporting mutual awareness of vehicles as well as for improving the efficiency of transportation through smart traffic management by intelligent transport systems (ITSs). Governments around the world have set ambitious goals for road fatality reduction in the near future; e.g., the European Union targets a 50% reduction of road fatalities by 2020 as compared to the year 2010. Furthermore, traffic telematic systems aim to minimize the environmental impact of transportation and maximize the utilization of available road infrastructure by adaptive traffic management. To realize these challenging targets, autonomous wireless information exchange among vehicles—vehicle to vehicle (V2V)—and with roadside infrastructure—vehicle to infrastructure (V2I)—are central ingredients. In addition to traffic efficiency and safety-related issues, vehicular communications is increasingly recognized as an important revenue driver by car manufacturing companies since it enables wirelessly connected in-vehicle entertainment systems that support on-demand video streaming and online Internet access for passengers. Also, in the future, machine-type communication is expected to play a major role in vehicular environments, with more sensors that monitor the internal state of vehicles and autonomously exchange service and maintenance information with cloud servers of manufacturers. Depending on the considered use-case, distinct quality of service (QoS) requirements come into play [1]: infotainment applications for in-car users require high bandwidth and network capacity, active road safety relies on delay- and outage-critical data transmission, whereas information exchange for road traffic efficiency management typically comes without strict QoS requirements and exhibits graceful degradation of performance with increasing latency.

Digital Object Identifier 10.1109/MSP.2016.2637938
Date of publication: 3 March 2017

Introduction

In recent years, specific vehicular communication systems, such as the European Telecommunications Standards Institute (ETSI) ITS G5 based on IEEE 802.11p dedicated short-range communication (DSRC), have been developed to enable timely and reliable exchange of information in so-called vehicular ad hoc networks (VANETs). These systems are commonly based on dedicated infrastructure and employ licensed transmission bands to avoid interference with other existing systems. However, interest in mobile communications technology to support VANETs has been recently increasing, because this technology is available off-the-shelf and, therefore, enables cost-effective implementations. Due to the virtual ubiquity of cellular networks, vehicular communications strategies of many companies are shaped by the development of mobile network technologies. In response, the progression of the universal mobile telecommunications system (UMTS) long-term evolution (LTE) toward vehicular communications is currently pushed within the Third Generation Partnership Project (3GPP) to meet the requirements of vehicular environments.

Efficient and reliable wireless communication with users at high mobility, however, comes with several unique challenges that the LTE standard cannot yet stand up to, as we demonstrate in this article by some selected examples. Yet, advanced signal processing at the transmitters and the receivers has the potential to alleviate the shortcomings of LTE. In this article, we provide an overview of efforts ongoing within the development of LTE Release 14 to enhance the support of vehicular communications. We further give insights into promising signal processing methods for efficient wireless connectivity at high mobility. While some of the described techniques can readily be employed without modification of the LTE standard, others require additional standardization efforts to harmonize the operation of communication devices.

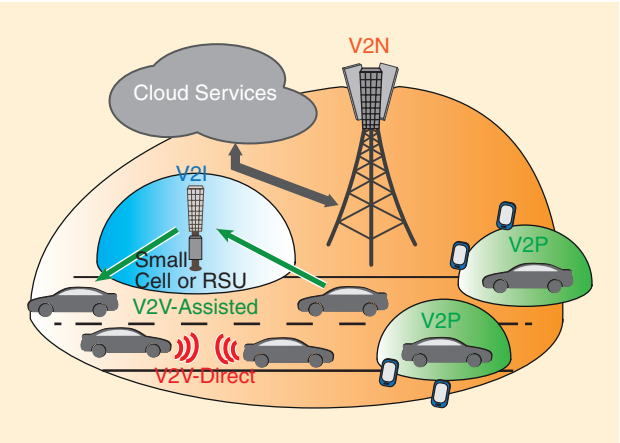


FIGURE 1. Vehicular communications scenarios considered within 3GPP: V2N provides connectivity to the cloud; V2I covers connections to RSUs; V2V deals with direct and assisted intervehicle transmission; V2P enables information exchange with other devices in the proximity of vehicles (e.g., pedestrians and cyclists).

LTE contenders for vehicular communications

The 3GPP recently initiated study items within Release 14 of UMTS LTE on vehicular communications; see [31] and [32]. The goal of these study items is to develop a set of LTE specifications for vehicular environments [LTE-based vehicle to X (V2X)] [2]. Currently, the focus of standardization is on exploring the support of active road safety applications, either in a dedicated transmission band or shared with conventional mobile subscribers. The different use cases considered within the 3GPP are illustrated in Figure 1. Connectivity in between vehicles, e.g., to enable the V2V exchange of location and trajectory information for driver assistance systems, is supported directly via device-to-device (D2D) transmission, as well as indirectly, employing base stations and dedicated roadside units (RSUs) as transmission hubs to facilitate information sharing over larger distances. The local exchange of traffic management information between ITS infrastructure and vehicles is covered under the umbrella of V2I communication. Connectivity to the Internet and to cloud services falls in the category of vehicle-to-network (V2N) transmission; finally, the exchange of mutual awareness information between vehicles and pedestrians (or cyclists) is supported by vehicle-to-pedestrian (V2P) transmission. Commonly, all of these different types of vehicular communication scenarios are summarized under the term *V2X* communication. Even though they appear very similar from a basic physical layer (PHY) perspective, the different use cases require individual treatment within standardization to enable efficient incorporation into the existing specifications [3].

Currently three technologies are considered as central for LTE based vehicular communications:

- *Dual connectivity to support high user mobility in dense heterogeneous networks* (*HetNets*): Mobile communication connections at high mobility suffer from frequent interruptions due to handovers in between cells; this is especially problematic in dense HetNets with coverage areas of base stations of mere tens to few hundreds of meters [4]. To alleviate such issues, the 3GPP has incorporated dual connectivity in LTE Release 12. With dual connectivity, robustness with respect to mobility is enhanced by connecting subscribers to two cells—master and secondary—simultaneously. Thereby, critical control and signaling information, as provided by the control plane of the mobile network, is kept in the master cell of the macro base-station layer, providing reliable coverage over large geographic areas with a minimal amount of handovers. The capacity and rate of the actual data transmission is enhanced by concurrently providing data over the user plane of the network from macro base stations and small cells [5]. For this data exchange, connection interruptions are less critical since they do not cause connectivity failures.
- *LTE-based broadcast services, as supported by the public warning system* (*PWS*) *and enhanced multimedia broadcast/multicast service* (*eMBMS*)*, for efficient distribution of messages among vehicles*: In cooperative ITS (C-ITS), two basic types of messages are specified by ETSI for exchanging

information among vehicles: 1) The periodic cooperative awareness message (CAM) is employed to share vehicle status information (location, direction, speed) with cars in the vicinity, and 2) the event-driven sporadic decentralized environmental notification message is used to exchange critical warnings, e.g., in case of traffic accidents (in the U.S. Society of Automotive Engineers J2735 standard, both types are covered by the basic safety message). Since such messages are shared with all other vehicles within a certain geographic region, broadcasting/multicasting them over the cellular network can be much more efficient than unicast transmission to each vehicle individually. The distribution of messages in certain geographic notification areas is supported by the PWS, which allows consideration of geographic packet routing information. With Multimedia Broadcast/Multicast Single Frequency Network (MBSFN), the LTE standard even supports multicasting in potentially very large geographic areas by synchronously transmitting the same information from multiple base stations.

- *Proximity services, including D2D communication to realize connectivity in between vehicles as well as between connected cars and handheld terminals*: Especially in safety-critical situations, latency of data exchange can be a limiting factor. Lowest latency is achieved by minimizing the amount of traffic nodes that need to be traversed between source and destination; hence, direct D2D communication is most promising for short-range information exchange with stringent latency requirements. Furthermore, D2D transmission can be designed to enable autonomous VANET operation in areas that are not covered by cellular infrastructure, thus enhancing the reliability and availability of vehicular communications. Besides these operational advantages, D2D offloading of traffic from the cellular network also helps to reduce the overhead caused by vehicular communications.

Our focus in this article is on infrastructure-based vehicular communications, i.e., on V2I and cellular-assisted V2V transmission. We present exemplary performance results of LTE-compliant V2X transmissions to demonstrate important potential shortcomings of the standard that may limit the applicability of LTE for vehicular communications. Based on these insights, we then present signal processing techniques to mitigate these weaknesses.

Dual connectivity-enhanced V2I transmission

Dual connectivity is intended to solve the handover problem of subscribers moving at high mobility through dense HetNets. To investigate the performance of this approach, we conducted LTE-compliant system-level simulations of such dense HetNets, employing the Vienna LTE system-level simulator [6]; Figure 2 shows the macroscopic signal-to-interference-plus-noise ratio (SINR) distribution of one snapshot of the evaluated HetNet. The network contains seven hexagonally arranged macro base stations (BS 1,…,BS 7) that each serve three sectors, with nominal coverage regions illustrated as white hexagons. Additionally, a large number of small cells is randomly distributed over the network area; these devices are equipped with omnidirectional antennas. Subscribers are served via dual connectivity from the strongest macro base station and small cell simultaneously, with the control plane being kept at the macro layer to minimize necessary handovers.

In our numerical investigation, we assume that all base stations operate with 5-MHz system bandwidth at the same carrier frequency $f_c = 2$ GHz, i.e., we consider a spatial frequency reuse-factor of one. Without dual connectivity, this leads to low SINR values at the cell edge as illustrated in Figure 2. Most contemporary fourth-generation cellular networks nevertheless utilize all available bandwidth at all base stations to enable harvesting the spatial reuse gain promised by network densification; this will hold even more for future fifth-generation (5G) networks. The considered situation corresponds to the so-called cochannel dual connectivity mode of LTE, where small cells and macro base stations employ the same carrier frequency for data transmission. Additionally, the LTE standard supports interfrequency dual connectivity, where different frequencies are employed on macro base stations and small cells. Notice that the SINR distribution in Figure 2 shows the behavior without cochannel dual connectivity. If dual connectivity is activated for a user, the cell edge between the corresponding macro base station and small cell vanishes, since data transmission to the user occurs from both base stations simultaneously. We evaluate the empirical cumulative distribution function (ECDF) of the downlink throughput of subscribers in Figure 3. We compare the performance of low- and high-mobility users moving at $v = 5$ km/h and $v = 150$ km/h, respectively. The system applies transmission rate adaptation according to the instantaneous SINR of subscribers, such as to match the spectral efficiency of the employed modulation and coding scheme (MCS) to the current channel quality. The required channel state information at the transmitter (CSIT) is provided as feedback information from the subscribers to the base stations using the LTE standard defined channel

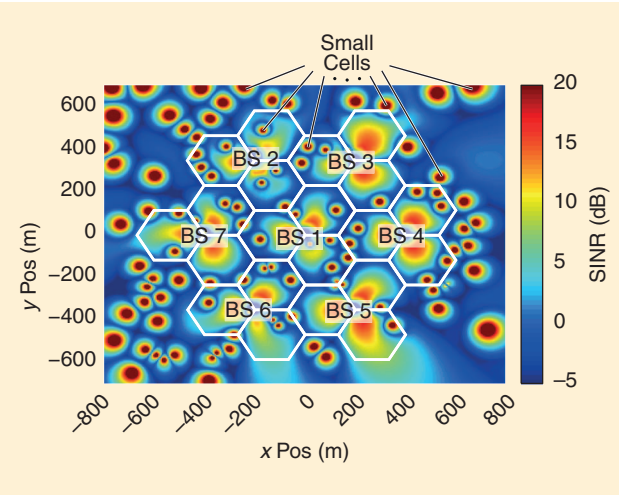


FIGURE 2. The macroscopic path loss-based SINR distribution of a random HetNet realization with seven macro base stations (each serving three sectors) and a small-cell density of 50 per km^2 (Poisson point process). The white hexagons illustrate the nominal assignment regions of macro base stations.

quality indicator (CQI) and the feedback algorithms described in [7]. More specifically, the CQI is utilized by the receiver to signal via a dedicated feedback link to the transmitter which MCS should be employed for transmission to achieve reliable as well as efficient transmission; this information is derived from the SINR currently experienced by the receiver. We observe from Figure 3 that, under ideal circumstances of delayless noncausal feedback (delay 0 ms), high-mobility users achieve almost the same performance as those at low mobility. There is only a small performance degradation, which is caused by the need to apply a slightly more conservative transmission rate adaptation, since the channel of high-mobility users varies even within one transmission time interval (TTI). This observation confirms the expected mobility enhancement achieved by dual connectivity. Yet, as soon as we consider certain delay in the CQI feedback link, we observe strong throughput degradation at high mobility, with more than 25% of subscribers obtaining zero throughput even with only 1 ms feedback delay. This performance degradation occurs because transmission rate adaptation with feedback delay is based on outdated channel state information (CSI) leading to signal outages due to mismatch between the utilized transmission rate and the rate supported by the channel. This loss in throughput goes hand in hand with an increase in latency, since each lost packet must be retransmitted. Notice, the TTI length of LTE is $T_s = 1$ ms. Hence, a delay of the feedback processing below this value is infeasible; even the value of 5 ms considered in Figure 3 may be hard to achieve in practice.

To gauge the expected impact of feedback delay τ in milliseconds, with respect to the system parameters carrier frequency f_c, TTI length T_s, and user velocity v, we define the normalized feedback delay as

$$\tau_n = \tau f_d = \tau_r T_s f_d = \tau_r T_s f_c \frac{v}{c}, \tag{1}$$

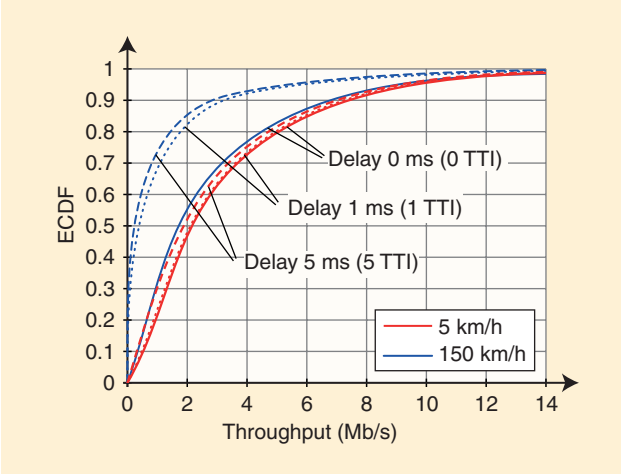


FIGURE 3. The throughput in dependence of user velocity and CSI feedback delay in a dense HetNet supporting dual connectivity.

CSI feedback delay is a significant issue for high-mobility users, and it can be a strongly limiting factor for the rate performance and reliability of wireless transmission in vehicular scenarios.

with f_d denoting the maximum Doppler frequency shift in hertz, $\tau_r = \tau/T_s$ being the relative feedback delay in multiples of TTIs, and c representing the speed of light. The feedback delay has to be seen in relation to the temporal variability of the channel; if the channel is quasi static (as in many indoor scenarios), a feedback delay is irrelevant. It only matters if the channel changes significantly within the duration of the feedback delay, which is gauged by the normalized feedback delay. More specifically, the Doppler shift f_d is inversely proportional to the coherence time of the channel; hence, the product τf_d is proportional to the number of coherence intervals elapsing during the feedback delay duration. The exact proportionality constant depends on the observed Doppler spectrum [8]. In our aforementioned example, we have for $\tau = 1$ ms with $v = 5$ km/h and $v = 150$ km/h, $\tau_n \approx 0.009$, and $\tau_n \approx 0.27$, respectively. We will encounter the normalized feedback delay again later in the section "Future Enhancements and Challenges."

CSI feedback delay is a significant issue for high-mobility users, and it can be a strongly limiting factor for the rate performance and reliability of wireless transmission in vehicular scenarios. The problem is, of course, not specific to dual connectivity, but occurs whenever rate adaptation is based on outdated CSI. In the section "CSI Feedback Enhancement," we discuss possible signal processing approaches to mitigate this performance degradation.

MBSFN-based V2V communication

When active road safety support is implemented in C-ITS, common information must be delivered to many vehicles within certain geographic regions; e.g., status information, such as position, velocity and direction, of one vehicle is shared with all other vehicles in the vicinity via CAMs to enhance mutual awareness of road traffic participants. Prior studies have shown that both a direct exchange of periodically generated CAMs using DSRC as well as an indirect distribution over dedicated roadside infrastructure, requires a substantial amount of bandwidth to guarantee timely packet delivery. In fact, with a growing number of vehicles within the geographic region of interest, the currently foreseen 5.9-GHz band for road safety critical communication can easily be overloaded by CAM distribution [9].

In such situations, mobile cellular networks can be helpful to offload some traffic from a dedicated ITS infrastructure. Employing eMBMS/PWS features of LTE, the broadcast nature of cellular systems can be utilized to efficiently deliver the same message to many users within certain geographic regions formed by MBSFN areas. To demonstrate the capabilities of such an approach, we conduct system-level simulations, comparing message distribution via unicast and multicast transmission [10]. Both approaches have certain advantages/disadvantages:

- Unicasting in LTE enables per-user scheduling on favorable time/frequency resources to utilize channel and multiuser

diversity. Furthermore, it supports dynamic link adaptation (multiantenna precoding and transmission rate), as well as selective PHY repetition of lost packets using hybrid automatic repeat request (HARQ). These features facilitate efficient and reliable data transmission to single users with low retransmission latency of lost packets.

- Multicasting in LTE is restricted to employ certain reserved subframes (multicast subframes) and, hence, cannot freely exploit channel diversity. In general, channel and multiuser diversity in multicasting are more difficult to harvest since many users are served in parallel, and, therefore, chances of finding time/frequency resources that are favorable for all users simultaneously are small. Nevertheless, scheduling/resource allocation gains are possible by avoiding the worst time/frequency resources of all users, such that signal outages become less likely. Multicast transmission in LTE does not support PHY retransmission of lost packets and, thus, has to rely on slow (in terms of latency) higher-layer protocols. Furthermore, transmission rate adaptation must be performed according to the channel quality of the worst users to guarantee reliable delivery to all multicast users, reducing the spectral efficiency of the system; indeed, currently existing proposals do not utilize rate adaptation at all but rather rely on fixed rate transmission to avoid CSI feedback from the users. Finally, multicasting in LTE also does not support multiantenna transmission, even though multicast-specific beamforming/precoding and space-time coding has the potential to substantially improve efficiency and reliability of data transmission.

In the following, we investigate the performance of CAM distribution in cellular networks serving vehicles that move at $v = 150$ km/h. We consider fixed-rate transmission (employing CQIs 3, 6, and 9), since rate adaptation in multicasting is not yet supported by the LTE standard. We conduct system-level simulations for an MBSFN area consisting of three macro base stations embedded within a larger cellular network. Within the MBSFN area, CAMs of size 300 bytes are exchanged among 21 vehicles (seven per base station) with a periodicity of 100 ms. Hence, for multicast transmission within the MBSFN area, the total traffic load is 300·8 bit/user·21 users/100 ms $\approx$ 0.5 Mbit/s, since each packet is synchronously broadcast from all three base stations. In case of unicast transmission, on the other hand, each vehicle individually receives the CAMs generated by the other 20 vehicles. Since seven vehicles are attached to each base station of the MBSFN area, this implies a unicast traffic load of 300·8 bit/user·20 users·7/100 ms $\approx$ 3.4 Mbit/s. We determine the overhead for the cellular network caused by CAM distribution, as well as the operationally critical parameters latency and message loss probability for supporting active road safety in C-ITS. We assume transmission with extended cyclic prefix for unicast and multicast operation.

The results of the simulation are summarized in Table 1. We observe that a unicast transmission with CQI 3 is unable to support the generated data traffic, i.e., the network overhead caused by CAM distribution is equal to 270% of the capacity (assuming 5-MHz bandwidth); thus, the message loss probability is very high ($> 63\%$) since many CAMs have to be dropped. With multicasting, however, the overhead is reduced to 60% and the message loss probability is below 1%. Yet, even with multicasting in our simulation scenario, it is not possible to sustain the generated CAM network load with CQI less than three, because the transmission efficiency would be too low. Notice that, at most, six subframes per radio frame (consisting of ten subframes) can be reserved for MBSFN operation in LTE. A very important metric for active road safety is latency. Since eMBMS does not support retransmission of lost packets, latency accumulates in multiples of the message generation period (100 ms) in case of packet loss. Thus, even short signal outages can severely increase latency; correspondingly, latency of multicasting deteriorates with increasing transmission rate. For active road safety, latency below 100 ms has to be achieved [11]. Notice that Table 1 only presents radio-link latency for downlink transmissions, neither accounting for uplink from vehicles to base stations nor for CAM distribution within the MBSFN area; hence, we present the percentage of users with downlink latency below 50 ms to incorporate a safety margin with respect to the prescribed 100 ms. We observe in Table 1 that LTE is not able to sustain this latency with sufficiently high probability in our simulation. Hence, performance improvements of multicasting in LTE are required to enable dependable support of road safety applications.

In the section "Multicast Enhancements," we discuss potential enhancements of LTE to improve multicast transmission, employing dynamic link adaptation and coordination of multipoint transmission. In addition to enhancing the wireless transmission, we also see potential in optimizing the C-ITS protocol itself. As mentioned previously,

Table 1. A comparison of unicast and multicast transmission for CAM distribution in cellular networks.

Rate	Metric	Unicast	Multicast
CQI3	Overhead	270%	60%
(0.377 bit/sym)	Message loss prob.	> 63%	< 1%
	Latency < 50 ms	n/a	93%
	Latency < 250 ms	n/a	98%
CQI6	Overhead	86%	20%
(1.176 bit/sym)	Message loss prob.	7%	3%
	Latency < 50 ms	9.5%	80%
	Latency < 250 ms	85%	93%
CQI9	Overhead	42%	10%
(2.406 bit/sym)	Message loss prob.	11%	11%
	Latency < 50 ms	49%	54%
	Latency < 250 ms	74%	76%

cellular networks can easily be overloaded by CAM distribution, especially when robust MCSs with low spectral efficiency are employed. In such overload situations, reliable distribution of C-ITS messages cannot be guaranteed and, thus, they must be avoided. One possible method could be to adapt CAM generation periodicity according to user mobility; i.e., at low mobility, less frequent vehicle status updates are required than at high mobility. Also, the spatial area in which CAMs are distributed, i.e., the number of base stations involved in CAM exchange, should be carefully chosen and potentially adapted according to traffic load.

Future enhancements and challenges

In this section, we present promising improvements for wireless vehicular communications in cellular networks and highlight associated signal processing challenges. Our focus is on two topics: in the section "CSI Feedback Enhancement," we discuss CSI feedback enhancements to improve throughput performance of mobile users, employing channel predictive approaches to partly compensate for feedback delay and temporal channel variation. As we demonstrated in the section "Dual Connectivity-Enhanced V2I Transmission," inaccurate and outdated CSIT can be the limiting factor for the achievable efficiency and reliability of wireless data transmission. Hence, enhancing CSI feedback for high-mobility situations is an essential prerequisite of wireless vehicular communications. In the section "Multicast Enhancements," we consider techniques for efficient multicast transmission, putting scope not only on advanced multiple-input, multiple-output (MIMO) and coordinated multipoint (CoMP) transmission schemes, but also highlighting issues related to scheduling and resource assignment. Such techniques are important for vehicular scenarios to avoid the network overload problem discussed in the section "MBSFN-Based V2V Communication" by improving the capacity of multicast transmissions. The presented methods, again, require accurate CSIT for beamformer/precoder calculation and, hence, rely on efficient CSI feedback techniques. Finally, we present further important topics for progressing wireless vehicular communications in the section "Further Research Topics to Enhance Wireless Vehicular Communications."

CSI feedback enhancement

CSI is useful for achieving the highest performance in multiple antenna wireless communications by enabling transmission rate adaptation, adaptive MIMO beamforming, and spatial multiplexing. While CSI at the receiver (CSIR) is comparatively easy to obtain through pilot-aided channel estimation, CSIT in frequency-division duplex systems is only available if the receiver provides feedback information to the transmitter; Figure 4 illustrates the situation. (In time-division duplex systems, channel reciprocity can be exploited to estimate the CSI on both sides of the link.) Since the feedback link between the transmitter and receiver is of limited capacity, quantization of CSI is necessary to enable signaling with a finite number of bits; this is known as *limited feedback operation*. In general, we distinguish between providing explicit and implicit CSI feedback. In the former case, the channel, as estimated by the receiver, is directly quantized and fed back, while in the latter case derived information, such as the optimal MIMO precoder and the supported transmission rate, is provided as side information. In LTE, the second type of feedback information is currently supported with the standard-defined CQI, the precoding matrix indicator (PMI), and the rank indicator (RI). More advanced multiuser MIMO and CoMP schemes, however, mostly rely on explicit CSIT.

Implicit CSI feedback

In the section "Dual Connectivity-Enhanced V2I Transmission," we observed that CQI feedback, as employed for transmission rate adaptation in LTE, is sensitive to feedback delay in high-mobility scenarios, since outdated CQI feedback can cause signal outages, thereby reducing transmission efficiency and increasing latency. A similar behavior exists for the MIMO-specific feedback information PMI and RI, yet with reduced sensitivity since their mismatch does not directly cause signal outages.

To reduce the impact of feedback delay, either the transmitter or the receiver should attempt to compensate for it. At the receiver side, one natural approach is to apply channel prediction. The simplest method for OFDM-based LTE is a subcarrier-wise linear extrapolation of the channel transfer function between pairs of transmit and receive antennas. If channel statistics are known, the prediction performance can be improved by linear minimum mean squared

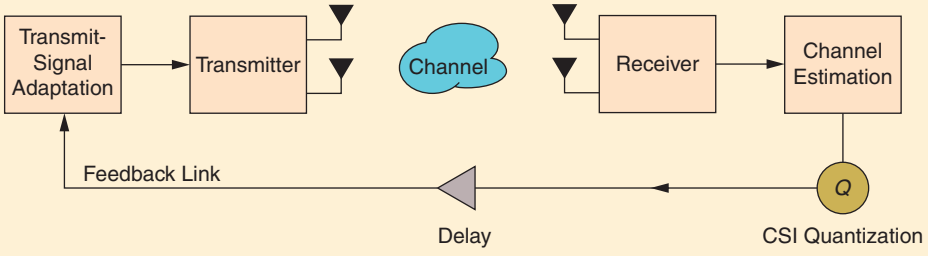


FIGURE 4. An illustration of limited feedback operation in wireless communications assuming error-free but delayed feedback.

error (MMSE) filtering. Without explicit knowledge of the channel statistics, iterative algorithms, such as least mean squares (LMS) or recursive least squares (RLS), can be applied to estimate the MMSE filter coefficients during operation. The linear MMSE filter can further be extended to jointly estimate the channel between all transmit and receive antenna pairs to exploit potential correlations [12]; similarly, the correlation between subcarriers can be utilized by joint estimation of the filters applied on multiple subcarriers.

Alternatively, instead of predicting the channel transfer function, one can predict the channel impulse response and transform the predicted impulse response to the frequency domain for feedback calculation. Efficient prediction of the channel impulse response is possible by applying the minimum-energy band limited discrete prolate spheroidal Slepian sequences as basis functions [13], which are good representatives for the finite support delay-Doppler spreading function of the wireless channel.

If channel prediction at the receiver is too costly, the transmitter can alternatively process the CQI feedback from the user before rate adaptation. One possibility for CQI selection at the transmitter is to maximize the expected throughput at time instant k

$$\mathrm{CQI}[k] = \underset{\mathrm{CQI}_i,\, i\in\{1,\ldots,C\}}{\operatorname{argmax}} \; T(\mathrm{CQI}_i)\, \mathbb{P}\left\{T(\mathrm{CQI}_i) \le r[k]\right\}, \qquad (2)$$

with $T(\mathrm{CQI}_i)$ denoting the transmission rate associated with CQI_i and C being the number of available CQIs as specified by the standard. The term $\mathbb{P}\{T(\mathrm{CQI}_i) \le r[k]\}$ is the probability that the currently supported rate $r[k]$ of the channel sustains the selected rate. This probability is commonly not known a priori by the transmitter; yet, assuming that the channel statistics do not vary too quickly, it can easily be learned over time from the user CQI feedback, which signals the highest instantaneously supported MCS [7]. Such a rate adaptation approach can potentially cause unacceptably high block-error ratio and signal outage probability. This can be avoided by adding an additional constraint to problem (2) to account for the acceptable block-error ratio.

In Figure 5, we compare the performance of the described methods in dependence of the normalized feedback delay. We observe that the predictive schemes achieve close to optimal performance up to a certain critical delay, whose exact value depends on the sophistication of the prediction method. Compensation at the transmitter, employing (2) with probabilities estimated from the user feedback, does not achieve optimal performance; yet, it exhibits more robust behavior at very high normalized feedback delay. A further improvement of predictive schemes can be expected by underlying realistic parametric radio channel models to reduce the number of free parameters to estimate/predict [14].

If channel prediction at the receiver is too costly, the transmitter can alternatively process the CQI feedback from the user before rate adaptation.

Explicit CSI feedback

Advanced multiuser MIMO and CoMP transmission schemes, such as block diagonalization (BD) or regularized BD (RBD) precoding, interference-leakage-based precoding [15], and interference alignment (IA), require explicit CSIT for the calculation of precoders. These precoding techniques can enhance the network capacity and reliability by reducing interference among subscribers and base stations. For many such precoding techniques, the underlying CSI can efficiently be represented as a point on a topological manifold. Let $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denote the channel matrix containing the complex-valued channel gains between the N_t transmit antennas and the N_r receive antennas; we assume $N_t \ge N_r$, which is commonly fulfilled in cellular downlink transmission. For channel-subspace-based precoding techniques, such as BD precoding and IA, the N_r-dimensional subspace spanned by $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is required as CSIT for precoder calculation; this information can be represented as a point on the complex Grassmann manifold $\mathcal{G}(N_t, N_r)$ of N_r-dimensional subspaces in the N_t-dimensional Euclidean space [16]. For other precoding schemes, such as RBD and the interference-leakage-based schemes presented in the section "Multicast Enhancements," the directions and magnitudes of the channel eigenmodes are necessary CSIT; i.e., the eigenvectors and eigenvalues of the channel Gramian $\mathbf{H}^*\mathbf{H}$, with $\mathbf{H}^*$ denoting the conjugate-transpose of $\mathbf{H}$. The Gramian is a point on the manifold of symmetric positive semidefinite matrices [17]. Furthermore, the matrix of eigenvectors corresponding to the N_r nonzero eigenvalues of $\mathbf{H}^*\mathbf{H}$ can be interpreted as a point on the compact Stiefel manifold $\mathcal{S}t(N_t, N_r)$ over the complex numbers [18]. Such CSI representation on manifolds enables dimensionality reduction, which allows efficient quantization with reduced feedback overhead as compared to quantization of the entire channel matrix.

To further reduce the feedback overhead, temporal channel correlation can be exploited through low-rate predictive

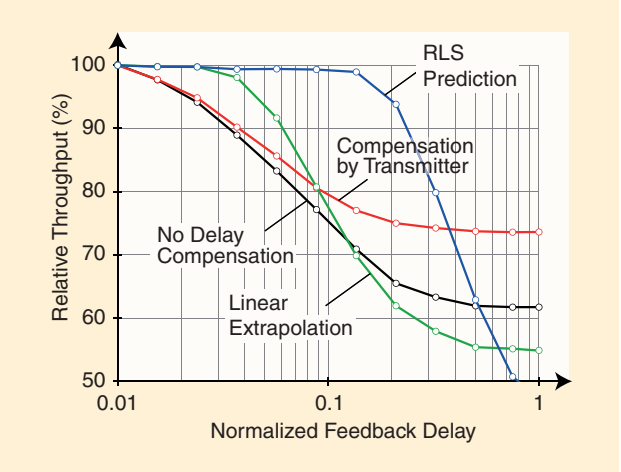


FIGURE 5. The sensitivity of different CQI feedback calculation methods with respect to feedback channel delay and user mobility.

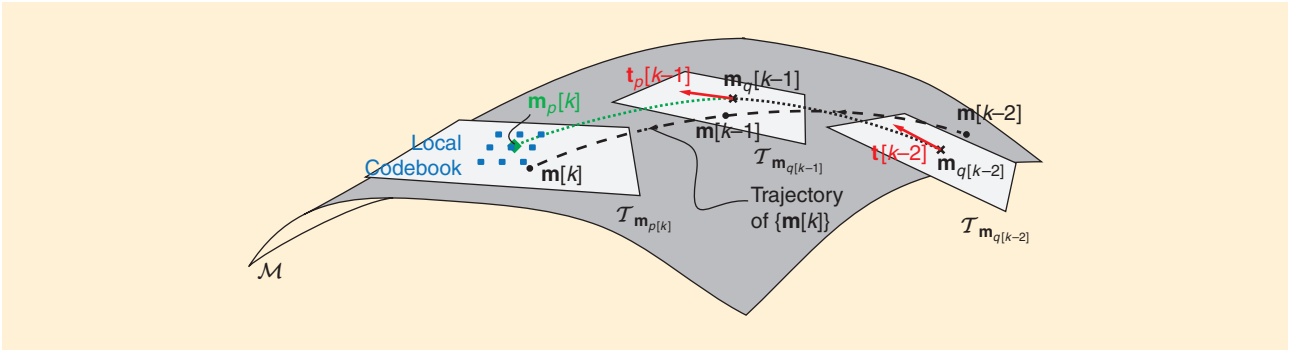


FIGURE 6. An illustration of predictive manifold quantization of a temporally correlated process $\{m[k]\}$ on the manifold $\mathcal{M}$.

manifold quantization [17]–[19]. Currently, the best rate-distortion performance is achieved by adaptive quantizers that adjust the applied quantization codebook to the temporal evolution of the channel. We summarize this adaptive quantization principle, as introduced for the Grassmannian in [12], for general Riemannian manifolds $\mathcal{M}$ and highlight the most important differential geometric concepts required. For this purpose, we consider a general temporally correlated process of points $\{\mathbf{m}[k]\} \in \mathcal{M}$ that is to be quantized; this process $\{\mathbf{m}[k]\}$ evolves on the manifold $\mathcal{M}$ as illustrated in Figure 6. The basic idea of predictive quantization is to utilize previously quantized observations $\mathbf{m}_q[k-1] = Q\{\mathbf{m}[k-1]\}, \mathbf{m}_q[k-2] = Q\{\mathbf{m}[k-2]\}, \ldots$to predict the current point $\mathbf{m}[k]$, where $Q\{\cdot\}$ denotes the employed quantization function. Given the prediction $\mathbf{m}_p[k]$, the method then generates a local quantization codebook to determine $\mathbf{m}_q[k] = Q\{\mathbf{m}[k]\}$. Manifold prediction and codebook adaptation are based on quantized CSI, since the decoder at the back end of the feedback link must be able to reproduce these steps. Realizing such a scheme on a manifold requires some additional intermediate actions as detailed next.

First, we notice that a linear prediction, as in the section "Implicit CSI Feedback," is not meaningful on a general manifold, since an addition of points on the manifold or scalar multiplication with values from an underlying field, as we have on a linear vector space, is in general not defined. Yet, to each point $m \in \mathcal{M}$, an entire linear vector space, the tangent space $\mathcal{T}_{\mathbf{m}}$, is associated, which locally represents the geometry of the manifold in the Euclidean space. Tangent vectors $\mathbf{t} \in \mathcal{T}_{\mathbf{m}}$ are induced by curves $\gamma(t) \in \mathcal{M}$, such that,

$$\mathbf{t} = \frac{\partial}{\partial t}\gamma(t)\Big|_{t=0}, \quad \gamma(0) = \mathbf{m} \tag{3}$$

with scalar $t \in \mathbb{R}$ parameterizing the curve; we assume without loss of generality that the curve traverses $\mathbf{m}$ at $t = 0$. In general, multiple curves can induce the same tangent, since it only depends on the first-order derivative.

For our manifold predictor, we need a one-to-one relationship between pairs of points $\mathbf{m}_1, \mathbf{m}_2$ on the manifold and tangent vectors $\mathbf{t} \in \mathcal{T}_{\mathbf{m}_1}$, such that,

$$\mathbf{t} = \lambda(\mathbf{m}_1, \mathbf{m}_2) \in \mathcal{T}_{\mathbf{m}_1}, \quad \mathbf{m}_2 = \rho(\mathbf{m}_1, \mathbf{t}) \in \mathcal{M}, \tag{4}$$

$$\mathbf{m}_2 = \rho(\mathbf{m}_1, \lambda(\mathbf{m}_1, \mathbf{m}_2)), \forall \mathbf{m}_1, \mathbf{m}_2 \in \mathcal{M}. \tag{5}$$

This is known as a *compatible lifting-retraction pair*, with $\lambda(\cdot,\cdot)$ being the lifting map and $\rho(\cdot,\cdot)$ the retraction map [18]. Ideally, we favor the exponential and logarithmic maps associated with the geodesic curve between $\mathbf{m}_1, \mathbf{m}_2$ for this purpose; yet, these are difficult to evaluate for certain cases, such as the Stiefel manifold. The geodesic represents the shortest path between $\mathbf{m}_1, \mathbf{m}_2$ on the Riemannian manifold.

Given a lifting-retraction pair, we can perform prediction in the linear tangent space, where we can reuse well-known linear algorithms, such as linear prediction and MMSE filtering, and translate the tangent prediction onto the manifold using the retraction. That is, given a set of previous observations $\{\mathbf{m}_q[k-1], \ldots, \mathbf{m}_q[k-L-1]\} \in \mathcal{M}$, we calculate

$$\mathbf{t}_i = \lambda(\mathbf{m}_q[k-1], \mathbf{m}_q[k-1-i]) \in \mathcal{T}_{\mathbf{m}_q[k-1]}, \quad \forall i \in \{1, \ldots, L\}, \tag{6}$$

$$\mathbf{t}_p[k-1] = \mathcal{P}(\mathbf{t}_1, \ldots, \mathbf{t}_L) \in \mathcal{T}_{\mathbf{m}_q[k-1]}, \tag{7}$$

$$\mathbf{m}_p[k] = \rho(\mathbf{m}_q[k-1], \mathbf{t}_p[k-1]) \in \mathcal{M}, \tag{8}$$

with $\mathcal{P}(\cdot)$ denoting the applied prediction function. Finally, we need to construct a local quantization codebook around $\mathbf{m}_p[k]$ for the quantization of $\mathbf{m}[k]$. This can again be achieved most easily by generating a codebook for the Euclidean tangent space and translating it onto the manifold; in [16] and [18], we provide corresponding codebook constructions for the Grassmannian and the Stiefel manifold, respectively.

In Figure 7(a), we demonstrate the performance of such predictive quantization on the Stiefel and Grassmann manifolds in dependence of the normalized sampling interval $\tau_s = t_s f_d$, with t_s being the time in between two samples of the manifold process $\{\mathbf{m}[k]\}$. Similar to the normalized feedback delay introduced previously, the normalized sampling interval accounts for the temporal variation of consecutive points of the manifold process $\{\mathbf{m}[k]\}$ due to movement. We assume $N_t \times N_r = 6 \times 2$ and employ a quantization codebook of size 256; i.e., the quantization resolution is

8 bits/sample. The performance of the quantization scheme is gauged in terms of the signal distortion introduced by the quantization process. We employ the chordal distance on the corresponding manifold as distortion metric; see [18] for details. In Figure 7(a), we compare the quantization distortion of predictive quantization to memoryless and differential quantization. Memoryless quantization refers to a quantization scheme that considers each sample $\mathbf{m}[k]$ individually without accounting for the past of the process $\{\mathbf{m}[k]\}$. Differential quantization refers to the case $\mathbf{m}_p[k] = \mathbf{m}_q[k-1]$.

Multicasting plays an important role in vehicular communications, since it enables efficient sharing of common information, such as traffic management advices and vehicle status updates, among vehicles.

Both differential and predictive quantization achieve an improvement over memoryless quantization for $\tau_s \leq 0.1$; at large τ_s, the distortion of differential/predictive quantization saturates at the performance of memoryless quantization. Hence, the considered differential/predictive quantization schemes achieve a significant gain only at low mobility; e.g., assuming $t_s = 1$ ms, $f_c = 800$ MHz, the point $\tau_s = 0.01$ corresponds to $v = 13.5$ km/h. Since CSI feedback in LTE is currently at most foreseen once every subframe (once every millisecond), this implies that currently available differential/predictive manifold quantization schemes are only advantageous at low mobility and not in vehicular scenarios. Performance can be improved by extending the codebook size; however, complexity issues will ultimately put limits on the supported size.

In Figure 7(b), we show the throughput of BD and RBD precoding when applying predictive Grassmannian and Stiefel manifold quantization to obtain CSIT. We show the performance relative to RBD with perfect CSIT. Imperfect CSIT due to quantization distortion causes residual multiuser interference, degrading the achievable throughput. We observe a significantly better performance of RBD at low signal-to-noise ratio (SNR), since this method implicitly reduces the number of served users with decreasing SNR via power allocation; BD precoding, on the other hand, always serves three users in parallel each with two spatial streams. At high SNR, BD performs slightly better due to the lower quantization distortion of its Grassmannian quantizer, causing less residual multiuser interference as compared to RBD with Stiefel manifold quantization. Notice that, in this simulation, we assume CSI feedback every millisecond and negligible feedback delay between the users and the base station. In the considered low-mobility scenario with $f_d = 10$ Hz, this is, however, not a significant restriction, because a reasonable feedback delay can be compensated via channel prediction at the receiver; see [12] for a breakdown of the contributions of quantization and delay compensation onto the overall distortion.

At high mobility, the presented CSI feedback method fails to achieve significant multiuser MIMO gains due to insufficient accuracy of CSIT. To support high-mobility scenarios, it would be necessary to reduce the sampling and feedback interval τ_s at least by a factor of ten. Additional gains might also be possible by improving the prediction function in (7); in our simulations, we employed linear prediction in combination with LMS to optimize filter coefficients. Alternatively, precoding schemes that are robust with respect to outdated CSIT, such as retrospective interference alignment, can be utilized. In general, though, such schemes require coding over many

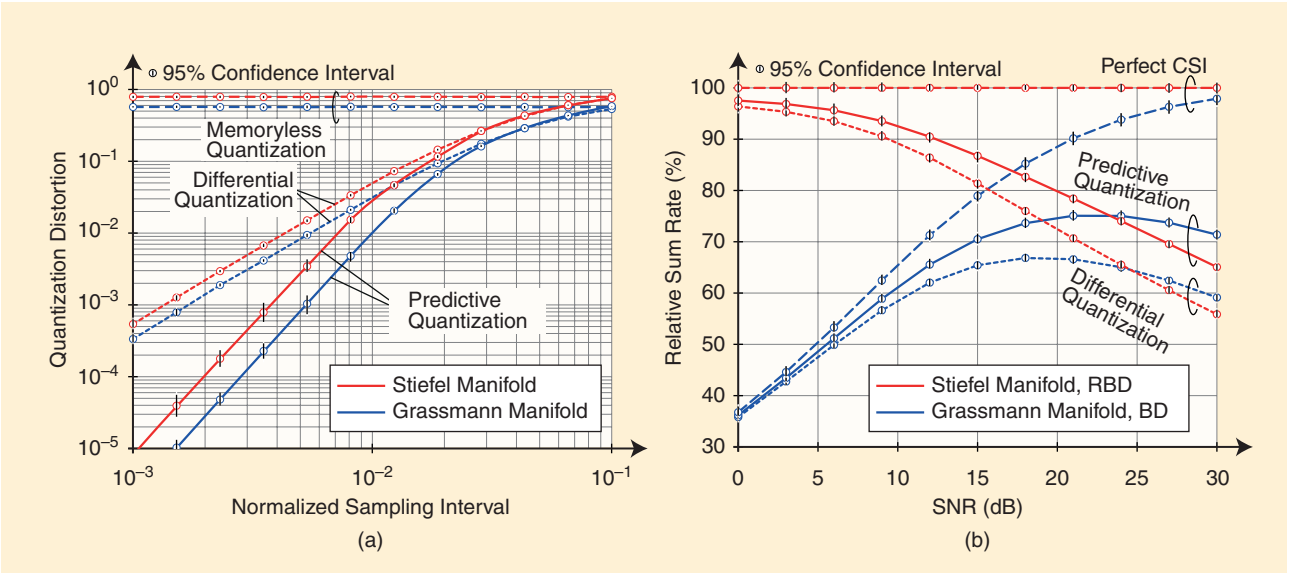


FIGURE 7. The quantization error of (a) manifold quantizers and (b) throughput comparison of RBD and BD with limited feedback at normalized sampling interval $\tau_s = 0.01$ and 8-b quantization codebook. We consider antenna arrays of size $N_t \times N_r = 6 \times 2$.

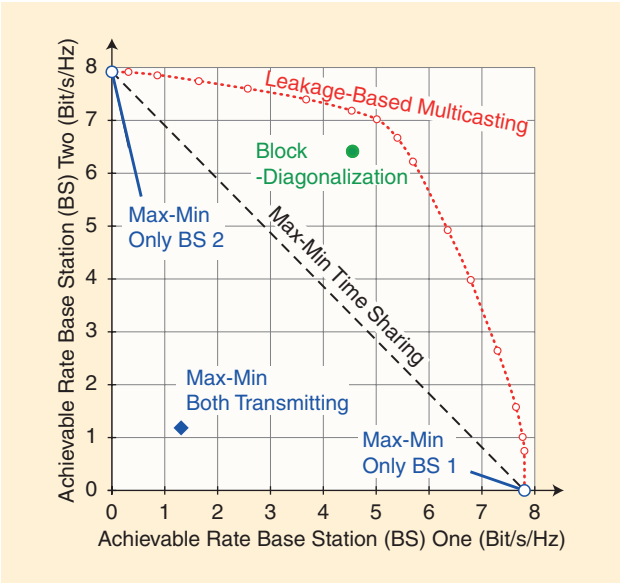


FIGURE 8. The achievable rate-region of the multicast interference channel with two base stations each multicasting common information to six users over $N_t = 8$ transmit antennas. Without beamforming, the SINR is 0 dB.

time frames for efficient operation, which may not be tolerable in terms of latency.

Multicast enhancements

Multicasting plays an important role in vehicular communications, since it enables efficient sharing of common information, such as traffic management advices and vehicle status updates, among vehicles. As we have seen in the section "MBSFN-Based V2V Communication," though, multicasting of CAMs can easily overload cellular networks, leading to unreliable operation of C-ITS. Next, we discuss improvements of the LTE standard to enhance the efficiency of cellular multicasting. We first consider multicasting-specific MIMO beamforming techniques, which will especially be relevant in future 5G cellular networks employing large-scale full-dimension (FD) MIMO arrays. We then discuss scheduling and resource allocation issues associated to multicast users in LTE networks.

MIMO and coordination

Currently, data transmission in LTE's eMBMS does not support multiantenna techniques. Already, early investigations on MIMO multicast-beamforming in single-cell networks, however, demonstrate promising capacity gains through max-min beamforming [20]. This approach attempts to maximize the minimum SNR of the multicast users without accounting for interference caused to other users. Even this seemingly simple problem, however, turns out to be nonconvex, and requires semidefinite relaxation (SDR) to determine, in general, suboptimal solutions; only for a small number of transmit antennas and users SDR provides a globally optimal beamforming solution.

Max-min beamforming becomes even more intricate when interference between multiple transmitters is considered, with each of them transmitting information to a different set of multicast users. This scenario is known as the *multicast interference channel*. In cellular networks, it represents a situation where several neighboring base stations serve disjoint sets of multicast users; in vehicular environments, it can come up along highways, where consecutive stretches of the road are served by multiple transmitters. In [21], an interference-leakage-based approach for coordinated beamformer optimization in the multiple-input, single-output multicast interference channel is proposed, where only transmitters are equipped with multiple antennas and not receivers. The method applies max-min beamforming optimization at each base station individually, while restricting the interference leakage caused to users of other base stations. Additionally, it employs an iterative exchange of interference-leakage parameters among base stations in combination with dual gradient optimization over leakage parameters to determine a locally optimal operating point of the multicast interference network. Figure 8 demonstrates the performance of this approach in a multicast interference network with two transmitters each serving six users over $N_t = 8$ transmit antennas. We consider an SNR of 20 dB and assume that, without beamforming, all users receive both transmitters, on average, equally strong; i.e., the macroscopic channel gain with respect to both transmitters is equal. This corresponds to a cell-edge situation where we can expect the largest gains from transmitter coordination. Notice, though, that the method itself is not restricted to equal gain scenarios. Figure 8 shows that, compared to max-min beamforming without considering interference, the rate region can substantially be extended. Hence, the method can significantly improve the cell-edge performance, thereby reducing connectivity failures due to extensive intercell interference. In general, we cannot claim that the method achieves the maximal rate region, since it is only guaranteed to converge to a local optimum; for the shown result, however, we have confirmed global optimality through monotonic optimization utilizing the outer-polyblock algorithm. In the considered high-SNR situation, BD precoding also performs reasonably well. If users are uniformly distributed in the network, the gain over time sharing diminishes with a growing number of users, since the optimal beam pattern tends to be isotropic. Yet, in practical situations, such as vehicles on a road, users are often spatially clustered, which may help mitigating such performance degradation. Early results to support multiple antennas at transmitters and receivers are reported in [22], leading to an iterative alternating optimization of transmit and receive filters. Extensions from beamforming to multiple data stream transmission are yet to be developed.

Beamformer designs for joint multicasting within MBSFN areas formed by multiple base stations, which account for interarea-interference in between neighboring MBSFN areas, are not available so far. Compared to the leakage-based multicasting scheme of [21], this implies additional per base station power constraints for the joint MBSFN beamformer

$$\max_{\mathbf{f}\in\mathbb{C}^{N_{\text{tot}}\times 1}} \min_{u\in U_{\text{MBSFN}}} \text{SINR}_u \tag{9}$$

subject to:

$$\|\mathbf{S}_b\mathbf{f}\|_2^2 \le P_b, \forall b \in B_{\text{MBSFN}}$$
$$\text{Leak}_{\bar{u}} \le L_{\bar{u}}, \forall \bar{u} \in \bar{U}_{\text{MBSFN}},$$

with f denoting the joint beamformer applied over all $N_{\text{tot}} = \sum_{b\in B_{\text{MBSFN}}} N_{t,b}$ transmit antennas $N_{t,b}$ of the set of B_{MBSFN} base stations within the MBSFN area. The parameter SINR_u represents the SINR of user u from the set U_{MBSFN} of users served via multicasting in the MBSFN area and $\text{Leak}_{\bar{u}}$ denotes the interference leakage caused to user $\bar{u}$ from the set $\bar{U}_{\text{MBSFN}}$ of users outside the MBSFN area; it is upper bounded by the leakage constraint $L_{\bar{u}}$. Finally, $\mathbf{S}_b$ represents a selector matrix, which chooses the entries of $\mathbf{f}$ corresponding to the transmit antennas of base station b, and P_b is the power constraint of base station b. This problem is nonconvex but, similar to [20], is amenable to SDR. The operating point of this multicast interference network with spatially distributed multicast transmitters can be optimized by coordinating the interference-leakage constraints of interfering base stations. Such coordination can be achieved by applying the dual gradient optimization proposed in [21].

Notice that the discussed beamforming approaches are all based on explicit CSIT, and, hence, the CSI quantization problematic discussed in the section "CSI Feedback Enhancement" applies to these methods as well. Furthermore, since the methods are iterative in nature, they exhibit intrinsic delay to achieve convergence. In vehicular environments, it might be necessary to employ suboptimal heuristics that are derived from optimal schemes; such heuristics are also provided in [21]. Also, especially for road safety applications, outage probability and latency minimization might be of higher importance than rate maximization. Optimization problems for such targets, however, are hard to formulate due to potential nonstationarity of wireless channels in vehicular environments [23]. Still, for limited time frames, assuming quasi-stationary conditions may be valid to formulate outage-constrained optimization problems.

Scheduling and resource allocation

A major weakness of eMBMS in the LTE standard is its inflexibility of multicast transmission; i.e., a fixed amount of subframes within each radio frame is reserved for multicasting. One reason for this is that multicast transmissions within MBSFN areas suffer from a larger delay spread due to single-frequency transmission of the same signal from multiple spatially distributed base stations. To compensate for the increased delay spread, multicast subframes employ the extended cyclic prefix (16.7μs) of LTE, whereas unicast subframes mostly use the shorter normal cyclic prefix (4.7μs) to minimize waste of bandwidth. According to our work on link adaptation, however, utilizing the extended cyclic prefix may not pay off, even if intersymbol interference (ISI) occurs; in many cases of realistic SNR, it is more efficient to simply accept the remaining ISI and use a more robust transmission rate to compensate for it [24]. Hence, the LTE standard should be extended to support multicasting in MBSFN areas with a short cyclic prefix. This then also allows the incorporation of multicast transmissions into normal subframes, facilitating a better dynamic resource assignment according to traffic requirements, as well as the exploitation of channel diversity in multicasting. Within the 3GPP, such extensions are considered in the study item on single-cell point-to-multipoint transmission; within this study, single-cell multicasting over the downlink shared channel is evaluated, which until now has only handled unicast data transmissions.

A big challenge from a signal processing perspective is efficient scheduling and coordination of multicast transmissions. In the context of vehicular communications, this involves dynamic formation of MBSFN areas, to limit the amount of CAMs that must be distributed while still providing mutual awareness over sufficiently large geographic areas. Furthermore, it implies determining groups of multicasting users that must share CAMs due to close spatial proximity. The multicast scheduler should also be able to decide to offload users with very poor channel quality to unicast transmissions, to support selective PHY retransmissions via HARQ, such as to reduce latency and to improve reliability. First results in this direction are provided in [25], where the authors consider optimizing proportional fairness of concurrent multicast and unicast transmissions. Heterogeneity of user channel conditions is considered in [25] by optimally partitioning multicast users into groups, so that users with good signal strength do not suffer by being grouped together with users of poor signal strength. A further challenge comes up when the system supports direct V2V transmission in addition to cellular-assisted transmission; then the scheduler must further group vehicles according to direct and assisted transmission.

Further research topics to enhance wireless vehicular communications

Channel estimation at high mobility

At very high mobility, accurate CSIR estimation using pilot signals can become challenging due to strong temporal variation of the wireless channel. Prior investigations have shown that LTE does not achieve MIMO spatial multiplexing gains at high mobility due to insufficient density of pilot symbols in the time domain [26]. This low density of pilot symbols can cause poor accuracy of channel estimation and, as a consequence, unreliable symbol detection. A remedy can be provided by adapting the pilot pattern to the Doppler- and delay-spread of the channel. These two properties characterize the coherence time and bandwidth of the channel, respectively, i.e., the time/frequency intervals over which the channel stays approximately constant. Such an adaptive pilot scheme can efficiently be realized with minimal feedback information about coherence time and bandwidth from the users, requiring an update whenever channel statistics vary significantly [26].

In vehicular environments, especially on highways, the wireless channel behaves markedly different as compared to other common situations in mobile communications. More specifically, the propagation is characterized by shadowing through other vehicles, high Doppler shifts with often sparse Doppler spectrum due to few dominant scatterers (e.g., road signs, highway overpasses), and possibly inherent nonstationarity of the channel statistics. Such effects can cause substantial performance degradation of common least squares and MMSE channel estimators if they are not properly considered in the algorithmic design. In [23], the impact of the shape of the Doppler spectrum on the performance of channel estimation at the receiver is discussed in more detail. Similarly, timing and carrier synchronization in vehicular scenarios with few dominant scatterers of similar strength that are strongly delay- and Doppler-shifted with respect to each other can be challenging.

A big challenge from a signal processing perspective is efficient scheduling and coordination of multicast transmissions.

ICI mitigation

OFDM multicarrier modulation, as employed in LTE, suffers at high mobility from intercarrier interference (ICI) due to the Doppler spread of the transmit signal. This effect, however, can, to a large extent, be mitigated through iterative ICI estimation and cancellation at the receivers. In [27], the authors propose ICI mitigation algorithms that enable achieving the performance of interference-free transmission. Alternatively, optimal pulse shaping at the transmitter, as currently pushed by many research groups and companies for 5G mobile communications, can be applied to maximize achievable data rates by trading off residual ICI/ISI for spectral efficiency [28]. Such an approach appears especially interesting when waveform parameters, such as subcarrier spacing, prototype pulse shape, and TTI length, can be adapted to the time/frequency-dispersion characteristics of the channel. A major challenge then is to optimize sets of compatible parameters that enable efficiently serving users with strongly different channel properties in parallel, e.g., static and highly mobile users. Novel multicarrier transmit waveforms, employing filter banks [filter-bank multicarrier modulation (FBMC)] or subband filters (universal filtered multicarrier/OFDM), support the necessary flexibility to adjust waveform parameters over subbands and, thus, provide the basis for channel-adaptive modulation.

Multiconnectivity

To enhance the robustness of the wireless transmission link, so as to support highly reliable communication, macro diversity can be exploited by extending the dual connectivity concept of LTE to a multitude of radio network access points. Maintaining multiple parallel connections to several macro base stations and/or small cells promises increased data throughput over multiple parallel data streams, improved reliability due to a reduced outage probability and enhanced robustness with respect to mobility, since hard handovers can be avoided. Yet, it implies that single users occupy resources at multiple transmission points, causing a reduction of the overall network capacity. To maintain the efficiency of the network, it is therefore necessary to dynamically decide whether the gain of utilizing multiconnectivity for a user outweighs the overhead caused on network capacity. Hence, multiconnectivity comes at the cost of requiring a sophisticated coordination of multiple transmission points, potentially implying a substantial backhaul signaling overhead. Such enhanced self-organizing network features can, e.g., be realized by cloud radio access network architectures. In [29], the authors investigate the tradeoff between radio-link failures and user throughput in dependence of the size of the set of active multiconnectivity transmission points, showing that both the cell-edge user throughput as well as the mobility performance can be improved simultaneously.

FD MIMO beamforming for high mobility

FD MIMO refers to wireless transmission systems that support active two-dimensional antenna arrays with a large number of antenna elements. This enables high-resolution adaptive beamforming in both the elevation and the azimuth domain, to achieve space-division multiple access gains through spatial separation of users. Within LTE standardization, work is ongoing to implement FD MIMO within Release 14. Currently, hybrid transceiver architectures are of interest, where part of the signal processing is performed in base band and part in the analog domain to limit the number of required radio-frequency chains. Analog beamforming approaches are mostly based on signal azimuth and elevation arrival/departure angles. To implement such schemes at high mobility, it is necessary to account for uncertainty in the signal arrival/departure angles due to user movement and estimation errors. In [30], the authors propose a corresponding robust beamformer optimization problem and demonstrate improved robustness at high mobility.

Conclusions

Vehicular communications is an integral part of innovative transport telematics systems for traffic management and active road safety. It plays a key role in making public and private transportation faster, more reliable, more efficient, and safer. Realizing the necessary information exchange among roadside infrastructure and vehicles efficiently and reliably can be challenging. In that respect, cellular networks can provide valuable support to dedicated vehicular communication systems, since today's cellular base stations are almost ubiquitously accessible and supply high bandwidth wireless connectivity. In this article, we have surveyed ongoing efforts and developments within the 3GPP to implement vehicular communications over LTE. We have discussed research challenges associated with wireless communications at high mobility, and we have provided an overview of promising signal processing techniques to tackle important hurdles. Even though significant progress in enhancing

wireless vehicular communications has been made in the past, there is room left for improvement.

Acknowledgments

The financial support by the Austrian Federal Ministry of Science, Research, and Economy and the National Foundation for Research, Technology, and Development is gratefully acknowledged. The research has been cofinanced by the Czech Science Foundation, Project Number 17-18675S "Future Tranceiver Techniques for the Society in Motion," and by the Czech Ministry of Education in the frame of the National Sustainability Program under grant LO1401.

Authors

Stefan Schwarz (sschwarz@nt.tuwien.ac.at) received his Dipl.-Ing. degree in electrical engineering in 2009 from the Vienna University of Technology (TU Wien). He received his Dr. techn. degree in telecommunications engineering in 2013 from TU Wien. He has a postdoctoral position at the Institute of Telecommunications of TU Wien, heading the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion.

Tal Philosof (tal.philosof@gm.com) received the M.Sc. and Ph.D. degrees from Tel-Aviv University, Israel, in 2003 and 2010, respectively. During 2004–2009, he served as a research fellow in several academia-industry scientific programs supported by the Israel Ministry of Industry. He is currently with the Wireless Enablers Lab, General-Motors R&D Advanced Technical Center Israel.

Markus Rupp (mrupp@nt.tuwien.ac.at) received his Dipl.-Ing. degree in 1988 from the University of Saarbrücken, Germany, and his Dr.-Ing. degree in 1993 at Technische Universität Darmstadt, Germany. Until 1995, he had a postdoctoral position at the University of Santa Barbara, California. From 1995 to 2001, he was with the Wireless Technology Research Department of Bell-Labs, New Jersey. Since October 2001, he has been a full professor of digital signal processing in mobile communications at TU Wien.

References

[1] G. Araniti, C. Campolo, M. Condoluci, A. Iera, and A. Molinaro, "LTE for vehicular networking: A survey," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 148–157, May 2013.

[2] H. Seo, K. D. Lee, S. Yasukawa, Y. Peng, and P. Sartori, "LTE evolution for vehicle-to-everything services," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 22–28, June 2016.

[3] Y. L. Tseng, "LTE-advanced enhancement for vehicular communication," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 4–7, Dec. 2015.

[4] Y. Zhou, Z. Lei, and S. H. Wong, "Evaluation of mobility performance in 3GPP heterogeneous networks," in *Proc. IEEE 79th Vehicular Technology Conf.*, May 2014, pp. 1–5.

[5] S. Jha, K. Sivanesan, R. Vannithamby, and A. Koc, "Dual connectivity in LTE small cell networks," in *Proc. Globecom Workshops*, Dec. 2014, pp. 1205–1210.

[6] M. Rupp, S. Schwarz, and M. Taranetz, *The Vienna LTE-Advanced Simulators: Up and Downlink, Link and System Level Simulation*. New York: Springer-Verlag, 2016.

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

[8] T. Rappaport, "Wireless communications: Principles and practice," in *Prentice Hall Communications Engineering and Emerging Technologies Series*. Upper Saddle River, NJ: Dorling Kindersley, 2009.

[9] L. Shi and K. W. Sung, "Spectrum requirement for vehicle-to-vehicle communication for traffic safety," in *Proc. IEEE 79th Vehicular Technology Conf.*, May 2014, pp. 1–5.

[10] I. Safiulin, S. Schwarz, and M. Rupp, "System level simulation of LTE MBSFN networks with high mobility users," in *Proc. 22nd Int. Conf. Systems, Signals and Image Processing*, London, 2015, pp. 21–24.

[11] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, and T. Weil, "Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions," *IEEE Commun. Surv. Tutor.*, vol. 13, no. 4, pp. 584–616, 2011.

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

[13] T. Zemen, C. F. Mecklenbrauker, and B. H. Fleury, "Time-variant channel prediction using time-concentrated and band-limited sequences," in *Proc. IEEE Int. Conf. Communications*, June 2006, vol. 12, pp. 5660–5665.

[14] A. Heidari, A. K. Khandani, and D. Mcavoy, "Adaptive modelling and long-range prediction of mobile fading channels," *IET Commun.*, vol. 4, no. 1, pp. 39–50, Jan. 2010.

[15] J. Park, G. Lee, Y. Sung, and M. Yukawa, "Coordinated beamforming with relaxed zero forcing: The sequential orthogonal projection combining method and rate control," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3100–3112, June 2013.

[16] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems," in *Proc. 38th Int. Conf. Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013, pp. 5021–5025.

[17] D. Sacristan-Murga and A. Pascual-Iserte, "Differential feedback of MIMO channel Gram matrices based on geodesic curves," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3714–3727, 2010.

[18] S. Schwarz and M. Rupp, "Predictive quantization on the Stiefel manifold," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 234–238, 2015.

[19] R. Pitaval and O. Tirkkonen, "Joint Grassmann-Stiefel quantization for MIMO product codebooks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 210–222, 2014.

[20] N. Sidiropoulos, T. Davidson, and Z. Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, June 2006.

[21] S. Schwarz and M. Rupp, "Transmit optimization for the MISO multicast interference channel," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4936– 4949, 2015.

[22] S. Schwarz, T. Philosof, and M. Rupp, "Leakage-based multicast transmit beamforming," in *Proc. Int. Conf. Communications*, London, U.K., June 2015, pp. 2405–2410.

[23] C. Mecklenbrauker, A. Molisch, J. Karedal, F. Tufvesson, A. Paier, L. Bernado, T. Zemen, O. Klemp, and N. Czink, "Vehicular channel characterization and its implications for wireless system design and performance," *Proc. IEEE*, vol. 99, no. 7, pp. 1189–1212, July 2011.

[24] E. Zöchmann, S. Pratschner, S. Schwarz, and M. Rupp, "Limited feedback in OFDM systems for combating ISI/ICI caused by insufficient cyclic prefix length," in *Proc. 48th Asilomar Conf. Signals, Systems and Computers*, Asilomar, CA, Nov. 2014, pp. 1–5.

[25] J. Chen, M. Chiang, J. Erman, G. Li, K. K. Ramakrishnan, and R. K. Sinha, "Fair and optimal resource allocation for LTE multicast (eMBMS): Group partitioning and dynamics," in *Proc. IEEE Conf. Computer Communications*, Apr. 2015, pp. 1266–1274.

[26] M. Simko, P. Diniz, Q. Wang, and M. Rupp, "Adaptive pilot-symbol patterns for MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4705–4715, Sept. 2013.

[27] R. Nissel and M. Rupp, "Doubly-selective MMSE channel estimation and ICI mitigation for OFDM systems," in *Proc. IEEE Int. Conf. Communications*, London, U.K., June 2015, pp. 6295–6300.

[28] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 80–93, Nov. 2014.

[29] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *Proc. IEEE Globecom Workshops*, Dec. 2015, pp. 1–6.

[30] S. Schwarz and M. Rupp, "Limited feedback based double-sided full-dimension MIMO for mobile backhauling," in *Proc. 50th Conf. Signals Systems and Computers*, Asilomar, CA, 2016.

[31] 3GPP. "Technical specification group services and system aspects; Study on LTE support for V2X services (Release 14). [Online]. Available: http://www.3gpp.org/DynaReport/22885.htm

[32] 3GPP. "Technical specification group radio access network; Study on LTE-based V2X services (Release 14). [Online]. Available: http://www.3gpp.org/DynaReport/36885.htm

SP

2. Link and System Level Modeling and Simulation

Mobile communications specifications become increasingly complex, in order to be able to cope with the challenging requirements imposed on the achievable capacity, throughput, latency and the reliability. This implies that purely mathematical analysis as well as measurement based investigations and field trials of link and system level techniques frequently encounter feasibility limitations, in terms of their cost, time consumption and tractability. Analytic investigations have the great potential to mathematically reveal relationships amongst key parameters; yet, often analytic tractability requires the application of coarse simplifications and restrictive assumptions, limiting analysis to information-theoretic values, such as the achievable transmission rate in terms of the mutual information. However, practical mobile communication systems do not achieve the information-theoretic mutual information; in fact they are mostly still far off [8, 9]. Therefore, realistic performance investigations of novel link and system level techniques need to account for the holistic performance of the entire transceiver chains and, even more importantly, for the complex interactions amongst network nodes. This entails that computer simulations are indispensable to validate or disprove analytic insight under realistic conditions.

Chapter Outline

In this chapter, I introduce the basic base band system model and notation employed throughout this thesis and I discuss the simulation methodology utilized in my research work to complement mathematical analysis. Our mobile communications research group has a long and successful history of developing standard-compliant cellular communications simulators [10] and sharing the program code with fellow researchers, with the goal of enhancing reproducibility in wireless communications research [11]. These tools are well received by the scientific community and are employed by a number of 3GPP-affiliated companies to benchmark their proposals.

2.1. Challenges in the Implementation of Mobile Communications Simulators

Implementing an entire communication standard in software implies a very large amount of program code. Our Vienna Cellular Communications Simulators (VCCS) [10], for example, contain more than a hundred-thousand lines of code implemented over years by dozens of PhD and Master-level students. Managing such a platform therefore comes with a number of challenges:

1. Maintaining readability, usability and modifiability of the code.
2. Verifying the functionality and validating the correctness.
3. Supporting replicability, reproducibility and reusability.

Code maintenance: Achieving the first bullet mostly requires adhering to well-established disciplined programming practices. This includes consistency in style of variable-, function- and object-definitions, proper commenting of program code, modular and object-oriented structuring of processing steps, well-defined interfaces in-between functions and objects, as well as, accompanying documentation of the program structure and the functionality. Especially object-oriented structuring is crucial to facilitate easy modifiability by newcomers, since it enables focusing on specifics parts of interest of the program code, without having to comprehend the entire simulation environment.

Verification and validation: Verifying the functionality and validating the correctness of the implementation is a difficult task, considering the large amount of program code involved. Here an incremental approach is expedient, starting with verification and validation of individual processing blocks (e.g. channel coding, modulation, equalization) and proceeding to the interaction of these building blocks, for which, once-again, an object-oriented structure is indispensable. Often, theoretical results for the performance of these individual blocks are available, which can be utilized as benchmarks.[1] Yet, this is not always possible, especially when it comes to more complex program blocks. In our VCCS, for example, we support the sophisticated stochastic-geometric 3GPP three dimensional (3D) spatial channel model (SCM) [12], for which no theoretical benchmarks exist. Then it is instrumental to compare against numerical evaluations conducted by other independent groups. This approach is also pursued by standardization bodies, such as 3GPP, which commonly provide

[1] As an example, it is well-known how the bit error probability (BEP) of uncoded OFDM transmission behaves as a function of the signal to noise ratio (SNR) for additive white Gaussian noise (AWGN) and Rayleigh fading channels.

online documents comparing numerical evaluations of several groups that can be utilized for validation. However, even with the utmost rigor, small mistakes can never be entirely avoided; to eliminate these, we provide the source of our program code to an online community of fellow engineers and researchers, which willingly utilize our simulator for their work and hint to problems they thereby encounter. With this approach, we get feedback from a much larger group of testers, as compared to limiting code distribution within our own research group, and we can thereby enhance the code integrity.

Replicability, reproducibility and reusability: One of the main goals of sharing our simulators with other researchers is to establish a common platform to support replicability, reproducibility and even reusability in wireless communications research [11].[2] This is often difficult, since the performance of novel techniques within wireless communication systems is impacted by a large number of parameters, which are in many cases not all stated in publications, already for reasons of space constraints. Our simulators support achieving these reproducibility targets, by providing a platform on which researchers can benchmark and exchange their algorithms in a well-defined (standard-complaint if so desired) environment. To further facilitate this exchange, we additionally administrate a web-forum where people can easily communicate their issues and find solutions to common pitfalls, which also promotes the first discussed challenge of maintaining usability.

2.2. Link Level Modeling of MIMO-OFDMA Transmissions

Many state-of-the-art wireless communication systems, such as, 3GPP 4G long term evolution (LTE) and 5G new radio (NR) [14, 15] as well as Institute of Electrical and Electronic Engineers (IEEE) 802.11 [16], employ cyclic prefix orthogonal frequency division multiplexing (CP-OFDM) as the basic physical layer (PHY) modulation scheme. Below, I present the mathematical base band description of such systems, which builds the basis of link level simulations and provides the underlying system model for later discussions in this thesis.

In link level simulations, the entire base band signal processing chain is implemented in software. Hence, signals are processed on a sample basis, allowing to realistically represent channel introduced distortions, such as, inter-carrier interference (ICI) and inter-symbol interference (ISI), without requiring approximations or abstractions. Link level simulations

[2] I follow the terminology of [13] – replicability: other researchers reproduce results with the same experimental setup – reproducibility: other researchers reproduce results with a different setup – reusability: repurposing of research results is facilitated.

thus provide a very low level of abstraction and are therefore well suited for rapid prototyping of novel PHY signal processing methods.

2.2.1. Link Level Base Band Description

I focus here on MIMO-OFDMA multi-carrier transmissions, because the techniques presented in later sections of this thesis are benchmarked within LTE-compliant simulation environments. Notice, though, that we also discuss similar base band descriptions for other multi-carrier schemes, such as, filter bank multi-carrier (FBMC), single carrier frequency division multiple access (SCFDMA) and universal filtered multi-carrier (UFMC), in our publications [17–19].

SISO-OFDM: The single-input single-output (SISO) input-output relationship of CP-OFDM considering transmission between a transmitter i and a receiver u, both equipped with a single antenna, can be stated as [10]:

$$r_u[n,k] = H_{u,i}[n,k]x_i[n,k] + z_u[n,k]. \tag{2.1}$$

Here, $[n,k]$ denotes the pair of time-frequency indices, also known as the resource element (RE) index. The channel coefficient, describing the CP-OFDM transmission between transmitter i and receiver u on RE $[n,k]$ in equivalent base band notation, is represented by the complex-valued scalar $H_{u,i}[n,k] \in \mathbb{C}$. The complex-valued scalar $z_u[n,k]$ comprises all additive distortions present in the transmission, such as: 1) thermal receiver noise, commonly modeled as complex-Gaussian random variable; 2) ISI, caused by excess delay of the channel power delay profile (PDP) $\tilde{h}_{u,i}(t,\tau)$ over the CP duration [20]; 3) ICI, due to the Doppler-spread introduced by the time-variant channel [21–23]; and others. If not stated otherwise, I assume that $z_u[n,k]$ can be represented as AWGN of variance $\sigma^2_{z,u}$.

The channel coefficient $H_{u,i}[n,k]$ is obtained by sampling the time-variant channel transfer function $H_{u,i}(t,f)$, which characterizes the physical transmission channel at time t and frequency f, on a time-frequency lattice with time-spacing $\Delta t = \frac{1}{B} = \frac{T}{N_{\text{FFT}}}$ and frequency-spacing $\Delta f = \frac{1}{T} = \frac{B}{N_{\text{FFT}}}$.[3] Here, T denotes the OFDM symbol duration (excluding CP), N_{FFT} is the length of the fast Fourier transform (FFT) applied by the transmitter and the receiver, and B is the system bandwidth. Notice, the channel PDP $\tilde{h}_{u,i}(t,\tau)$ and the channel transfer function $H_{u,i}(t,f)$ form a Fourier-transform pair over the delay variable τ [24].

[3] Notice, NR provides the option to employ additional pulse-shaping on top of OFDM to reduce out-of-band (OOB) emissions; such pulse-shaping can be considered as part of the channel transfer function.

SISO-OFDMA: In OFDMA, multi-user multiple-access is achieved simply by sharing the REs amongst users; hence, whether the transmit signal $x_i[n,k]$ is intended for receiver u or a different receiver depends on the multi-user schedule $\mathcal{S}_i[n,k] \subseteq \{1,\ldots,U\}$, where U denotes the total number of receivers present in the system. This schedule is in practice signaled to the receivers over dedicated signaling channels (e.g. the physical downlink control channel (PDCCH) in LTE downlink). In case of single antenna transmission, I assume that at most a single user is served on each RE: $|\mathcal{S}_i[n,k]| \leq 1$. Considering non-orthogonal multiple access (NOMA) transmission [25], $x_i[n,k]$ could also be a superposition of symbols intended for multiple receivers, i.e. $|\mathcal{S}_i[n,k]| > 1$. This then requires successive interference cancellation (SIC) at the receivers to eliminate multi-user interference. However, NOMA is not addressed further in this thesis. I consider the optimization of multi-user scheduling strategies in more detail in Chapter 4.

MIMO-OFDMA: The extension of the base band system model towards multiple antennas is straightforward:

$$\mathbf{r}_u[n,k] = \mathbf{H}_{u,i}[n,k]\mathbf{x}_i[n,k] + \mathbf{z}_u[n,k]. \tag{2.2}$$

The received signal $\mathbf{r}_u[n,k]$ is now an M_u dimensional vector, with M_u denoting the number of receive antennas of receiver u, and $\mathbf{x}_i[n,k]$ is an N_i dimensional vector, with N_i denoting the number of transmit antennas of transmitter i. Matrix $\mathbf{H}_{u,i}[n,k]$ is the MIMO channel matrix, containing the complex-valued channel gains of all pairs of transmit and receive antennas. It is obtained by sampling the time-variant MIMO transfer function $\mathbf{H}_{u,i}(t,f)$, which in turn represents the Fourier-transform of the time-variant MIMO impulse response $\tilde{\mathbf{H}}_{u,i}(t,\tau)$.

Employing multi-user MIMO transmission based on linear precoding, the transmit signal vector $\mathbf{x}_i[n,k]$ can be written as:

$$\mathbf{x}_i[n,k] = \sum_{p \in \mathcal{S}_i[n,k]} \mathbf{F}_{i,p}[n,k]\mathbf{s}_p[n,k], \tag{2.3}$$

with $\mathbf{s}_p[n,k] \in \mathbb{C}^{\ell_p[n,k]\times 1}$ being the transmit symbols intended for receiver p and $\ell_p[n,k]$ denoting the number of spatial streams (or layers in LTE notation) transmitted to receiver p. Matrix $\mathbf{F}_{i,p}[n,k] \in \mathbb{C}^{N_i \times \ell_p[n,k]}$ is the precoder employed for transmission to receiver p. I consider optimization of the precoding matrices in more detail in Chapters 4 and 5.

Multi-cell MIMO-OFDMA: If multiple transmitters are present in the systems, as is commonly the case in mobile communications, a further generalization step is required:

$$\mathbf{r}_u[n,k] = \sum_{j=1}^{J} \mathbf{H}_{u,j}[n,k]\mathbf{x}_j[n,k] + \mathbf{z}_u[n,k],, \tag{2.4}$$

with J denoting the total number of transmitters. The model incorporates multi-point transmission, simply by including receiver u in the schedules $\mathcal{S}_j[n,k]$ of multiple transmitters. I consider optimization of multi-point transmissions in more detail in Chapter 5.

2.2.2. Channel Coding and Symbol Mapping

The transmit symbols $\mathbf{s}_p[n,k]$ commonly originate from the actual information data bits, via a combination of channel coding and mapping onto a symbol constellation. The specifics of this process depend on the considered transmission standard; e.g., LTE employs a BICM structure [26] with a selection of modulation and coding schemes (MCSs), combining variable rate Turbo coding with 4/16/64/256 quadrature amplitude modulation (QAM) over a powerful bit-interleaver. In link level simulations, these processing steps are explicitly implemented in software. However, for system level simulations, as discussed below, the performance of the modulation and coding stage is commonly abstracted to reduce computational complexity.

2.2.3. Modeling of RF-Hardware Impairments

To represent impairments caused by the radio frequency (RF) processing part of real-world hardware, RF distortion models are required in link level simulations. Such models exist for a number of impairments:

- Finite resolution of analog to digital converters (ADCs) and digital to analog converters (DACs): Modeling of such distortions commonly relies on an additive quantization noise model and the Bussgang theorem [27] to represent quantization errors as an uncorrelated additive distortion whose variance depends on the ADC/DAC resolution [28–30]. In our publication [30], we investigate the trade-off between the achievable energy-efficiency and the spectral-efficiency of a point-to-point link as a function of the ADC/DAC resolution. Our results show that, over a realistic SNR range, the optimal ADC/DAC resolution is relatively low (2-4 bit); it also depends on the exact values of the power consumption and loss of other RF front-end devices.
- Up/down-conversion to/from the carrier frequency by a local oscillator: This part mainly causes phase-noise due to variations of the local oscillator frequency around the actual carrier frequency and can therefore not be modeled as an additive distortion. It has been shown that the phase-noise can be accurately modeled in base band as a discrete random walk of the signal phase with correlated Gaussian increments [31].
- Non-linearity of power amplifiers: Such a non-linearity causes amplitude as well as phase distortions [32]. Several models for the power amplifier non-linearity exist,

naming the well-established Rapp model as just one example for the amplitude distortion [33]. Evaluating the impact of the power amplifier non-linearity recently regained interest in the context of millimeter wave (mmWave) transmissions, due to the relatively high expected non-linearity of cost-effective mmWave power amplifiers [34].

- Phase accuracy and power dissipation of the phase shifting network (PSN), connecting antenna elements to RF chains: Accurate modeling of these RF devices is indispensable for investigation of analog beamforming and hybrid precoding architectures [30, 35–37]. Such architectures employ RF phase-shifters and signal combiners, realized in the PSN, to achieve directional beamforming in multi-antenna systems. They are currently popular for mmWave transmissions, to reduce the number of required RF chains and thereby the hardware complexity [38]. In our publication [36], we show that antenna selection architectures can achieve a higher throughput than hybrid precoding solutions, when realistic assumptions on the power dissipation of RF devices are considered.

2.3. System Level Abstraction of MIMO-OFDMA Transmissions

While the major focus of link level simulations in wireless communications is on the evaluation of PHY processing techniques, system level simulations support the network-wide performance investigation and behavioral prediction of entire mobile networks. They are unchallenged for understanding the potential of novel technologies on a large scale with many interacting network elements.

2.3.1. Abstraction of PHY Processing

Abstractions of PHY details are required to enable large-scale simulations with manageable computational complexity. This is commonly achieved in close orchestration with a link level simulator, by applying link abstraction models to directly map from relatively easily computable metrics, such as, the signal to interference and noise ratio (SINR), to the actual metrics of interest, such as, the block error ratio (BLER) and the corresponding throughput of data transmission, without having to simulate the entire PHY processing chain [11, 39–41].

The PHY abstraction can efficiently be achieved by so-called effective SINR mapping (ESM) techniques [42–44], which transform a time-frequency selective channel, corresponding to the REs scheduled to a user, to an equivalent AWGN channel with appropriate effective SNR. In case of mutual information ESM (MIESM), the effective SNR is calculated such that an

equivalence is established in terms of the mutual information. Thereby, a careful calibration is required to match the performance achievable by the MCSs of the considered transmission standard [45–47], which necessitates an underlying link level simulator. In a second step, precomputed SNR-to-BLER look-up tables are then utilized to map the effective AWGN SNR to the BLER and, ultimately, accounting for the number of scheduled REs and the employed MCS, to the throughput.

2.3.2. Macro- and Microscopic Channel Models

The accuracy and realism of system level simulations is very much dependent on the employed macro- and microscopic channel models, which determine the SINR at the receivers. Currently, so-called stochastic-geometric channel models are the state-of-the-art for system level simulations. Stochastic-geometric channel models parametrize scattering objects, which determine the multi-path propagation of the wireless channel, in terms of probability distributions of the propagation angles (azimuth and elevation), delays and powers of signals scattered by these objects [48]. Prominent examples are the 3GPP SCM, the Wireless World Initiative New Radio (WINNER) channel model, as well as, the recently proposed 3GPP 3D SCM [12, 49–51].

Spatially Consistent Channel Modeling

An important research issue in the context of channel modeling is the consideration of spatial consistency in propagation conditions. Propagation conditions at two geometrically close positions are bound to be similar, due to the geometric correlation of the relevant scattering objects for such close positions. However, such a correlation is not provided by existing stochastic-geometric channel models in the first place, because scattering objects are independently drawn for each position. As discussed in *[Articel 1]*, this leads to a highly unrealistic behavior, especially in combination with FD-MIMO systems, which apply spatial beamforming to serve multiple users in parallel, as well as, when considering macroscopic movement of users. Early work to enhance the spatial consistency of system level simulations only considered a correlation of the macroscopic shadow fading [52, 53]. We recently extended these methods to additionally allow for a spatial correlation of the line-of-sight/non-line-of-sight state of users[4], as well as, the microscopic fading parameters (signal angles, delays, powers) [54, 55]. These contributions enable a realistic consideration of spatial consistency in system level simulations, as required for the evaluation of technologies developed to support the requirements of the *Society in Motion*. The models have been validate through extensive Ray tracing simulations of urban environments [56].

[4]Notice, this impacts the path-loss exponent and the microscopic fading distribution of the 3GPP 3D SCM.

2.4. Scientific Contributions and Publications

As mentioned above, our mobile communications research group has years-long experience in developing standard-compliant cellular communications simulators, applying them to reveal strength and weaknesses of the respective standard, as well as, in developing standard-enhancements to mitigate these weaknesses. My involvement in these developments started in 2008 as a Master-level student, with the first release of the Vienna LTE simulators [11], progressed to lead-developer of the Vienna LTE-Advanced (LTE-A) simulators [10] as a PhD student, and lead to project coordinator of our 5G NR simulators [57, 58] as a post-doc University assistant. During that period, our simulation methodologies and abstraction techniques became ever more sophisticated, to improve the realism of our simulators and to account for an increasing number of impairments caused by the channel and by RF-hardware components.

2.4.1. Selected Publications

We summarize our simulation methodologies and our insights on the performance of novel signal processing techniques within standard-compliant environments in the three journal articles:

[Article 3] 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

[Article 4] M. Taranetz, T. Blazek, T. Kropfreiter, M. Müller, S. Schwarz, and M. Rupp, "Runtime precoding: Enabling multipoint transmission in LTE-advanced system-level simulations," *IEEE Access*, vol. 3, pp. 725–736, 2015

[Article 5] E. Zöchmann, S. Schwarz, S. Pratschner, L. Nagel, M. Lerch, and M. Rupp, "Exploring the physical layer frontiers of cellular uplink – The Vienna LTE-A uplink simulator," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 118, 2016

2.4.2. Summary of Scientific Contributions

Within the scope of the *Society in Motion*, the most relevant scientific contributions of these articles are:

1. Optimization of the pilot signal power allocation as a function of the delay- and Doppler-spread of the channel in *[Article 3]*.

2. Development of ICI abstraction models for realistic system level simulations of high-mobility users in *[Article 4]*.
3. Derivation of a compact vector-matrix base band system model for SCFDMA transmission over doubly (time-frequency) selective channels in *[Article 5]*.

[Article 3] Pushing the limits of LTE: A survey on research enhancing the standard: In this paper, we discuss research work that is based on LTE standard-compliant assumptions, implying that the results are with minor effort immediately applicable within the LTE standard. We thereby attempt to bridge the gap between researchers and engineers working in this area. We furthermore introduce the Vienna LTE simulators as a vehicle to support reproducible wireless communications research in a well-defined environment, and showcase our own efforts to enhance the LTE standard. Highly relevant for the *Society in Motion* are thereby our results on optimal power allocation for LTE pilot signals, which show that a significant amount of pilot signal power can be saved especially at high mobility. The intuition behind these results is as follows: at high velocity interpolation errors between pilot positions are dominant over estimation errors on pilot positions; hence, the estimation quality on pilot positions can be reduced without significantly impairing the overall channel estimation performance, allowing for power reduction of pilot signals. These results have later on been extended to optimizing not only the power allocation, but also the density of pilot positions in the OFDM lattice of REs (the resource grid), revealing the significant potential of channel adaptive pilot patterns to enhance the throughput performance and reliability even in strongly doubly-selective channels; see [61] and our discussion in *[Articel 1]* for more details. In our recent work [62], we take such channel adaptive approaches even a step further and additionally optimize the numerology (i.e., the subcarrier spacing, symbol- and CP-duration) of the OFDM transmission with respect to the time-frequency selectivity of the channel.

[Article 4] Runtime precoding: Enabling multipoint transmission in LTE-advanced system-level simulations: In this contribution, we discuss enhancements of the Vienna LTE system level simulator to support the important LTE-A feature of multipoint transmission. This allows to efficiently simulate coordinated multi-point transmission (CoMP) techniques and multimedia broadcast single frequency network (MBSFN) operation (see Chapter 5 for details). As discussed in *[Article 2]*, MBSFN is an enabling technology for cellular-assisted vehicular communications, because it facilitates an efficient distribution of common information (e.g. road-traffic status updates) to many vehicles in parallel. When simulating vehicular communications scenarios, or other high-mobility applications, it is important to account for the distortions introduced by ICI due to the Doppler-spread of the signal. Such additive distortions can be accounted for in system level

simulations relatively easily, if they can be well approximated as being Gaussian distributed. In this case, it suffices to add an additional noise term of appropriate power in the SINR calculation. As we show in *[Article 4]*, this approach for ICI modeling works very well in OFDM up to significant Doppler spreads lying in the order of 10 % of the subcarrier spacing, while still assuming block-fading (i.e., a temporally constant channel) within each OFDM symbol. Beyond this point, it becomes necessary to account for the channel variation within each OFDM symbol, requiring fast-fading simulations. This is because very strong temporal fading of the channel impairs the performance of the forward error correction (FEC) Turbo channel code of LTE. Increased temporal channel variations also lead to a performance deterioration of channel estimation methods. As we shown in [63], the distortion caused by imperfect channel estimation can also be accounted for as additional noise, whose variance depends on the correlation properties of the channel and the placement of pilot symbols in the resource grid. Such modeling approaches thus enable realistic investigations of high-mobility scenarios on system level.

[Article 5] Exploring the physical layer frontiers of cellular uplink – The Vienna LTE-A uplink simulator: In this journal article, we introduce our last member of the Vienna LTE simulator suite, the Vienna LTE-A uplink simulator. Since the uplink of LTE is based on SCFDMA, the link level model described in Section 2.2 is not directly applicable. SCFDMA applies a discrete Fourier transform (DFT)-precoding step to spread the user signals over the entire allocated bandwidth. This provides peak to average power ratio (PAPR) advantages over OFDMA; yet, it entails a throughput performance loss when linear zero forcing (ZF) or minimum mean squared error (MMSE) receivers are employed. In *[Article 5]*, we provide a vector-matrix input-output relationship of SCFDMA, derive SINR expressions for ZF and MMSE receivers, and conduct an asymptotic analysis of the achievable rate with growing bandwidth. The uplink of LTE employs frequency-spreading of pilot signals to conserve the good PAPR properties of SCFDMA. In *[Article 5]*, we investigate the mean squared error (MSE) of several linear channel estimation techniques as a function of the user velocity, and show that the LTE uplink is more sensitive to temporal channel variations than the downlink, since the pilot-distance in time is larger.

Received March 5, 2013, accepted April 16, 2013, published May 10, 2013.

Digital Object Identifier 10.1109/ACCESS.2013.2260371

Pushing the Limits of LTE: A Survey on Research Enhancing the Standard

STEFAN SCHWARZ, JOSEP COLOM IKUNO, MICHAL ŠIMKO, MARTIN TARANETZ, QI WANG, AND MARKUS RUPP

Institute of Telecommunications, Vienna University of Technology, Vienna 1040, Austria

Corresponding author: S. Schwarz (sschwarz@nt.tuwien.ac.at)

This work was supported in part by the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility, KATHREIN-Werke KG, and A1 Telekom Austria AG.

ABSTRACT Cellular networks are currently experiencing a tremendous growth of data traffic. To cope with this demand, a close cooperation between academic researchers and industry/standardization experts is necessary, which hardly exists in practice. In this paper, we try to bridge this gap between researchers and engineers by providing a review of current standard-related research efforts in wireless communication systems. Furthermore, we give an overview about our attempt in facilitating the exchange of information and results between researchers and engineers, via a common simulation platform for 3GPP long term evolution (LTE) and a corresponding webforum for discussion. Often, especially in signal processing, reproducing results of other researcher is a tedious task, because assumptions and parameters are not clearly specified, which hamper the consideration of the state-of-the-art research in the standardization process. Also, practical constraints, impairments imposed by technological restrictions and well-known physical phenomena, e.g., signaling overhead, synchronization issues, channel fading, are often disregarded by researchers, because of simplicity and mathematical tractability. Hence, evaluating the relevance of research results under practical conditions is often difficult. To circumvent these problems, we developed a standard-compliant open-source simulation platform for LTE that enables reproducible research in a well-defined environment. We demonstrate that innovative research under the confined framework of a real-world standard is possible, sometimes even encouraged. With examples of our research work, we investigate on the potential of several important research areas under typical practical conditions, and highlight consistencies as well as differences between theory and practice.

INDEX TERMS Heterogeneous networks, distributed antenna systems, frequency synchronization, pilot power allocation, multiuser gains, LTE, MIMO, reproducible research.

I. INTRODUCTION

Life without ubiquitous possibilities to connect to the Internet is hard to imagine nowadays. Cellular networks play a central role in our global networking and communication infrastructure. To ensure and even enhance availability, the standardization process of new communication systems is governed by concerns about reliability, interoperability and security, besides trying to improve the performance of current technology. Still, the ever-increasing demand for higher data-rates forces the consideration of novel research results during standardization. Mobile data traffic is predicted to increase 13-fold between 2012 and 2017, culminating in a monthly global data traffic of more than 10 exabytes by 2017 [1]. To sustain such a traffic growth, standardization experts ceaselessly improve wireless networks. In this process, however, innovative research results are frequently met with skepticism, because assumptions made by researchers are sometimes too simplistic and idealistic to reflect the performance under practical conditions. Consider, e.g., spatial interference management techniques like linear multi-user MIMO (MU-MIMO) transmission [2], [3] and interference alignment [4], [5]. While theory predicts tremendous spectral efficiency gains under the assumption of perfect channel state information at the transmitter (CSIT), heavy losses are reported if this assumption is only slightly violated [6], [7], e.g., due to delayed or quantized CSI feedback. Although tailored solutions exist to resolve these problems, e.g., [8]–[11], standardization is still reluctant about investing in the

CSI feedback, possibly because communication and information exchange between academia and industry is insufficient.

The world's leading cellular networking technology these days is standardized by the 3rd Generation Partnership Project (3GPP), a collaboration between telecommunication associations spread all over the world. Technical specifications for the radio access network technology, the core network and the service architecture are released every few years, constantly evolving the cellular system with a major focus on compatibility between releases. With the introduction of Long Term Evolution (LTE) in Release 8 (2008) [12], an entirely new air interface based on Orthogonal Frequency Division Multiplexing (OFDM) was implemented setting the basis for a 4G capable mobile communication technology, and first LTE networks went on-air in 2009/10. Since then, work on LTE advanced (LTE-A) (i.e., LTE Rel. 10) and beyond is ongoing in the standardization groups to enhance the transmission capabilities of LTE. The basis for downlink MU-MIMO and other non-codebook based precoding schemes was set by extending the reference symbol structure of legacy LTE for user-specific reference symbols [13]. Network densification, captured under the keywords heterogeneous networks, small- and femto-cells and distributed antenna systems (DASs), is recognized as an important means to circumvent the capacity crunch [14]. This has to go hand in hand with improved interference management, i.e., coordinated multi-point transmission and reception (CoMP), to cope with the increased interference between cells [15].

Despite all of these technologies being considered in the standard, their implementation is largely based on very simple concepts. Hence, research work is ongoing in parallel to develop more sophisticated solutions. We provide an overview of such standard related research during the course of this article, summarizing results of other research groups and providing a more detailed view on our own work. Our main focus is thereby on LTE networks, although the same concepts are mostly applicable for WiMAX [16] and WiFi [17] as well.

In trying to classify research work with respect to its practical applicability, theoretical results mostly lag behind due to the coarse abstraction required to facilitate analytical tractability. Such results have their significance in providing upper bounds on system performance, analyzing the potential of new technologies, and opening/identifying new fields for research and engineering activities. Simulations, on the other hand, enable investigations of much more complex, detailed and realistic scenarios, and facilitate comparison of different algorithms under identical conditions. Simulations can give a qualitative understanding of the interplay of different components of the systems, which may not be obtainable otherwise for complexity reasons. Measurements and field trials, finally, avoid all kinds of assumptions and models, thus reflecting reality most closely. Still, the involved expenditure of time, labor and money, and the lack of generality in the obtained results prohibits their application in early stages of research.

The simulation approach is hence adopted by many researchers in combination with theoretical investigations, due to its flexibility and efficiency. Though standardized simulation models exist for parts of the environment, e.g., the wireless channel in cellular communications [18], [19], there is still lots of ambiguity in many simulation parameters left, making it often difficult to reproduce results of other researchers and hampering cross-comparison of different techniques, a circumstance that has been complained about quite openly in [20], [21]. Moreover, the code of such highly complex systems may contain more than 100.000 lines, making thorough testing practically impossible in a reasonably short time. Only working in parallel with many independent research groups and communicating publicly via a web-based forum makes it possible to identify programming bugs and have the code checked independently several times. These facts were the motivation for our group to develop standard-compliant open-source Matlab-based link- and system-level simulation environments for LTE to facilitate reproducibility and information exchange via a common platform.

A. Organization

In Section II, we introduce our link- and system-level simulators and highlight the multitude of scenarios that can be investigated with these tools. Then, in Section III, we provide a review of current physical layer research that is closely related to LTE and LTE-A. We also consider in more detail the impact of imperfect frequency synchronization on the system throughput in this section, and present a pilot power allocation algorithm that efficiently distributes the available transmit power among reference- and data-symbols. Section IV is dedicated to interference in wireless networks. A literature survey gives an overview of state of the art research considering interference management techniques and algorithms for wireless communications. A very simple technique for mitigating the detrimental effects of interference is opportunistic scheduling, i.e., exploiting the interference dynamics to serve users whenever they experience favorable channel conditions. This is considered in Section IV-B. In sections IV-C and IV-D, we give our view on heterogeneous networking architectures by investigating the performance of distributed antenna systems and macro-femto overlay networks.

II. FACILITATING REPRODUCIBILITY

The *Vienna LTE simulators* [22] are a simulation suite for link- and system-level simulation of LTE networks that is developed and extended by our research group since the first specifications of LTE were published in 2008. The simulators are publicly available for download (www.nt.tuwien.ac.at/ltesimulator) under an academic non-commercial use license. They facilitate reproducibility of research results, and contribute to bridge the gap between researchers and standardization experts in LTE and LTE-A.

Since its first release in 2009, the LTE downlink link-level simulator was downloaded more than 16 000 times and is currently (February 2013) in its eighth release. It was also extended to LTE-A and augmented with an uplink version in 2011 [23]. The LTE system-level simulator experienced even more attention, with more than 22 000 downloads, thus confirming the demand for a consistent simulation environment. Details about the physical layer abstraction models employed for system-level simulation can be found in [24], [25].

Other research groups provide similar tools, such as the *Open Wireless Network Simulator* [26] developed at RWTH Aachen University, or the system-level simulator *LTESim* [27] provided by the Telematics lab of Politecnico di Bari, but we are neither aware of the sophistication and accuracy of these tools nor of how active research is conducted around them.

In our research work, the simulators are utilized to investigate cellular networks in varying degrees of abstraction. The simulators play a central role in the collaboration with our industry partners. Consider the example cellular network shown in Fig. 1. It consists of three macro base stations with sectorized antennas, plus additional radio access equipment. Users 1–3 are served in the "classical" way, by attaching the user equipment (UE) to the strongest macro base station and treating other base stations as interferers. The data-transmission can be optimized by focusing on the radio link between a base station and a single UE, which relies on detailed modeling of the physical-layer and requires link-level simulations. Two examples for link-optimization conducted in our group are treated in some detail in sections III-B and III-C, on the subjects of pilot power allocation for channel estimation and carrier frequency synchronization.

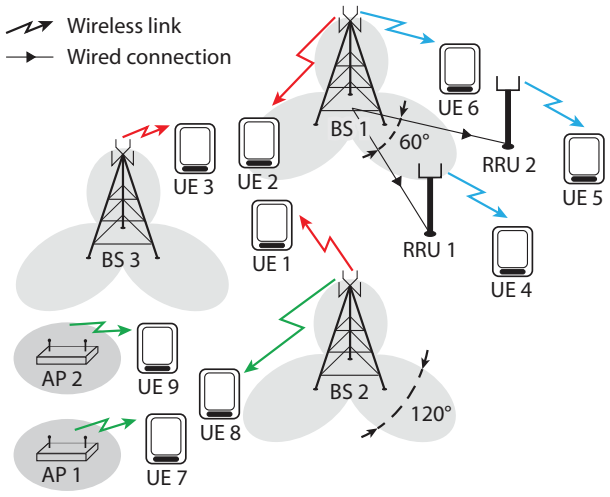


FIGURE 1. Example cellular network consisting of three sectorized macro base stations and additional radio access equipment, visualizing different scenarios considered in our research work.

An alternative perspective for the optimization of cellular communication systems is the network viewpoint. Here, a large network consisting of a multitude of base stations and UEs is considered. To keep the computational complexity of the associated system-level simulations tractable, abstraction of the physical-layer details is necessary. Section IV-B treats multi-user scheduling as an example, confirming the theoretically well-known double-logarithmic growth of the sum-rate with the number of users under the practical constraints that are introduced by the LTE standards, but also showing that the spatial degrees of freedom are not fully utilized by the system.

Extensions of the classical sectorized cellular network architecture to heterogeneous networks, containing different types of radio access equipment, are in the scope of many recent research activities. Two examples are shown in Fig. 1. Users 4–6 are jointly served by a single base station whose transmission capabilities are enhanced by remote radio units (RRUs), forming a so called distributed antenna system (DAS). The performance of different transmission strategies in such distributed antenna systems is evaluated in Section IV-C. In this case, we cannot abstract the physical layer details, because they are necessary to calculate the precoders and beamformers of the considered transmission strategies. To circumvent computational complexity issues in this case, we employ a hybrid of link- and system-level simulations, by augmenting the link-level simulator with an appropriate inter-cell-interference model [28], [29].

Femto cell access points are a popular technique for increasing the spatial reuse of existing cellular networks. Users close to access points are offloaded from the macro cellular network and served by the femto cells (see Users 7–9 in Fig. 1). The benefits of LTE femto cell-enhanced macro networks in terms of user throughput and fairness of the resource allocation are investigated in Section IV-D, by means of system-level simulations.

The link-level simulator can also be beneficially employed to simplify measurement campaigns. In our research group it serves as a front-end for a measurement testbed, generating the base band transmit signal and detecting the received signal. Some recent measurement results investigating the impact of different antenna configurations on the LTE downlink throughput can be found in [30]. Also, the system-level simulator can be used in combination with real measured network data by loading measured channel pathloss maps into the simulator. We successfully employed this feature to confirm observed trends of our industry partners in the real world measured throughput.

III. ENHANCING THE PHYSICAL-LAYER

Recent measurement and simulation based investigations of current cellular communication systems (HSDPA, WiMAX, LTE) have revealed a large performance gap between the throughput achieved in such systems and the theoretical upper bounds determined by channel capacity [31], [32]. Although it is often believed that the potential of the physical-layer is already largely exploited, these investigations show that

there is still lots of space for improvement left, even when considering only a single transmitter-receiver pair. This was also observed in [20] who pointed out that innovative PHY research is possible by investigating more realistic mathematical models that consider constraints and suboptimalities of practical systems, e.g., channel estimation errors, power amplifier non-linearities, synchronization errors, realistic CSI feedback algorithms,... Notice though that the accumulated throughput loss of a single link, although significant, is made up of a multitude of small contributions [33]. Hence, large performance gains from a single PHY research front only may not be obtained.

Conducting research on the physical-layer of a wireless communication system requires detailed modeling of the wireless channel and the signal-processing algorithms applied at the transmitter and receiver. In this section, we firstly provide a survey on standard-related PHY research in Section III-A, focusing on LTE and LTE-A. Notice that there exists a huge amount of literature on that topic; hence we can only provide our restricted view and by far not an exhaustive description. We also provide two examples of our research work to establish the utility of the Vienna LTE link-level simulator for PHY investigations. In Section III-B substantial power savings are demonstrated at high user velocities by exploiting the mean squared error (MSE) saturation of well-known channel estimation algorithms in temporally weakly correlated channels. Furthermore, the impact of a frequency estimation error, causing a carrier frequency offset between transmitter and receiver, on the throughput of an LTE system is considered in Section III-C.

A. LITERATURE SURVEY

One important physical-layer research topic for wireless communication systems in the uplink as well as the downlink is channel estimation. The channel estimation error not only impacts the post-equalization signal to interference and noise ratio (SINR) [25], [34] but also the accuracy of CSI feedback estimation and hence the achievable throughput of the system [35]. Important aspects of channel estimation covered in the literature are high-velocity fast-fading scenarios [36]–[38], iterative channel estimation [39], [40], special designs for multi-user and multi base station situations [41], [42] as well as low-complexity solutions [43], [44].

Another topic of practical relevance is the timing- and frequency-synchronization of transmitter and receiver or of multiple transmission points. A frequency synchronization error between transmitter and receiver results in inter-carrier-interference after OFDM signal processing at the receiver [45]. Similarly a synchronization error in the symbol timing leads to inter-symbol-interference [46]. Both effects significantly reduce the achievable throughput of the system, hence accurate synchronization algorithms are required, e.g., [47]–[49].

Also, CSI feedback estimation is important for LTE to achieve the highest throughput in MIMO systems. Here, one has to distinguish between CSI feedback for codebook based precoding and feedback for non-codebook based precoding. The former consists of a precoding matrix indicator and a rank indicator, to select the preferred precoder from a given codebook, and a channel quality indicator to adapt the transmission rate [50]–[52]. Non-codebook based precoding enables more sophisticated transmission strategies like zero forcing (ZF) beamforming [3], block-diagonalization precoding [2] or interference alignment [4], [5]. In this case, mostly explicit CSI feedback is employed, quantizing the wireless channel in some form directly and feeding it back to the base station [8]–[11], [53].

Other topics of interest include advanced receiver designs [54], consideration of power-amplifier non-linearities [55], and efficient reference signal designs [56]–[58].

B. PILOT POWER ALLOCATION

Modern standards for wireless communication systems such as LTE and WiMAX exclusively rely on coherent transmission techniques. Detection of coherently transmitted data symbols requires knowledge of the channel experienced during transmission, which is obtained by channel estimation. The estimates are calculated from known symbols, so called pilot symbols, that are multiplexed within the data symbols. The amount of power assigned to the pilot symbols has a crucial impact on the quality of the channel estimation, which in turn significantly impacts the throughput performance of the system. The channel estimation error leads to an additional noise term in the SINR of the transmission, whose variance is determined by the applied channel estimator and the pilot symbol density and power.

Consider an LTE transmitter which has a certain amount of power available for transmission. The available power is divided between pilot and data symbols. The quality of the channel estimates improves with the amount of power invested into the pilot symbols, thus reducing the noise contribution of the channel estimator. But in turn the power of the data symbols has to be decreased to keep the power budget balanced, degrading the received signal power. Therefore, an equilibrium point in the power assignment between pilot and data symbols exists, in which the SINR is maximized.

The precisely elaborated physical-layer of the Vienna LTE link-level simulator enables an extensive investigation of different pilot symbol power allocation algorithms. The optimal distribution of the available transmit power between pilot and data symbols was derived in [59] and [60] for time-invariant and time-variant channels, respectively. The solution turned out independent of the operating point (signal to noise ratio (SNR)) and of the actual channel realization, making it very robust and applicable in practical systems. In [60] it was realized that in time-variant channels state-of-the-art channel estimators (least-squares and linear minimum mean squared error (MMSE)) exhibit an error-floor, which increases with decreasing temporal channel correlation. Thus, at a given user velocity (which determines the temporal channel correlation in the link-level simulator according to Clarke's model [61]) and operating point, a further increase of the pilot symbol

power does not necessarily lead to an improvement in the quality of the channel estimate. Therefore, one might think that more power should be allocated to the data symbols. This, however, does not improve the post-equalization SINR either, because the interlayer interference increases with the data symbol power, due to the imperfect channel knowledge. Based on this insight, a power efficient power allocation algorithm was proposed in [62]. In this algorithm the total transmit power of pilot and data symbols is minimized at a given user velocity and noise power, while constraining the post-equalization SINR not to decline with respect to the case that all available transmit power is used. The case in which all available transmit power is used and equally distributed among pilot- and data-symbols is referred to as unit power allocation.

Fig. 2 demonstrates the performance of the power efficient power allocation algorithm in comparison to unit power allocation, as obtained by link-level simulations. The average user throughput versus user speed for three different antenna configurations $N_t \times N_r \in \{1 \times 1, 2 \times 2, 4 \times 4\}$ is shown in the upper part of the figure. It can be seen that the power-efficient power allocation algorithm achieves virtually

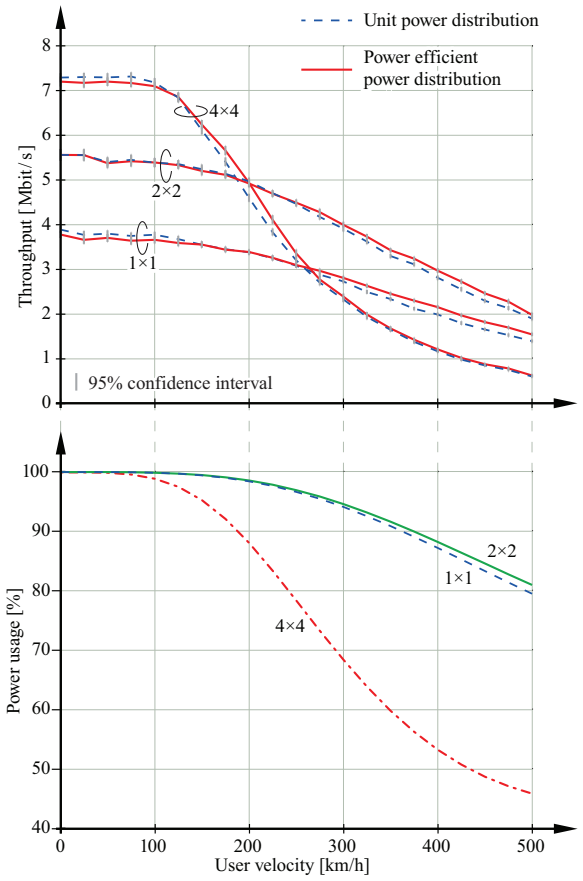


FIGURE 2. Comparison of the average user throughput of an LTE system with unit power allocation for pilot and data symbols, and with power efficient power allocation. The SNR is set to 20 dB. The lower part indicates the power savings of the power efficient power allocation versus unit power allocation.

the same throughput as the unit power allocation algorithm. The power consumption utilizing the power efficient power allocation algorithm with respect to unit power allocation is depicted in the lower part of Fig. 2. The figure shows that at high user velocities substantial power savings are possible without degrading the throughput performance of the system. Note that LTE is defined to operate up to user velocities of 500 km/h.

C. IMPACT OF IMPERFECT FREQUENCY SYNCHRONIZATION

One crucial issue that a novel technique encounters in real world applications is to cope with the physical impairments which are usually not taken into account in simulation-based experiments. For a communication system, typical such examples are an offset between the local oscillators at the transmitter and the receiver, oscillator phase noise or an imbalance between the in-phase and quadrature-phase branches in the front end processing.

Taking the carrier frequency offset (CFO), i.e., the offset between the carrier frequencies at the transmitter and the receiver, as an example, plenty of literature can be found on how to estimate the CFO in the digital signal processing domain. The estimator's performance is usually evaluated in terms of the estimation error, in other words, the MSE. This is shown in the center part of Fig. 3 for two specific examples: (i) the time domain and (ii) the frequency domain estimators of [63], and two different transmit-receive antenna configurations $N_t \times N_r \in \{1 \times 1, 2 \times 2\}$. However, the performance of a communication system is evaluated in terms of the overall coded throughput, encompassing all the signal processing steps applied at the transmitter and receiver, e.g., coding, modulation, equalization, and detection. Therefore, a link performance prediction model is desirable, providing a direct mapping from the residual CFO to the coded throughput. In the following, we show how to utilize such a mapping to estimate the throughput loss caused by the carrier frequency estimation error.

In [45], the authors derive an analytic expression for the post-equalization SINR achieved on a resource element[1] of the LTE downlink with imperfect frequency synchronization. This expression can be exploited to estimate the throughput loss of the LTE system, and thus obtain the desired relation:

1) For a given CFO estimation scheme, determine the estimation performance in terms of the MSE. As shown, e.g., in [45] the MSE is theoretically given by a function which depends on the SNR expressed as MSE(SNR).
2) Calculate the residual CFO $\varepsilon = \sqrt{\text{MSE(SNR)}}$. Utilizing the model of [45], the post-equalization SINR on a resource element r can then be expressed as a function of the CFO, namely $\text{SINR}_r(\varepsilon)$.
3) Estimate the throughput of the system from $\text{SINR}_r(\varepsilon)$.

[1] In LTE a resource element denotes the basic unit of physical OFDM time-frequency resources.

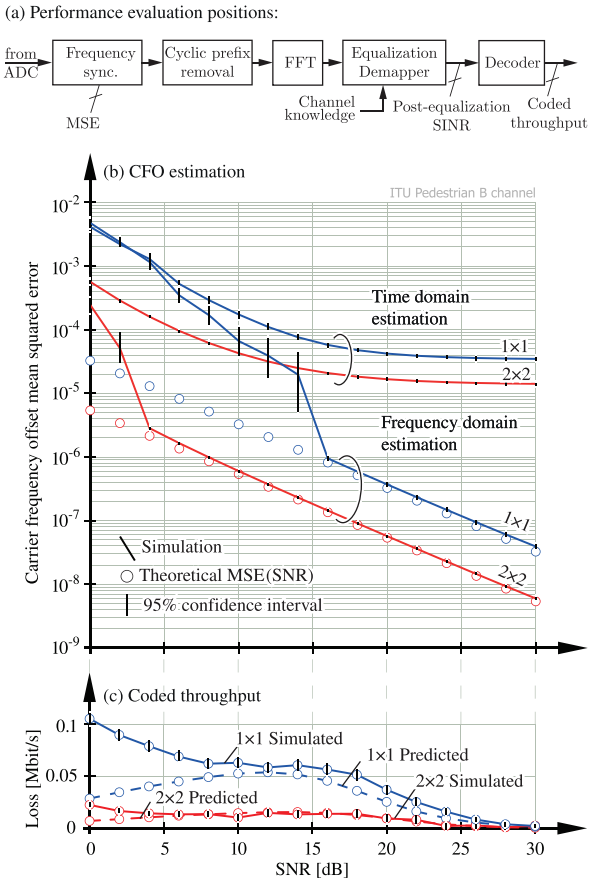


FIGURE 3. **(a) The three performance evaluation positions in the receiver signal processing chain. (b) MSE performance of the carrier frequency synchronization scheme in [63]. (c) The predicted and simulated coded throughput loss resulting from the residual estimation error in (b).**

In general, pre-computed mapping tables valid for the considered system (e.g., obtained from link-level simulations) can be employed to map the post-equalization SINR to a corresponding throughput value. Here, we are only interested in the throughput loss compared to the case of perfect synchronization. In this case, we can employ the bit-interleaved coded modulation (BICM) capacity to estimate the throughput $f\ (\mathrm{SINR}_r(\varepsilon))$, since LTE is based on a BICM architecture. The imperfect channel code will cause an offset in the absolute value of the estimated throughput, but this offset approximately cancels out when calculating the throughput loss $\Delta B = \sum_r f\ (\mathrm{SINR}_r(0)) - \sum_r f\ (\mathrm{SINR}_r(\epsilon))$.

The Vienna LTE link-level simulator enables such investigations and greatly facilitates a standard compliant validation. In the bottom part of Fig. 3, the predicted loss in terms of the coded throughput is compared to the results obtained by means of extensive link-level simulations. The figure confirms that the prediction model performs well as soon as the MSE follows the theoretical MSE(SNR) relation. The large deviation of the 1×1 system at low SNR can be explained by the deviating performance of the employed synchronization algorithm from the theoretical curve, as shown in the center part of Fig. 3.

IV. TREATING INTERFERENCE IN CELLULAR NETWORKS

Many research efforts currently concentrate on the interference between multiple transmitter and receiver pairs. We review some of this work with focus on applicability in practical systems in the near future in Section IV-A.

A robust way to deal with interference is opportunistic scheduling, which exploits the interference dynamics to serve users whenever they experience good channel conditions. This is investigated in Section IV-B using system-level simulations.

With sufficient CSIT, sophisticated signal-processing algorithms can be applied before transmission to avoid/minimize interference between several nodes. In Section IV-C multi-user beamforming in DASs with perfect and quantized CSIT is considered, revealing large throughput gains in comparison to single-user MIMO (SU-MIMO) systems.

Finally, in Section IV-D the impact of interference caused by femto-cell deployments on existing macro cellular networks is explored with the aid of the Vienna LTE system-level simulator, in terms of user throughput and resource allocation fairness.

A. LITERATURE SURVEY

An effective way to deal with the exponential growth in wireless data traffic is the deployment of small-size low-power base stations, so called femto-cells, within the usual network of macro base stations. Special attention must then be paid to the management of interference between different layers of the resulting heterogeneous network. Also, to improve coverage and to mitigate shadowing- and penetration losses, the network can be extended by low-cost relay nodes and RRUs, further complicating the inter-cell-interference situation. There are many approaches for solving the interference problematic. Following the notation introduced with CoMP in LTE-A, these methods can basically be classified as coordinated scheduling, coordinated beamforming or joint processing techniques.

With coordinated scheduling, base stations exchange control information via the X2 interface, to coordinate the assignment of time-frequency resources. Some recent examples of coordinated scheduling algorithms can be found in [64], [65]; a thorough overview of multi-cell scheduling in LTE is presented in [66]. Coordinated beamforming also relies on the exchange of control information between base stations. Here the goal is to form transmit beams such as to minimize the power of interfering signals and maximizing the power of the intended signals. Typically, this leads to optimization problems involving SINR or signal to leakage ratios [67]–[69]. Also, interference alignment falls into this category. Joint processing, finally, requires not only an exchange of control information, but also an exchange of the user data. With these techniques, the transmit antennas of several base

stations form a virtual MIMO system and jointly transmit data to users, effectively eliminating cell boundaries. Although such methods can achieve the best performance, they also suffer from the highest CSI feedback overhead and backhaul load. Some of the latest proposals can be found in [70], [71]. A more comprehensive overview of several CoMP techniques is presented in [72].

B. SCHEDULING AND MULTI-USER GAIN

When a single transmission link is considered, time and frequency diversity can be exploited to increase the throughput and reliability of the data transmission. In a practical cellular network that serves not only a single user but a multitude of them, multi-user diversity can additionally be utilized to increase the total throughput of the cell. In combination with the spatial degrees of freedom added by a MIMO system, theory states that the throughput gain due to multi-user diversity follows an $N_t \log\log k$ rule with the number of concurrent users k [73], where N_t is the number of transmit antennas.

The theoretical $N_t \log\log k$ rule, although useful as an upper bound on the achievable diversity, is not directly applicable to the throughput of an LTE link, because other parasitic effects encountered in a practical system diminish parts of the promised gains [32], e.g., a growing pilot symbol overhead with increasing number of transmit antennas and a limited choice of possible precoding matrices in MIMO systems. Hence, there is a need for realistic throughput simulations.

Simulations of large cellular networks, with many cells and users being present, are very computationally demanding. If implemented via link-level simulations, a single simulation of that kind could last several months. By applying physical-layer abstraction models, it is possible to reduce the complexity of such system-level simulations, without significantly compromising the accuracy of the results [24].

Fig. 4 shows simulation results obtained with the Vienna LTE system-level simulator, comparing the performance of several scheduling algorithms in an LTE network. The left-hand side results are obtained in a SISO system, while the right-hand side performance is achieved in a MIMO system with $N_t \times N_r = 2 \times 2$ employing LTE's closed-loop spatial multiplexing (CLSM) mode, using the CSI feedback algorithms of [50]. The following scheduling schemes, implemented according to [74], are employed:

1) Best CQI scheduling assigns resources to the users with the best channel conditions only. This algorithm is the practical counterpart to the theoretical cell throughput upper bounds, but it achieves the lowest fairness in terms of distributing resources among users.
2) Proportional fair scheduling aims at increasing fairness and avoiding the user starving issue encountered in the best CQI scheduler, by scheduling users whenever their channel conditions are good with respect to their own average channel quality.
3) The round robin scheduler equally distributes resources among users. While the former two algorithms are opportunistic in nature, serving users whenever they experience good channel and interference conditions, this latter algorithm ignores any channel state information and thus does not make use of the available diversity.

An adaptation of the $N_t \log\log k$ rule is employed to quantify the spatial multiplexing and multi-user gains of a practical system, by introducing a scaling factor a for the multiplexing gain and a diversity gain loss factor b for the multi-user diversity. The results for the considered LTE system show a multiplexing gain factor (N_t/a) of 1.56 for an $N_t \times N_r = 2 \times 2$ system, and a gain of 2.66 for a 4×4 antenna configuration compared to the SISO case, thus considerably below the theoretical values of 2 and 4, respectively. A similar analysis

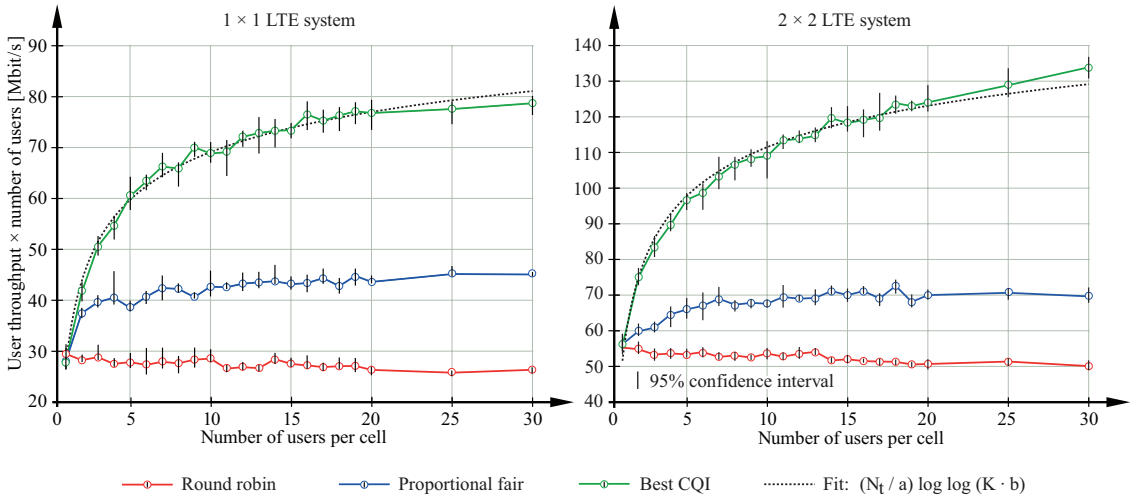


FIGURE 4. Multi-user gain simulation results for a SISO (left) and 2 × 2 closed-loop spatial multiplexing LTE MIMO (right) configuration. Solid line: throughput results. Dashed line: log log *k* approximation.

applied to proportional fair scheduling shows the same gains relative to the SISO case.

Such an analysis of the multi-user gains of LTE can serve as the basis for the extension to more complex setups, such as ones including distributed antennas or femto cells.

C. DISTRIBUTED ANTENNA SYSTEMS

A DAS is a cellular networking architecture in which several transmission points, controlled by a single base station, are geographically distributed throughout the network. DASs make use of RRUs to extend the base stations' central antenna ports (cf. BS 1 in Fig. 1). Coherent data transmission from all antennas is enabled by a high-bandwidth low-latency connection between the RRUs and the base station, thus making spatial multiplexing of several data-streams and/or users possible. Several publications have established the theoretical potential of DASs for improving the network capacity, reducing the outage probability and improving the area spectral efficiency (e.g. [28]), but investigations taking into account the constraints imposed by a practical system, e.g., channel coding, limited feedback, are scarce in literature (e.g. [29], [75]).

Simulations of advanced transceivers, especially for MU-MIMO transmission, as well as limited feedback algorithms require detailed knowledge of the users' channels, and are thus hardly amenable to system-level abstraction. Therefore link-level simulations appear as the appropriate choice, but are complicated by the fact that multiple base stations should be simulated to account for the changes in the out-of-cell interference environment caused by RRUs. In our simulations, we strike a compromise between computational complexity and accuracy, by employing the link-level simulator to explicitly simulate three cells and considering interference from more distant base stations with the out-of-cell interference model of [28]. For that purpose, the Vienna LTE-A link-level simulator is extended with a distance-dependent pathloss model and a macroscopic fading model, determining the SNR of a user based on its position (see [75] for details).

Fig. 5 shows simulation results obtained with this hybrid link/system-level simulation environment. A network of base stations arranged in a regular hexagonal grid is considered. Each base station employs 120° sectorized transmit antennas and thus serves three cells. Additionally, each cell contains two RRUs at a distance of 2/3 the cell radius (see BS 1 in Fig. 1). The throughput performance with and without RRUs is shown in the left and right parts of Fig. 5, respectively. Without RRUs $N_t = 8$ transmit antennas are collocated at the base station, and with RRUs two antennas are placed at each RRU leaving four collocated antennas for the base station. Each receiver is equipped with $N_r = 4$ antennas.

In the simulations, different SU-MIMO and MU-MIMO transceivers are compared, assuming either perfect or quantized CSIT. In SU-MIMO, users are served on separate time/frequency resources, thus avoiding in-cell interference between users. In a MU-MIMO system, users can additionally be multiplexed in the spatial domain. In this case, interference can be avoided by appropriate pre-processing at the transmitter, e.g., employing the simple linear precoding technique known as ZF beamforming [3]. The advantage of such techniques is that the potential spatial multiplexing gain is only limited by the number of transmit antennas, whereas in SU-MIMO the minimum of N_t and N_r is the decisive factor. Additionally, high receive antenna correlation often limits the spatial multiplexing capabilities of hand held devices strictly below N_r, a problem that is altogether circumvented in MU-MIMO because different users typically experience uncorrelated channel conditions. But there is also a downside to MU-MIMO: Perfect interference-cancellation is only achieved with perfect CSIT, otherwise residual interference impairs the transmission. Note that we restrict the MU-MIMO system to transmit at most one stream per user for simplicity.

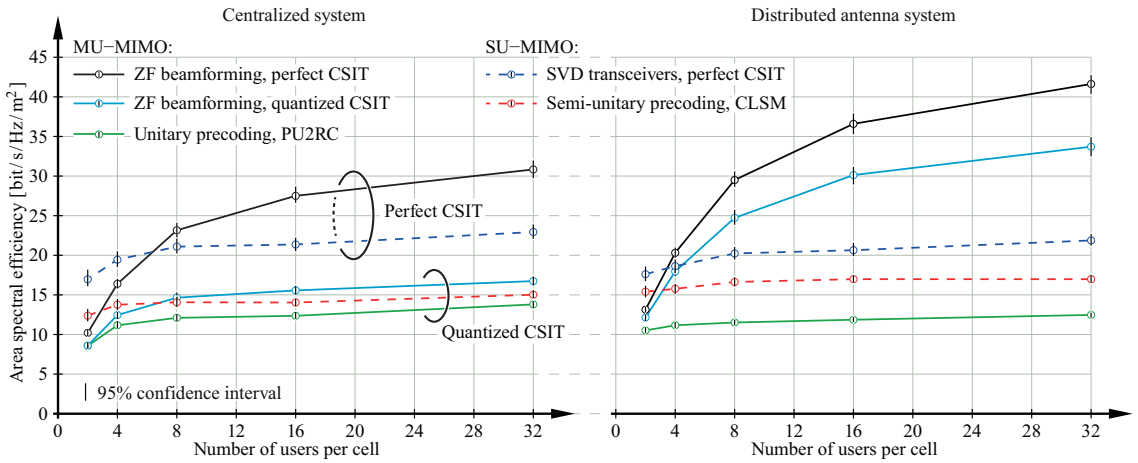


FIGURE 5. Comparison of the area spectral efficiency obtained in a cellular network without (left-hand side) and with (right-hand side) distributed antennas versus the number of users per cell. The performance with different transceivers and CSI feedback algorithms is compared. The total amount of transmit antennas per cell equals eight. The users are equipped with fourss receive antennas.

With perfect CSIT, Fig. 5 shows that ZF beamforming based MU-MIMO outperforms SU-MIMO with capacity achieving singular value decomposition (SVD) based transceivers, as soon as there are enough users per cell (≥ 8) to exploit the spatial multiplexing capabilities of the base station. Considering quantized CSIT, ZF beamforming performs similar to LTE's CLSM in the centralized system, while in the DAS a large throughput gain is achieved. This gain is enabled by investing the available feedback bits in those antennas of the distributed antenna array that experience a small macroscopic pathloss, thus exploiting the macro-diversity of the DAS. On the other hand, MU-MIMO with unitary precoding based on per-user unitary beamforming and rate control (PU2RC) achieves a lower throughput than SU-MIMO in both systems, because it cannot exploit the macro-diversity. Note that all considered algorithms have the same feedback overhead (an 8 bit memoryless quantization codebook is used in all cases). For details on the considered transceiver architectures and feedback algorithms the interested reader is referred to [29], [76].

D. MACRO-FEMTO OVERLAY NETWORKS

One of the most efficient methods to enhance capacity in a macro cellular network is to minimize the distance between transmitter and receiver [77]. This can be realized with smaller cell sizes and achieves the twofold benefits of increased spatial reuse and improved link quality. However, it comes at the cost of additional interference and required infrastructure. Femto-cells are user-deployed low-power base stations, which offer an economical way to realize small cells in existing macro cellular networks. Since they primarily belong to the unplanned part of the network, it is one of the network providers' major concerns, how the link quality of macro-cell attached users is influenced by a femto-cell deployment. We investigate this question by enforcing two approaches:

1) **Stochastic system model**: In order to carry out system-level simulations, it has been agreed in standardization meetings on models, such as the dual-stripe or the 5×5 grid model [78]. Although these models improve reproducibility, they do not meet scientific researchers' claim for analytical treatability. For this reason, we incorporate femto-cell deployments in our LTE system-level simulator based on *stochastic geometry*. We utilize Poisson point- as well as cluster-processes, which not only reflect the opportunistic placement of femto-cells, but also provide analytically tractable expressions for performance metrics like outage probability and SINR [77], [79]. The results obtained with these "simple" models indicate the same trends as the more elaborated environments mentioned above.
2) **Fairness metric**: Based on the stochastic models, we investigate *how many femto-cells can be beneficially deployed in an existing macro-cell*, which arises the prior question: *Beneficial in which sense*, which relates to the question posed in [20], if capacity is really the only metric that should be considered during system optimization.

   In our setup, clusters of users are spread homogeneously over the macro-cell area (according to a Poisson cluster process). Then, one by one, femto-cell access points are added to the network and placed at the center of these clusters. Thus, an increasing amount of users is in coverage of a femto-cell while the total number of users remains constant. Fig. 6 depicts simulation results for the average user throughput (middle solid line) plotted versus the femto-cell density, i.e., the number of employed femto-cell access points per macro-cell. The curve indicates performance improvements for an increasing number of femto-cells. However, it conceals the imbalance of femto-cell and macro-cell user throughput, as shown by the upper- and lower solid line in Fig. 6, respectively. Therefore, we emphasize the importance of a fairness metric to quantify the distribution of the throughput values among the users. In our work, we utilize Jain's fairness index [80], as shown by the dashed line in the figure.

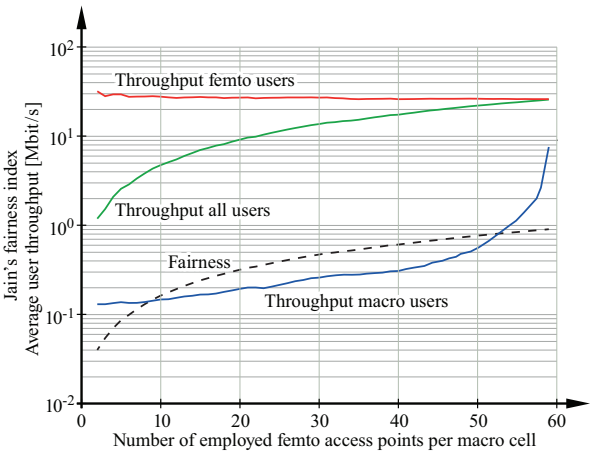


FIGURE 6. Comparison of the average user throughput achieved by macro-cell users, femto-cell users and both types combined versus the number of femto access points per macro-cell. Also shown is the fairness of the resource allocation in terms of Jain's fairness index.

Succinctly, we stress the significance of incorporating various metrics into the performance assessment of heterogeneous cellular networks, and encourage to apply stochastic geometry for the system models of such networks. A more detailed investigation of femto-cell deployments in user hotspot scenarios can be found in [81].

V. CONCLUSIONS

To sustain the exponential growth in mobile data-traffic currently experienced by network operators all over the world, a close cooperation between researchers and standardization/industry experts is required, to identify key technologies that can cope with this demand in the near future. In this article, we present our approach to bridge the gap

between researchers and practitioners working on LTE cellular networks by means of the *Vienna LTE simulators*, a standard-compliant open-source Matlab-based simulation platform. The simulators facilitate a standard-compliant validation of novel research results and simplify the evaluation of such results in terms of their significance for practical systems. Furthermore, a unifying platform greatly improves the reproducibility and comparability of different algorithms, by removing uncertainties in the multitude of simulation parameters encountered in such complex systems.

Throughout the article we provide an overview of LTE standard-related research efforts pursued by scientists and engineers world-wide. We also demonstrate the capabilities of the *Vienna LTE simulators* with examples of the research-work conducted in our group with the aid of the simulators. Topics such diverse as pilot-power allocation and frequency synchronization, multi-user scheduling and beamforming, and heterogeneous networking architectures can be effectively treated and investigated with the simulator suite.

We thus encourage researches and engineers, whose field of work is related to LTE/LTE-A, to take a closer look at the *Vienna LTE simulators* and find out whether they can benefit from using this platform.

REFERENCES

[1] *Cisco Visual Networking Index: Forecast Update, 2012–2017*, Cisco Systems, Inc., San Jose, CA, USA, Feb. 2013.
[2] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
[3] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[4] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
[5] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.
[6] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 1–5, Nov. 2006.
[7] B. Nosrat-Makouei, J. Andrews, and R. Heath, Jr., "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2783–2794, Jun. 2011.
[8] T. Wild, "A rake-finger based efficient channel state information feedback compression scheme for the MIMO OFDM FDD downlink," in *Proc. IEEE 71st Veh. Technol. Conf.*, May 2010, pp. 1–5.
[9] T. Inoue and R. Heath, Jr., "Grassmannian predictive coding for limited feedback multiuser MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 3076–3079.
[10] O. El Ayach and R. Heath, Jr., "Grassmannian differential limited feedback for interference alignment," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6481–6494, Dec. 2012.
[11] S. Schwarz and M. Rupp, "Adaptive channel direction quantization–enabling multi user MIMO gains in practice," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 6947–6952.
[12] "E-UTRA: Physical channels and modulation," 3GPP, Austin, Texas, Tech. Rep. TS 36.211, Dec. 2008.
[13] Y.-H. Nam, Y. Akimoto, Y. Kim, M. I. Lee, K. Bhattad, and A. Ekpenyong, "Evolution of reference signals for LTE-advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 132–138, Feb. 2012.
[14] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.
[15] T.-T. Tran, Y. Shin, and O.-S. Shin, "Overview of enabling technologies for 3GPP LTE-advanced," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–12, 2012.
[16] *IEEE Standard for Air Interface for Broadband Wireless Access Systems*, IEEE Standard 802.16, Jun. 2012.
[17] *IEEE Standard for Information Technology—Specification Requirements Part 11: Wireless LAN MAC and PHY Specifications*, IEEE Standard 802.11, Mar. 2012.
[18] (2008, Dec.). *Technical Specification Group Radio Access Networks: Deployment Aspects (Release 8)* [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25943.htm
[19] L. Hentilä, P. Kyösti, M. Käske, M. Narandzic, and M. Alatossava. (2007, Dec.). *MATLAB Implementation of the WINNER Phase II Channel Model Ver1.1* [Online]. Available: https://www.ist-winner.org/phase_2_model.html
[20] M. Dohler, R. Heath, Jr., A. Lozano, C. Papadias, and R. Valenzuela, "Is the PHY layer dead?" *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 159–165, Apr. 2011.
[21] L. Hanzo, H. Haas, S. Imre, D. O'Brien, M. Rupp, and L. Gyongyosi, "Wireless myths, realities, and futures: From 3G/4G to optical and quantum wireless," *Proc. IEEE*, vol. 100, no. 13, pp. 1853–1888, 2012.
[22] C. Mehlführer, J. C. Ikuno, M. Simko, S. Schwarz, and M. Rupp, "The Vienna LTE simulators—Enabling reproducibility in wireless communications research," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 1–14, 2011.
[23] J. Blumenstein, J. C. Ikuno, J. Prokopec, and M. Rupp, "Simulating the long term evolution uplink physical layer," in *Proc. 53rd Int. Symp.*, Sep. 2011, pp. 141–144.
[24] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in *Proc. IEEE 71st Veh. Technol. Conf.*, May 2010, pp. 1–5.
[25] J. Ikuno, S. Pendl, M. Simko, and M. Rupp, "Accurate SINR estimation model for system level simulation of LTE networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 1471–1475.
[26] D. Bultmann, M. Muhleisen, K. Klagges, and M. Schinnenburg, "OpenWNS—Open wireless network simulator," in *Proc. Eur. Wireless Conf.*, May 2009, pp. 205–210.
[27] G. Piro, L. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: An open-source framework," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 498–513, Feb. 2011.
[28] R. Heath, Jr., T. Wu, Y. H. Kwon, and A. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4885–4899, Oct. 2011.
[29] S. Schwarz, R. Heath, Jr., and M. Rupp, "Single-user MIMO versus multi-user MIMO in distributed antenna systems with limited feedback," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 54, Mar. 2013.
[30] M. Lerch and M. Rupp, "Measurement-based evaluation of the LTE MIMO downlink at different antenna configurations," in *Proc. 17th Int. ITG Workshop Smart Antennas*, Mar. 2013, pp. 1–6.
[31] S. Caban, J. Garcia Naya, and M. Rupp, "Measuring the physical layer performance of wireless communication systems: Part 33 in a series of tutorials on instrumentation and measurement," *IEEE Instrum. Meas. Mag.*, vol. 14, no. 5, pp. 8–17, Oct. 2011.
[32] S. Caban, C. Mehlführer, M. Rupp, and M. Wrulich, *Evaluation of HSDPA and LTE: From Testbed Measurements to System Level Performance*. New York, NY, USA: Wiley, 2012.
[33] S. Schwarz, M. Simko, and M. Rupp, "On performance bounds for MIMO OFDM based wireless communication systems," in *Proc. IEEE 12th Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2011, pp. 311–315.
[34] T. Weber, A. Sklavos, and M. Meurer, "Imperfect channel-state information in MIMO transmission," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 543–552, Mar. 2006.
[35] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.
[36] X. Dai, W. Zhang, J. Xu, J. Mitchell, and Y. Yang, "Kalman interpolation filter for channel estimation of LTE downlink in high-mobility environments," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, p. 232, 2012.
[37] L. Yang, G. Ren, B. Yang, and Z. Qiu, "Fast time-varying channel estimation technique for LTE uplink in HST environment," *IEEE Trans. Veh. Technol.*, vol. 61, no. 9, pp. 4009–4019, Nov. 2012.
[38] M. Meidlinger, M. Šimko, Q. Wang, and M. Rupp, "Channel estimators for LTE-A downlink fast fading channels," in *Proc. Int. ITG Workshop Smart Antennas*, Mar. 2013, pp. 1–5.

[39] Y. Liu and S. Sezginer, "Iterative compensated MMSE channel estimation in LTE systems," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 4862–4866.
[40] T. Peng, X. Lou, J. Xu, Q. Zhou, X. Li, and Y. Chen, "An improved iterative channel estimation based LTE downlink," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, May 2012, pp. 1721–1724.
[41] W. J. Hwang, J.-H. Jang, and H.-J. Choi, "An enhanced channel estimation method for MU-MIMO based LTE-Advanced system," in *Proc. 17th Asia-Pacific Conf. Commun.*, Oct. 2011, pp. 163–167.
[42] Y. Ohwatari, N. Miki, T. Abe, S. Nagata, and Y. Okumura, "Investigation on improvement in channel estimation accuracy using data signal muting in downlink coordinated multiple-point transmission and reception in LTE-Advanced," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2011, pp. 1288–1293.
[43] B. Deng and W. Wang, "Low complexity channel estimation method of MIMO-OFDM for LTE downlink," in *Proc. 7th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Jul. 2011, pp. 1–4.
[44] X. Ma and W. Xu, "Low-complexity channel estimation for 3GPP LTE terminals," in *Proc. 18th Eur. Wireless Conf.*, Apr. 2012, pp. 1–7.
[45] Q. Wang and M. Rupp, "Analytical link performance evaluation of LTE downlink with carrier frequency offset," in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2011, pp. 1959–1962.
[46] Q. Wang, M. Simko, and M. Rupp, "Performance analysis of LTE downlink under symbol timing offset," in *Proc. 16th Int. ITG Workshop Smart Antennas*, Mar. 2012, pp. 41–45.
[47] X. Li and J. Wang, "A novel time and frequency synchronization algorithm for LTE system," in *Proc. 7th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2011, pp. 1–4.
[48] H. Hwang and D. Kim, "Practical time synchronization for OFDM systems on mobile channel," in *Proc. 21st Int. Conf. Comput. Commun. Netw.*, Aug. 2012, pp. 1–5.
[49] S. Huang, Y. Su, Y. He, and S. Tang, "Joint time and frequency offset estimation in LTE downlink," in *Proc. 7th Int. ICST Conf. Commun. Netw. China*, Aug. 2012, pp. 394–398.
[50] S. Schwarz and M. Rupp, "Throughput maximizing feedback for MIMO OFDM based wireless communication systems," in *Proc. 12th Int. Conf. Signal Process. Adv. Wireless Commun.*, Jun. 2011, pp. 316–320.
[51] L. Cui, C. Qi, Y. Chen, and Z. Shao, "Adaption feedback of precoding matrix indicator for 3GPP LTE/LTE-A system," in *Proc. 4th Int. High Speed Intell. Commun. Forum*, May 2012, pp. 1–4.
[52] W. Fu, F. Zhang, G. Yantao, and Z. Cui, "Research on the selection algorithm of precoding matrix in TDD LTE," in *Proc. 3rd Int. Conf. Commun. Softw. Netw.*, May 2011, pp. 365–368.
[53] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems," in *Proc. 38th Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 1–6.
[54] H. Kwon, J. Lee, and I. Kang, "Successive interference cancellation via rank-reduced maximum a posteriori detection," *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 628–637, Feb. 2013.
[55] R. Dallinger, H. Ruotsalainen, R. Wichman, and M. Rupp, "Adaptive pre-distortion techniques based on orthogonal polynomials," in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1945–1950.
[56] S. Wang, X. Lou, T. Peng, R. Ma, and Q. Wei, "A novel design of reference signal scheme for LTE," in *Proc. 7th Int. Conf. Natural Comput.*, vol. 3. Jul. 2011, pp. 1795–1799.
[57] M. Simko, P. Diniz, Q. Wang, and M. Rupp, "New insights in optimal pilot symbol patterns for OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2013, pp. 1–4.
[58] M. Simko, Q. Wang, and M. Rupp, "Optimal pilot patterns under time-variant channels," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2013, pp. 1–6.
[59] M. Šimko, S. Pendl, S. Schwarz, Q. Wang, J. C. Ikuno, and M. Rupp, "Optimal pilot symbol power allocation in LTE," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2011, pp. 1–5.
[60] M. Šimko, Q. Wang, and M. Rupp, "Optimal pilot symbol power allocation under time-variant channels," *EURASIP J. Wireless Commun. Netw.*, 2012, pp. 1–5.
[61] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, Jul.–Aug. 1968.
[62] M. Šimko, P. S. R. Diniz, Q. Wang, and M. Rupp, "Power efficient pilot symbol power allocation under time-variant channels," in *Proc. IEEE Veh. Tech. Conf.*, Sep. 2012, pp. 1–5.
[63] Q. Wang, C. Mehlführer, and M. Rupp, "Carrier frequency synchronization in the downlink of 3GPP LTE," in *Proc. 21st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2010, pp. 939–944.
[64] G. Liu, X. Su, G. Wu, and S. Li, "Coordinated scheduling with adaptive feedback in cooperative multi-cell downlink system," in *Proc. 1st IEEE Int. Conf. Commun. China*, Aug. 2012, pp. 647–652.
[65] B. Clerckx, H. Lee, Y. Hong, and G. Kim, "A practical cooperative multicell MIMO-OFDMA network based on rank coordination," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, pp. 1–11, Feb. 2013.
[66] E. Pateromichelakis, M. Shariat, A. Quddus, and R. Tafazolli, "On the evolution of multi-cell scheduling in 3GPP LTE/LTE-A," *IEEE Commun. Surv. Tut.*, vol. PP, no. 99, pp. 1–17, Jul. 2012.
[67] L. Qiang, Y. Yang, F. Shu, and W. Gang, "Coordinated beamforming in downlink CoMP transmission system," in *Proc. 5th Int. Conf. Commun. Netw. China*, Aug. 2010, pp. 1–5.
[68] C.-B. Chae, I. Hwang, R. Heath, Jr., and V. Tarokh, "Interference aware-coordinated beamforming in a multi-cell system," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3692–3703, Oct. 2012.
[69] H. Sun, W. Fang, J. Liu, and Y. Meng, "Performance evaluation of CS/CB for coordinated multipoint transmission in LTE-A downlink," in *Proc. IEEE 23rd Int. Symp. Pers. Indoor Mobile Radio Commun.*, Sep. 2012, pp. 1061–1065.
[70] B. Wang, B. Li, and M. Liu, "A novel precoding method for joint processing in CoMP," in *Proc. Int. Conf. Netw. Comput. Inf. Security*, vol. 1. May 2011, pp. 126–129.
[71] N. El-Din, E. Sourour, K. Seddik, and I. Ghaleb, "A modified joint processing technique for coordinated multi-point transmission in LTE-advanced," in *Proc. IEEE Int. Conf. Wireless Inf. Technol. Syst.*, Nov. 2012, pp. 1–4.
[72] X. Tao, X. Xu, and Q. Cui, "An overview of cooperative communications," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 65–71, Jun. 2012.
[73] R. Heath, Jr., M. Airy, and A. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," in *Proc. 35th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2001, pp. 1194–1199.
[74] S. Schwarz, C. Mehlführer, and M. Rupp, "Low complexity approximate maximum throughput scheduling for LTE," in *Proc. 44th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2010, pp. 1563–1569.
[75] S. Schwarz, R. Heath, Jr., and M. Rupp, "Multiuser MIMO in distributed antenna systems with limited feedback," in *Proc. 4th Int. Workshop Heterogeneous Small Cell Netw.*, Dec. 2012, pp. 546–551.
[76] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on a Grassmann-manifold for limited feedback beamforming systems," in *Proc. IEEE 20th Eur. Signal Process. Conf.*, Nov. 2013, pp. 2536–2540.
[77] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Found Trends Netw.*, vol. 3, no. 2, pp. 127–248, Feb. 2009.
[78] *Simulation Assumptions and Parameters for FDD HeNB RF Requirements*, Standard 3GPP RAN4, Mar. 2009.
[79] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.
[80] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Estern Research Lab., Hudson, MA, USA, Tech. Rep. TR-301, Sep. 1984.
[81] M. Taranetz and M. Rupp, "Performance of femtocell access point deployments in user hot-spot scenarios," in *Proc. Australasian Telecommun. Netw. Appl. Conf.*, Nov. 2012, pp. 1–5.

STEFAN SCHWARZ was born in Neunkirchen, Austria, in 1984. He received the B.Sc. degree in electrical engineering and the Dipl.-Ing. degree (M.Sc. equivalent) in telecommunications engineering from the Vienna University of Technology, Vienna, Austria, in 2007 and 2009, respectively, and with highest distinctions. In 2010, he received the Honorary Price of the Austrian Minister of Science and Research 2010, for excellent graduates of scientific and artistic universities. Since 2008, he is a Project Assistant with the Mobile Communications Group, Institute of Telecommunications, Vienna University of Technology. His current research interests include limited feedback single- and multi-user MIMO-OFDM communication systems and multi-user scheduling algorithms. He is currently pursuing the Ph.D. degree.

JOSEP COLOM IKUNO was born 1984 in Barcelona, Spain. He received the Bachelors degree in telecommunications engineering from Universitat Pompeu Fabra, Barcelona, in 2005, and the Master's degree in telecom engineering from Universitat Politècnica de Catalunya, Catalunya, Spain, after a stay in Vienna for the Master's thesis in 2008, both with honours. Since 2008, he is with the Vienna University of Technology, Mobile Communications Group, Vienna, Austria, where he works on the Ph.D. thesis in LTE link and system level simulation, modelling and optimization. His current research interests include video coding, OFDM systems, system level simulation, and modelling of wireless networks with emphasis on the integration of real operator data in the simulations and models.

MICHAL ŠIMKO was born in Bratislava, Slovakia, in October 1985. He received the Dipl.-Ing. degree (Hons.) in electrical engineering from the Vienna University of Technology, Vienna, Austria, in June 2009. Since August 2009, he is employed as a Project Assistant with the Institute of Telecommunications, Vienna University of Technology, and he is currently pursuing the Ph.D. degree in telecommunications engineering. In June 2011, he received the Bachelors degree in industrial–organizational and social psychology from Comenius University, Bratislava. He has been awarded the Best Student Paper Award by IEEE VTS Society at IEEE VTC 2011 Spring in Budapest.

MARTIN TARANETZ received the B.Sc. degree in electrical engineering and Dipl.-Ing degree (M.Sc. equivalent) in telecommunications from the Vienna University of Technology, Vienna, Austria, in 2008 and 2011, respectively. His Diploma thesis focused on capacity optimization with fractional frequency reuse. He is currently pursuing the Ph.D. degree in telecommunications engineering and he is employed as a University Assistant with the Institute of Telecommunications, Vienna University of Technology. His current research interests include heterogeneous cellular networks and statistical characterization of network interference.

QI WANG was born in Beijing, China, in November 1982. She received the Bachelors degree in telecommunication engineering from the Beijing University of Posts and Telecommunications, Beijing, in 2005. In 2007, after finishing the thesis ''Study of Channel Estimation in MIMO-OFDM for Software Defined Radio'', she received the Master degree in computer science and engineering from Linköping University, Linköping, Sweden. In November 2007, she joined the Institute of Telecommunications, Vienna University of Technology, as a Project Assistant and Ph.D. Student. She received the Dr.-Ing. degree from the Vienna University of Technology in 2013. Currently, she is an R&D Engineer with AKG Acoustics GmbH.

MARKUS RUPP received the Dipl.-Ing. degree from the University of Saarbruecken, Saarbruecken, Germany, in 1988, and the Dr.-Ing. degree from Technische Universitaet Darmstadt, Darmstadt, Germany, in 1993, where he was with Eberhardt Haensler, designing new algorithms for acoustical and electrical echo compensation. From November 1993 to July 1995, he held a Post-Doctoral position with the University of Santa Barbara, Santa Barbara, CA, USA, with S. Mitra, where he worked with A. H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 to August 2001, he was a Technical Staff Member with the Wireless Technology Research Department, Bell-Labs, Crawford Hill, NJ, USA, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11 and UMTS. Since October 2001, he is a Full Professor for Digital Signal Processing in Mobile Communications, Vienna University of Technology, Vienna, Austria, where he founded the Christian-Doppler Laboratory for Design Methodology of Signal Processing Algorithms in 2002 at the Institute of Telecommunications. He was an Associate Editor of the IEEE Transactions on Signal Processing from 2002 to 2005, is currently an Associate Editor of *EURASIP Journal of Advances in Signal Processing* and *EURASIP Journal on Embedded Systems*. He is elected AdCom member of EURASIP since 2004 and served as a President of EURASIP from 2009 to 2010. He has authored or co-authored more than 400 scientific papers and patents on adaptive filtering, wireless communications, and rapid prototyping, as well as automatic design methods.

IEEE Access
The journal for rapid open access publishing

Received April 27, 2015, accepted May 13, 2015, date of publication June 1, 2015, date of current version June 9, 2015.

Digital Object Identifier 10.1109/ACCESS.2015.2437903

Runtime Precoding: Enabling Multipoint Transmission in LTE-Advanced System-Level Simulations

MARTIN TARANETZ, (Student Member, IEEE), THOMAS BLAZEK, THOMAS KROPFREITER, MARTIN KLAUS MÜLLER, (Student Member, IEEE), STEFAN SCHWARZ, (Member, IEEE), AND MARKUS RUPP, (Fellow, IEEE)

Institute of Telecommunications, Vienna University of Technology, Vienna 1040, Austria

Corresponding author: M. Taranetz (martin.taranetz@nt.tuwien.ac.at)

This work has been funded by the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility, the A1 Telekom Austria AG, and the KATHREIN-Werke KG. The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged. The authors would like to thank the TU Wien for funding the publication fees.

ABSTRACT System-level simulations have become an indispensable tool for predicting the behavior of wireless cellular systems. As exact link-level modeling is unfeasible due to its huge complexity, mathematical abstraction is required to obtain equivalent results by less complexity. A particular problem in such approaches is the modeling of multiple coherent transmissions. Those arise in multiple-input-multiple-output transmissions at every base station but nowadays so-called coordinated multipoint (CoMP) techniques have become very popular, allowing to allocate two or more spatially separated transmission points. Also, multimedia broadcast single frequency networks (MBSFNs) have been introduced recently in long-term evolution (LTE), which enables efficient broadcasting transmission suitable for spreading information that has a high user demand as well as simultaneously sending updates to a large number of devices. This paper introduces the concept of runtime-precoding, which allows to accurately abstract many coherent transmission schemes while keeping additional complexity at a minimum. We explain its implementation and advantages. For validation, we incorporate the runtime-precoding functionality into the Vienna LTE-A downlink system-level simulator, which is an open source tool, freely available under an academic noncommercial use license. We measure simulation run times and compare them against the legacy approach as well as link-level simulations. Furthermore, we present multiple application examples in the context of intrasite and intersite CoMP for train communications and MBSFN.

INDEX TERMS Link abstraction, link quality model, runtime-precoding, 3GPP, LTE-A, Vienna LTE-A downlink system level simulator, MIESM, Vienna LTE-A downlink link level simulator, LTE transmission modes, coordinated multipoint, multimedia broadcast single frequency networks, high-user mobility.

I. INTRODUCTION

3gpp LTE-Advanced (LTE-A) is a key technology to meet the requirements of fourth generation mobile communication systems (4G), as specified by the ITU-R [1], [2]. Today, its limits are continuously pushed by academia, industry and standardization bodies. Typically, the potential performance of novel contributions is evaluated by extensive *simulations* under realistic conditions before gaining acceptance. To significantly reduce the complexity of such process, evaluations are often separated into two stages or levels of abstraction, referred to as *link level* and *system level*, respectively [3].

Link level simulations are applied for assessing the performance of the physical layer as well as those higher layer aspects that are directly related to the radio interface. Typically, only a single link is evaluated, including features such as synchronization, modulation and coding, channel fading, channel estimation and multi-antenna processing. It is by far too complex to simulate a substantial number of such links.

The focus of the present work is on system level, where performance evaluation requires to encompass a large number of network elements and upscales the number of interconnecting links. In this case, the main interest lies in network-related issues such as resource allocation, mobility management and network planning. Hence, computational complexity needs to be decreased substantially in order to make the problem feasible. A widely accepted solution is the application of *link abstraction models* [3] that specify the interaction between

link- and system level simulators and are further detailed in Section II. This structure is also expected to persist in simulation tools for the fifth generation of wireless cellular networks (5G) [4].

A. SURVEY ON EXISTING SIMULATION TOOLS

There are several ways to categorize existing system-level simulation tools. First, we may distinguish between simulators, which are implemented as 'modules' of a larger suite, and those, which are specifically designed for LTE-A [5]. Examples for the former include the *Riverbed SteelCentral NetModeler* (formerly *OPNET Modeler Suite*) [6], *OMNeT++* [7], *IT++* [8], *ns-2* [9], [10], *GNS3* [11], *openWNS* [12] and *Hurricane II* [13]. The main drawback of these solutions is their low level of detail, hence leaving most implementation work to the user. Consequently, results often lack accuracy and verification. On the other hand, technology-specific simulators are mainly developed by network operators and vendors, and are typically not intended for commercial distribution [5], [14]–[16]. Such tools yield a broad range of possibilities for parameter calibration and statistical evaluation. Thus, they are key instruments for the standardization process and the development of new technologies.

While these two classes of simulators largely vary in complexity, scalability and usability, probably the most relevant difference for scientific research is their accessibility. The authors strongly believe that *open access* is a key prerequisite for reproducible simulation studies. The short list of openly available, technology-specific approaches includes *LTE-Sim* [17], the tool presented in [16] and the *Vienna LTE-A downlink system level simulator*. While the first lacks of detailed Multiple-Input-Multiple-Output (MIMO) modeling, the second provides a rather limited set of features. In this work, we employ the *Vienna LTE-A downlink system level simulator* [18] (latest version: v1.8 r1375), subsequently referred to as *Vienna LTE-A simulator*, which is briefly introduced in the following.

B. THE VIENNA LTE-A DOWNLINK SYSTEM LEVEL SIMULATOR

The Vienna LTE-A simulator is implemented in object-oriented MATLAB[1] and is made openly available for download under an academic, non-commercial use license. Its rich set of features and easy adaptability has led to numerous publications from researchers all over the globe, including studies on energy-efficient cell-coordination schemes [20], handover algorithms in self-optimizing networks [21], and resource allocation techniques for femtocell networks [22] as well as for machine-to-machine communication [23]. On top of that, the *open accessibility* warrants the reproducibility of these contributions. Today (May 2015), the simulator counts more than 30 000 downloads and undergoes permanent peer-review from a substantially large online community. With some 100 000 lines of code, employing a large forum with active users is the only method to guarantee its quality. For a comprehensive description, the interested reader is referred to [24].

[1]For further information, see [19].

C. CONTRIBUTIONS AND ORGANIZATION

This work is organized as follows. Section II briefly summarizes concepts for LTE-A physical layer modeling on system level. The centerpiece of this contribution is the presentation of a new *runtime-precoding* concept in Section III. Its innovation for openly available system level simulation tools lies in enabling the evaluation of scenarios where signals are coherently transmitted from multiple spatially separated transmission points while not compromising the complexity reduction as achieved by well established link abstraction models. For verifying the efficiency of our solution, we extend the link quality model of the Vienna LTE-A simulator by a runtime-precoding feature. We compare simulation run times as obtained with the enhanced- and the legacy model as well as with link level simulations. Section IV presents two example applications, showing the efficiency of intra- and inter-site Coordinated Multipoint (CoMP) schemes in high-speed train scenarios, and the implementation of eNodeB-coordinating entities in Multimedia Broadcast Single Frequency Networks (MBSFNs). Their investigation represent a novelty in the field of open-source LTE-A simulators, since both only became feasible with the proposed runtime-precoding concept. Section V concludes the work and outlines directions for future enhancements. The simulation assumptions throughout the paper are largely based on recommendations from 3GPP [25]–[27] and ITU-R [28], respectively. The source files for reproducing the results are openly available for download on our website www.nt.tuwien.ac.at/ltesimulator. Note that the proposed *runtime-precoding* concept is applicable to *any* system level simulation tool that employs the well accepted separation between link quality- and link performance model, as explained in Section II. Thus, the Vienna LTE-A simulator is just one out of many possible tools, which has been chosen to demonstrate the capabilities of the method.

II. PHYSICAL LAYER MODELING

This section provides a brief introduction to modeling concepts of the physical layer of LTE-A on system level. The LTE-A PHY procedures can conceptually be described as a Bit-Interleaved Coded

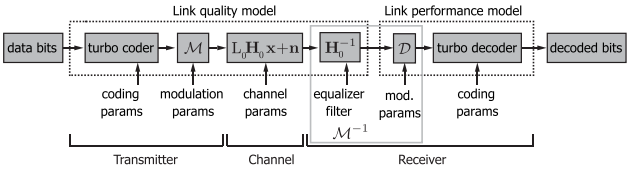


FIGURE 1. Separation of an LTE link into link quality- and link performance model. The link can equivalently be described as an LTE BICM transmitter-receiver chain [24].

Modulation (BICM)-system [24], as shown in Figure 1. It comprises a transmitter including channel coder, bit interleaver and modulator ($\mathcal{M}$). In LTE-A, coding and interleaving is achieved by a turbo-coder in combination with rate matching. The symbol mapping employs 4-, 16- and 64-QAM with Gray mapping, respectively. Signal propagation over an $N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}$ MIMO channel is commonly modeled by slowly-varying, position-dependent macro-scale fading L_0, small-scale fading $\mathbf{H}_0$ and Additional White Gaussian Noise (AWGN). The matrix representation follows from the assumption that the cyclic prefix exceeds the channel length, hence omitting inter-symbol interference. The channel coefficients are typically calculated from a power-delay profile or a ray-based spatial channel model, such as the Winner model [29] or 3GPP's 3D model [30]. The receiver encompasses an equalizer filter and a demodulator ($\mathcal{M}^{-1}$) as well as a turbo decoder, which provides de-interleaving and channel decoding.[2]

The objective of the *link abstraction model* is to predict the performance of the presented LTE-A link, given a parameterization of the inputs. For simplification, the model can be divided into a *link quality*- and a *link performance model*, as indicated in Figure 1. The link quality model measures the quality of the received signal after equalization.[3] The link performance model translates this measure into Block-Error Ratio (BLER) and further into (area) spectral efficiency and effective throughput, based on the employed Modulation and Coding Scheme (MCS).

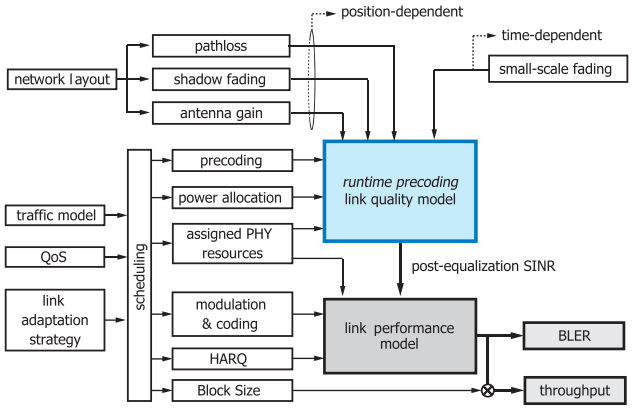


FIGURE 2. LTE link abstraction model with new *link quality model* as employed in the Vienna LTE-A downlink system level simulator v1.8 r1375.

The model in Figure 1 is a simplification of the actual link abstraction model, as it does not account for interference from other base stations. Its expansion to the whole network is illustrated in Figure 2. The figure identifies the main components of the model as network layout, time-variant fading and scheduling. It also illustrates the corresponding input-output relations to the link quality- and link performance model, respectively.

The Vienna LTE-A simulator employs a Mutual Information based exponential SNR Mapping (MIESM) for the SINR-to-BLER mapping [32], [33], which already proved beneficial in Release 5 of UMTS [34], and was shown to outperform all other approaches (e.g., Exponential effective SINR mapping (EESM) [35]) in both complexity and performance. This method compresses the Signal-to-Interference-plus-Noise Ratio (SINR) values of the assigned Resource Blocks (RBs) for each User Equipment (UE) and 1 ms-long subframe (subsequently also denoted as Transmission Time Interval (TTI)) into an *effective SINR*, yielding an AWGN-equivalent representation in terms of mutual information. These SINR values are then mapped to a BLER by means of an AWGN BLER curve of the corresponding MCS. The curves are obtained from LTE link level simulations, thus forming the *only* computationally costly physical layer evaluation, which is required for the link abstraction model.

III. RUNTIME-PRECODING

In existing open-source system level simulation tools, UE association is limited to a single eNodeB with all antennas being mounted at the same site. In such scenarios, the fading as experienced over a MIMO link can be decomposed into a slowly varying, position-dependent macro-scale component and a faster changing small-scale component, as shown in Figure 2. Macro-scale fading is determined by the network layout and comprises antenna directivity, path loss and shadowing. Small-scale fading represents fast, frequency-selective channel variations over time. As explained in Section II, it is commonly modeled by a normalized[4] $N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}$ channel matrix $\mathbf{H}_0$, where N_{Tx} and N_{Rx} denote the number of transmit- and receive antennas, respectively.

In the single-eNodeB-single-site case, the macro-scale parameter L_0 is a scalar, which is applied on all entries of $\mathbf{H}_0$. Thus, both L_0 and $\mathbf{H}_0$ can be computed *off-line* and *independently* from each other. Such separation further enables to determine the optimal precoder for each transmission rank a-priori with minimum loss of accuracy [24]. The *effective channel* $\mathbf{H} = \mathbf{G}\mathbf{H}_0\mathbf{W}$, which encompasses precoder $\mathbf{W}$ and receive filter $\mathbf{G}$, can be stored in *channel traces* and may be reused in all simulations with the *same MIMO setting*.

If the desired signal is received from *multiple* eNodeBs (e.g., in certain Coordinated Multipoint (CoMP) schemes, as indicated in Figure 3a) or from a single eNodeB with geographically separated antennas (e.g., in Distributed Antenna System (DAS)- and Remote Radio Head (RRH) deployments, as shown in Figure 3b), a-priori computation of the optimal

[2]In the current version of the Vienna LTE-A simulator, low complexity models for Zero Forcing (ZF)- and Minimum Mean Square Error (MMSE) receivers are available. The former approaches the average performance of an optimal receiver by exploiting Multi-User (MU) diversity, which is typically present in system level scenarios [31].

[3]Since the metric has to represent the quality of the input to the turbo decoder, the post-equalization Signal-to-Interference-plus-Noise Ratio (SINR) is a straightforward choice [24].

[4]All entries of H_i have unit mean power in ensemble average.

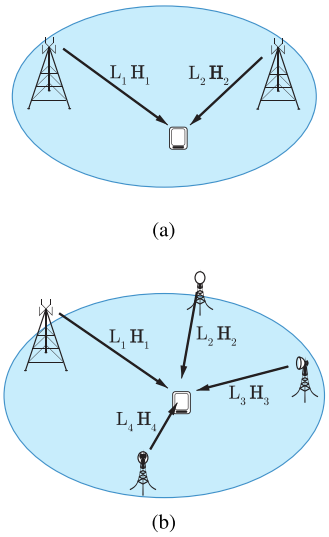


FIGURE 3. Spatially distributed transmission points. (a) CoMP scenario. (b) eNodeB with RRHs.

precoders and the corresponding receive filters is no longer possible. In this case, only the normalized small-scale fading matrices $\mathbf{H}_i$ for each transmitter site can be pre-generated. The *composite channel* of the desired signal does not become available until *runtime*. It is obtained by stacking the matrices $\mathbf{H}_i$ from each transmitter-site and weighting them with the corresponding macro-scale losses, L_i, i.e.,

$$\mathbf{H}'_0 = [L_1\mathbf{H}_1\ L_2\mathbf{H}_2\ L_3\mathbf{H}_3\ \ldots]. \tag{1}$$

Then, the *effective channel* is calculated as $\mathbf{H} = \mathbf{G}\,\mathbf{H}'_0\,\mathbf{W}$. In this case, the optimal precoder, $\mathbf{W}$, and the corresponding receive filter $\mathbf{G}$ are determined *at runtime*. The challenge is to enable coherent signal reception from spatially distributed transmission points while keeping additional computational complexity at a minimum. Subsequently, we introduce the so called *runtime-precoding* method.

A. IMPLEMENTATION

The functionality of *runtime-precoding* is implemented in the UE's link quality model, as indicated in Figure 2. Its major building blocks are outlined in Algorithm 1. The model enables coherent signal reception from spatially distributed sources, *which can be selected at runtime*. At first, it collects the normalized small-scale fading channels and macro-scale losses for both desired and interfering signals. Then, the channels are stacked according to (1).

In the next step, the corresponding precoders are determined. By default, they may be chosen from a 3GPP standard-compliant codebook[5] [25]. Nonetheless, the availability of the *full channel at runtime* allows researchers to apply arbitrary precoders and beamformers,[6] yielding a profound novelty in open-source LTE-A system level simulation tools.[7]

Algorithm 1: Proposed UE link Quality Model With Runtime-Precoding Functionality.

```
Result: post equalization SINR
collect macroscopic path losses and normalized channel
matrices from all transmitting sites;
calculate composite desired channel by stacking channel
matrices of desired signal;
determine precoder or beamformer;
if there are interferers then
    calculate composite interfering channels by stacking
    channel matrices of interferers;
    determine precoders or beamformers for interfering
    channels;
else
    noise power only;
end
calculate receive filters and effective channel matrices;
determine post equalization SINR and store for link
performance model;
calculate feedback based on actual channel;
```

B. PERFORMANCE EVALUATION

In this section, we evaluate the price to pay for enabling coherent multi-point transmission in system level simulations. For this purpose, we extend the UE link quality model of the Vienna LTE-A simulator by the runtime-precoding functionality according to Algorithm 1. Then we measure simulation run times with the new- and the legacy model. The results are compared with run times as obtained with the Vienna LTE-A Downlink Link Level simulator [39]. For a meaningful comparison, all simulations were carried out on the same hardware, an Intel(R) Core(TM) i7-3930K CPU @ 3.20 GHz, equipped with 32 GB of DDR3 1333 quad-channel RAM.

The common setup, which is employed in both link- and system level simulations, is summarized in Table 1. We carry out simulations with the LTE bandwidths $B = \{1.4, 3, 5, 10, 20\}$ MHz and the $N_{\mathrm{Tx}} \times N_{\mathrm{Rx}}$ antenna configurations $\{2 \times 2, 4 \times 2, 4 \times 1\}$ for various simulation lengths (measured in multiples of 1 TTI), in particular $N_{\mathrm{TTI}} = \{100, 500, 1000\}$ on system level and $N_{\mathrm{TTI}} = \{10, 100, 500\}$ on link level, respectively. Moreover, we perform system level simulations with $K = \{1, 10, 100\}$ UEs at a simulation length of 100 TTI. A round robin scheduler is employed.[8]

Figure 4 shows the obtained simulation run times. Each point was computed by averaging over ten simulation runs.

[5]There is no standardized method to determine the Precoding Matrix Indicator (PMI). The Vienna LTE-A downlink system level simulator employs a scheme that maximizes the mutual information between transmitted and received symbols [36].

[6]In general, beamforming strategies are found by solving multi-objective optimization problems [37].

[7]The implementation of 3GPP's 3D channel model will enable the investigation of *elevation beamforming* and full-dimension MIMO [38].

[8]Note that the round robin scheduler does not increase in complexity with the number of physical RBs (i.e., increasing B). Other scheduling algorithms may have a considerable impact on the simulation run time in multi-user scenarios [24].

TABLE 1. Simulation parameters as employed for the simulation run time evaluation.

Parameter	Value
Frequency	2.14 GHz
Number of eNodeBs	1
Transmit power	5 W
eNodeB antenna gain pattern	omni-directional
LTE transmission mode	CLSM
Path loss in dB	$\max\left(10\log_{10}\left(\frac{4\pi d f}{c_0}\right)^2, 0\right)$
Shadow fading	none
Channel model	ITU-R Ped-A, block fading
Receiver type	zero forcing
Noise power spectral density	−160 dBm/Hz
UE position	random in circle with $R = 250$ m
UE antenna gain pattern	omni-directional
Traffic model	full buffer
Channel knowledge	perfect
Feeback	AMC: CQI, MIMO: PMI and RI
Feedback delay	3 TTI

It is observed that on system level the results scale approximately linearly with the simulation length N_{TTI}, the bandwidth B and the number of UEs K. Compared to this, the link level results exhibit a slightly non-linear scaling with B (note that on link level, only a single link is evaluated, i.e., $K = 1$). Both link- and system level run times show a non-linear dependence on the number of transmit- and receive antennas, N_{Tx} and N_{Rx}, respectively. From these observations, we can derive the following generic run time estimator (for a better understanding of the scaling with N_{Tx} and N_{Rx}, additional simulations with the MIMO configurations $\{2 \times 1, 4 \times 1\}$ were carried out):

$$\begin{aligned} &T_{[\mathrm{s}]}(N_{\mathrm{TTI}}, B, N_{\mathrm{Tx}}, N_{\mathrm{Rx}}, K) \\ &= (c_0 + N_{\mathrm{TTI}} \cdot K \cdot B_{[\mathrm{MHz}]} \cdot (c_1 N_{\mathrm{Tx}} + c_2 N_{\mathrm{Rx}} + c_3 N_{\mathrm{Tx}} N_{\mathrm{Rx}}) \\ &\quad + \ldots + N_{\mathrm{TTI}} \cdot K \cdot B_{[\mathrm{MHz}]}^2 \cdot (c_4 N_{\mathrm{Tx}} + c_5 N_{\mathrm{Rx}} + c_6 N_{\mathrm{Tx}} N_{\mathrm{Rx}})) \\ &\quad \times \ldots \times 1\,\mathrm{s}, \end{aligned} \tag{2}$$

where $B_{[\mathrm{MHz}]} = B/10^6$ Hz. Next, we compute the coefficients $c_0, \ldots, c_6$ by linear least squares. The results are summarized in Table 2. Compared to the legacy link quality model, simulations with the new model require 1.25× longer for initialization (represented by the coefficient c_0), which is still 52.5× faster than link level simulations. It slightly decreases the scaling with $B_{[\mathrm{MHz}]} \cdot N_{\mathrm{Rx}}$ (referring to c_2) by 1.4× while increasing c_3 (according to the scaling with $B_{[\mathrm{MHz}]} \cdot N_{\mathrm{Tx}} \cdot N_{\mathrm{Rx}}$) by 2.7×. The latter term is of particular relevance for investigating massive MIMO scenarios with a large number of transmit antennas. On link level, c_2 (referring to a scaling proportional to $B_{[\mathrm{MHz}]} \cdot N_{\mathrm{Rx}}$) is 41.2× larger than on system level, while $c_3 = 0$ (corresponding to the scaling proportional to $B_{[\mathrm{MHz}]} \cdot N_{\mathrm{Rx}} \cdot N_{\mathrm{Tx}}$). On the other hand, the link level simulation run times scale with $N_{\mathrm{Tx}} \cdot B_{[\mathrm{MHz}]}^2$ and $N_{\mathrm{Rx}} \cdot B_{[\mathrm{MHz}]}^2$ (referring to c_4 and c_5), while system level simulations with both legacy- and new link quality model exhibit no dependency on $B_{[\mathrm{MHz}]}^2$, i.e., $c_4 = 0$, $c_5 = 0$ and $c_6 = 0$, respectively.[9]

IV. EXAMPLE APPLICATIONS

This section presents two examples for the application of the *runtime-precoding* enhanced UE link quality model. The simulation studies are carried out with the Vienna LTE-A simulator and, to the best of our knowledge, are the first of their kind with an openly-available LTE-A system level simulation tool.

A. HIGH-SPEED TRAIN SCENARIOS

This section investigates intra-site Joint Processing (JP) and inter-site Coordinated Scheduling (CS) [26] in the context of train communications. The particular scenario represents wireless access at high user mobility, which is becoming an increasingly important topic for commuting and traveling [40]–[42]. High-speed train scenarios have the specific feature of UEs moving in a deterministic manner and imposing a short but heavy traffic demand on the currently traversed cell. To support such traffic distribution, the application of RRHs is well suited [43], [44]. We apply the *runtime-precoding* enabled RRH feature of the Vienna LTE-A simulator to elaborate the impact of various RRH collaboration schemes on the UE performance. We assume direct links between the RRHs and the UEs, i.e., the trains are not equipped with roof-mounted relay nodes. This is a particularly realistic scenario for smaller countries such as most European, where mobile providers and train operators refrain from collaborating across country boarders.

1) SIMULATION SETUP

For simulations, we consider a representative segment of a railroad track, as shown in Figure 5. It comprises four equidistantly spaced sites, which are located along the tracks. Each site employs two RRHs pointing in opposite directions. For simplicity, only RRHs directed towards the train are taken into account. We assume the train to move between RRH_0 and RRH_1, and denote its location by the position of its center. Nodes RRH_{-1} and RRH_2 serve as dominant interferers.

Three ways of associating eNodeBs with RRHs are investigated:

- *Baseline:* All RRHs employ different cell IDs, i.e., are associated with different eNodeBs.
- *Coordination:* All RRHs are associated with different eNodeBs. Nodes RRH_0 and RRH_1 coordinate their transmission such that they do not interfere each other (e.g., by CS, thus effectively representing an example for *inter-site CoMP* [26]).
- *Cooperation:* Nodes RRH_0 and RRH_1 belong to the same eNodeB, thus acting as a DAS (as an example for *intra-site CoMP* [26]).

[9] A more detailed analysis of the link level simulation times shows that the non-linearity mainly arises from the symbol demapping and decoding, as described in Section II.

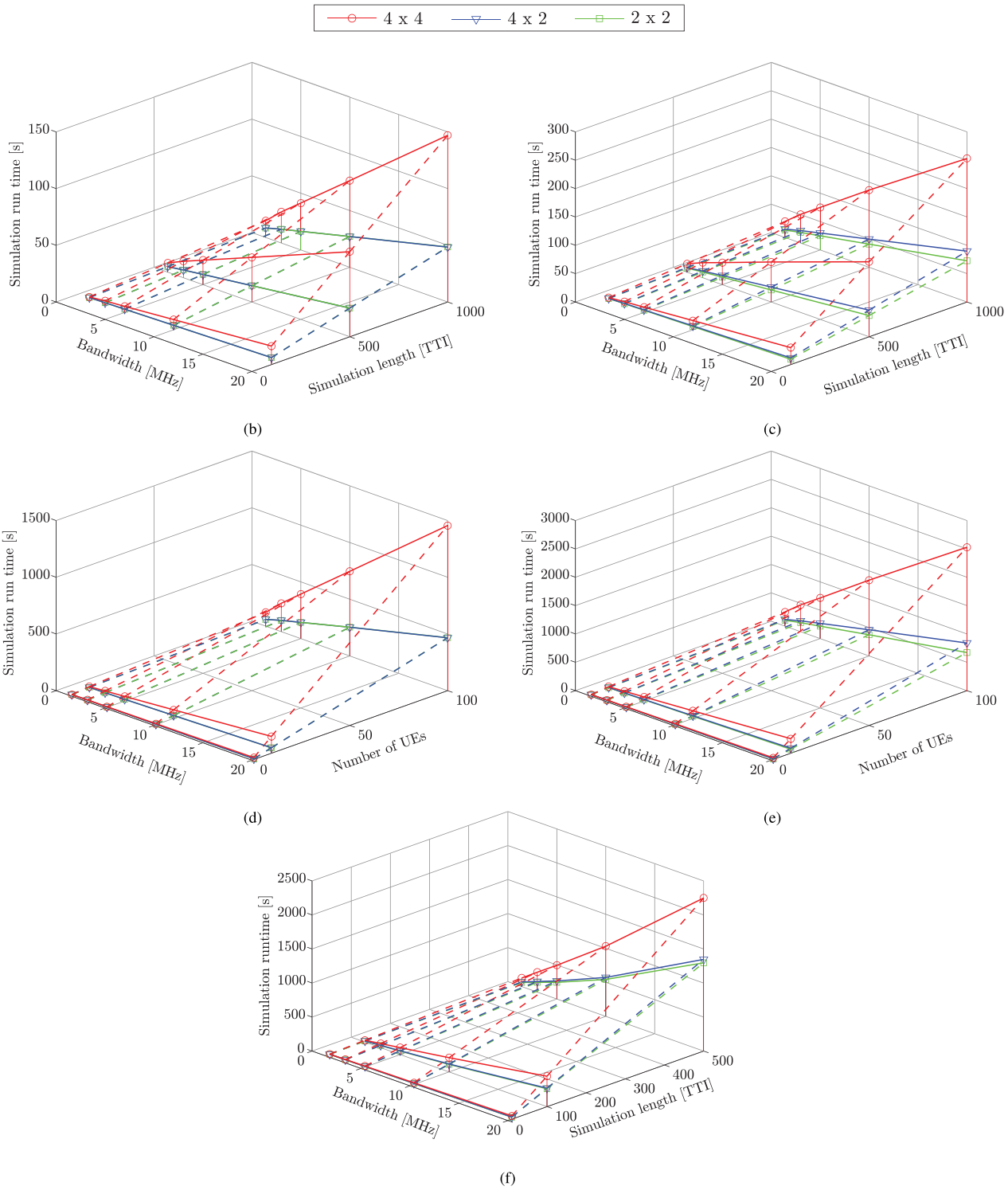


FIGURE 4. Simulation run times [s] for various LTE bandwidths [MHz], antenna configurations $\{N_{TX} \times N_{RX}\}$ and number of UEs. (b) system level, *old* model, 1 UE. (c) system level, *new* model, 1 UE. (d) system level, *old* model, 100 TTIs. (e) system level, *new* model, 100 TTIs. (f) link level, 1 UE.

For a fair comparison, the total transmit power per eNodeB is limited to 40 W. Elaborated power allocation are omitted. In the cooperation case, serving- as well as interfering RRHs transmit with a power of 20 W. The parameters for simulation are summarized in Table 3. Transceiver impairments due to the high speeds, such as Inter-Carrier Interference (ICI), are taken into account by a *short block fading* model, which is explained in Appendix A.

TABLE 2. Coefficients for the runtime estimator in (2) as obtained by linear least squares. Values are provided for system level simulations with the legacy- and the new link quality model as well as for link level simulations.

Simulation type	c_0	c_1	c_2	c_3	c_4	c_5	c_6
Legacy link quality model	0.8	0	$8.2 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$	0	0	0
New link quality model	1.0	0	$5.7 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$	0	0	0
Link level simulations	52.5	0	$235.0 \cdot 10^{-4}$	0	$2.3 \cdot 10^{-4}$	$11.0 \cdot 10^{-4}$	0

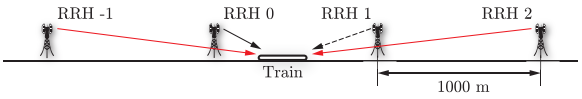


FIGURE 5. Representative segment for simulation of high-speed train scenario. RRHs are equidistantly spaced out along the tracks. The train moves between RRH_0 and RRH_1, RRH_{-1} and RRH_2 serve as interferers.

TABLE 3. Parameters for high-speed train scenario simulations.

Parameter	Value
Frequency	2.14 GHz
LTE bandwidth	B = 20 MHz
Inter-RRH distance	1000 m
eNodeB transmit power	40 W
Antennas per RRH	2
LTE transmission mode	CLSM
Antenna gain in dB	$A(\theta) = -\min\left(12\left(\frac{\theta}{70^\circ}\right), 20\,\mathrm{dB}\right)$
Minimum coupling loss	70 dB
Path loss model	$128.1 + 37.6 \log_{10}(R)$, R in km
Shadow fading model	correlated log normal, σ = 10 dB
Channel model	ITU-R Veh-A, *short* block fading
Receiver type	zero forcing
Noise power spectral density	-174 dBm/Hz
Receiver noise figure	9 dB
Train speed	250 km/h
Train length	200.84 m
Train penetration loss	constant, 30 dB
Active UEs	K = 230
UE distribution within train	uniform
Antennas per UE	1
Traffic model	full buffer
Scheduler	proportional fair
Channel knowledge	perfect
Feedback	AMC: CQI, MIMO: PMI and RI
Feedback delay	3 TTI
RRH backhaul connection	radio over fiber, no delay
Simulation length in TTI	N_{TTI} = 100
Number of simulation runs	24 per train position

2) SIMULATION RESULTS

Figure 6 provides results in terms of train average spectral efficiency, i.e., the mean over all passengers at a certain train position.[10] It is observed that the cooperation scheme universally achieves the highest performance while completely avoiding hand-overs in the region of interest. Coordination among RRHs yields a lower performance, particularly in the middle between two RRHs. Note, however, that in practical scenarios such scheme is typically far less complex than cooperation. This stems from the fact that only

[10]We evaluated 24 train locations between 600 m and 1400 m. For each location, 24 simulation runs with a length of 100 TTIs were carried out.

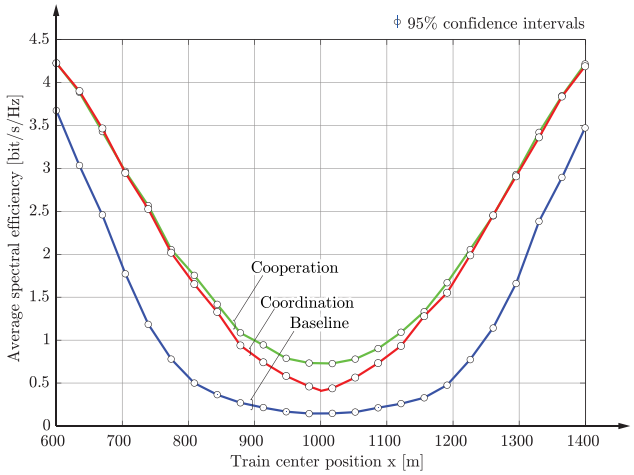


FIGURE 6. Train average spectral efficiency [bit/s/Hz] versus train position for three RRH collaboration schemes.

control data has to be exchanged among the eNodeBs. The baseline scenario yields the worst performance due to the presence of a strong interferer. Such studies are valuable for network providers, to minimize the cost of necessary installations along the tracks. Due to the short simulation length of 100 TTI (0.1 s), effects of hand-overs are not considered. They may completely be prevented by the concept of *moving cells*, as reported in [45].

3) SIMULATION RUN TIMES

Employing the same hardware as utilized for the performance analysis in Section III-B, we measure the system level simulation run times for each of the three RRH association scenarios. The results are provided in Table 4. They were obtained by averaging over 24 train positions and 24 simulation runs for each position. It is found that the new link quality slows down simulation by a factor of 3.7 compared to the legacy model. On the other hand, it enables to evaluate the cooperation scheme in the first place.

TABLE 4. System level simulation run times [s] as required for evaluating various RRH association schemes in train scenarios.

Simulation type	*Baseline*	*Coord.*	*Coop.*
Legacy link quality model	66	66	–
New link quality model	242	242	277

TABLE 5. Simulation parameters for MBSFN scenario.

Parameter	Value
Frequency	2.14 GHz
LTE bandwidth	$B = 5$ MHz
Macroscopic path loss	fixed
Shadow fading	none
Channel model	ITU-R Veh-A, block fading
Receiver type	zero forcing
Average noise power	−13 dB
Transmission rate	fixed, 1.2 bit per channel use (CQI 6)
Scheduler	round robin multicast group scheduler
Cyclic prefix	extended
Simulation length in TTI	$N_{\mathrm{TTI}} = 10\,000$

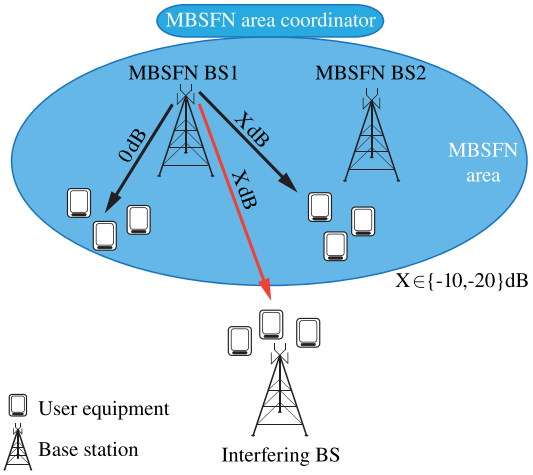


FIGURE 7. Example scenario for comparison of MBSFN simulations on link level and system level.

B. MULTIMEDIA BROADCAST SINGLE FREQUENCY NETWORKS

Employing cellular networks to broadcast information to a multitude of UEs in parallel, enabling services such as mobile radio and TV, is a desire that persists already since the introduction of Multimedia Broadcast Multicast Services (MBMS) with Release 6 of UMTS (March 2005). In LTE, the concept has been extended to enhanced MBMS (eMBMS), including Multimedia Broadcast Single Frequency Networks (MBSFNs) as a main feature. By grouping several eNodeBs to broadcast the same information over potentially large geographic areas, MBSFNs allow to form so-called *multicast/broadcast areas*, as indicated in Figure 7. Within such area, interference between eNodeBs is not only avoided, but even exploited as useful signal. The resulting peak data rates and network capacities enable high-quality video broadcasting and low latency transmissions [2]. Currently, interest in enhanced MBMS (eMBMS) and MBSFNs is increasing within the 3GPP, as they allow to efficiently spread information that has a high user demand, e.g., at large events (venue casting) or during peak hours, as well as to simultaneously update a vast amount of devices, which is expected to become a common use case with the advent of the Internet of Things [46]. Moreover, they enable public safety services, road safety applications, as well as tasks of fleet control with LTE-A.

1) IMPLEMENTATION

The support of MBSFN in the Vienna LTE-A simulator required some significant modifications of the program code. Signals that previously counted as interference, now have to be considered as useful signals in the SINR calculation. However, it is not expedient to simply add up the received powers from all the eNodeBs within the MBSFN area, since these base stations broadcast the same information synchronously. Thus, instead of adding signal powers, it is necessary to add the channel matrices $\mathbf{H}_i$ of all eNodeBs $i \in \mathcal{A}$ within the MBSFN area $\mathcal{A}$, i.e., $\mathbf{H}_{\mathrm{MBSFN}} = \sum_{i\in\mathcal{A}} \mathbf{H}_i$. This became feasible with the *proposed runtime-precoding concept* as outlined in Section III. In an additional step, the MBSFN base stations are filtered out from the interfering eNodeBs and the corresponding effective MBSFN channel matrices for multicast UEs are determined.

The support of MBSFN further requires a new element in the networking hierarchy, termed *MBSFN area coordinator*. It determines the resource allocation for all eNodeBs within the corresponding MBSFN area during subframes that are utilized for multicast transmission. UEs within the MBSFN area can subscribe to multicast groups in order to receive certain multicast messages. The task of the MBSFN area coordinator is to decide on the schedule of these multicast groups within the MBSFN subframes. This group schedule is then forwarded to the corresponding MBSFN eNodeBs, which signal the information to the multicast UEs. As an example, a round robin multicast group scheduler has been implemented in the Vienna LTE-A simulator.

To enable LTE-A standard-compliant MBMS/MBSFN simulations, a few further restrictions have to be considered in the simulator. Multicast transmissions within MBSFN areas always apply the extended cyclic-prefix of LTE-A, in order to avoid inter-symbol interference due to the increased delay spread of the effective MBSFN channel. Furthermore, reference-symbols are more densely placed in MBSFN subframes, to compensate for the larger frequency selectivity of the effective channel due to the increased delay spread. Basically, on system level, both of these changes cause a difference in the number of available resource elements compared to unicast transmission and are taken into account when calculating the throughput.

2) COMPARISON OF MBSFN SIMULATIONS ON LINK LEVEL AND SYSTEM LEVEL

For validation of the system level MBSFN implementation, we cross-compare with results from link level simulations. A simple example scenario as illustrated in Figure 7, is employed. A larger network cannot be simulated on link level due to computational complexity issues. In this scenario, we consider a single MBSFN area consisting of two MBSFN eNodeBs and serving a total number of six UEs. An additional unicast base station serves three UEs and acts as out-of-area interferer. The direct link between a UE and

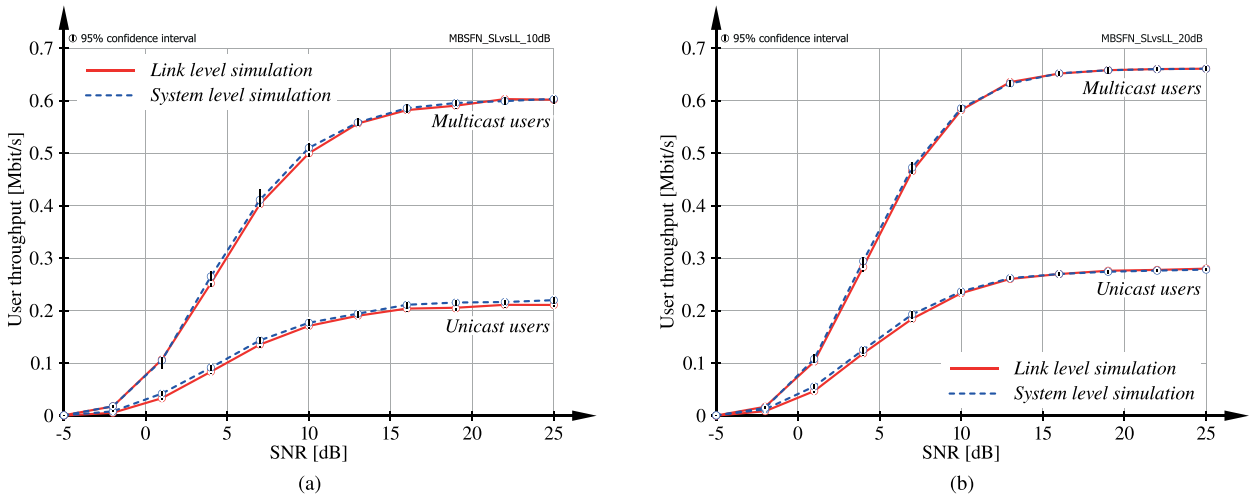


FIGURE 8. **Throughput [Mbit/s] as achieved by multicast and unicast UEs. Results are obtained from link- and system level simulations with 95 % confidence intervals. (a) Signal to interference ratio 10 dB. (b) Signal to interference ratio 20 dB.**

its associated eNodeB has an average gain of 0 dB, whereas the link to other UEs experiences an average gain of X dB, with $X \in \{-10, -20\}$. We employ a fixed transmission rate of approximately 1.2 bit per channel use (corresponding to Channel Quality Indicator (CQI) 6), to avoid the impact of transmission rate adaptation in the results.

The results of the link level simulation (red solid) and the system level simulation (blue dashed) are shown in Figure 8. At high interference power, i.e., $X = -10$ dB, we observe that there is a marginal mismatch between link and system level results, as the system level simulator slightly overestimates the performance. This is caused by the fact that, on system level, interference is considered Gaussian distributed, which is an over-simplification since the interference signals are taken from a finite symbol alphabet (4-QAM). However, when simulating larger networks, as is commonly the case on system level, the law of large numbers validates the application of this Gaussian approximation. Despite the small size of the scenario, cross-comparing the simulation times of system- and link level exhibit a speed-up of about 3.5×.

V. CONCLUSION AND FUTURE WORK

This work presented the concept of *runtime-precoding* that enables the simulation of coherent signal transmission from spatially separated transmission points on system level. Our approach only alters the link quality model while preserving the complexity gains as achieved by state of the art link abstraction models. The price to pay for enabling coherent multi-point transmission turned out to be an additional upscaling of the simulation run time which is proportional to the product of the number of transmit- and receive antennas, respectively. On the other hand, the run times showed a linear- rather than a quadratic growth with the bandwidth, the latter being observed in link level simulations. Two examples for the application of runtime precoding, inter- and intra-site CoMP in fast train scenarios and MBSFNs, were studied. While the results for a simple MBSFN scenario were highly consistent with link level simulations, achieving a simulation run time speed-up of about 3.5×, such comparison was not possible at all for the CoMP scenarios, as the large number of network elements restricted evaluations to system level. The runtime-precoding feature slowed down the simulations of the coordinated-scheduling scenario by 3.7×. On the other, it enabled the investigation of the joint transmission scheme in the first place. Thus, enhancing the link quality model by runtime-precoding provides a convenient tool for simulating large-scale coherent multi-point transmission scenarios, where evaluation on link level is no longer feasible due to complexity issues. The source code, including all the presented new features and examples, is freely available for download under an academic, non-commercial use license. Our future work is directed towards 3-dimensional channel models [30], which will enable the investigation of elevation beamforming and full-dimensional MIMO [38].

APPENDIX
MODELING HIGH USER-MOBILITY IN THE VIENNA LTE-A SYSTEM LEVEL SIMULATOR

The LTE-A standard is designed to support reliable communication with users moving at velocities as high as 500 km/h [2]. At such high speeds, several transceiver impairments such as channel estimation errors [47] and ICI between Orthogonal Frequency Division Multiplexing (OFDM) subcarriers [48], become non-negligible. These and similar imperfections were, however, not considered in legacy versions of the Vienna LTE-A simulator.

A. IMPLEMENTATION

In this section, we explain the extension of the Vienna LTE-A simulator to account for the suboptimal transceiver operation. In a first step, we investigate the impact of ICI on the performance of the system. Then, we show that a similar

approach can be applied to consider other imperfections as well. The basic requirement is that the impairments can be modeled sufficiently well as *additional Gaussian noise* and are treated as such by the receivers. Although such assumptions do not hold in general, they were found to accurately resemble link level simulations as well as measurements for the considered LTE-A scenarios.

The impact of ICI on the performance of OFDM systems has been reported in several publications. In [48], a relatively simple expression for the average ICI power as experienced by an OFDM transmission over a fast and frequency selective Rayleigh fading channel is derived. This expression is particularly suitable for the Vienna LTE-A simulator, as it only depends on the maximum Doppler shift of the channel and can hence be evaluated easily. However, the model is not very precise as it employs an average over the fading channel rather than using the current realization.

In the Vienna LTE-A simulator, the ICI power can be taken into account by modifying the post-equalization SINR calculation in the UE link quality model (conf. Algorithm 1). In particular, we consider the ICI power as additional Gaussian noise and add it on top of the already existing thermal noise. Before, however, it is necessary to account for the macro-scale fading as well as the UE power allocation by multiplying the ICI power as obtained from [48] with the corresponding fading parameters (pathloss, shadow fading, antenna gain) as well as the transmit power. In case that RRHs are employed, we assume the ICI power contributions from the antenna arrays to be statistically independent and simply accumulate the individual values.

B. SIMULATIONS

The performance of this simple ICI model can be investigated and verified by comparing the outcome of system level simulations to the results as obtained with the Vienna LTE-A link level simulator [39]. The link level simulator supports a fast fading channel model, in which the channel varies in between OFDM samples (0.52 μs sampling rate at 1.4 MHz bandwidth) and causes ICI. In contrast, the system level simulator applies a block fading channel model, where the channel varies only from one LTE subframe (14 OFDM symbols, 1 ms duration) to the next. With both simulators, we evaluate a single-UE-single-eNodeB scenario as specified in Table 1. Both eNodeB and UE are equipped with one antenna and the transmission takes place with a fixed transmission rate of approximately 5.5 bit per channel use (corresponding to CQI 15). To highlight the impact of ICI, an unrealistically high average Signal-to-Noise Ratio (SNR) of 50 dB is employed.

Throughput and BLER results are visualized in Figure 9. The curves denoted as *fast fading* represent the link level results. The straight line in Figure 9a, denoted as *BF 14 OFDM symbols without ICI*, has been obtained from system level simulations without adding ICI noise. In this case, the performance is *independent* of the user velocity. Adding ICI noise yields the curve denoted as *BF 14* (block fading over 14 OFDM symbols), which lies significantly above the fast fading link level result. Thus, ICI noise alone is *not sufficient* to realistically represent the performance of a fast fading channel.

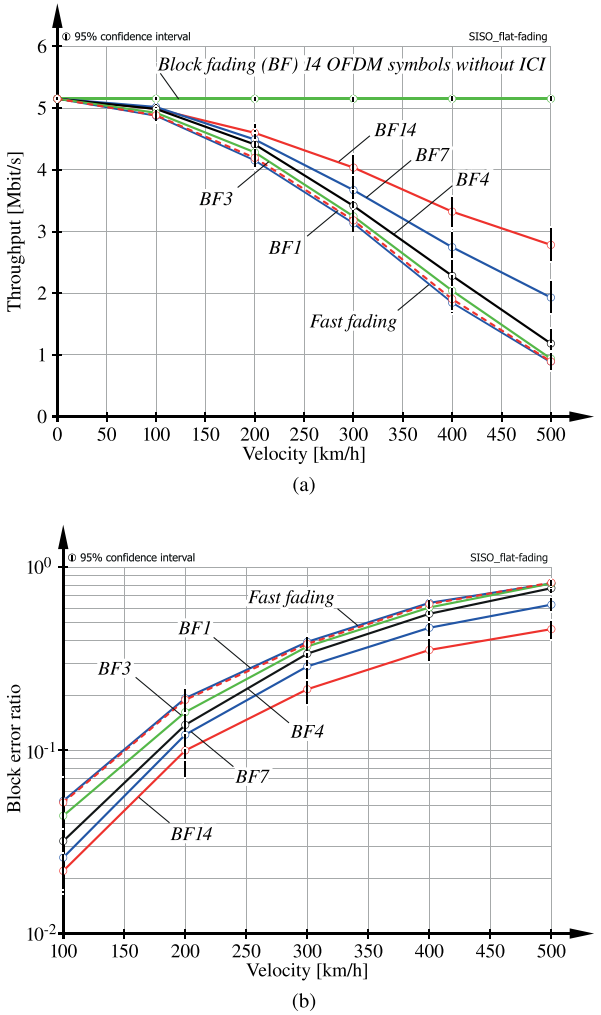


FIGURE 9. Throughput and BLER for fast fading channels and block fading channels with varying fading block-length: BFx corresponds to block fading over x OFDM symbols. ICI is considered as additional Gaussian noise. (a) Throughput reduction with growing user velocity. (b) BLER degradation with growing user velocity.

C. SHORT BLOCK FADING

It is necessary to consider the increased temporal diversity of the channel by reducing the block-length of the block fading channel model. The results are illustrated by the remaining curves in Figure 9a. Figure 9b shows the corresponding BLER curves. It is observed that a fading block-length of at most three OFDM symbols should be employed to accurately reproduce the link level results. The shortening of the block-length is denoted as *short block fading*.

Activating ICI noise and selecting a fading block-length of three OFDM symbols for the system level simulation, we finally compare the throughput performance in a more complex scenario, where the eNodeB is equipped with

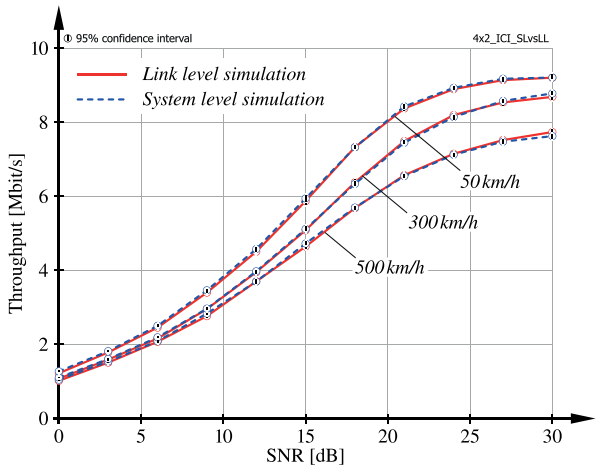


FIGURE 10. Comparison of the throughput for different user velocities as obtained from link- and system level simulations, respectively.

four transmit antennas and the UE has two receive antennas. The SNR is varied from 0 to 30 dB. We evaluate the throughput reduction with increasing user velocity, assuming that transmission rate adaptation is activated. The results of the link level simulation (red solid) and the system level simulation (blue dashed) are shown in Figure 10. We observe an accurate match between link level and system level, substantiating the validity of this simple ICI model. In this setup, the system level simulator was roughly 35× faster.

Hence, the Vienna LTE-A system level simulator was extended to enable a reduced fading block-length, facilitating high-mobility simulations. This *short block fading* option, however, should be utilized deliberately, as it significantly increases the computational complexity of the simulations.

REFERENCES

[1] (Jan. 2015). *ITU-R Working Party 5D (WP 5D) Contributions*. [Online]. Available: http://www.itu.int/md/R07-WP5D-C/en

[2] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Amsterdam, The Netherlands: Elsevier, 2011.

[3] S. Ahmadi, *LTE-Advanced: A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies* (ITPro Collection). Amsterdam, The Netherlands: Elsevier, 2013.

[4] Y. Wang, J. Xu, and L. Jiang, "Challenges of system-level simulations and performance evaluation for 5G wireless networks," *IEEE Access*, vol. 2, pp. 1553–1561, 2014.

[5] M. Gerasimenko *et al.*, "Performance comparison of system level simulators for 3GPP LTE uplink," in *Internet of Things, Smart Spaces, and Next Generation Networking* (Lecture Notes in Computer Science), vol. 7469, S. Andreev, S. Balandin, and Y. Koucheryavy, Eds. Berlin, Germany: Springer-Verlag, 2012, pp. 186–197.

[6] (Jan. 2015). *Riverbed*. [Online]. Available: http://www.riverbed.com/products/performance-management-control/opnet.html?redirect=opnet

[7] (Jan. 2015). *OMNeT++*. [Online]. Available: http://www.omnetpp.org

[8] (Mar. 2015). *IT++*. [Online]. Available: http://itpp.sourceforge.net/4.3.1/

[9] (Jan. 2015). *ns-2*. [Online]. Available: http://www.isi.edu/nsnam/ns

[10] (Jan. 2015). *ns-3*. [Online]. Available: http://www.nsnam.org/

[11] (Jan. 2015). *GNS3*. [Online]. Available: http://www.gns3.com/

[12] S. Max, D. Bültmann, R. Jennen, and M. Schinnenburg, "Evaluation of IMT-advanced scenarios using the open wireless network simulator," in *Proc. Int. ICST Conf. Simulation Tools Techn.*, Torremolinos, Spain, Mar. 2010, p. 26.

[13] (Jan. 2015). *Hurricane II WAN Emulation and Network Simulation*. [Online]. Available: http://packetstorm.com/packetstorm-products/hurricane-ii-software

[14] (Jan. 2015). *Nomor System Level Simulation*. [Online]. Available: http://www.nomor.de/home/solutions-and-products/system-level-simulation

[15] D. Martín-Sacristán, J. F. Monserrat, V. Osa, and J. Cabrejas, "LTE-advanced system level simulation platform for IMT-advanced evaluation," *Waves*, 2011.

[16] Y. Li, F. Yu, S.-L. Zheng, and C.-L. Yang, "LTE system level simulation with MATLAB," in *Proc. Int. Conf. Internet Technol. Appl. (iTAP)*, Aug. 2011, pp. 1–4.

[17] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: An open-source framework," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 498–513, Feb. 2011.

[18] (Feb. 2015). *The Vienna LTE Simulators*. [Online]. Available: http://www.nt.tuwien.ac.at/ltesimulator

[19] Mathworks. (Jul. 2014). *MATLAB Documentation*. [Online]. Available: http://www.mathworks.com/help/matlab/index.html

[20] K. Abdallah, I. Cerutti, and P. Castoldi, "Energy-efficient coordinated sleep of LTE cells," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 5238–5242.

[21] M. Carvalho and P. Vieira, "An enhanced handover oscillation control algorithm in LTE self-optimizing networks," in *Proc. IEEE Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Brest, France, Oct. 2011, pp. 1–5.

[22] M. M. Selim, M. El-Khamy, and M. El-Sharkawy, "Enhanced frequency reuse schemes for interference management in LTE femtocell networks," in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, Aug. 2012, pp. 326–330.

[23] S. Y. Shin and D. Triwicaksono, "Radio resource control scheme for machine-to-machine communication in LTE infrastructure," in *Proc. Int. Conf. ICT Converg. (ICTC)*, Jeju Island, Korea, Oct. 2012, pp. 1–6.

[24] J. C. Ikuno, "System level modeling and optimization of the LTE downlink," Ph.D. dissertation, Vienna Univ. Technol., Vienna, Austria, 2013.

[25] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*, 3rd Generation Partnership Project (3GPP), document Rec. TR 36.814, Mar. 2010.

[26] *Coordinated Multi-Point Operation for LTE Physical Layer Aspects*, 3rd Generation Partnership Project (3GPP), document Rec. TR 36.819, Sep. 2013.

[27] *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) System Scenarios*, 3rd Generation Partnership Project (3GPP), document Rec. TR 36.942, Oct. 2014.

[28] *Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000*, ITU Radiocommunication Sector (ITU-R), document Rec. ITU-R M.1225.

[29] P. Kyösti *et al.*, "WINNER II channel models," Tech. Rep., Sep. 2007. [Online]. Available: http://www.ist-winner.org/deliverables.html

[30] *Study on 3D Channel Model for LTE*, 3rd Generation Partnership Project (3GPP), document Rec. TR 36.873, Sep. 2014.

[31] R. W. Heath, Jr., M. Airy, and A. J. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," in *Proc. Conf. Rec. 35th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2001, pp. 1194–1199.

[32] I. Latif, F. Kaltenberger, and R. Knopp, "Link abstraction for multi-user MIMO in LTE using interference-aware receiver," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 842–846.

[33] M. Döttling *et al.*, "Assessment of advanced beamforming and MIMO technologies," Tech. Rep. D2.7, May 2005.

[34] S. Caban, M. Rupp, C. Mehlführer, and M. Wrulich, *Evaluation of HSDPA and LTE: From Testbed Measurements to System Level Performance*. New York, NY, USA: Wiley, 2011.

[35] Nortel Networks, *OFDM Exponential Effective SIR Mapping Validation, EESM Simulation Results*, 3rd Generation Partnership Project (3GPP), document Rec. TR R1-040089, Jan. 2004.

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

[37] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Found. Trends Commun. Inf. Theory*, vol. 9, nos. 2–3, pp. 113–381, 2013.

[38] *Study on Elevation Beamforming/Full-Dimension (FD) MIMO for LTE*, 3rd Generation Partnership Project (3GPP), document Rec. TR 36.897, Dec. 2014.

[39] S. Schwarz, J. C. 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.

[40] B. Ai *et al.*, "Challenges toward wireless communications for high-speed railway," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2143–2158, Oct. 2014.
[41] R. He, Z. Zhong, B. Ai, G. Wang, J. Ding, and A. F. Molisch, "Measurements and analysis of propagation channels in high-speed railway viaducts," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 794–805, Feb. 2013.
[42] L. Gao, Z. Zhong, B. Ai, and L. Xiong, "Estimation of the Ricean factor in K the high speed railway scenarios," in *Proc. Int. ICST Conf. Commun. Netw. China*, Aug. 2010, pp. 1–5.
[43] M. K. Müller, M. Taranetz, and M. Rupp. (2015). "Providing current and future cellular services to high speed trains." [Online]. Available: http://arxiv.org/abs/1505.04557
[44] M. K. Müller, M. Taranetz, and M. Rupp, "Performance of remote unit collaboration schemes in high speed train scenarios," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Boston, MA, USA, Sep. 2015.
[45] C. D. Gavrilovich, Jr., "Broadband communication on the highways of tomorrow," *IEEE Commun. Mag.*, vol. 39, no. 4, pp. 146–154, Apr. 2001.
[46] H. Kopetz, "Internet of Things," in *Real-Time Systems* (Real-Time Systems Series). New York, NY, USA: Springer-Verlag, 2011, pp. 307–323.
[47] M. Simko, P. S. R. Diniz, Q. Wang, and M. Rupp, "Adaptive pilot-symbol patterns for MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4705–4715, Sep. 2013.
[48] Y.-S. Choi, P. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1375–1387, Aug. 2001.

MARTIN TARANETZ (S'07) received the B.Sc. degree in electrical engineering and Dipl.-Ing degree (M.Sc. equivalent) in telecommunications with highest honors from the TU Wien, Vienna, Austria, in 2008 and 2011, respectively. He is currently pursuing the Ph.D. degree in telecommunications engineering and he is employed as a University Assistant with the Institute of Telecommunications, TU Wien. From 2014 to 2014, he was a Visiting Researcher with the Wireless Networking and Communications Group, The University of Texas at Austin. His research interests lie in the fields of wireless communications and signal processing. In his Ph.D. thesis, he focuses on system level modeling and evaluation of heterogeneous cellular networks. He is a Reviewer for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

THOMAS BLAZEK received the B.Sc. degree in electrical engineering from the TU Wien, Vienna, Austria, in 2013. He is currently pursuing the M.Sc. degree in telecommunications engineering. He is employed as a Project Assistant at the Institute of Telecommunications, TU Wien, focusing on vehicular connectivity. His current research topics include modeling and analysis of vehicular channels.

THOMAS KROPFREITER received the B.Sc. degree in electrical engineering and the Dipl.-Ing degree (M.Sc. equivalent) in telecommunication engineering from the TU Wien, Vienna, Austria, in 2012 and 2014, respectively. His Diploma thesis dealt with a quantization of soft information in distributed hypothesis testing scenarios. From 2013 to 2015, he was a member of the Mobile Communications Group, TU Wien, where he worked on the topic of beamforming in multicell scenarios. Since 2015, he has been employed as a Project Assistant at TU Wien, where he is currently pursuing the Ph.D. degree. His current research interests are random finite sets in the field of target tracking and self localization.

MARTIN KLAUS MÜLLER (S'13) received the B.Eng. degree in electrical and telecommunication engineering from the DHBW Ravensburg, Friedrichshafen, in cooperation with Rohde and Schwarz, in 2009, and the Dipl.-Ing degree (M.Sc. equivalent) in telecommunications from the TU Wien, Vienna, Austria, in 2013. His Diploma thesis focused on feedback based LTE-A measurements. He is currently pursuing the Ph.D. degree in telecommunications engineering and he is employed as a Project Assistant with the Institute of Telecommunications, TU Wien. His current research interests include mobile cellular access in train- and highway scenaros and wireless communication in indoor-scenarios.

STEFAN SCHWARZ (S'10–M'14) received the B.Sc. degree in electrical engineering and the Dipl.-Ing. degree (M.Sc. equivalent) in telecommunications engineering with highest distinctions in 2007 and 2009, respectively, both at the TU Wien, and the Dr.techn. degree (Ph.D. equivalent) in telecommunications engineering with highest distinctions at the TU Wien, in 2013. In 2010, he received the honorary price of the Austrian Minister of Science and Research, for excellent graduates of scientific and artistic universities, and in 2014, he received the INiTS Award in the category Information and Communication Technologies for his dissertation on Limited Feedback Transceiver Design for Downlink MIMO OFDM Cellular Networks. Since 2008, he has been a Project Assistant with the Mobile Communications Group, Institute of Telecommunications, TU Wien. His research interests are located in the broad fields of wireless communications and signal processing. In his dissertation, he focused on limited feedback single- and multiuser MIMO-OFDM communication systems, and multi-user scheduling algorithms. He actively serves as a Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS, SIGNAL PROCESSING, WIRELESS COMMUNICATIONS AND VEHICULAR TECHNOLOGY, *EURASIP Journal on Signal Processing*, and *Journal on Wireless Communications and Networking*.

MARKUS RUPP received the Dipl.-Ing. degree at the University of Saarbrücken, Germany, in 1988, and the Dr.-Ing. degree from the Technische Universität Darmstadt, Germany, in 1993, where he worked with Eberhardt Hänsler on designing new algorithms for acoustical and electrical echo compensation. From 1993 to 1995, he had a post-doctoral position with the University of Santa Barbara, California, with S. Mitra, where he worked with A. H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From 1995 to 2001, he was a member of Technical Staff with the Wireless Technology Research Department, Bell-Labs, Crawford Hill, NJ, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11, and UMTS. Since 2001, he has been a Full Professor of Digital Signal Processing in Mobile Communications at the TU Wien, where he served as the Dean from 2005 to 2007. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005, where he is currently an Associate Editor of *JASP EURASIP Journal of Advances in Signal Processing*, and *JES EURASIP Journal on Embedded Systems*. He has been elected as a AdCom Member of EURASIP since 2004 and serving as the President of EURASIP from 2009 to 2010. He has authored or co-authored over 500 scientific papers, including 15 patents on adaptive filtering, and wireless communications.

Zöchmann *et al. EURASIP Journal on Wireless Communications and Networking* (2016) 2016:118
DOI 10.1186/s13638-016-0609-1

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Exploring the physical layer frontiers of cellular uplink

The Vienna LTE-A Uplink Simulator

Erich Zöchmann[1,2*], Stefan Schwarz[1,2], Stefan Pratschner[1], Lukas Nagel[1], Martin Lerch[1] and Markus Rupp[1]

Abstract

Communication systems in practice are subject to many technical/technological constraints and restrictions. Multiple input, multiple output (MIMO) processing in current wireless communications, as an example, mostly employs codebook-based pre-coding to save computational complexity at the transmitters and receivers. In such cases, closed form expressions for capacity or bit-error probability are often unattainable; effects of realistic signal processing algorithms on the performance of practical communication systems rather have to be studied in simulation environments. The Vienna LTE-A Uplink Simulator is a 3GPP LTE-A standard compliant MATLAB-based link level simulator that is publicly available under an academic use license, facilitating reproducible evaluations of signal processing algorithms and transceiver designs in wireless communications. This paper reviews research results that have been obtained by means of the Vienna LTE-A Uplink Simulator, highlights the effects of single-carrier frequency-division multiplexing (as the distinguishing feature to LTE-A downlink), extends known link adaptation concepts to uplink transmission, shows the implications of the uplink pilot pattern for gathering channel state information at the receiver and completes with possible future research directions.

Keywords: 4G mobile communication, Cellular uplink, Computer simulation

1 Introduction

Current cellular wireless communications employs Universal Mobile Telecommunications System (UMTS) Long Term Evolution (LTE) as the high data rate standard [1]. The increasing demand of high data traffic in up- and downlink forces engineers to push the limits of LTE [2], e.g. through enhanced multi-user multiple input, multiple output (MIMO) support [3, 4], coordinated multipoint (CoMP) transmission/reception [5, 6] as well as improved channel state information (CSI) feedback algorithms [7]. The authors of [8] predict further evolution of existing LTE/LTE-Advanced (LTE-A) systems in parallel to the development of new radio-access technologies operating at millimetre wave frequencies even beyond the expected roll-out of 5G technologies by 2020. Fair comparison of novel signal processing algorithms and transceiver designs has to assure equal testing and evaluation conditions to enable reproducibility of results by independent groups of researchers and engineers [9]. For performing system-level simulations, [10, 11] or [12] are freely accessible options. For link level, multiple commercial products are available that facilitate reproducible research, such as, *is-wireless* LTE PHY LAB [13] or *Mathworks* LTE System Toolbox [14] and some non-commercial projects which were introduced in [15] and [16]. To the best of the authors' knowledge, however, the Vienna LTE simulators are the only MATLAB-based suite of simulation tools including LTE system and link level, publicly available under an academic use licence, thus, free of charge for academic researchers all over the world. The software suite consists of three simulators. The downlink link and system level simulators are comprehensively studied in [2, 9, 17]. In this paper, we introduce the latest member of the family of Vienna LTE Simulators, that is, the Vienna LTE-A Uplink Link Level Simulator, downloadable at [18], and highlight our research conducted by means of this simulator.

*Correspondence: ezoechma@nt.tuwien.ac.at
[1]Institute of Telecommunications, TU Wien, Gußhaustraße 25, 1040 Vienna, Austria
[2]Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion, Vienna, Austria

1.1 Outline and contributions

We start with a brief re-capitulation of the LTE-A specifics and introduce the modulation and multiple access scheme and the employed MIMO signal processing of LTE-A uplink in Section 2. We then develop a matrix model describing the input-output relationship of the LTE-A uplink and present signal-tointerference-and-noise ratio (SINR) expressions for single-carrier frequency-division multiplexing (SC-FDM) as well as orthogonal frequency-division multiplexing (OFDM). The OFDM SINR expression and the performance of OFDM will serve as reference to study the effects of discrete Fourier transform (DFT) spreading imposed by SC-FDM.

In Section 3, we investigate the physical layer performance of SC-FDM and OFDM, comparing bit error ratio (BER) and peak-to-average power ratio (PAPR). The BER for LTE single input single output (SISO) transmissions was already analysed in link-level simulations by [19–21] and semi-analytically by [22, 23]. By means of our simulator, we reproduce these results and provide bounds to predict the performance of SC-FDM with respect to OFDM. The insights gathered by the BER simulations allow us to interpret the difference in throughput obtained by OFDM and SC-FDM, as discussed in Section 4.

Based on the SINR expressions developed in Section 2, we present a limited feedback strategy for link adaptation in Section 4 and contrast the performance of LTE uplink with channel capacity and other performance upper bounds that account for practical design restrictions [24]. Until Section 5, we assume perfect CSI at the receiver. The remaining sections will describe methods to obtain CSI at the receiver.

In Section 5, we highlight and describe the demodulation reference signal (DMRS) structure employed in LTE-A uplink to facilitate channel estimation of the time-frequency selective wireless channel.

Based on the obtained insights, we elaborate on the basic concept of DFT-based time domain channel estimation in Section 6 and review alternative code/frequency domain methods that can outperform DFT-based schemes [25].

Due to the increasing number of mobile users that stay connected while travelling in cars or (high speed) trains, we then shift our focus to high velocity scenarios. Such scenarios entail high temporal selectivity of the wireless channel, rendering accurate channel interpolation very important to sustain reasonable quality of service. We introduce and investigate basic concepts of channel interpolation in Section 7.

We briefly discuss open questions for future research in Section 8 and conclude in Section 9. Details to the handling of the simulator are provided in [26].

1.2 Notation

Matrices are denoted by bold uppercase letters such as $\boldsymbol{H}$ and vectors by bold lowercase letters such as $\boldsymbol{h}$. The entries of vectors and matrices are accessed by brackets and subscripts, e.g. $[\boldsymbol{h}]_k$ and $[\boldsymbol{H}]_{k,n}$. Spatial layers or receive antennas are denoted by superscripts in braces, e.g. $\boldsymbol{x}^{(l)}$. The superscripts $(\cdot)^T$ and $(\cdot)^H$ express transposition and conjugate transposition. $\|\cdot\|_2$, $\|\cdot\|_\infty$ and $\|\cdot\|_F$ symbolize the Euclidean norm, the Maximum norm and the Frobenius norm, respectively. The entrywise (Hadamard) product is denoted by $\odot$ and the Kronecker product by $\otimes$. The all ones vector/matrix is denoted by $\mathbb{1}$. The operator $\boldsymbol{X} = \mathrm{Diag}(\boldsymbol{x})$ places the vector $\boldsymbol{x}$ on the main diagonal of $\boldsymbol{X}$, and conversely, the operator $\boldsymbol{x} = \mathrm{diag}(\boldsymbol{X})$ returns the vector $\boldsymbol{x}$ from the main diagonal of $\boldsymbol{X}$. A block-wise Toeplitz (circulant, diagonal) matrix is a block matrix with each matrix of Toeplitz (circulant, diagonal) shape. The size of matrices is expressed via their subscripts, whenever necessary.

2 LTE-specific system model and SINR

$$
\begin{aligned}
\hat{\boldsymbol{x}} &= \left(\boldsymbol{I}_L \otimes \boldsymbol{D}_{N_{SC}}^H\right) \boldsymbol{F} \left(\boldsymbol{I}_{N_R} \otimes \boldsymbol{M}^H \boldsymbol{D}_{N_{\mathrm{FFT}}} \boldsymbol{P}_{\mathrm{remCP}}\right) \\
&\quad \times \boldsymbol{H} \left(\boldsymbol{I}_{N_T} \otimes \boldsymbol{P}_{\mathrm{addCP}} \boldsymbol{D}_{N_{\mathrm{FFT}}}^H \boldsymbol{M}\right) \left(\boldsymbol{W} \otimes \boldsymbol{I}_{N_{\mathrm{SC}}}\right) \left(\boldsymbol{I}_L \otimes \boldsymbol{D}_{N_{\mathrm{SC}}}\right) \boldsymbol{x} \\
&\quad + \underbrace{\left(\boldsymbol{I}_L \otimes \boldsymbol{D}_{N_{\mathrm{SC}}}^H\right) \boldsymbol{F} \left(\boldsymbol{I}_{N_R} \otimes \boldsymbol{M}^H \boldsymbol{D}_{N_{\mathrm{FFT}}} \boldsymbol{P}_{\mathrm{remCP}}\right) \boldsymbol{n}}_{\tilde{\boldsymbol{n}}} \\
&= \overbrace{\left(\boldsymbol{I}_L \otimes \boldsymbol{D}_{N_{\mathrm{SC}}}^H\right)}^{\boldsymbol{I}_{LN_{\mathrm{SC}}} \text{ for OFDM}} \boldsymbol{F} \boldsymbol{H}_{\mathrm{eff}} \overbrace{\left(\boldsymbol{I}_L \otimes \boldsymbol{D}_{N_{\mathrm{SC}}}\right)}^{\boldsymbol{I}_{LN_{\mathrm{SC}}} \text{ for OFDM}} \boldsymbol{x} + \tilde{\boldsymbol{n}} \\
&= \boldsymbol{K}\boldsymbol{x} + \tilde{\boldsymbol{n}} = \underbrace{\boldsymbol{I} \odot \boldsymbol{K}\boldsymbol{x}}_{\text{desired signal}} + \underbrace{(\boldsymbol{K} - \boldsymbol{I} \odot \boldsymbol{K})\boldsymbol{x}}_{\text{intra- and interlayer interference}} \\
&\quad + \tilde{\boldsymbol{n}}\,.
\end{aligned}
\tag{1}
$$

LTE operates on a time-frequency grid as shown in Fig. 1. The number of subcarriers is always a multiple of 12; 12 adjacent subcarriers over 7(or 6—in case of extended CP) successive OFDM symbols are called resource block (RB). Each RB thus consists of 12×7 (12×6) resource elements (REs), corresponding to the different time-frequency bins. A detailed description of LTE up- and downlink is available, e.g. in [27].

We focus on those details necessary to describe our system model at time n[1]. LTE employs OFDM(A)[2] as physical layer modulation and multiple access scheme in the downlink and SC-FDM(A), i.e. DFT-spreaded OFDM, in the uplink. In a SC-FDM model, OFDM can be considered a special case. The major difference is an additional spreading and de-spreading stage at the transmitter and receiver, highlighted via dashed boxes in Fig. 2. The common parts of the system model will be described from left to right.

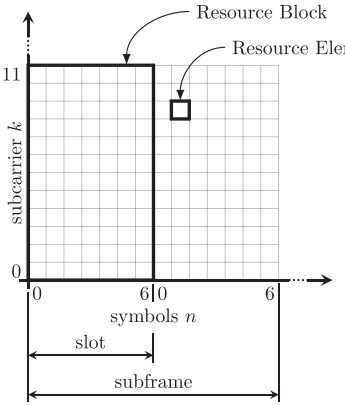


Fig. 1 The LTE-A uplink resource grid

Right after the DFT spreading, the DMRS is inserted. The DMRS will be considered later for the purpose of channel estimation (CE). Next, MIMO precoding is carried out, exploiting a set of semi-unitary precoding matrices $\boldsymbol{W}$, pooled in the precoder codebook $\mathcal{W}$, as defined in [1]. For LTE-A uplink transmission, the precoding matrix applied for a given user is equal for all RBs assigned to this user. In case of spatial multiplexing, each spatial layer is transmitted with equal power.

Each antenna is equipped with its own OFDM modulator, consisting of subcarrier mapping, inverse fast Fourier transform (IFFT) and a CP addition. To cope with the channel dispersion and to avoid Intersymbol Interference (ISI), LTE employs a CP. As a result of multipath propagation, a previous symbol may overlap with the present symbol, introducing ISI and impairing the orthogonality between subcarriers, i.e. causing Intercarrier Interference (ICI) [28]. Normal and extended CP lengths, with a respective duration of 4.7 and 16.7 μs, are standardized, enabling a simple trade-off between ISI immunity and CP overhead.

At the transmitter, the processing occurs in a reversed order. First, the OFDM demodulation/FFT takes place to get back into the frequency domain. The immunity to multipath propagation (stemming from the CP) allows to employ one-tap frequency domain equalizers $\boldsymbol{F}$ without performance loss. At last, de-spreading delivers the data estimates.

All this previously informally described processing is linear, and we are able to formulate a matrix-vector input-output relationship between a (stacked) data-vector $\boldsymbol{x}$ and its estimate $\hat{\boldsymbol{x}}$. For simplicity, we assume that the channel stays constant during one OFDM symbol. A detailed system description based on [29] can be found in [30].

In order to adapt the data transmission to the current channel state, LTE-A applies limited feedback; a comprehensive specification follows in Section 4. Limited feedback is depicted via the feedback arrow in Fig. 2. The data vector $\boldsymbol{x}^{(l)} \in \mathbb{C}^{N_{\mathrm{SC}} \times 1}$ of layer $l \in \{1, \ldots, L\}$ contains modulated symbols for each of the N_{SC} subcarriers. The number of transmit layers depends on the LTE-A specific rank indicator (RI) feedback. The data symbols are coded with a punctured turbo code whose rate is determined by the channel quality indicator (CQI). Subsequently,

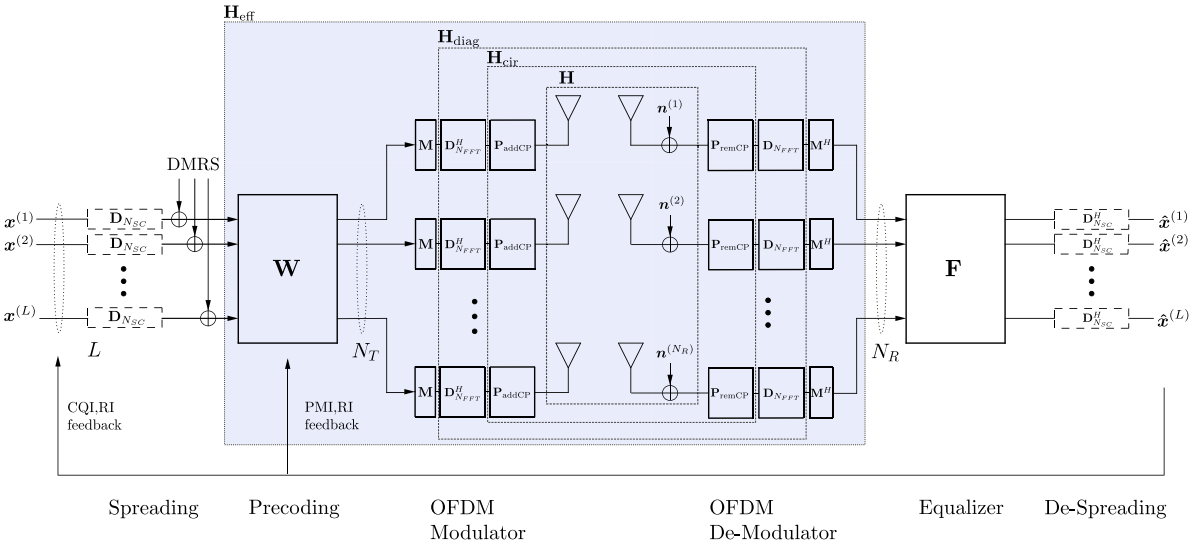


Fig. 2 The LTE-A uplink transceiver

the codewords are mapped onto a quadrature amplitude modulation (QAM) alphabet (4/16/64 QAM), where the size of the alphabet depends on the CQI as well. All $\boldsymbol{x}^{(l)}$ are stacked into one vector $\boldsymbol{x} \in \mathbb{C}^{N_{\text{SC}}L\times 1}$ on which layer-wise spreading and joint precoding—according to the precoding matrix indicator (PMI)—of all subcarriers take place. The subsequent OFDM modulator consists of the localized subcarrier mapping $\boldsymbol{M}$, mapping N_{SC} subcarriers to the centre of an N_{FFT} point IFFT and the addition of the CP.

Depending on the level of abstraction, our system model can be described via different channel matrices. The physical baseband time domain channel is described by a block-wise Töplitz matrix $\boldsymbol{H} \in \mathbb{C}^{(N_{\text{FFT}}+N_{\text{CP}})N_{\text{R}}\times(N_{\text{FFT}}+N_{\text{CP}})N_{\text{T}}}$, with N_{T} transmit and N_{R} receive antennas, which turns block-wise circulant ($\boldsymbol{H}_{\text{cir}}$) after addition ($\boldsymbol{P}_{\text{addCP}}$) and removal ($\boldsymbol{P}_{\text{remCP}}$) of an appropriately chosen CP of length N_{CP}. Finally, it turns diagonal after the IFFT and FFT on the transmitter and receiver, respectively. An example of the Töplitz and diagonal structured channel is demonstrated in Fig. 3a, b, respectively.

$$\boldsymbol{H}_{\text{diag}} = \left(\boldsymbol{I}_{N_R} \otimes \boldsymbol{D}_{N_{\text{FFT}}}\boldsymbol{P}_{\text{remCP}}\right)\boldsymbol{H}\left(\boldsymbol{I}_{N_T} \otimes \boldsymbol{P}_{\text{addCP}}\boldsymbol{D}^{H}_{N_{\text{FFT}}}\right) \tag{2}$$

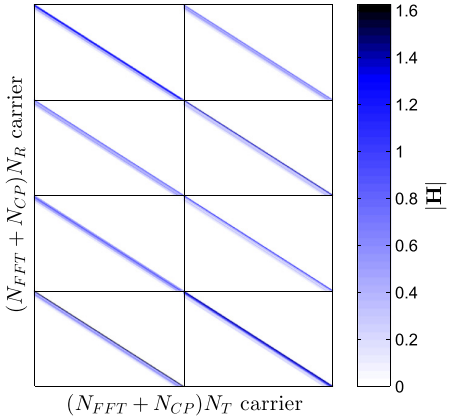
The last step of the OFDM de-modulator is the reversal of the localized subcarrier mapping $\boldsymbol{M}^H$. The effective MIMO channel from L transmit layers to N_{R} receive antennas, incorporating the precoder, the OFDM modulator, the time-domain MIMO channel $\boldsymbol{H}$ and the OFDM de-modulator, is abstracted to one block matrix $\boldsymbol{H}_{\text{eff}}$. This greatly facilitates the readability of all formulas later on.

$$\boldsymbol{H}_{\text{eff}} = \left(\boldsymbol{I}_{N_R} \otimes \boldsymbol{M}^H\right)\boldsymbol{H}_{\text{diag}}\left(\boldsymbol{I}_{N_T} \otimes \boldsymbol{M}\right)\left(\boldsymbol{W} \otimes \boldsymbol{I}_{N_{\text{SC}}}\right) \tag{3}$$

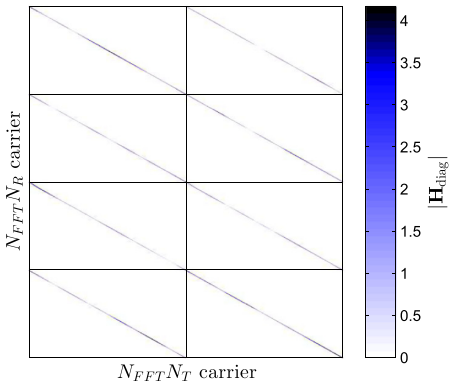
The additive noise is assumed independent across antennas and is distributed zero mean, white Gaussian $\boldsymbol{n}^{(i)} \sim \mathcal{CN}\{\boldsymbol{0}, \sigma_n^2\boldsymbol{I}\}, \quad i \in \{1,\ldots,N_{\text{R}}\}$. The stacked noise vector $\boldsymbol{n} = \left(\left(\boldsymbol{n}^{(1)}\right)^T,\ldots,\left(\boldsymbol{n}^{(N_{\text{R}})}\right)^T\right)^T$ is thus zero mean, white Gaussian as well.

The frequency domain one-tap equalizer[3] $\boldsymbol{F}$ is chosen conforming to different criteria, either the zero forcing (ZF) criterion, which removes all channel distortions at risk of noise enhancement, or the minimum mean squared error (MMSE) criterion, which tries to minimize the effects of noise enhancement and channel distortion.

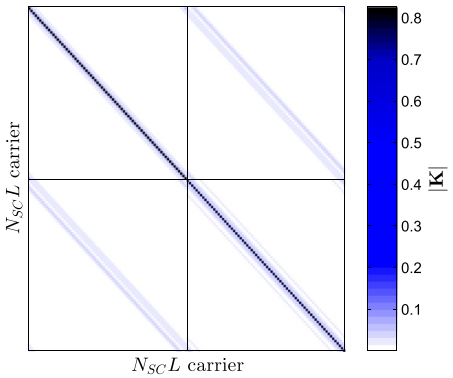
After the de-spreading operation, the data estimates $\hat{\boldsymbol{x}}$ of the noisy, received signal are given in Eq. (1), with the beforementioned convenient abbreviation (3), and $\boldsymbol{D}_{N_{\text{FFT}}}$ is the DFT matrix of size N_{FFT}.



(a) Magnitude values of the Töplitz structured MIMO time domain channel matrix with 2 transmit antennas and 4 receive antennas.



(b) Diagonalized MIMO channel matrix after CP addition/removal and IFFT/FFT at the transmitter and receiver, respectively.



(c) Input-Output matrix $\boldsymbol{K}$ employing an MMSE equalizer. The non-diagonal elements represent the intra- and interlayer interference.

Fig. 3 Examples of different channel abstractions

2.1 SC-FDM SINR

The special structure of Eq. (1), due to the frequency domain one-tap equalizer and the DFT spreading, yields a block-wise circulant input-output matrix, cf. Fig. 3c,

$$\boldsymbol{K} = \left(\boldsymbol{I}_L \otimes \boldsymbol{D}_{N_{\mathrm{SC}}}^H\right) \boldsymbol{F}\boldsymbol{H}_{\mathrm{eff}} \left(\boldsymbol{I}_L \otimes \boldsymbol{D}_{N_{\mathrm{SC}}}\right) . \tag{4}$$

This block-wise circulant structure produces a constant post equalization and post spreading SINR over all subcarriers within one layer [30]. The detailed derivation is provided in the Appendix.

$$\mathrm{SINR}^{\text{SC-FDM, }(l)} = \frac{\frac{\sigma_x^2}{N_{\mathrm{SC}}}\left|\mathbb{1}_{N_{\mathrm{SC}}}^T \boldsymbol{S}^{(l)} \mathrm{diag}(\boldsymbol{F}\boldsymbol{H}_{\mathrm{eff}})\right|^2}{\sigma_x^2\|\boldsymbol{S}^{(l)}\boldsymbol{F}\boldsymbol{H}_{\mathrm{eff}}\|_F^2 - \frac{\sigma_x^2}{N_{\mathrm{SC}}}\left|\mathbb{1}_{N_{\mathrm{SC}}}^T \boldsymbol{S}^{(l)} \mathrm{diag}(\boldsymbol{F}\boldsymbol{H}_{\mathrm{eff}})\right|^2 + \sigma_n^2\|\boldsymbol{S}^{(l)}\boldsymbol{F}\|_F^2}, \tag{5}$$

where

$$\boldsymbol{S}^{(l)} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{N_{\mathrm{SC}}} & \boldsymbol{0} \end{pmatrix} , \tag{6}$$

selects that part of $\boldsymbol{F}\boldsymbol{H}_{\mathrm{eff}}$ effecting the lth layer. The second moment (power) of the zero mean transmit symbols is depicted by σ_x^2.

2.2 OFDM SINR

In contrast to SC-FDM, no spreading takes place for OFDM. The dashed boxes in Fig. 2 are replaced by identity matrices; they are simply omitted. Thus, different subcarriers k are orthogonal/independent and the equalizer treats the corresponding subcarrier channel $\boldsymbol{H}_k$ only. We use the subscript k to denote the relevant part of the full channel matrix $\boldsymbol{H}_{\mathrm{eff}}$ for the kth subcarrier. The corresponding indices within the diagonal matrix $\boldsymbol{H}_{\mathrm{diag}}$ are $\mathbb{1}_{N_R \times N_T} \otimes \mathrm{Diag}\,(\boldsymbol{e}_k)$, with the canonical base vectors $\boldsymbol{e}_k$. Using this notation, the effective subcarrier channel $\boldsymbol{H}_k \in \mathbb{C}^{N_R \times L}$ is

$$\boldsymbol{H}_k = \left[\boldsymbol{H}_{\mathrm{diag}}\right]_{\mathbb{1}_{N_R \times N_T} \otimes \mathrm{Diag}(\boldsymbol{e}_k)} \boldsymbol{W} , \tag{7}$$

and $\boldsymbol{F}_k$ is its linear one-tap equalizer. The SINR formula is quite similar to the SC-FDM case, except that the SINR shows subcarrier dependency now. The SINR vector at layer l reads

$$\left[\mathbf{SINR}^{\text{OFDM, }(l)}\right]_k = \frac{\sigma_x^2\left|\boldsymbol{s}^{(l)}\mathrm{diag}(\boldsymbol{F}_k\boldsymbol{H}_k)\right|^2}{\sigma_x^2\|\boldsymbol{s}^{(l)}\boldsymbol{F}_k\boldsymbol{H}_k\|_2^2 - \sigma_x^2\left|\boldsymbol{s}^{(l)}\mathrm{diag}(\boldsymbol{F}_k\boldsymbol{H}_k)\right|^2 + \sigma_n^2\|\boldsymbol{s}^{(l)}\boldsymbol{F}_k\|_2^2} , \tag{8}$$

with the selection vector

$$\boldsymbol{s}^{(l)} = \begin{pmatrix} 0 \ldots 0 & 1 & 0 \ldots 0 \end{pmatrix} , \tag{9}$$

with appropriate number of zeros and a one at the lth position.

3 SC-FDM features

We first discuss the main reason to apply SC-FDM at uplink transmissions, namely PAPR. Then, we look at the expenses of employing it. We will see a worse performance of the coded transmission.

3.1 Peak-to-average-power ratio

SC-FDM is employed as the physical layer modulation scheme for LTE uplink transmission, due to its lower PAPR compared to OFDM [31]. Lower PAPR, or similarly lower crest factor, leads to reduced linearity requirements for the power amplifiers and to relaxed resolution specifications for the digital-to-analogue converters at the user equipments, entailing higher power efficiency.

The Vienna LTE-A uplink simulator calculates the discrete-time baseband PAPR with the default oversampling factor $J = 4$ [32]. The discrete time signal on transmit antenna $t \in \{1, \ldots, N_T\}$ is therefore calculated as

$$\left[\boldsymbol{s}_{\mathrm{tx}}^{(t)}\right]_m = \frac{1}{\sqrt{N_{\mathrm{FFT}}}} \sum_{k=0}^{N_{\mathrm{FFT}}-1} \left[\boldsymbol{x}_{\mathrm{pre}}^{(t)}\right]_k e^{j\frac{2\pi mk}{JN_{\mathrm{FFT}}}} , \quad 0 \le m \le JN_{\mathrm{FFT}} - 1 , \tag{10}$$

where $\boldsymbol{x}_{\mathrm{pre}}^{(t)}$ is the transmit vector right after precoding and before the IFFT at transmit antenna t. The PAPR of the stacked vector $\boldsymbol{s}_{\mathrm{tx}} = \left(\left(\boldsymbol{s}_{\mathrm{tx}}^{(1)}\right)^T, \ldots, \left(\boldsymbol{s}_{\mathrm{tx}}^{(N_T)}\right)^T\right)^T$ is calculated as

$$\begin{aligned} \mathrm{PAPR}\{\boldsymbol{s}_{\mathrm{tx}}\} &= \frac{\max\limits_{1\le t\le N_T} \max\limits_{0\le m\le JN_{\mathrm{FFT}}-1} \left(\left|\left[\boldsymbol{s}_{\mathrm{tx}}^{(t)}\right]_m\right|^2\right)}{\mathbb{E}_t\left\{\mathbb{E}_n\left\{\left|\left[\boldsymbol{s}_{\mathrm{tx}}^{(t)}\right]_m\right|^2\right\}\right\}} \\ &\approx N_T N_{\mathrm{FFT}} \|\mathrm{diag}\left(\boldsymbol{s}_{\mathrm{tx}}\boldsymbol{s}_{\mathrm{tx}}^H\right)\|_\infty / \|\boldsymbol{s}_{\mathrm{tx}}\|_2^2 , \end{aligned} \tag{11}$$

where the Euclidean norm in the denominator serves as an estimate for the ensemble average.

Figure 4 depicts the PAPR of OFDM and SC-FDM obtained for different system bandwidths. Already for a

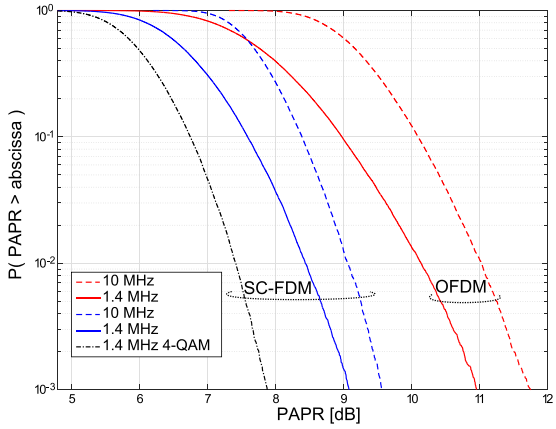


Fig. 4 PAPR for SC-FDM and OFDM for different bandwidths (1.4 and 10 MHz) and modulation alphabets (4/64 QAM)

small bandwidth (1.4 MHz), there is a significant reduction for SC-FDM over OFDM. With increasing bandwidth, OFDM's PAPR grows and the gains obtained by SC-FDM become more and more pronounced. The PAPR also depends on the modulation alphabet; the smaller the alphabet, the smaller the PAPR. This effect is illustrated in dotted lines in Fig. 4, where we have shown the PAPR of 4-QAM, exemplarily.

3.2 BER comparison over frequency selective channels

The additional spreading of SC-FDM leads to an SINR expression that is constant on all subcarriers as for single-carrier transmission, legitimating its name. The aim of this subsection is to analyse the SINR expression more carefully for the SISO case[4] and draw conclusions on BER performance.

We focus on the two most prominent equalizer concepts and start with the ZF equalizer, for whom the SC-FDM signal-to-noise ratio (SNR) expression (5) reduces to the harmonic mean

$$\mathrm{SNR}_{\mathrm{ZF}}^{\text{SC-FDM}} = \frac{\sigma_x^2}{\sigma_n^2} \frac{1}{\frac{1}{N_{\mathrm{SC}}} \sum_{k=1}^{N_{\mathrm{SC}}} \frac{1}{|\boldsymbol{H}_k|^2}}, \tag{12}$$

whereas the OFDM expression (8) is sub-carrier dependent and becomes proportional to the channel transfer function

$$\left[\mathbf{SNR}_{\mathrm{ZF}}^{\mathrm{OFDM}}\right]_k = \frac{\sigma_x^2}{\sigma_n^2} |\boldsymbol{H}_k|^2 . \tag{13}$$

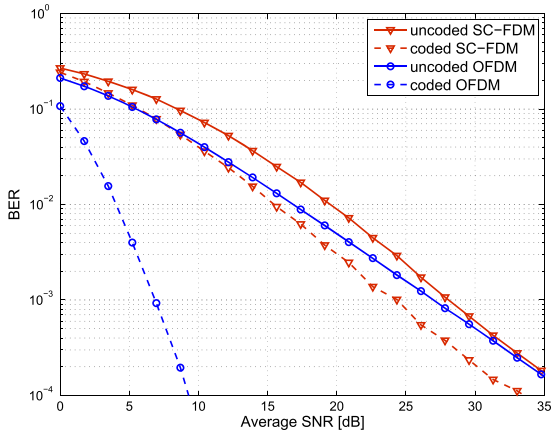
The average OFDM SNR

$$\overline{\mathrm{SNR}_{\mathrm{ZF}}^{\mathrm{OFDM}}} = \frac{\sigma_x^2}{\sigma_n^2} \frac{1}{N_{\mathrm{SC}}} \sum_{k=1}^{N_{\mathrm{SC}}} |\boldsymbol{H}_k|^2 \tag{14}$$

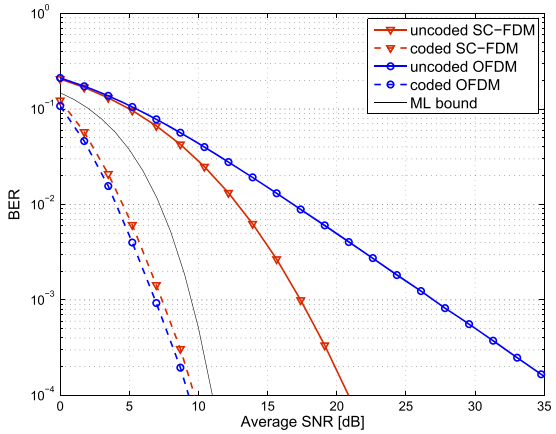
yields an upper bound on the SC-FDMA SNR due to the harmonic mean—arithmetic mean inequality [33].

$$\mathrm{SNR}_{\mathrm{ZF}}^{\text{SC-FDM}} \leq \overline{\mathrm{SNR}_{\mathrm{ZF}}^{\mathrm{OFDM}}} \tag{15}$$

Equality in Eq. (15) holds if and only if the channel is frequency flat. The difference between the harmonic mean and the arithmetic mean gets increasingly pronounced, the more selective the channel becomes. We therefore expect the (uncoded) BER of SC-FDM and ZF equalization to perform worse than OFDM, which is also validated by simulations. The BER simulations were carried out with CQI = 4 on a PedB channel [34]. This modulation and coding scheme (MCS) employs 4-QAM and has an effective code-rate of 0.3008. As expected, the BER performance of SC-FDM is worse than OFDM, both shown in Fig. 5a in solid lines. Due to the spreading, SC-FDM already expends all channel diversity and coding does not increase the SNR slope of the BER curve. This manifests in an almost parallel shift of the BER curve for SC-FDM,



(a) ZF receiver



(b) MMSE receiver

Fig. 5 BER comparison between OFDM and SC-FDM for a SISO PedB channel with 5 MHz bandwidth and fixed CQI = 4 transmission

as visual in Fig. 5a in red dashed lines. None exploited diversity allows coded OFDM to increase the BER slope considerably, cf. Fig. 5a blue dashed line.

The MMSE SINR expression is less intuitive and for the purpose of comparison, similar mathematical transformations as in [35] and [23] are required to arrive at

$$\begin{aligned}\mathrm{SINR}_{\mathrm{MMSE}}^{\text{SC-FDM}} &= \frac{\sigma_x^2}{\sigma_n^2} \frac{1 - \frac{\sigma_n^2}{\sigma_x^2} \frac{1}{N_{\mathrm{SC}}} \sum_{k=1}^{N_{\mathrm{SC}}} \frac{1}{\frac{\sigma_n^2}{\sigma_x^2} + |\boldsymbol{H}_k|^2}}{\frac{1}{N_{\mathrm{SC}}} \sum_{k=1}^{N_{\mathrm{SC}}} \frac{1}{\frac{\sigma_n^2}{\sigma_x^2} + |\boldsymbol{H}_k|^2}} \\ &= \frac{\sigma_x^2}{\sigma_n^2} \left(\frac{1}{\frac{1}{N_{\mathrm{SC}}} \sum_{k=1}^{N_{\mathrm{SC}}} \frac{1}{\frac{\sigma_n^2}{\sigma_x^2} + |\boldsymbol{H}_k|^2}} - \frac{\sigma_n^2}{\sigma_x^2} \right) .\end{aligned} \tag{16}$$

The detailed derivation is shown in the Appendix. The denominator of Eq. (16) is regularized and less sensitive to spectral notches.

An upper bound on the SINR can be obtained via the maximum of the transfer function $\boldsymbol{H}_k$

$$\begin{aligned}\mathrm{SINR}_{\mathrm{MMSE}}^{\text{SC-FDM}} &\leq \frac{\sigma_x^2}{\sigma_n^2}\left(\frac{1}{\frac{1}{\frac{\sigma_n^2}{\sigma_x^2}+\max_k|\boldsymbol{H}_k|^2}}-\frac{\sigma_n^2}{\sigma_x^2}\right)\\ &= \frac{\sigma_x^2}{\sigma_n^2}\max_k|\boldsymbol{H}_k|^2\,. \end{aligned} \quad (17)$$

In the low SNR regime $\frac{\sigma_n^2}{\sigma_x^2} \gg |\boldsymbol{H}_k|^2$, this bound becomes tight. The higher the inverse SNR $\frac{\sigma_n^2}{\sigma_x^2}$ in relation to the maximum of the transfer function, the tighter the bound becomes. The average OFDM SNR can never be larger to its maximum entry and is only equal for frequency flat channels. At low SNR, a lower BER is thus expected. Again, this presumption is validated by our simulation, showing that the uncoded BER is lower for SC-FDM as for ODFM, cf. Fig. 5b in solid lines. Although the uncoded BER shows superior performance, the coded BER is lower for OFDM due to the coding gains stemming from channel diversity, cf. Fig. 5b dashed lines.

A bound for the maximum likelihood (ML) detection performance was derived in [36]. As the bandwidth increases, the slope of the BER curve achieved with MMSE receivers tends to the slope of ML detection, demonstrating the full exploitation of channel diversity by the MMSE equalizer, cf. Fig. 5b black line.

4 Link adaptation

In the previous section, we investigated BER performance of OFDM and SC-FDM transmission with different channel models and receivers. We observed significant BER degradation of SC-FDM as compared to OFDM when ZF detection is employed, whereas coded BER is very similar when MMSE detection is used. In this section, we evaluate how such BER differences impact the actual throughput performance of LTE-A uplink when transmission rate adaptation is employed. We first consider ideal rate adaptation and compare SC-FDM transmission to OFDM with ZF and MMSE receivers. Then, we extend our single-user MIMO CSI feedback algorithms proposed for LTE downlink in [37] to LTE uplink and evaluate their performance comparing to the throughput bounds developed in [24]. We also highlight some important basic differences between link adaptation in LTE up- and downlink transmissions.

4.1 Performance with ideal rate adaptation

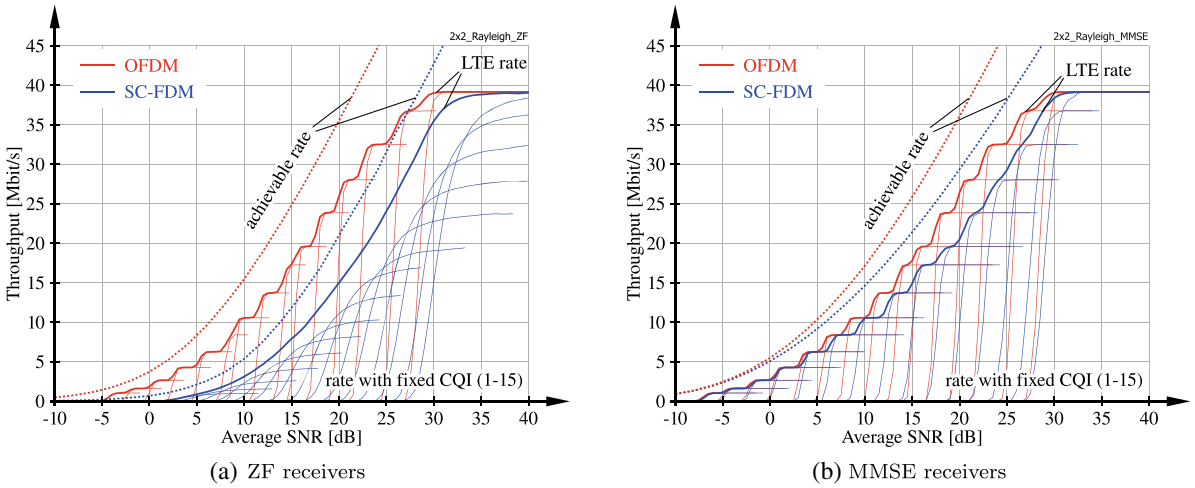
As demonstrated in the previous section, SC-FDM provides a significant advantage in terms of PAPR over OFDM, thus relaxing linearity requirements of radio frequency power amplifiers for user equipments. Yet, this comes at the cost of coded BER degradation since channel diversity is lost and the performance is mostly dominated by the weakest subcarrier of a user, especially with ZF receivers; c.f. (12). This diversity loss cannot be recovered by forward-error-correction channel coding, since the DFT-spreading applied with SC-FDM effectively causes an averaging over SINR observed on all scheduled subcarriers according to (5). As a consequence, SC-FDM transmission over frequency selective channels achieves worse throughput than OFDM. This is demonstrated in Fig. 6, where we cross-compare the achievable rate, as defined in (18) and (19), and the actual throughput of SC-FDM and OFDM transmission as obtained by the Vienna LTE-A Uplink Simulator. We consider single-user transmission over 5 MHz bandwidth assuming $N_{\mathrm{T}} = N_{\mathrm{R}} = 2$ antennas at the user and the base station and $L = 2$ spatial layers. The precoder is selected as a scaled identity matrix: $\boldsymbol{W} = 1/\sqrt{L}\,\boldsymbol{I}_L$. We consider transmission over independent and identically distributed frequency-selective Rayleigh fading channels, emphasizing the difference between OFDM and SC-FDM. The achievable rate in bits per OFDM/SCFDM symbol with Gaussian signalling and equal power allocation over subcarriers and spatial layers is calculated as

$$R^{\text{OFDM}} = \sum_{k=1}^{N_{\text{SC}}}\sum_{l=1}^{L}\log_2\left(1+\left[\mathbf{SINR}^{\text{OFDM, }(l)}\right]_k\right), \quad (18)$$

$$R^{\text{SC-FDM}} = N_{\text{SC}}\sum_{l=1}^{L}\log_2\left(1+\text{SINR}^{\text{SC-FDM, }(l)}\right), \quad (19)$$

with the receiver-specific post-de-spreading (post-equalization) SINRs from (5) and (8), respectively.

We observe a significant loss of achievable rate of SC-FDM transmission compared to OFDM in Fig. 6, which is especially pronounced with ZF receivers due to noise enhancement. In Fig. 6, we also show the actual rate achieved by LTE uplink SC-FDM transmission with ideal rate adaptation and compare to the performance obtained by OFDM transmission; the corresponding curves are denoted by *LTE rate*. We determine the performance of ideal rate adaptation by simulating all possible transmission rates, corresponding to CQI1 to CQI15, and selecting at each subframe the largest rate that achieves error free transmission. The figure also shows the throughput of the individual CQIs. We observe a gap between the LTE throughput with OFDM and SC-FDM that is similar to the gap in terms of achievable rate. Notice that the performance loss with MMSE receivers is significantly smaller than with ZF detection, since MMSE avoids excessive noise enhancement.



(a) ZF receivers (b) MMSE receivers

Fig. 6 Throughput comparison of OFDM and SC-FDM with rate adaptation and 2 × 2 Rayleigh fading channels of 5-MHz bandwidth

We also observe in Fig. 6a that the gain achieved by instantaneous rate adaptation, as compared to rate adaptation based on the long-term average SNR, is much larger for ZF SC-FDM than for ZF OFDM; this is evident from the distance between the curves with rate adaptation (LTE rate) and the curves with *fixed CQI*. The reason for this behaviour is that the SNR of ZF SC-FDM shows strong variability around its means, since it is dominated by the worst-case per-subcarrier SNR according to (12); the average SNR over subcarriers of ZF OFDM, however, approximately coincides with its mean value. This implies that the optimal CQI of ZF SC-FDM can vary significantly in-between subframes, as reflected by the large average SNR variation required to increase the rate with fixed CQI from zero to its respective maximum. Yet, for ZF OFDM, the throughput of the individual CQIs follows almost a step function; hence, rate adaptation can be based on the long-term average SNR without substantial performance degradation.[5]

In case $N_\mathrm{R} > L$, we can easily estimate the achievable rate of SC-FDM transmission: The per-layer SNR with ZF receivers is governed by the harmonic mean of the channel responses on the individual subcarriers, similar to (12)

$$\mathrm{SNR}_\mathrm{ZF}^{\mathrm{SC-FDM},\,(l)} = \frac{\sigma_x^2/\sigma_n^2}{\frac{1}{N_\mathrm{SC}}\sum_{k=1}^{N_\mathrm{SC}}\left[\left(\boldsymbol{H}_k\boldsymbol{W}\right)^\mathrm{H}\left(\boldsymbol{H}_k\boldsymbol{W}\right)\right)^{-1}\Big]_{l,l}}, \tag{20}$$

with $\boldsymbol{H}_k \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{T}}$ denoting the OFDM channel matrix on subcarrier k. Assuming constant precoding and semi-correlated Rayleigh fading

$$\boldsymbol{H}_k = \tilde{\boldsymbol{H}}_k\boldsymbol{C}_T^{\frac{1}{2}}, \quad \left[\tilde{\boldsymbol{H}}_k\right]_{i,j} \sim \mathcal{CN}\left\{0,1\right\}, \tag{21}$$

with $\boldsymbol{C}_T \in \mathbb{C}^{N_\mathrm{T}\times N_\mathrm{T}}$ determining the spatial correlation at the user equipment side, the matrix in the denominator of (20) follows a complex inverse Wishart distribution with N_R degrees of freedom and scale matrix $\boldsymbol{C} = \left(\boldsymbol{W}^\mathrm{H}\boldsymbol{C}_T\boldsymbol{W}\right)^{-1}$

$$\overline{\boldsymbol{H}} = \left(\left(\boldsymbol{H}_k\boldsymbol{W}\right)^\mathrm{H}\left(\boldsymbol{H}_k\boldsymbol{W}\right)\right)^{-1} \sim \mathcal{CW}_L^{-1}\left\{N_\mathrm{R},\boldsymbol{C}\right\}. \tag{22}$$

Letting $N_\mathrm{SC} \to \infty$, we can replace the term in the denominator of (20) with its expected value

$$\frac{1}{N_\mathrm{SC}}\sum_{k=1}^{N_\mathrm{SC}}\left[\,\overline{\boldsymbol{H}}\,\right]_{l,l} \overset{N_\mathrm{SC}\to\infty}{\longrightarrow} \mathbb{E}\left(\left[\,\overline{\boldsymbol{H}}\,\right]_{l,l}\right). \tag{23}$$

This expected value only exists in case $N_\mathrm{R} > L$ [38]. For $N_\mathrm{R} = L$, the diagonal elements of $\overline{\boldsymbol{H}}$ follow a heavy-tailed inverted Gamma distribution [39, 40] with non-finite first moment. Yet, for $N_\mathrm{R} > L$, which is a common situation in cellular networks since the base station is mostly equipped with far more antennas than the users, the expected value is

$$\mathbb{E}\left(\left[\,\overline{\boldsymbol{H}}\,\right]_{l,l}\right) = \frac{1}{N_\mathrm{R}-L}\left[\boldsymbol{C}\right]_{l,l}. \tag{24}$$

Hence, we can estimate the achievable rate of SC-FDMA transmission over semi-correlated Rayleigh fading channels

$$R^{\mathrm{SC-FDM}} \approx N_\mathrm{SC}\sum_{l=1}^{L}\log_2\left(1+\frac{\sigma_x^2/\sigma_n^2}{[\boldsymbol{C}]_{l,l}}(N_\mathrm{R}-L)\right) \tag{25}$$

$$\approx N_\mathrm{SC}L\left[\log_2\left(\frac{\sigma_x^2/\sigma_n^2}{\left(\prod_{l=1}^{L}[\boldsymbol{C}]_{l,l}\right)^{1/L}}\right)+\log_2\left(N_\mathrm{R}-L\right)\right]. \tag{26}$$

Here, (26) resembles the high SNR approximation of the achievable rate of OFDM transmission with ZF detection as proposed in ([41] Eq. (14)); even more, for fixed L and letting N_R grow to infinity, (26) and ([41] Eq. (14)) tend to the same limit, due to channel hardening on each subcarrier with growing number of receive antennas.

In Fig. 7, we investigate the performance of the rate estimate (25) for $N_\mathrm{T} = L = 4$ and varying number of receive antennas. We assume $\boldsymbol{W} = 1/\sqrt{L}\,\boldsymbol{I}_L$ and

$$\boldsymbol{C}_T = \begin{bmatrix} 1 & 0.9 & \dots & 0.9 \\ 0.9 & \ddots & & \vdots \\ \vdots & & & 0.9 \\ 0.9 & \dots & 0.9 & 1 \end{bmatrix},$$

and consider the smallest LTE bandwidth of $N_\mathrm{SC} = 72$ subcarriers. We observe that the proposed estimate performs very well even at this small bandwidth; notice, though, that a more realistic channel model with correlation over subcarriers may require larger bandwidth to validate the proposed estimate. Figure 7 also confirms the observation that single-user MIMO OFDM and SC-FDM with ZF detectors tend to the same limiting performance with increasing number of receive antennas.

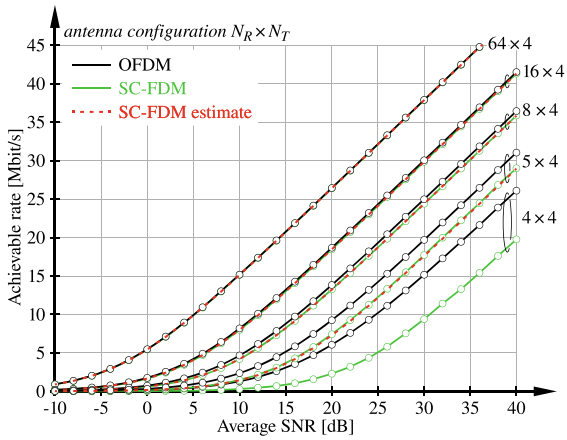
This statement, however, will not hold true if the total number of layers grows proportionally with the number of receive antennas. For example, multi-user MIMO transmission with ZF equalization and single-antenna users achieves only a diversity order of $N_\mathrm{R} - L + 1$ [42], with L denoting the total number of layers being equal to the number of spatially multiplexed users. Hence, if L scales proportionally with N_R, channel hardening on each subcarrier will not occur and thus the performance of OFDM and SC-FDM will not coincide.

4.2 Performance with realistic link adaptation

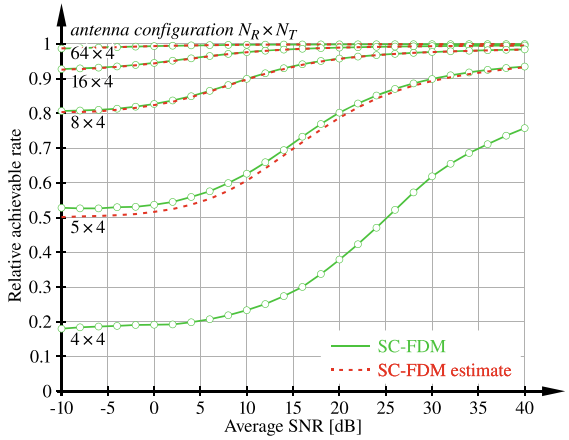
Instantaneous rate adaptation is an important tool for exploiting diversity of the wireless channel in LTE, by adjusting the transmission rate according to the current channel quality experienced by a user. LTE specifies a set of 15 different MCSs; the selected MCS is signalled by the CQI.

LTE additionally supports spatial link adaptation by means of codebook-based precoding with variable transmission rank. With this method, the precoding matrix $\boldsymbol{W} \in \mathbb{C}^{N_\mathrm{T} \times L}$ satisfying $\boldsymbol{W}^\mathrm{H}\boldsymbol{W} = 1/L\,\boldsymbol{I}_L$ is selected from a standard-defined codebook $\mathcal{W}_L$ of scaled semi-unitary matrices; furthermore, the number of spatial layers L can be adjusted to achieve a favourable trade-off between beamforming and spatial multiplexing. The selected precoder and transmission rank are signalled, employing the PMI and the RI. In single-user MIMO LTE uplink transmission, the same precoder is applied on all RBs that are assigned to a specific user, whereas frequency-selective precoding is supported in LTE downlink.

There is a basic difference between the utilization of CQI, PMI and RI in up- and downlink directions of frequency division duplex (FDD) systems. In the downlink, the base station is reliant on CSI feedback from the users for link adaptation and multi-user scheduling [43], since channel reciprocity cannot be exploited in FDD. CQI, PMI and RI can be employed to convey such CSI from the users to the base station via dedicated feedback channels [44]. In the uplink, on the other hand, the base station can by itself



(a) Comparison of achievable rate and the rate estimate of Section 4.1.



(b) Relative achievable rate of SC-FDM with respect to OFDM.

Fig. 7 Achievable rate of OFDM and SC-FDM with ZF equalizers and growing number of receive antennas at fixed number of streams $L = 4$

determine CSI exploiting the sounding reference signals (SRSs) transmitted by the users. In this case, CQI, PMI and RI are employed by the base station to convey to the users its decision on link adaptation that has to be applied by the users during uplink transmission.

In principal, link adaptation must be jointly optimized with multi-user scheduling to optimize the performance of the system, since the effective SC-FDM SINR (and thus the rate) of a user depends on the assigned RBs according to (5). For reasons of computational complexity, however, we assume that the multi-user schedule is already fixed and determine link adaptation parameters based on this resource allocation. We modify the approach proposed in [37] for LTE downlink transmission to determine the link adaptation parameters in four steps:

1. Determine the optimal precoder for each transmission rank $L \leq \min(N_{\mathrm{T}}, N_{\mathrm{R}})$ by maximizing transmission rate

$$\hat{\boldsymbol{W}}(L) = \underset{\boldsymbol{W}\in\mathcal{W}_L}{\arg\max} \sum_{l=1}^{L} f\left(\mathrm{SINR}^{\text{SC-FDM, }(l)}(\boldsymbol{W})\right). \tag{27}$$

   Here, function $f(\cdot)$ maps SINR to rate; this could be either an analytical mapping, such as (19), or a mapping table representing the actual performance of LTE. In our simulations, we employ the bit interleaved coded-modulation (BICM) capacity as proposed in [37], since LTE is based on a BICM architecture.
2. Determine the optimal LTE transmission rates per layer for each L and $\hat{\boldsymbol{W}}(L)$. We employ a target block error ratio (BLER) mapping in our simulations to determine the highest rate that achieves BLER ≤ 0.1.
3. Select the transmission rank $\hat{L}$ that maximizes the sum rate over spatial layers, utilizing the LTE transmission rates determined above.
4. Set the RI and PMI according to $\hat{L}$ and $\hat{\boldsymbol{W}}(L)$, respectively, and set the pCQI conforming to the corresponding LTE transmission rates.

In Fig. 8, we evaluate the performance of single-user MIMO LTE uplink transmission over $N_{\mathrm{T}} = N_{\mathrm{R}} = 4$ antennas with link adaptation, 1.4 MHz system bandwidth and ZF receiver. We do not consider signalling delays between the base station and the user. We employ the VehA channel model [34] and compare the absolute and relative (to channel capacity) throughput to the performance bounds proposed in [24].[6] *Channel capacity* is obtained by applying singular value decomposition (SVD)-based transceivers and water-filling power allocation over subcarriers and spatial streams. Notice that we do not account for guard band and CP overheads when calculating the channel capacity; that is, we only consider subcarriers that are available for data transmission. The *achievable channel capacity* takes overhead for pilot symbols (DMRS and SRS) into account, corresponding to a loss of 16.7 % in our simulation. The *achievable BICM bound* additionally accounts for equal power allocation, codebook-based precoding and ZF detection as well as the applied BICM architecture as detailed in [24].

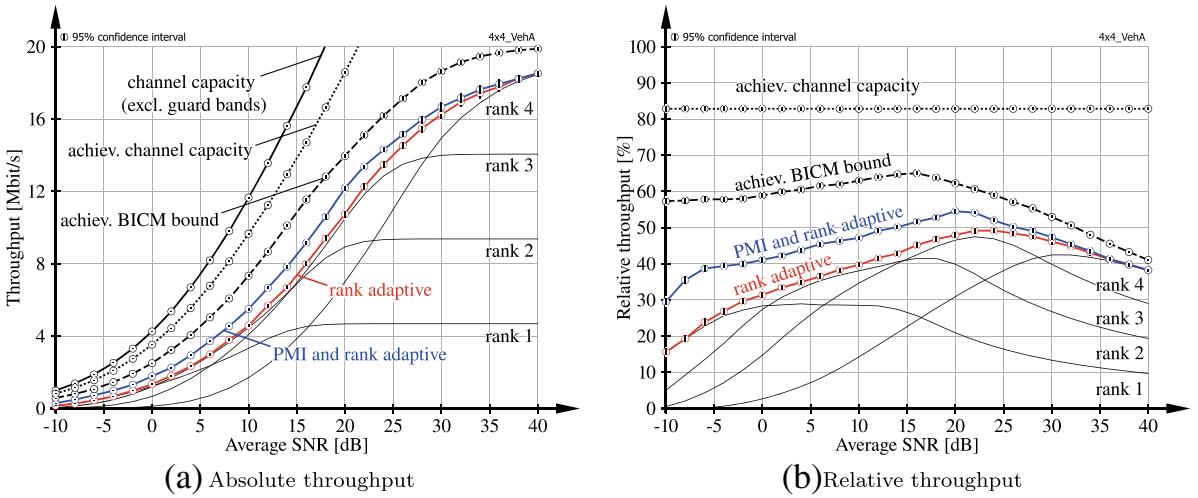


Fig. 8 Absolute and relative throughput of LTE uplink transmission over 4 × 4 VehA channels of 1.4-MHz bandwidth employing rate adaptation. We compare the performance of fixed rank, rank adaptive and PMI + rank adaptive transmission to the performance bounds proposed in [24]

The performance of LTE uplink transmission with full link adaptation (*PMI and rank adaptive*) is similar to the achievable BICM bound but shifted by approximately 3 dB. Notice that the saturation value is not the same because the highest CQI of LTE achieves 5.55 bit/channel use, whereas the BICM bound saturates at 6 bit/channel use. We also show the performance of LTE uplink when restricted to fixed precoding (*rank adaptive*) and fixed rank transmission (*ranks 1, 2, 3, 4*). We observe that rank adaptive transmission even outperforms the envelope of the fixed rank transmission curves, since instantaneous rank adaptation selects the optimal rank in each subframe independently. In terms of relative throughput, we see that LTE uplink with ZF receivers achieves around 40–50 % of channel capacity; remember, though, that this does not include CP and guard band overheads.

5 Reference symbols

In LTE uplink, two types of reference signals are standardized. For CE and coherent detection, DMRS are exploited, while SRS are employed for channel sounding to enable frequency selective scheduling. For the purpose of CE, we will consider DMRS only. The reference symbols are defined in [1] and are explained in more detail in [45, 46]. As shown in Fig. 9, DMRS are multiplexed in the resource grid at OFDM symbol time $n = 3$ in every slot. In a physical uplink shared channel (PUSCH) transmission of the LTE-A uplink, a DMRS occupies all scheduled subcarriers. We assume that the user is assigned to all N_{SC} subcarriers starting at 0, i.e., $k \in \{0, 1, \ldots, N_{\mathrm{SC}} - 1\}$. We denote the Zadoff-Chu (ZC) base sequence on N_{SC} subcarriers for one slot by $\bar{\boldsymbol{r}} \in \mathbb{C}^{N_{\mathrm{SC}} \times 1}$. The base sequences $\bar{\boldsymbol{r}}$ are complex exponential sequences lying on the unit circle fulfilling

$$|[\bar{\boldsymbol{r}}]_k| = 1 \,. \tag{28}$$

In LTE-A, the DMRS of different transmission layers in the same slot are orthogonal in terms of frequency domain code division multiplexing (FD-CDM) [45]. This

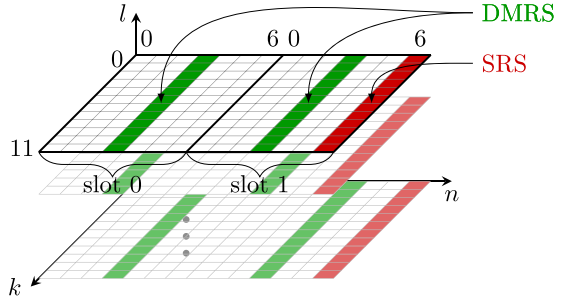


Fig. 9 The LTE-A uplink reference symbol allocation in two slots (one subframe)

is obtained by cyclically shifting the base sequence. Similar to [47], DMRS on layer l for one slot are given by

$$\boldsymbol{R}^{(l)} = \mathrm{Diag}\left(\boldsymbol{r}^{(l)}\right) = \boldsymbol{T}^{(l)}\mathrm{Diag}(\bar{\boldsymbol{r}}) \,, \tag{29}$$

with the cyclic shift operator

$$\boldsymbol{T}^{(l)} = \mathrm{Diag}\left(e^{j0}, \ldots, e^{j\alpha_l k}, \ldots, e^{j\alpha_l (N_{\mathrm{SC}}-1)}\right) \,, \tag{30}$$

and the layer dependent cyclic shift α_l. We further conclude from (28) to (30) that $(\boldsymbol{R}^{(l)})^H = (\boldsymbol{R}^{(l)})^{-1}$ which implies $\left(\boldsymbol{R}^{(l)}\right)^H \boldsymbol{R}^{(l)} = \boldsymbol{I}_{N_{\mathrm{SC}}}$. Exploiting (28), the product of two DMRS from layers l and u with $l, u \in \{1, \ldots, L\}$ becomes

$$\begin{aligned}\left(\boldsymbol{R}^{(l)}\right)^H \boldsymbol{R}^{(u)} &= \left(\boldsymbol{T}^{(l)}\right)^H \boldsymbol{T}^{(u)}\mathrm{Diag}\,(\bar{\boldsymbol{r}})^H \mathrm{Diag}\,(\bar{\boldsymbol{r}}) \\ &= \mathrm{Diag}\left(e^{j0} \;\ldots\; e^{j\Delta\alpha k} \;\ldots\; e^{j\Delta\alpha (N_{\mathrm{SC}}-1)}\right)\boldsymbol{I} \,,\end{aligned} \tag{31}$$

with $\Delta\alpha = \alpha_u - \alpha_l$ being the cyclic phase shift between DMRS of two different spatial layers. The FD-CDM orthogonality can therefore be exploited as

$$\mathrm{trace}\left(\left(\boldsymbol{R}^{(u)}\right)^H \boldsymbol{R}^{(l)}\right) = \left(\boldsymbol{r}^{(u)}\right)^H \boldsymbol{r}^{(l)} = \begin{cases} N_{\mathrm{SC}} & \text{for } u = l \\ 0 & \text{for } u \neq l \,. \end{cases} \tag{32}$$

After transmission over a frequency selective channel, this orthogonality has to be exploited to separate all effective MIMO channels at the receiver.

6 Channel estimation

For channel estimation we exploit the system model only at symbol times, where reference signals are allocated. For a normal CP length, this is the 4th symbol in each slot, i.e. $n = 3$ as shown in Fig. 9. Since we estimate the channel only at this single symbol time per slot, interpolation in time has to be carried out to obtain channel estimates for the whole resource grid. The effects of interpolation will be studied in Section 7. As illustrated in Fig. 2, the DMRS are added after DFT spreading, right before precoding. As the channel estimation takes place after the receiver's DFT, just before equalization, the system model for CE amounts to an OFDM system. The system model (1) therefore reads as

$$\boldsymbol{y} = \boldsymbol{H}_{\mathrm{eff}}\boldsymbol{r} + \boldsymbol{n}' \,, \tag{33}$$

with (pre-equalization) noise

$$\boldsymbol{n}' = \left(\boldsymbol{I}_{N_R} \otimes \boldsymbol{M}^H \boldsymbol{D}_{N_{\mathrm{FFT}}} \boldsymbol{P}_{\mathrm{remCP}}\right)\boldsymbol{n} \,, \tag{34}$$

and the stacked vector $\boldsymbol{r}$ consisting of DMRS $\boldsymbol{r}^{(l)} \in \mathbb{C}^{N_{\mathrm{SC}} \times 1}$ from all active spatial layers $l \in \{1, \ldots, L\}$, i.e. $\boldsymbol{r} = \left(\left(\boldsymbol{r}^{(1)}\right)^T, \ldots, \left(\boldsymbol{r}^{(L)}\right)^T\right)^T$. To consider the received signal separately for each receive antenna i, we can select the

according part from $\boldsymbol{y}$ by left multiplying with the selector matrix $\boldsymbol{S}^{(i)}$ from (6). The received signal $\boldsymbol{y}^{(i)} = \boldsymbol{S}^{(i)}\boldsymbol{y}$ on antenna i is given by

$$\begin{aligned}\boldsymbol{y}^{(i)} &= \left(\boldsymbol{H}_{\text{eff}}^{(i,1)}, \ldots, \boldsymbol{H}_{\text{eff}}^{(i,L)}\right)\boldsymbol{r} + \boldsymbol{n}'^{(i)} \\ &= \sum_{l=1}^{L} \boldsymbol{H}_{\text{eff}}^{(i,l)}\boldsymbol{r}^{(l)} + \boldsymbol{n}'^{(i)}\,,\end{aligned} \tag{35}$$

with the pre-equalization noise $\boldsymbol{n}'^{(i)} = \boldsymbol{S}^{(i)}\boldsymbol{n}'$ on receive antenna i and $\boldsymbol{H}_{\text{eff}}^{(i,l)} = \boldsymbol{S}^{(i)}\boldsymbol{H}_{\text{eff}}\left(\boldsymbol{S}^{(l)}\right)^T$ being the $(i,l)^{\text{th}}$ block of $\boldsymbol{H}_{\text{eff}}$. Since $\boldsymbol{H}_{\text{eff}}^{(i,l)}$ is diagonal, we exploit the relations $\boldsymbol{R}^{(l)} = \text{Diag}\left(\boldsymbol{r}^{(l)}\right)$ and $\boldsymbol{h}_{\text{eff}}^{(i,l)} = \text{diag}\left(\boldsymbol{H}_{\text{eff}}^{(i,l)}\right)$ to estimate a channel vector rather than a matrix and rearrange terms in (35) leading to

$$\begin{aligned}\boldsymbol{y}^{(i)} &= \sum_{l=1}^{L} \boldsymbol{R}^{(l)}\boldsymbol{h}_{\text{eff}}^{(i,l)} + \boldsymbol{n}'^{(i)} \\ &= \underbrace{\left(\boldsymbol{R}^{(1)}, \ldots, \boldsymbol{R}^{(L)}\right)}_{\boldsymbol{R}}\boldsymbol{h}_{\text{eff}}^{(i)} + \boldsymbol{n}'^{(i)}\,,\end{aligned} \tag{36}$$

with the stacked vector $\boldsymbol{h}_{\text{eff}}^{(i)} = \left(\left(\boldsymbol{h}_{\text{eff}}^{(i,1)}\right)^T, \ldots, \left(\boldsymbol{h}_{\text{eff}}^{(i,L)}\right)^T\right)^T$ of all effective channels from L active layers to receive antenna i for which we will drop the subscript in the following.

6.1 Minimum mean square error estimation

First, we present a MMSE estimator where we exploit (36) and estimate the stacked vector $\boldsymbol{h}^{(i)}$ consisting of effective channels from all L active layers to receive antenna i. The MMSE CE for receive antenna i is given by

$$\hat{\boldsymbol{h}}_{\text{MMSE}}^{(i)} = \underset{\hat{\boldsymbol{h}}^{(i)}}{\arg\min}\, \mathbb{E}\left\{\left\|\hat{\boldsymbol{h}}^{(i)} - \boldsymbol{h}^{(i)}\right\|_2^2\right\}\,, \tag{37}$$

which leads to the well-known solution [48]

$$\hat{\boldsymbol{h}}_{\text{MMSE}}^{(i)} = \left(\sigma_{\boldsymbol{n}^{(i)}}^2\boldsymbol{C}_{\boldsymbol{h}^{(i)}}^{-1} + \boldsymbol{R}^H\boldsymbol{R}\right)^{-1}\boldsymbol{R}^H\boldsymbol{y}^{(i)}\,, \tag{38}$$

with $\boldsymbol{C}_{\boldsymbol{h}^{(i)}} = \mathbb{E}\{\boldsymbol{h}^{(i)}\boldsymbol{h}^{(i)H}\}$.

6.2 Correlation-based estimation

As a low complexity approach, we correlate (matched filter) the received signal with the reference symbol of layer l to obtain a channel estimate for the effective channel $\boldsymbol{h}^{(i,l)}$ from layer l to receive antenna i

$$\tilde{\boldsymbol{h}}^{(i,l)} = \left(\boldsymbol{R}^{(l)}\right)^H\boldsymbol{y}^{(i)}\,. \tag{39}$$

In fact, this correlation approach is optimum in a least squares (LS) sense

$$\boldsymbol{h}_{\text{LS}}^{(i)} = \underset{\boldsymbol{h}}{\arg\min}\,\|\boldsymbol{y}^{(i)} - \boldsymbol{R}\boldsymbol{h}\|_2 = \left(\underbrace{\boldsymbol{R}\boldsymbol{R}^H}_{L\boldsymbol{I}}\right)^{-1}\boldsymbol{R}^H\boldsymbol{y}^{(i)}\,. \tag{40}$$

Inserting our system model (36) and exploiting (31), we obtain

$$\begin{aligned}\tilde{\boldsymbol{h}}^{(i,l)} &= \left(\boldsymbol{R}^{(l)}\right)^H\textstyle\sum_{u=1}^{L}\boldsymbol{R}^{(u)}\boldsymbol{h}^{(i,u)} + \left(\boldsymbol{R}^{(l)}\right)^H\boldsymbol{n}'^{(i)} \\ &= \boldsymbol{h}^{(i,l)} + \underbrace{\sum_{\substack{u=1\\u\neq l}}^{L}\left(\boldsymbol{T}^{(l)}\right)^H\boldsymbol{T}^{(u)}\boldsymbol{h}^{(i,u)}}_{\text{inter-layer interference}} + \tilde{\boldsymbol{n}}^{(i)}\,.\end{aligned} \tag{41}$$

Here, $\tilde{\boldsymbol{n}}^{(i)}$ has the same distribution as $\boldsymbol{n}'^{(i)}$ since $(\boldsymbol{R}^{(l)})^H$ is unitary and introduces phase changes only, cf. (29). Due to the allocation of DMRS on the same time and frequency resources on different spatial layers, the initial estimate $\tilde{\boldsymbol{h}}^{(i,l)}$ of one effective MIMO channel actually consists of a superposition of all L effective MIMO channels to receive antenna i. The unintentional contributions in (41), from layers $u \neq l$ are inter-layer interference, making it unsuited as initial estimate for coherent detection. Different methods to separate the different effective MIMO channels in (41) will be presented in the following.

6.2.1 DFT-based channel estimation

A well-known approach for CE in LTE-A uplink is DFT-based estimation [46], which aims to separate the MIMO channels contributing to (41) in time domain. For this, the individual cyclic shift of each DMRS is exploited. Applying a DFT on the receive signal, the individual phase shifts will translate into shifts in time domain. This makes a separation of channel impulse response (CIR)s from different MIMO channels possible by windowing. In our simulator, we implemented a DFT-based estimator as in [49] or [47].

6.2.2 Averaging

For physically meaningful channels, neighbouring subcarriers will be correlated within the coherence bandwidth [50]. We utilize this property and exploit the DMRS structure to perform frequency domain CE. As explained in [25], applying a sliding averaging on the initial estimate $\tilde{\boldsymbol{h}}^{(i,l)}$ from (41) over $\bar{\gamma}$ adjacent subcarriers ($\bar{\gamma}$ equals 1,2,4,4 for L equals 1, 2, 3, 4, respectively) cancels the inter-layer interference, assuming the channel to be frequency flat on these $\bar{\gamma}$ consecutive subcarriers. The sliding average is given by

$$\left[\hat{\boldsymbol{h}}_{\text{SAV}}^{(i,l)}\right]_k = \frac{1}{\bar{\gamma}^2}\sum_{t=k-\bar{\gamma}+1}^{k}\sum_{j=t}^{t+\bar{\gamma}-1}\left[\tilde{\boldsymbol{h}}^{(i,l)}\right]_j\,, \tag{42}$$

for $\bar{\gamma} \leq k \leq N_{\mathrm{SC}} - \bar{\gamma} + 1$. The second sum describes the averaging of $\bar{\gamma}$ elements while the first sum describes the shift of this averaging window.

6.2.3 Quadratic smoothing

Another method exploiting channel correlations to estimate the channel in frequency domain is quadratic smoothing (QS). This scheme cannot remove the inter-layer interference entirely, which manifests in a higher error floor, but shows improved performance at lower SNR in return. As explained in [25], this estimation method, exploiting the smoothing matrix $\boldsymbol{Q}$ and a smoothing factor γ, is given by

$$\hat{\boldsymbol{h}}_{\mathrm{QS}}^{(i,l)} = \left(\boldsymbol{I}_{N_{\mathrm{SC}}L} + \lambda \boldsymbol{Q}^H \boldsymbol{Q}\right)^{-1} \underbrace{\left(\boldsymbol{R}^{(l)}\right)^H \boldsymbol{y}^{(i)}}_{\tilde{\boldsymbol{h}}^{(i,l)}} . \tag{43}$$

Similar to (42), this can be interpreted as another way to cope with the inter-layer interference in (41) by post processing. This method does not use the DMRS structure explicitly but suppresses the interference by smoothing. It is therefore not able to cancel the complete inter-layer interference but shows an improved performance at low SNR.

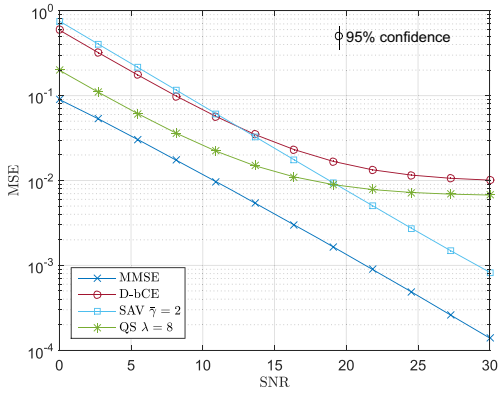
6.3 MSE and BER comparison

We assume a single user 2×2 MIMO transmission with $N_{\mathrm{SC}} = 72$ subarriers, a fixed number of layers $L = 2$ and a typical urban (TU) channel model [34] at zero speed. We perform a simulation with one-point extrapolation, cf. Section 7, and show the MSE curves of the proposed estimators in Fig. 10a. The DFT-based CE (*D-bCE*) shows the highest error flow of all estimators at high SNR while the *MMSE* estimator of course shows best performance over the whole SNR range. Compared to these two methods, the Sliding-Averaging estimator (42), denoted by *SAV*, encounters an 8-dB SNR penalty when compared to MMSE but comes closest to *MMSE* performance at high SNR. The quadratic smoothing estimation is denoted by *QS* and shows a significant improvement for low SNRs because it smooths over several observed channel coefficients. Quadratic smoothing performs uniformly better than *D-bCE* over the whole SNR range and comes close to 4 dB to MMSE at low SNR. The high error floor shows that QS is not able to cancel all the inter-layer interference.

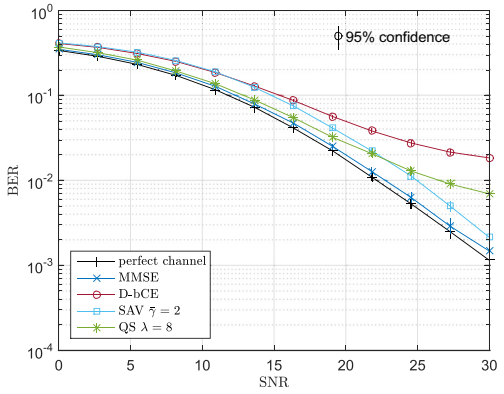
In terms of BER performance, at high SNR, naturally the estimation method with the lowest MSE leads to the smallest BER. At low SNR, the difference in CE MSE translates into very small differences in BER, meaning, we cannot gain too much from a good low SNR MSE performance of QS or MMSE estimation. Considering estimation complexity and that MMSE as well as QS require prior channel knowledge, *SAV* estimation is a good complexity performance trade-off.

7 Channel interpolation

Under fast fading conditions, additional effects influence the performance of LTE uplink transmissions. Doppler shifts degrade the SINR by introducing velocity dependent ICI [51] whereas the SINR increases with increasing subcarrier spacing. The subcarrier spacing of 15 kHz that is used in LTE makes transmissions quite robust against ICI. The impact of ICI becomes only evident at high velocities and high SNR. Figure 12b shows the BER for the case of perfect channel knowledge where the performance is only degraded by noise and ICI. At 200 km/h, the BER saturates due to ICI at high SNR whereas ICI mitigation techniques



(a) MSE curves of the proposed CE methods for a 2×2 transmission with $L = 2$ on a TU channel.



(b) Resulting uncoded BER curves for the proposed CE schemes for 4QAM.

Fig. 10 Channel estimation performance comparison for block fading

[52] show promising results to reduce this impact of ICI.

Another effect that hampers LTE transmissions at high velocities are temporal channel interpolation errors. While in the LTE downlink, the pattern used to multiplex data and reference symbols is a good trade-off between a small temporal and spectral spacing accounting for highly frequency selective channels and fast-fading channels and a rather small overhead, this is different in the uplink. As shown in Fig. 11a, uplink DMRSs occupy the whole subband. While there is no need for interpolation over frequency, the temporal spacing is about twice the spacing of the reference symbols in the downlink. Furthermore, if frequency hopping is performed, the number of adjacent pilots transmitted in the same subband is two for inter-subframe frequency hopping and only one for intra-subframe frequency hopping where frequency hopping is performed on a per-slot basis. Due to this special structure channel, interpolation in the LTE uplink is a challenging problem. Therefore, we investigated various channel interpolation techniques using a single, two or three consecutive pilot symbols. Figure 11b–e illustrates the channel interpolation techniques considered. The highest channel interpolation errors (Fig. 12a) are observed for *1 point extrapolation* where the channel estimate obtained in a certain slot is used to equalize the symbols within that slot and no interpolation is performed at all. The higher the number of pilots involved in channel interpolation, the lower the MSE gets. The results in terms of BER in Fig. 12b show a similar behaviour.

For a measurement-based comparison of interpolation techniques using channel estimates form both, the previous and the subsequent subframe, the reader is referred to [53].

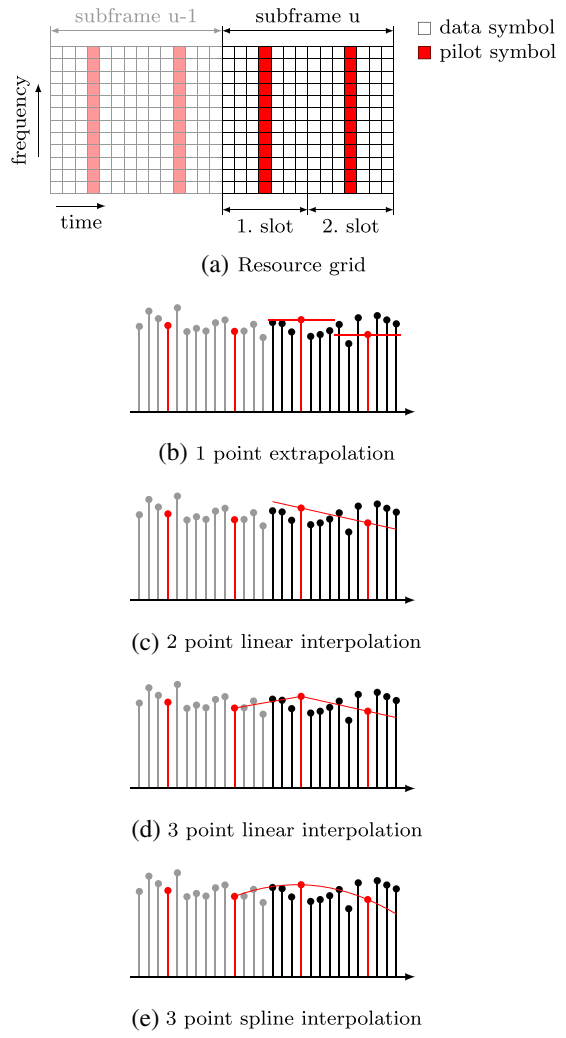


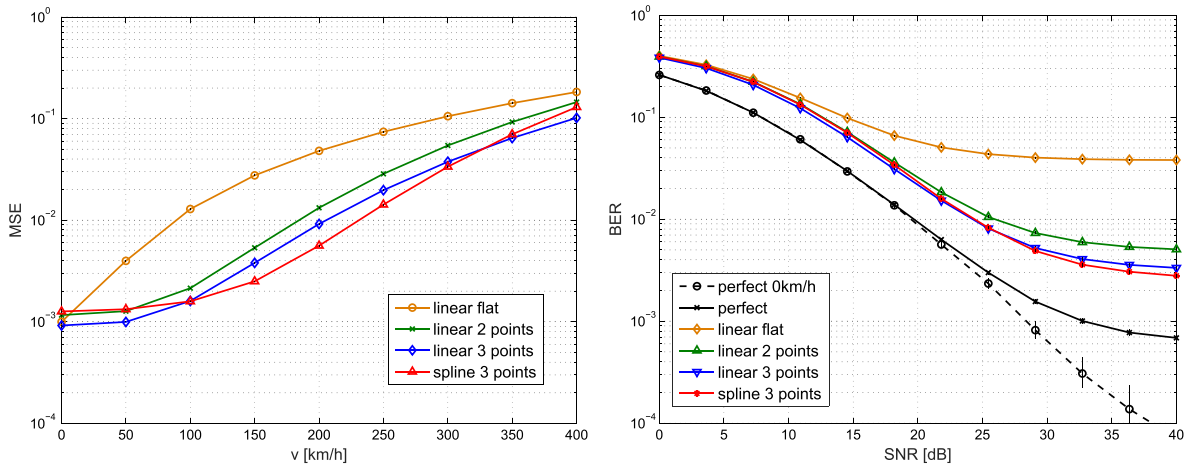
Fig. 11 Channel interpolation techniques for the LTE-A uplink pilot pattern **a** using estimates from **b** the actual slot, **c** the actual subframe and **d**–**e** the actual and previous subframe

8 Future research questions

Until now our research efforts on the Vienna LTE-A Uplink Simulator have been concentrated on single links between user and base station, focusing on basic transceiver issues such as link adaptation and channel estimation. Our treatment of the link performance analysis is not considered complete. There are still important parameters to investigate, such as different forms of channel coding, enhanced channel estimation and detection [54, 55] and analysis of SC-FDM sensitivity to synchronization mismatch, similar to our downlink investigations [56].

In the future, our scope will shift to multi-user multi-base station scenarios, enabling on one hand exploitation of multi-user diversity in space, time and frequency and, on the other hand, consideration of interference in-between simultaneous transmissions from multiple base stations. Even though, for reasons of computational complexity, simulations will be confined to comparatively small scenarios containing some few base stations, we still expect to extract valuable performance indicators for coordinated multipoint reception schemes [57], accounting for practical constraints, such as, limited back-haul capacity.

We will address cross-layer multi-user scheduling, jointly optimizing multi-user resource allocation and per-user link adaptation; this is an intricate issue in LTE, due to the non-linear relationship between the resources assigned to a user and its corresponding SC-FDM SINR (5); we have already addressed this issue for the downlink in [43]. Multi-user scheduling, furthermore, has to find a favourable trade-off between transmission efficiency and fairness of resource allocation. We will extend existing downlink schedulers, which enable Pareto-efficient transmission with arbitrary fairness, to



(a) Channel estimation and interpolation error at 30dB SNR.

(b) Uncoded BER for 4-QAM at 200 km/h.

Fig. 12 Comparison of channel interpolation techniques using different numbers of reference symbols and LS-SAV channel estimation

the uplink specifics and compare to other proposals, e.g. [58].

The realization of massive MIMO in LTE compliant systems is another highly important research topic, since it promises an order of magnitude network efficiency gains through spatial multiplexing of users [59–61]. Yet, many issues still need to be better understood and resolved to enable efficient massive MIMO transmission in practice. One important step towards reasonable performance investigation of massive antenna arrays is to employ realistic channel models, such as, the 3GPP three-dimensional channel model [62], which we plan to incorporate in future releases of our simulator.

9 Conclusions

For an LTE-A uplink transmission model, we derived SINR expressions, both with and without DFT precoding. We specialized these equations to ZF and MMSE receivers and showed that ZF performance is strongly affected by the worst subcarrier. Comparing the resulting BER we revealed that SC-FDM performance is generally inferior to OFDM and that applying MMSE equalization is crucial to get closer to OFDM performance.

Based on the system's SINR, we analysed the achievable rate. We also introduced a method to estimate the SC-FDM rate for $N_R > L$. Further, a possible calculation of LTE-A link adaptation parameters was proposed to achieve throughout close to performance bounds.

Lastly, we considered methods to gather CSI at the receiver. We compared the performance of various channel estimation and interpolation techniques. By incorporating the channel estimates of the previous subframe, we showed superior performance in terms of channel interpolation.

Endnotes

[1]Note that we use the symbol n as time index and the vector $\boldsymbol{n}$ for noise, the distinction should be clear from the context.

[2]Within this paper we focus on a single user's link performance. Multi user / multi basestation simulations are possible to perform, but come at very long simulation times. For sake of readability we use the non-standardized OFDM, SC-FDM notation in the remainder of this manuscript.

[3]A multi-tap equalizer applied on the intralayer interference visual in Fig. 3c could possibly enhance the link performance.

[4]The reduction to SISO is done to make our results comparable even to older frequency domain equalization (FDE) works, e.g., [63].

[5]Notice, however, that instantaneous rate adaptation for ZF OFDM can be advantageous in case of frequency-correlated channels [44].

[6]Notice that the simulation setup is the same as employed in [24] for the investigation of LTE downlink transmission, thus, facilitating the comparison of up- and downlink performance.

Appendix

General MIMO SC-FDMA SINR expression

The signal estimates are described via the input-output relationship in Eq. (1). We first slice out that part of $\boldsymbol{K}$

which acts on layer l by multiplying with the selector matrix $\boldsymbol{S}^{(l)}$ from the left. As indicated in (1), the signal estimate consists of three contributions.

$$\begin{aligned} \text{signal:} \quad & \hat{\boldsymbol{x}}_{\text{s}} = \boldsymbol{S}^{(l)} \left(\boldsymbol{I} \odot \boldsymbol{K}\right) \boldsymbol{x} \\ \text{interference:} \quad & \hat{\boldsymbol{x}}_{\text{i}} = \boldsymbol{S}^{(l)} \left(\boldsymbol{K} - \boldsymbol{I} \odot \boldsymbol{K}\right) \boldsymbol{x} \\ \text{noise:} \quad & \hat{\boldsymbol{x}}_{\text{n}} = \boldsymbol{S}^{(l)} \tilde{\boldsymbol{n}} \end{aligned}$$

As $\boldsymbol{x}$ and $\tilde{\boldsymbol{n}}$ are zero mean random quantities, their power is described by means of the second moment. To calculate the second moments, we take out the diagonal elements of the respective covariance matrices of each contribution.

$$\text{SINR}_l^{\text{SC-FDM}} = \left[\left(\boldsymbol{I} \odot \mathbb{E}\{\hat{\boldsymbol{x}}_{\text{s}} \hat{\boldsymbol{x}}_{\text{s}}^H\}\right) \left(\boldsymbol{I} \odot \mathbb{E}\{\hat{\boldsymbol{x}}_{\text{i}} \hat{\boldsymbol{x}}_{\text{i}}^H\} + \boldsymbol{I} \odot \mathbb{E}\{\hat{\boldsymbol{x}}_{\text{n}} \hat{\boldsymbol{x}}_{\text{n}}^H\}\right)^{-1} \right]_{1,1} \tag{44}$$

Before, we derive the different covariance matrices, we recapitulate a required property of circulant matrices. A circulant matrix $\boldsymbol{C} \in \mathbb{C}^{N \times N}$ is fully described by its first column $\boldsymbol{c}$, as its eigenvectors are the DFT basis-vectors and its eigenvalues are the DFT of $\boldsymbol{c} = (c_0, c_1, \ldots, c_{N-1})$.

$$\boldsymbol{C} = \begin{pmatrix} c_0 & c_{N-1} & \ldots & c_1 \\ c_1 & c_0 & & c_2 \\ \vdots & & \ddots & \vdots \\ c_{N-1} & \ldots & c_1 & c_0 \end{pmatrix} \tag{45}$$

$$= \boldsymbol{D}^H \text{Diag}\left(\boldsymbol{D}\boldsymbol{c}\right) \boldsymbol{D} = \boldsymbol{D}^H \boldsymbol{\Lambda} \boldsymbol{D} \tag{46}$$

The main diagonal elements c_0 of $\boldsymbol{C}$ are given by

$$c_0 = \frac{1}{N} \sum_{i=0}^{N-1} [\,\boldsymbol{D}\boldsymbol{c}]_i = \frac{1}{N} \sum_{i=0}^{N-1} [\,\boldsymbol{\Lambda}]_{i,i} = \frac{1}{N} \mathbb{1}^T \text{diag}\left(\boldsymbol{\Lambda}\right) . \tag{47}$$

$\mathbb{E}\{\hat{\boldsymbol{x}}_{\text{s}} \hat{\boldsymbol{x}}_{\text{s}}^H\}$:
The input-output matrix $\boldsymbol{K}$ is of block-circulant structure, as illustrated in Fig. 3c. The eigenvalues of the diagonal blocks are given by $\text{diag}\left(\boldsymbol{\Lambda}\right) = \boldsymbol{S}^{(l)} \text{diag}\left(\boldsymbol{F}\boldsymbol{H}_{\text{eff}}\right)$ and the diagonal elements of the lth diagonal block are then $\frac{1}{N} \mathbb{1}^T \boldsymbol{S}^{(l)} \text{diag}\left(\boldsymbol{F}\boldsymbol{H}_{\text{eff}}\right)$ as asserted by Eq. (47), thus

$$\boldsymbol{S}^{(l)} \left(\boldsymbol{I} \odot \boldsymbol{K}\right) = \frac{1}{N} \mathbb{1}^T \boldsymbol{S}^{(l)} \text{diag}\left(\boldsymbol{F}\boldsymbol{H}_{\text{eff}}\right) \boldsymbol{I} \tag{48}$$

Assuming zero mean, white data with variance, σ_x^2 the diagonal elements of $\mathbb{E}\{\hat{\boldsymbol{x}}_{\text{s}} \hat{\boldsymbol{x}}_{\text{s}}^H\}$ are given by $\sigma_x^2 |\frac{1}{N} \mathbb{1}^T \boldsymbol{S}^{(l)} \text{diag}\left(\boldsymbol{F}\boldsymbol{H}_{\text{eff}}\right)|^2$.

$\mathbb{E}\{\hat{\boldsymbol{x}}_{\text{i}} \hat{\boldsymbol{x}}_{\text{i}}^H\}$:
If $\boldsymbol{C}$ is circulant

$$\tilde{\boldsymbol{C}} = \boldsymbol{C} - c_0 \boldsymbol{I} = \begin{pmatrix} 0 & c_{N-1} & \ldots & c_1 \\ c_1 & 0 & & c_2 \\ \vdots & & \ddots & \vdots \\ c_{N-1} & \ldots & c_1 & 0 \end{pmatrix} \tag{49}$$

is circulant as well and the diagonal elements of $\tilde{\boldsymbol{C}} \tilde{\boldsymbol{C}}^H$ are the sum of the magnitude squares of $\tilde{\boldsymbol{c}} = (0, c_1, \ldots, c_{N-1})$. Using Parseval's theorem, we arrive at

$$\begin{aligned} \sum_{i=1}^{N-1} |c_i|^2 &= \tfrac{1}{N} \textstyle\sum_{j=1}^{N-1} |[\boldsymbol{\Lambda}]_{j,j}|^2 \\ &= \tfrac{1}{N} \textstyle\sum_{j=0}^{N-1} |[\boldsymbol{\Lambda}]_{j,j}|^2 - \left| \tfrac{1}{N} \textstyle\sum_{j=0}^{N-1} [\boldsymbol{\Lambda}]_{j,j} \right|^2 . \end{aligned} \tag{50}$$

The inter-layer interference consists of $L-1$ $\boldsymbol{C}$-type blocks, where we simply average the magnitude squares of the eigenvalues, i.e. the corresponding block-part of $\boldsymbol{F}\boldsymbol{H}_{\text{eff}}$. The intra-layer interference is described via a $\tilde{\boldsymbol{C}}$ block and is given in Eq. (50). Both contributions can be compactly written as

$$\sigma_x^2 \frac{1}{N} \|\boldsymbol{S}^{(l)} \boldsymbol{F}\boldsymbol{H}_{\text{eff}}\|_F^2 - \sigma_x^2 |\frac{1}{N} \mathbb{1}^T \boldsymbol{S}^{(l)} \text{diag}\left(\boldsymbol{F}\boldsymbol{H}_{\text{eff}}\right)|^2 . \tag{51}$$

$\mathbb{E}\{\hat{\boldsymbol{x}}_{\text{n}} \hat{\boldsymbol{x}}_{\text{n}}^H\}$:
The noise covariance matrix is circulant as well and the detailed derivations can be found in [30].

SISO MMSE SC-FDMA SINR expression

For a SISO system and a one-tap equalizer, the expression $\boldsymbol{F}\boldsymbol{H}_{\text{eff}}$ is of a diagonal shape. [30] has shown, that the MMSE equalizer for SC-FDM equals the OFDM expression, i.e. $\boldsymbol{F} = \left(\frac{\sigma_n^2}{\sigma_x^2} \boldsymbol{I} + \boldsymbol{H}_{\text{eff}}^H \boldsymbol{H}_{\text{eff}}\right)^{-1} \boldsymbol{H}_{\text{eff}}^H$. Thus, the elements on the main diagonal of $\boldsymbol{F}\boldsymbol{H}_{\text{eff}}$ are simply given by $|\boldsymbol{H}_k|^2 \left(\frac{\sigma_n^2}{\sigma_x^2} + |\boldsymbol{H}_k|^2\right)^{-1}$, and we rewrite (5) to (55), where we have used the identity

$$\frac{1}{N_{\text{SC}}} \sum_{k=1}^{N_{\text{SC}}} \frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2} + |\boldsymbol{H}_k|^2} = 1 - \frac{\sigma_n^2}{\sigma_x^2} \frac{1}{N_{\text{SC}}} \sum_{k=1}^{N_{\text{SC}}} \frac{1}{\frac{\sigma_n^2}{\sigma_x^2} + |\boldsymbol{H}_k|^2} \tag{52}$$

from [23].

$$\mathrm{SINR}_{\mathrm{MMSE}}^{\mathrm{SC\text{-}FDM}} = \frac{\frac{\sigma_x^2}{N_{\mathrm{SC}}}\left(\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}\right)^2}{\sigma_x^2\sum_{k=1}^{N_{\mathrm{SC}}}\left(\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}\right)^2-\frac{\sigma_x^2}{N_{\mathrm{SC}}}\left(\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}\right)^2+\sigma_n^2\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\left(\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2\right)^2}} \tag{53}$$

$$= \frac{\frac{1}{N_{\mathrm{SC}}}\left(\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}\right)^2}{\left(\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}\right)-\frac{1}{N_{\mathrm{SC}}}\left(\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}\right)^2}$$

$$= \frac{\frac{1}{N_{\mathrm{SC}}}\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}}{1-\frac{1}{N_{\mathrm{SC}}}\sum_{k=1}^{N_{\mathrm{SC}}}\frac{|\boldsymbol{H}_k|^2}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}} \tag{54}$$

$$= \frac{\sigma_x^2}{\sigma_n^2}\frac{1-\frac{\sigma_n^2}{\sigma_x^2}\frac{1}{N_{\mathrm{SC}}}\sum_{k=1}^{N_{\mathrm{SC}}}\frac{1}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}}{\frac{1}{N_{\mathrm{SC}}}\sum_{k=1}^{N_{\mathrm{SC}}}\frac{1}{\frac{\sigma_n^2}{\sigma_x^2}+|\boldsymbol{H}_k|^2}} \tag{55}$$

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This work has been funded by the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion and the Institute of Telecommunications, TUWien. The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged.

Received: 8 September 2015 Accepted: 6 April 2016
Published online: 27 April 2016

References

1. 3rd Generation Partnership Project (3GPP), Evolved Universal Terrestrial Radio Access (E-UTRA) physical channels and modulation. TS 36.211, 3rd Generation Partnership Project (3GPP) (2015). 3rd Generation Partnership Project (3GPP)
2. S Schwarz, JC Ikuno, M Simko, M Taranetz, Q Wang, M Rupp, Pushing the limits of LTE: a survey on research enhancing the standard. IEEE Access. **1**, 51–62 (2013)
3. Y Yan, H Yuan, N Zheng, S Peter, in *International Symposium on Wireless Communication Systems*. Performance of uplink multi-user MIMO in LTE-advanced networks, (2012), pp. 726–730
4. C-H Pan, Design of robust transceiver for precoded uplink MU-MIMO transmission in limited feedback system. Wireless Personal Commun. **77**(2), 857–879 (2014)
5. L Falconetti, S Landstrom, in *8th International Symposium on Wireless Communication Systems*. Uplink coordinated multi-point reception in LTE heterogeneous networks, (Aachen, Germany, 2011), pp. 764–768
6. Q Li, Y Wu, S Feng, P Zhang, Y Zhou, in *IEEE 77th Vehicular Technology Conference (VTC Spring)*. Cooperative uplink link adaptation in 3GPP LTE heterogeneous networks, (Dresden, Germany, 2013), pp. 1–5
7. Y Xue, Q Sun, B Jiang, X Gao, in *IEEE 79th Vehicular Technology Conference (VTC Spring)*. Link adaptation scheme for uplink MIMO transmission with turbo receivers, (Seoul, Korea, 2014), pp. 1–5
8. E Dahlman, G Mildh, S Parkvall, J Peisa, J Sachs, Y Selén, J Sköld, 5G wireless access: requirements and realization. Commun. Mag. IEEE. **52**(12), 42–47 (2014)
9. C Mehlführer, JC Ikuno, M Simko, S Schwarz, M Rupp, The Vienna LTE simulators—enabling reproducibility in wireless communications research. EURASIP Journal on Advances in Signal Processing (JASP) special issue on Reproducible Research (2011)
10. N Baldo, M Miozzo, M Requena-Esteso, J Nin-Guerrero, in *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. An open source product-oriented LTE network simulator based on ns-3 (ACM, Miami, USA, 2011), pp. 293–298
11. G Piro, LA Grieco, G Boggia, F Capozzi, P Camarda, Simulating LTE cellular systems: an open-source framework. IEEE Trans. Vehic. Technol. **60**(2), 498–513 (2011)
12. G Stea, A Virdis, G Nardini, in *SIMULTECH 2014*. SimuLTE: A modular system-level simulator for LTE/LTE-A networks based on OMNeT++ (insticc, Vienna, Austria, 2014), pp. 59–70
13. is-wireless, LTE PHY LAB (2015). http://is-wireless.com/4g-simulators/lte-phy-lab
14. H Zarrinkoub, *Understanding LTE with MATLAB: from Mathematical Modeling to Simulation and Prototyping*. (John Wiley & Sons, Hoboken, 2014)
15. G Gómez, D Morales-Jiménez, JJ Sánchez-Sánchez, JT Entrambasaguas, A next generation wireless simulator based on MIMI-OFDM: LTE Case Study. EURASIP J. Wireless Commun. Netw. **2010**, 15 (2010)
16. MV Lima, CM Gussen, BN Espíndola, TN Ferreira, WA Martins, PS Diniz, in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference On*. Open-source physical-layer simulator for LTE systems (IEEE, Kyoto, Japan, 2012), pp. 2781–2784
17. M Taranetz, T Blazek, T Kropfreiter, MK Müller, S Schwarz, M Rupp, Runtime precoding: enabling multipoint transmission in LTE-advanced system-level simulations. IEEE Access. **3**, 725–736 (2015)
18. ITC, Vienna LTE-A Uplink Simulator (2016). https://www.nt.tuwien.ac.at/research/mobile-communications/vienna-lte-a-simulators/lte-a-uplink-link-level-simulator/. [Online]
19. J Zhang, C Huang, G Liu, P Zhang, in *Communications and Networking in China, 2006. ChinaCom'06. First International Conference On*. Comparison of the link level performance between OFDMA and SC-FDMA (IEEE, Beijing, China, 2006), pp. 1–6
20. D Sinanović, G Šišul, B Modlic, in *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference On*. Comparison of BER characteristics of OFDMA and SC-FDMA in frequency selective channels (IEEE, Sarajevo, Bosnia and Herzegovina, 2011), pp. 1–4
21. M Suarez, O Zlydareva, in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2012 4th International Congress On*. LTE transceiver performance analysis in uplink under various environmental conditions (IEEE, St. Petersburg, Russia, 2012), pp. 84–88
22. JJ Sánchez-Sánchez, MC Aguayo-Torres, U Fernández-Plazaola, BER analysis for zero-forcing SC-FDMA over Nakagami-m fading channels. IEEE Trans. Vehic. Technol. **60**(8), 4077–4081 (2011)
23. JJ Sánchez-Sánchez, U Fernández-Plazaola, MC Aguayo-Torres, in *Broadband and Biomedical Communications (IB2Com), 2011 6th International Conference On*. BER analysis for SC-FDMA over Rayleigh fading channels (IEEE, Melbourne, Australia, 2011), pp. 43–47
24. S Schwarz, M Simko, M Rupp, in *Signal Processing Advances in Wireless Communications SPAWC, 2011*. On performance bounds for MIMO OFDM based wireless communication systems, (San Francisco, CA, 2011), pp. 311–315
25. S Pratschner, E Zöchmann, M Rupp, Low complexity estimation of frequency selective channels for the LTE-A uplink. submitted to Wireless Communication Letters (2015). doi:10.1109/LWC.2015.2481428
26. Vienna LTE-A Simulators – LTE-A Link Level Uplink Simulator Documentation, V1.4 (2015). https://www.nt.tuwien.ac.at/wp-content/uploads/2016/01/LTEAlinkDoc_v1_5.pdf
27. E Dahlman, S Parkvall, J Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*. (Elsevier Academic Press, Waltham, 2011)
28. VD Nguyen, H-P Kuchenbecker, in *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. Intercarrier and

intersymbol interference analysis of OFDM systems on time-invariant channels, vol. 4, (Lisboa, Portugal, 2002), pp. 1482–1487
29. A Wilzeck, Q Cai, M Schiewer, T Kaiser, in *Proceedings of the International ITG/IEEE Workshop on Smart Antennas*. Effect of multiple carrier frequency offsets in MIMO SC-FDMA systems, (Vienna, Austria, 2007)
30. E Zöchmann, S Pratschner, S Schwarz, M Rupp, in *19th International ITG Workshop on Smart Antennas (WSA)*. MIMO transmission over high delay spread channels with reduced cyclic prefix length, (Ilmenau, Germany, 2015)
31. HG Myung, J Lim, D Goodman, Single carrier FDMA for uplink wireless transmission. Vehic. Technol. Mag. IEEE. **1**(3), 30–38 (2006)
32. T Jiang, Y Wu, An overview: peak-to-average power ratio reduction techniques for OFDM signals. IEEE Trans. Broadcasting. **54**(2), 257 (2008)
33. PS Bullen, *Handbook of Means and Their Inequalities*. (Springer, Berlin, 2003)
34. Technical Specification Group Radio Access Network, *Deployment Aspects*. (3rd Generation Partnership Project (3GPP), 2014). TR 25.943 Version 12.0.0, http://www.3gpp.org/DynaReport/25943.htm
35. MD Nisar, H Nottensteiner, T Hindelang, in *Mobile and Wireless Communications Summit, 2007. 16th IST*. On performance limits of DFT spread OFDM systems (IEEE, Budapest, Hungary, 2007), pp. 1–4
36. M Geles, A Averbuch, O Amrani, D Ezri, Performance bounds for maximum likelihood detection of single carrier FDMA. IEEE Trans. Commun. **60**(7), 1945–1952 (2012)
37. S Schwarz, M Rupp. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (IEEE, San Francisco, CA, USA, 2011), pp. 316–320
38. D Maiwald, D Kraus, Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices. IEEE Proc. Radar, Sonar Navigation. **147**(4), 162–168 (2000)
39. AK Gupta, DK Nagar, *Matrix Variate Distributions*. Monographs and surveys in pure and applied mathematics, vol. 104. (Chapman & Hall/CRC, Oregon, 2000)
40. R Xu, Z Zhong, J-M Chen, B Ai, Bivariate Gamma distribution from complex inverse Wishart matrix. IEEE Commun. Lett. **13**(2), 118–120 (2009)
41. DA Gore, RW Heath, AJ Paulraj, Transmit selection in spatial multiplexing systems. IEEE Commun. Lett. **6**(11), 491–493 (2002)
42. A Hedayat, A Nosratinia, Outage and diversity of linear receivers in flat-fading MIMO channels. IEEE Trans. Signal Process. **55**(12), 5868–5873 (2007)
43. S Schwarz, C Mehlführer, M Rupp, in *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems, and Computers*. Low complexity approximate maximum throughput scheduling for LTE, (Pacific Grove, California, 2010), pp. 1563–1569
44. S Schwarz, C Mehlführer, M Rupp, in *Wireless Advanced (WiAD), 2010 6th Conference On*. Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE, (London, UK, 2010), pp. 1–6
45. X Hou, Z Zhang, H Kayama, in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*. DMRS design and channel estimation for LTE-advanced MIMO uplink, (Anchorage, USA, 2009), pp. 1–5
46. X Zhang, Y Li, in *International Conference on Connected Vehicles and Expo (ICCVE)*. Optimizing the MIMO channel estimation for LTE-advanced uplink, (Beijing, China, 2012), pp. 71–76
47. C-Y Chen, DW Lin, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Channel estimation for LTE and LTE-A MU-MIMO uplink with a narrow transmission band, (Florence, Italy, 2014), pp. 6484–6488
48. SM Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. (Prentice Hall, Upper Saddle River, 1993)
49. Q Zhang, X Zhu, T Yang, J Liu, An enhanced DFT-based channel estimator for LTE-A uplink. IEEE Trans. Vehicular Technol. **62**(9), 4690–4696 (2013)
50. B Sklar, Rayleigh fading channels in mobile digital communication systems. i. characterization. Commun. Mag. IEEE. **35**(7), 90–100 (1997)
51. P Robertson, S Kaiser, in *IEEE Vehicular Technology Conference, Fall*. The effects of Doppler spreads in OFDM (A) mobile radio systems, vol. 1 (IEEE, 1999), pp. 329–333
52. R Nissel, M Rupp, in *Proceedings of the IEEE International Conference on Communications (ICC'15)*. Doubly-selective MMSE channel estimation and ICI mitigation for OFDM systems, (London, UK, 2015)
53. M Lerch, in *Proceedings of Elmar 2015*. Experimental comparison of fast-fading channel interpolation methods for the LTE uplink, (Zadar, Croatia, 2015)
54. R Nissel, M Rupp, in *2015 IEEE International Conference on Communications (ICC)*. Doubly-selective MMSE channel estimation and ICI mitigation for OFDM systems (IEEE, 2015), pp. 4692–4697
55. G Berardinelli, CN Manchón, L Deneire, TB Sørensen, P Mogensen, K Pajukoski, in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*. Turbo receivers for single user MIMO LTE-A uplink (IEEE, 2009), pp. 1–5
56. Q Wang, C Mehlführer, M Rupp, in *2010 IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*. Carrier frequency synchronization in the downlink of 3GPP LTE (IEEE, 2010), pp. 939–944
57. M Sawahashi, Y Kishiyama, A Morimoto, D Nishikawa, M Tanno, Coordinated multipoint transmission/reception techniques for LTE-advanced [coordinated and distributed mimo]. Wireless Commun. IEEE. **17**(3), 26–34 (2010)
58. N Abu-Ali, A-EM Taha, M Salah, H Hassanein, Uplink scheduling in LTE and LTE-advanced: tutorial, survey and evaluation framework. Commun. Surveys Tutor. IEEE. **16**(3), 1239–1265 (2014)
59. F Rusek, D Persson, BK Lau, EG Larsson, TL Marzetta, O Edfors, F Tufvesson, Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)
60. L Lu, GY Li, AL Swindlehurst, A Ashikhmin, R Zhang, An overview of massive MIMO: benefits and challenges. IEEE J. Selected Topics Signal Process. **8**(5), 742–758 (2014)
61. E Larsson, O Edfors, F Tufvesson, T Marzetta, Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
62. 3GPP: TSG RAN; Study on 3D channel model for LTE (release 12) (2015). http://www.3gpp.org/DynaReport/36873.htm
63. T Shi, S Zhou, Y Yao, in *Emerging Technologies: Frontiers of Mobile and Wireless Communication, 2004. Proceedings of the IEEE 6th Circuits and Systems Symposium On*. Capacity of single carrier systems with frequency-domain equalization, vol. 2 (IEEE, 2004), pp. 429–432

3. Acquisition of Channel State Information at the Transmitter

Link adaptation is one of the most important PHY advances of wireless communications over the last decades, enabling to exploit the channel fading characteristics to enhance the spectral efficiency as well as the reliability of data transmissions. Link adaptation encompasses:

- Transmission rate adaptation via adaptive modulation and coding (AMC).
- Resource allocation and multi-user scheduling.
- Adaptive multicarrier waveform parametrization in flexible numerology use-cases.
- Beamforming and precoder optimization in multi-antenna systems.
- Multi-point transmission coordination and interference mitigation.

CSIT is instrumental to enable link adaptation. It provides the necessary information for the transmitter to select appropriate transmission parameters, in order to fulfill the QoS requirements of the users. There basically exist two approaches for obtaining CSIT:

- Direct estimation of the CSIT from the reverse link, exploiting the channel reciprocity in time division duplex (TDD) systems.
- Feedback of the CSI from the receiver over a dedicated limited capacity feedback link. This approach is mainly applied in frequency division duplex (FDD) systems.

Chapter Outline

In this chapter, I provide a discussion on CSIT estimation for TDD and FDD systems, focusing, however, mainly on the second group of limited feedback methods and my corresponding contributions in this area. In principle, acquisition of CSIT through reverse link pilot signals in TDD systems is more efficient than dedicated feedback in FDD systems. However, existing wireless networks are mostly based on FDD and converting them to TDD implies a significant financial burden for network operators. Even more, FDD systems are

easier to operate and more effective especially for delay-sensitive applications [64]. It is therefore expected that both, FDD and TDD, will find application in future mobile networks.

To simplify discussions and to avoid ambiguity, I denote in the following the transmitter as the base station (BS) and the receiver as the user equipment (UE); hence, the link between the transmitter and the receiver corresponds to the downlink of a mobile network and the reverse link corresponds to the uplink.[1]

3.1. Reverse Link CSIT Estimation in TDD Systems

Direct estimation of the CSIT by the BS is based on pilot signals in the uplink and relies on channel reciprocity between the up- and downlink. This approach has not received much attention in 4G and prior mobile communication systems, because these systems are mostly based on FDD and, hence, do not exhibit channel reciprocity. However, direct estimation of the CSIT by the BS gained significant attention in the context of TDD-based massive MIMO, because it enables the estimation of the large involved channel matrices with modest signaling overhead [65, 66]. It is therefore expected that TDD will be employed in 5G systems, at least in part of the associated spectrum.

Three main effects impact the quality of the CSIT achievable with this method:

1. Imperfect channel reciprocity due to non-symmetric characteristics of the transmit and receive front-ends [67–69].
2. Temporal channel variation during the up- and downlink duty cycle [70, 71].
3. Pilot-contamination in multi-cell systems [72, 73].

Channel reciprocity: The first issue of non-reciprocal front-ends arises from the fact that the transmit and receive chains of communication devices are not the same and, hence, the effective base band channels of the up- and downlink are also different. This issue can be mitigated by extra calibration efforts, as discussed e.g. in [67, 68].

Temporal channel variations: Temporal channel variations during the up- and downlink duty cycle, also known as channel aging, imply that the channel measured during the uplink transmission has changed until it comes to the downlink transmission. Hence, link adaptation based on uplink CSI estimates can be outdated by the time of downlink transmission, causing a mismatch between the selected transmission parameters and the actual channel state. This

[1] The same discussions also apply to the converse situation.

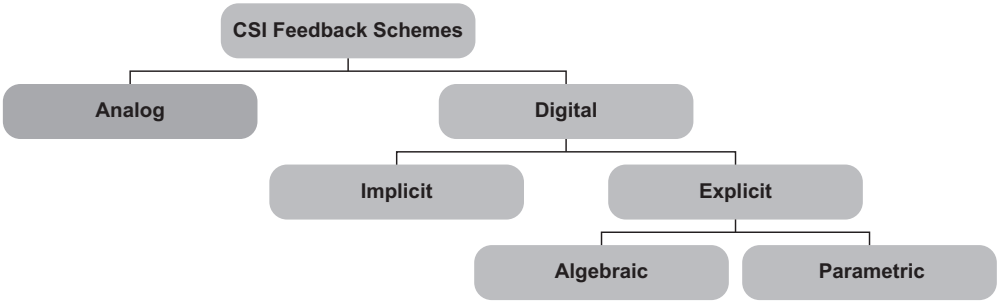


Figure 3.1.: Classification scheme for CSI feedback methods.

issue can partly be alleviated by predictive approaches [70]; yet, in practice, mostly more conservative link adaptation (e.g. in terms of transmission rate) is applied to compensate for the mismatch.

Pilot-contamination: The problem of pilot-contamination is closely related to temporal channel variations. These channel variations limit the coherence time of the channel and thereby also the total number of orthogonal pilot sequences that are available for channel estimation during the coherence time. This implies that the same pilot sequences have to be reused in multiple cells of the mobile communication network, causing interference in-between these cells and, as a consequence, the well-investigated pilot-contamination effect. A number of methods have been proposed to avoid significant pilot-contamination, such as, smart scheduling of pilots amongst cells [74] and pilot power allocation schemes [75]. These methods basically rely on significant path-loss differences between the signals received from multiple cells, to make sure that the intended signal experiences only negligible interference from other cells.

3.2. Classification of CSI Feedback Methods for FDD Systems

In FDD systems, the CSI of the downlink transmission can directly be estimated by means of pilot training signals only at the UE-side of the link. Acquisition of the CSIT at the BS, in contrast, requires dedicated feedback from the UEs, occupying resources of the uplink channel and thus reducing the uplink capacity available for data transmission. Correspondingly, efficient provisioning of the CSI feedback has received significant attention by the scientific community and many different methods have been proposed. In Figure 3.1, I present a classification scheme of CSI feedback methods, which builds the basis for further discussions below.

In general, we can distinguish between the large groups of analog and digital feedback methods. The initial step of either of these methods is to obtain a downlink channel estimate $\hat{\mathbf{H}}_{u,i}[n,k]$ at the UE-side through pilot signal based downlink training. Notice, I employ the notation of Equation (2.2) with the hat $\hat{}$ signifying that this is an estimate of the actual channel matrix $\mathbf{H}_{u,i}[n,k]$.

3.2.1. Analog CSI Feedback

In analog feedback methods, the individual elements of $\hat{\mathbf{H}}_{u,i}[n,k]$ are directly sent over the uplink channel from the UE to the BS by setting the uplink transmit signal equal to these elements [76, 77]. Hence, in general, several uplink REs are occupied to convey the CSI of a single downlink RE, in order to signal all elements of the matrix. To limit the feedback overhead, the CSI is therefore commonly sent only for a small subset of downlink REs and the remaining CSIT is obtained through interpolation at the BS. Provided the BS has an estimate of the uplink channel (through additional pilot signal based uplink training), it can obtain an estimate of the CSI sent by the UE through equalization of the channel distortions. The CSIT quality is thereby mainly impaired by five sources:

1. The channel estimation error at the UE.
2. The resolution of ADCs/DACs of the transceivers.
3. The linearity of the power amplifiers.
4. The distortions introduced by the uplink channel and the quality of the equalization.
5. The CSI estimation and interpolation error at the BS.

Even though it has been shown that analog feedback can theoretically achieve the optimal rate-distortion performance under certain idealistic conditions [76], practically the method has received little attention, due to the high linearity requirements imposed on the power amplifiers. Furthermore, the analog CSI feedback cannot be protected against transmission impairments and is therefore very sensitive to distortions caused by noise.

3.2.2. Digital CSI Feedback

Digital CSI feedback methods employ quantization at the UEs to obtain a digital representation of the relevant CSI. This digital CSI is then multiplexed with common data in the uplink channel, enjoying all the well-known advantages of digital over analog transmission. Additionally, digital CSI feedback enables an arbitrary rate-distortion trade-off by varying the number of bits of the CSI quantizer, and thus the feedback overhead can easily be

adjusted. For these reasons, only digital CSI feedback methods have been implemented so far in commercial systems and I restrict my further discussions below to digital methods.

Implicit Digital CSI Feedback

Digital CSI feedback methods can further be categorized as implicit or explicit schemes. In implicit CSI feedback methods, the estimated channel matrix $\hat{\mathbf{H}}_{u,i}[n,k]$ is utilized by the UE to derive optimal link adaptation parameters, such as, the supported transmission rate for AMC or the optimal precoder from a mutually agreed codebook, as I discuss in detail in [78]. This implicit information is then signaled to the BS. As my investigation in [8] demonstrates, implicit CSI feedback exhibits good performance in single-user MIMO transmission, where relatively small precoder codebook sizes are sufficient to achieve a performance close to the channel capacity. Implicit CSI feedback has been implemented in 4G LTE. I discuss such methods in more detail in Section 3.3.

Explicit Digital CSI Feedback:

In explicit digital CSI feedback methods, the channel matrix $\hat{\mathbf{H}}_{u,i}[n,k]$ is directly quantized and provided as feedback information. Such methods are necessary whenever the precoders at the BS are not selected from a codebook, but are rather directly calculated from the channel matrices (e.g., in most multi-user MIMO schemes, such as ZF beamforming [79]); such methods are denoted as non-codebook-based precoding methods. Efficient CSI quantization is crucial for such methods to achieve an adequate quality of the CSIT with acceptable feedback overhead. To realize this target, it is important to reduce the channel matrix to the essential information required by the respective transmission scheme, before applying quantization. Two philosophies exist in this regard within the scientific literature:

- **Algebraic methods:** These methods exploit algebraic properties of the essential CSIT to reduce the dimensionality of the CSI quantization problem. For example, the essential CSIT of many transmission schemes lies on a topological manifold, such as, the Grassmannian, the Stiefel manifold or the cone of positive semidefinite matrices of fixed rank; cf. *[Article 2]*.
- **Parametric methods:** These methods apply a parametric approach to achieve an efficient CSI representation. For example, a geometric parametrization, in terms of signal arrival and departure angles, can be utilized to achieve a sparse representation of the channel matrix in a suitably chosen basis; see *[Article 7]*.

Algebraic methods are mainly popular for relatively small-scale MIMO systems and for rich scattering environments (Rayleigh fading), where the channel matrices exhibit little structure.

Parametric methods, on the other hand, have gained significant interest in the context of massive and FD-MIMO systems, as such systems provide sufficient spatial resolution to differentiate individual scattering multi-path components. Additionally the propagation is often dictated by a few dominant specular scatterers, especially in the mmWave band as our measurement results in [80] reveal. I discuss both algebraic and parametric feedback methods in more detail in Section 3.4.

3.3. Implicit Digital CSI Feedback

Implicit CSI feedback represents the state-of-the-art in existing 4G wireless communication systems. It basically consists of two components:

- Channel quality feedback for transmission rate adaptation and multi-user scheduling.
- MIMO processing feedback for spatial preprocessing (beamforming/precoding) of transmit signals.

I follow below the LTE notation and denote the channel quality feedback information as channel quality indicator (CQI) and the MIMO processing feedback as precoding matrix indicator (PMI) and rank indicator (RI).

3.3.1. Channel Quality Feedback

The CQI provides time, frequency and spatially resolved information about the channel quality. It represents the achievable transmission rate of a UE at a given RE and spatial stream/layer as determined by the input-output relationship (2.4). This information is utilized by the BS to derive the multi-user MIMO-OFDMA resource allocation (see Chapter 4 for more details) and to set the transmission rate of the UEs. The CQI feedback is commonly calculated under the assumption that the BS follows the precoder recommendation of the PMI and RI feedback, which allows to accurately estimate the SINR of the precoded link.

Challenges in the CQI Feedback Calculation

Calculating the CQI feedback raises a number of challenges, especially when considering restrictions imposed by practical systems:

- Representing the realistic performance of systems: The information-theoretic mutual information is not sufficiently representative for this purpose, since existing wireless

communication systems do not achieve it [8]. It is necessary to account for the actual performance of the applied transceivers, as determined by the supported MCSs.

- Reducing feedback overhead: This requires quantization of the CQI, as well as, clustering of REs into larger blocks (clusters), for which a single CQI each is provided. The challenge is to optimize the CQI quantization regions and to accurately estimate the achievable rate within a cluster, which may exhibit substantial channel variations depending on the cluster size and the time-frequency selectivity of the channel.
- Temporal channel variations during the processing delay of the feedback loop: These variations cause outdated channel quality information, leading to signal outages.
- Channel quality estimation in case of non-codebook-based precoding: In this situation, the applied beamformer/precoder is not known at the time of feedback calculation and, hence, the SINR of the precoded link is not available. Therefore, SINR lower bounds are commonly applied to obtain robust CQI estimates.
- Compensating for a residual CQI feedback mismatch to avoid signal outages.

Realistic performance estimation and feedback overhead reduction: I address these two challenges in the publication *[Conf. 1]*; see Section 3.5. My investigations show that ESM techniques are applicable to accurately estimate the achievable rate of the supported MCSs within the feedback clusters.

Temporal channel variations: I discuss the issue of temporal channel variations during the feedback processing time in *[Article 2]* in more detail: up to a certain point, these channel variations can be compensated by predictive approaches. If the temporal channel variations become too strong, however, the prediction fails and more robust approaches are required, which are based on channel statistics rather than the instantaneous channel state.

Non-codebook-based precoding: Channel quality estimation for non-codebook-based precoding is dependent on the transmit strategy applied by the BS. In [81], CQI feedback for ZF beamforming based multi-user MIMO is derived. I generalize this result in *[Article 8]* to block diagonalization (BD) precoding, including feedback clustering. For other methods, similar approaches can be applied to derive SINR estimates/lower-bounds.

Mismatch compensation: When CQI feedback is employed for transmission rate adaptation, a residual mismatch of the CQI feedback is observable in the acknowledgment (ACK)/not ACK (NACK) feedback of the automatic repeat request (ARQ) protocol, which is present in most communication systems. In [82], the authors propose a CQI mismatch compensation algorithm, which utilizes this ACK/NACK feedback to improve the CQI accuracy and, thus, to reduce signal outages.

3.3.2. Implicit MIMO Processing Feedback

Codebook-Based Precoding

Implicit MIMO processing feedback builds on the underlying assumption that the possible precoders $\mathbf{F}_{u,i}[n,k]$, cf. Equation (2.3), are taken from a codebook $\mathcal{F}_{\ell}^{(N)}$ of supported precoding matrices. Here, the subscript ℓ refers to the number of spatial streams/layers to be transmitted in parallel and the superscript (N) indicates the number of transmit antennas. Multiple codebooks for different numbers of streams are commonly supported; often so-called nested codebooks are employed, where lower-rank precoding matrices are nested inside higher-rank precoders, which facilitates reducing the computational complexity of the feedback calculation [83]. The UEs and the BS both are informed of these codebooks. The UE signals the preferred number of streams ℓ^*, known as the transmission rank, with the RI, and the index of the chosen precoder from the corresponding codebook $\mathcal{F}_{\ell^*}^{(N)}$ with the PMI.

MIMO Processing Feedback Calculation

Existing wireless communication systems do not achieve the information-theoretic mutual information. As mentioned above, the practical performance of the supported MCSs can accurately be estimated by ESM techniques. This, though, causes a coupling between the selection of the optimal precoder (i.e. the RI and PMI) of a feedback cluster and the corresponding transmission rate (i.e. the CQI), as I show in my publication [84]. Therefore, a joint optimization of the PMI, RI and CQI is necessary, which, however, can cause an unacceptably high computational complexity, since the corresponding combinatorial optimization problem requires an exhaustive search. To reduce the complexity, it is possible to preselect the PMI and RI based on the information-theoretic mutual information, causing a slight performance degradation [84]. In case of LTE, which employs a BICM structure, a better performance is achieved by utilizing the BICM capacity [85] instead of the mutual information, since it provides a tighter upper bound on the actual system performance.

Determining the MIMO processing feedback in principle faces the same challenges as discussed above for the CQI feedback. However, selecting suboptimal precoders is not as

critical as choosing the wrong CQI, because it does not immediately cause signal outages; it rather implies a loss in spectral efficiency, since the channel state is not ideally exploited.

Optimal Precoder Codebook Design

Implicit MIMO feedback is mainly applied in the context of single-user MIMO transmissions. Efficient multi-user MIMO transmission schemes require a direct calculation of the precoders based on explicit CSIT. The issue of optimal codebook design for single-user MIMO transmission has received significant attention in the scientific literature. In [86, 87], the authors consider the optimal codebook design of semi-unitary precoding, restricting to uniform power-allocation over streams: $\mathbf{F}_{u,i}[n,k]^{\mathrm{H}}\mathbf{F}_{u,i}[n,k] = \frac{P_i}{\ell_u[n,k]}\mathbf{I}_{\ell_u[n,k]}$. Several results have been established for semi-unitary precoding:

- Infinite resolution semi-unitary precoding in combination with transmission rank adaptation achieves a performance close to the Shannon capacity. The feedback rate vs. capacity loss trade-off with finite resolution and randomly constructed unitary codebooks is investigated in [88] for isotropically distributed channel matrices (e.g. Rayleigh fading), showing that modest codebook sizes are sufficient to obtain practically perfect performance.
- Optimal semi-unitary codebook constructions for isotropically distributed channels correspond to maximally spaced Grassmannian subspace packings [89]. Such packings are difficult to obtain in general; yet, suboptimal codebooks with good performance can efficiently be constructed [90].
- Optimal codebook constructions for correlated channels are not known. The authors of [91] propose a heuristic systematic codebook design for correlated channels, by rotating and scaling spherical caps on the Grassmannian. In *[Article 8]*, I propose a heuristic random correlated codebook construction that is based on multiplying independent and identically distributed (iid) Gaussian matrices with a correlation matrix. Such constructions have been shown to provide a significant performance gain for transmission over correlated channels. A theoretical performance investigation of such schemes, though, is still an open research problem.

3.4. Explicit Digital CSI Feedback

The most straight-forward way to provide explicit CSI feedback is to apply scalar quantization to the individual elements of the channel matrix $\hat{\mathbf{H}}_{u,i}[n,k]$. Such an unstructured approach is adopted in IEEE 802.16d [92] by independently quantizing the amplitude and

phase of each matrix element. Scalar quantization exhibits a low complexity and is appropriate when channels are quasi-static, such that infrequent CSI feedback is sufficient and the overhead therefore is negligible [93]. In mobile communications, however, time-variant channels require more frequent CSI feedback and thus more efficient quantization.

As mentioned above, I discuss in this thesis two ways to reduce the feedback overhead: 1) Exploiting the algebraic structure of the underlying transmit precoding scheme to distill the essential CSIT; 2) Employing a parametric approach to achieve a sparse channel matrix decomposition in a suitably chosen basis.

3.4.1. Algebraic CSI Feedback

Algebraic CSI feedback methods utilize the fact that only a partial information about the channel matrix is required for many multi-antenna transmission schemes.

Motivating Example

Consider single-user MIMO transmission. It is well-known that the channel capacity can be achieved by matching the spatial signature of the transmit signal to the eigen-structure of the channel matrix. Specifically, the optimal precoder is equal to the matrix of right singular-vectors of the channel matrix times a power-allocation matrix that follows from the water-filling solution [94]:

$$\mathbf{F}_{u,i}[n,k] = \mathbf{V}_{u,i}[n,k]\mathbf{P}_{u,i}^{1/2}[n,k], \tag{3.1}$$

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

Here, (3.2) represents a singular value decomposition (SVD) of the channel matrix and $\mathbf{P}_{u,i}[n,k]$ is the diagonal power allocation matrix. Obviously, this precoding solution can be calculated from knowledge of the semi-unitary matrix $\mathbf{V}_{u,i}[n,k]$ and the singular values in $\mathbf{\Sigma}_{u,i}[n,k]$; both can be obtained from an eigen-decomposition of the **channel Gramian**:

$$\mathbf{H}_{u,i}^{\mathrm{H}}[n,k]\mathbf{H}_{u,i}[n,k] = \mathbf{V}_{u,i}[n,k]\mathbf{\Sigma}_{u,i}^{2}[n,k]\mathbf{V}_{u,i}^{\mathrm{H}}[n,k]. \tag{3.3}$$

If the transmission scheme is further restricted to semi-unitary precoding with equal-gain transmission, it suffices to feed back the first $\ell_u[n,k]$ columns of $\mathbf{V}_{u,i}[n,k]$, corresponding to the preferred transmission rank, to calculate the optimal semi-unitary precoder:

$$\mathbf{F}_{u,i}[n,k] = \frac{P_i}{\ell_u[n,k]} \left[\mathbf{V}_{u,i}[n,k]\right]_{1:\ell_u[n,k]}. \tag{3.4}$$

Notice, that the mutual information is invariant with respect to right-multiplication of the precoder with an arbitrary unitary matrix $\mathbf{Q}$:

$$\log_2 \det \left(\mathbf{I}_{M_u} + \frac{1}{\sigma_{z,u}^2} \mathbf{H}_{u,i}[n,k] \mathbf{F}_{u,i}[n,k] \left(\mathbf{H}_{u,i}[n,k] \mathbf{F}_{u,i}[n,k] \right)^{\mathrm{H}} \right) =$$
$$\log_2 \det \left(\mathbf{I}_{M_u} + \frac{1}{\sigma_{z,u}^2} \mathbf{H}_{u,i}[n,k] \mathbf{F}_{u,i}[n,k] \mathbf{Q} \left(\mathbf{H}_{u,i}[n,k] \mathbf{F}_{u,i}[n,k] \mathbf{Q} \right)^{\mathrm{H}} \right). \tag{3.5}$$

This establishes an equivalence relationship amongst precoders: $\mathbf{F}_{u,i}[n,k] \equiv \mathbf{F}_{u,i}[n,k]\mathbf{Q}$ for all unitary $\mathbf{Q}$. For semi-unitary precoding, this equivalence implies that any orthogonal basis that spans the same subspace as $\operatorname{span}\left(\left[\mathbf{V}_{u,i}[n,k] \right]_{1:\ell_u[n,k]} \right)$ is suitable for precoding.[2] Such an orthogonal basis can be calculated from any $\mathbf{W}_{u,i}[n,k] \in \mathbb{C}^{N_i \times \ell_u[n,k]}$ that satisfies:

$$\operatorname{span}\left(\mathbf{W}_{u,i}[n,k] \right) = \operatorname{span}\left(\left[\mathbf{V}_{u,i}[n,k] \right]_{1:\ell_u[n,k]} \right). \tag{3.6}$$

Correspondingly, subspace information is sufficient for semi-unitary precoding. Since subspaces can be represented as points on a **Grassmann manifold**, Grassmannian quantization has received significant attention by the scientific community.

Grassmannian CSI Feedback

Grassmannian CSI feedback is applicable to all transmit schemes that require information about the transmit subspace, i.e., the space spanned by (a subset of) the columns of $\mathbf{V}_{u,i}[n,k]$. As shown above, this includes semi-unitary precoding for single-user MIMO transmission, but also multi-user and multi-cell interference-cancellation schemes, such as, ZF beamforming [95], BD precoding [96] and interference alignment (IA) [97]. The common goal of such interference-cancellation schemes is to steer the transmit signal of a UE into the null-space of the channels of other UEs to avoid interference, requiring subspace information.

Performance: It has been shown for all of these interference-cancellation schemes that the residual multi-user interference with imperfect (quantized) CSIT is proportional to the Grassmannian chordal distance error [96–98]. Hence, minimizing the Grassmannian chordal distance is commonly applied as the quantization metric. To achieve the full multiplexing gain/degrees of freedom (DoF), the residual interference must scale inversely proportional to the SNR, which implies that the CSI feedback overhead in [bits] must grow logarithmically with the SNR. This is in contrast to single-user MIMO transmission, where the full multiplexing gain can be achieved even with a fixed codebook size.

[2] Notice, this does not hold for non-semi-unitary precoding, i.e. if non-equal power allocation is applied.

Quantization codebooks: Efficient memoryless Grassmannian quantization of isotropically distributed channel matrices requires quantization codebooks that resemble maximally spaced subspace packings on the Grassmannian; hence, the same codebook constructions as discussed in Section 3.3 for implicit codebook based semi-unitary precoding are applicable for explicit Grassmannian CSI feedback. Optimal codebook constructions for non-isotropically distributed channels are not known.

Receive antenna combining: If the number of receive antennas M_u is larger than the number of streams $\ell_u[n,k]$, the growth rate of the feedback overhead with the SNR, to achieve the full multiplexing gain, can be substantially reduced by smart receive antenna combining. In [99], the authors propose quantization based combining (QBC) for single-stream transmission $\ell_u[n,k]=1$, which achieves the minimal feedback overhead growth rate. I extend QBC in [100] to subspace quantization-based combining (SQBC), supporting multi-stream transmission. Both QBC and SQBC achieve the optimal multiplexing gain with minimal feedback overhead; however, at low SNR these receive antenna combiners exhibit a performance gap of several [dB] with respect to maximum eigen-mode combining (MEC). To alleviate this loss, I propose in [101] maximum expected achievable rate combining (MERC), which obtains the same performance as SQBC at high SNR but without the performance loss compared to MEC at low SNR.

Gramian CSI Feedback

Channel Gramian CSI feedback is applicable for all transmit schemes that require information about the transmit-side eigen-structure of the channel; that is, the matrix of right singular vectors and the corresponding singular values. Despite non-equal-gain single-user MIMO precoding, I show in [78,102] that this includes SINR based multi-user and multi-cell precoding schemes, such as, regularized block diagonalization (RBD) precoding [103] and max-SINR IA [104], as well as, methods that are interference-leakage based, such as, signal to leakage and noise ratio (SLNR) precoding [105] and the multicasting scheme presented in *[Article 11]*.

Performance: Currently, no theoretical performance results exist on Gramian based CSI feedback. In [78,102] and *[Article 6]*, I provide simulation based performance comparisons of Grassmannian and Gramian CSI feedback for BD and RBD. In Figure 3.2, I present results from *[Article 6]*, showing the relative throughput of BD and RBD precoding with imperfect CSIT, with respect to RBD precoding with perfect CSIT. I consider a CSI feedback overhead of 6 bits and 10 bits per transmission time interval (TTI). The results show that the Gramian scheme performs better at low SNR, since it enables a more favorable

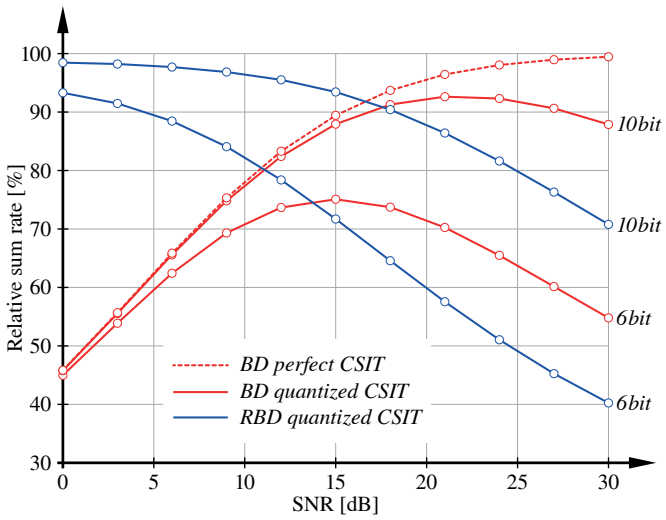


Figure 3.2.: Throughput of BD and RBD precoding with imperfect CSIT relative to RBD precoding with perfect CSIT. The results are taken from *[Article 6]* and are presented here in relative terms.

trade-off between the intended signal power and the residual interference. However, at high SNR the Grassmannian scheme performs better, because it achieves a lower quantization error for a given feedback overhead and thereby a lower residual multi-user interference.

Quantization codebooks: There exist two different possibilities for quantizing Gramian CSI: 1) direct quantization of the channel Gramian, which corresponds to a point on the positive semi-definite cone of fixed rank [106–108]; 2) separate quantization of the singular values and the matrix of right singular vectors, which corresponds to a point on the compact Stiefel manifold [78, 102]. The latter approach has two advantages: 1) it can employ exactly the same codebooks as Grassmannian quantization; only the quantization metric is different, see *[Article 6]*. This enables a seamless transition between Gramian and Grassmannian feedback depending on the SNR as I discuss in [102]. 2) Low rate quantization of the singular values is sufficient to still provide close-to-optimal performance, as I show in [78]. When applying separate quantization of the singular values and the singular vectors, this can be exploited by optimizing the feedback bit allocation between both.

Differential and Predictive CSI Feedback

Until now I have discussed memoryless quantization approaches. Such approaches do not account for the past when quantizing the current CSIT. Memoryless approaches exploit the channel correlation in time or frequency only in so far as the quantization interval can be adapted with respect to such correlation.

Predictive approaches, contrarily, exploit the channel correlation directly, by quantizing a residual with respect to a prediction, rather than the channel itself. Differential approaches are a special case of predictive approaches, where the previously quantized CSI simply acts as the prediction. Differential and predictive CSI feedback methods have been developed for several manifolds, including the Grassmannian, see [109] and my work [110, 111], the Stiefel manifold, see *[Article 6]*, and the positive semi-definite cone [107]. In *[Article 2]*, I discuss a general framework and requirements of predictive quantization on manifolds. Such approaches can provide significant performance gains over memoryless quantization, provided the channel correlation is sufficiently high. Depending on the dimensions of the involved manifold and the operating carrier frequency, gains are achieved in low to moderate mobility scenarios (up to urban vehicular speed) [110, 111], *[Article 6]*.

3.4.2. Parametric CSI Feedback

The basic idea of parametric CSI feedback approaches is to decompose the channel matrix in a possibly over-complete basis[3], commonly denoted as the dictionary $\mathcal{D}$, which facilitates a sparse approximation with a few dominant basis expansion terms:

$$\mathbf{H}_{u,i}[n,k] = \sum_{s=1}^{S} c_{u,i}^{(s)}[n,k]\mathbf{D}_{u,i}^{(s)}[n,k] + \mathbf{R}_{u,i}^{(S)}[n,k]. \tag{3.7}$$

Here, $\mathbf{D}_{u,i}^{(s)}[n,k] \in \mathcal{D}$ denotes the s-th selected entry/atom from the dictionary, $c_{u,i}^{(s)}[n,k] \in \mathbb{C}$ is the expansion coefficient and $\mathbf{R}_{u,i}^{(S)}[n,k]$ is the residual achieved with S expansion terms. Given a dictionary, such a decomposition can be calculated, e.g. by applying the orthogonal matching pursuit (OMP) algorithm [112]. Alternatively to decomposing the time-variant channel transfer function $\mathbf{H}_{u,i}[n,k]$, it is also possible to apply a similar approach to decompose the time-variant channel impulse response $\tilde{\mathbf{H}}_{u,i}(t,\tau)$, as discussed below.

Sparse channel decomposition: The goal is to obtain a good sparse approximation of $\mathbf{H}_{u,i}[n,k]$ with a very small number S of expansion terms by means of a dictionary that supports easy parametrization and therefore efficient CSI quantization/feedback. This can be achieved if the channel matrix is compressible in the chosen dictionary [113], which basically means that the magnitude of the expansion coefficients $c_{u,i}^{(s)}[n,k]$ decreases exponentially with the index s.

Since the decomposition at the UE is calculated from the estimated channel $\hat{\mathbf{H}}_{u,i}[n,k]$, it is furthermore desired that the basis provides robustness with respect to channel estimation

[3] A basis of a vector space is said to be over-complete if it is complete even after the removal of a vector from the basis.

errors to achieve a denoising gain. Moreover, the decomposition should be relatively stable over time and frequency, such that infrequent CSI feedback is sufficient.

Sparse channel decomposition is especially relevant for massive MIMO systems, to reduce the number of parameters required for channel representation [114].

Spatially Correlated Channels

In [115–117], the authors consider spatially correlated massive MIMO systems with correlation matrix:

$$\mathbf{C}_H = \mathbb{E}\left(\mathrm{vec}\left(\mathbf{H}_{u,i}[n,k]\right)\left(\mathrm{vec}\left(\mathbf{H}_{u,i}[n,k]\right)\right)^{\mathrm{H}}\right). \tag{3.8}$$

The authors propose to employ the unitary matrix $\mathbf{\Xi}_{\mathrm{KLT}}$ derived from the Karhunen Loeve transform (KLT) $\mathbf{C}_H = \mathbf{\Xi}_{\mathrm{KLT}}\mathbf{\Lambda}\mathbf{\Xi}_{\mathrm{KLT}}^{\mathrm{H}}$ as dictionary for decomposition of the vectorized channel $\mathrm{vec}\left(\mathbf{H}_{u,i}[n,k]\right)$. The contributions demonstrate that the KLT dictionary achieves a highly accurate CSI recovery with high compression ratios.

The KLT decomposition, however, requires knowledge of the spatial channel correlation matrix, which may not be available in practice. As an alternative, the authors also investigate application of the two dimensional (2D) discrete cosine transform (DCT), which is less efficient yet signal independent.

Limited Scattering Environments

Sparsity of MIMO channels naturally arises in situations where the signal propagation between the transmitter and the receiver is dominated by a relatively small number of scattering components. Consider the finite-scatterer doubly-directional time-variant impulse response of a MIMO system under the planar-wave narrowband array assumption [118, 119]:

$$\tilde{\mathbf{H}}_{u,i}(t,\tau) = \sum_{\ell=1}^{L_{u,i}} \alpha_{u,i,\ell}(t) f_{u,i}(\tau - \tau_{u,i,\ell}(t)) \mathbf{a}_u(\mathbf{\Psi}_{u,\ell}(t)) \mathbf{a}_i(\mathbf{\Psi}_{i,\ell}(t))^{\mathrm{T}}. \tag{3.9}$$

Here, $L_{u,i}$ denotes the number of multi-path components, and $\alpha_{u,i,\ell}(t) \in \mathbb{C}$ is the complex-valued amplitude of the ℓ-th path between transmitter i and receiver u. The corresponding propagation delay is $\tau_{u,i,\ell}(t)$. The scalar function $f_{u,i}(\tau)$ is the effective overall base band impulse response of the transmit and receive filters. The two vectors $\mathbf{a}_u(\mathbf{\Psi}) \in \mathbb{C}^{M_u \times 1}$, $\mathbf{a}_i(\mathbf{\Psi}) \in \mathbb{C}^{N_i \times 1}$ denote the receive and transmit antenna array response vectors with respect to a plane wave from direction $\mathbf{\Psi} = [\phi, \theta]^{\mathrm{T}}$ in azimuth ϕ and elevation θ.

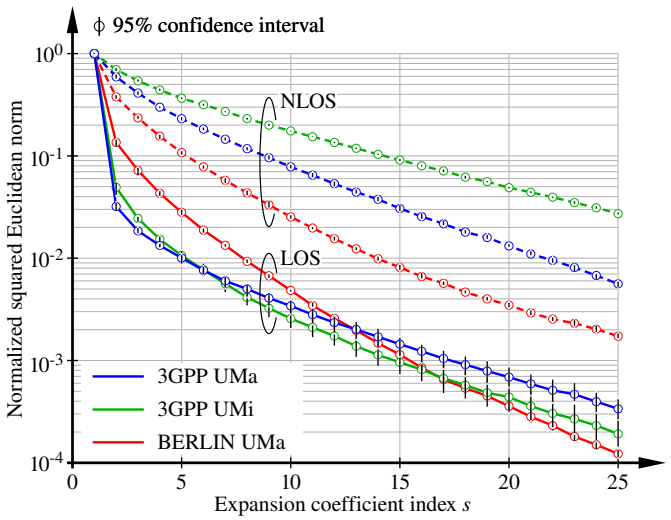


Figure 3.3.: Magnitude of the expansion coefficients as a function of the expansion coefficient index s, for different channel models and parameters. The results are taken from *[Article 7]*.

Angular channel decomposition: Provided $L_{u,i}$ is relatively small, a dictionary matched to the compound antenna array response matrix $\mathbf{A}_{u,i}(\mathbf{\Psi}_u, \mathbf{\Psi}_i) = \mathbf{a}_u(\mathbf{\Psi}_u)\mathbf{a}_i(\mathbf{\Psi}_i)^{\mathrm{T}}$ with varying $\mathbf{\Psi}_u$, $\mathbf{\Psi}_i$ is suitable for a sparse decomposition of Equation (3.9) [120]. To enable closely matching arbitrary signal angles $\mathbf{\Psi}_{u,\ell}(t)$, $\mathbf{\Psi}_{i,\ell}(t)$, the dictionary generally needs a high angular resolution and is therefore highly over-complete. This can be problematic in terms of the computational complexity of the applied compressed sensing (CS) algorithm for channel decomposition.

In [121] and *[Article 7]*, I show that such an angular decomposition can also be successfully applied to the OFDM channel matrix $\mathbf{H}_{u,i}[n,k]$, provided the propagation is dominated by a direct line of sight (LOS) path. In Figure 3.3, I show a result taken from *[Article 7]* to demonstrate this capability. The figure exhibits the magnitude of the expansion coefficients as a function of the expansion coefficient index s, for different channel parameters of the 3GPP 3D SCM [12] and the Quadriga channel model of the Heinrich Hertz Institute [122]. A good compression with a small number of expansion terms is achieved in the LOS situations, whereas in non line of sight (NLOS) a much larger number of expansion terms is required for the same approximation quality.

An angular channel decomposition is especially suitable in the mmWave band. This is because mmWave transmissions commonly rely on highly directive antennas to compensate for the increased path-loss. Such antennas act as spatial filters and effectively reduce the number of relevant multi-path components, as we demonstrate, e.g., in [80] through measurements.

3.5. Scientific Contributions and Publications

Calculation of CSI feedback was already the main topic of my PhD thesis [47]. My early focus was on provisioning of standard-compliant implicit CSI feedback, as summarized in *[Conf 1]*, and focused later on predictive Grassmannian CSI feedback, to support efficient multi-user MIMO transmission in mobile scenarios [110]. Following graduation, I generalized these predictive schemes to general topological manifolds; see *[Article 2]* and *[Article 6]*. Predictive quantization on manifolds is, however, computationally demanding; the methods are therefore not well suited for massive MIMO implementations, due to the high involved dimensionality. I therefore developed explicit CSI feedback algorithms that are suitable for massive MIMO in *[Article 7]*.

3.5.1. Selected Publications

The following publications within the broad field of CSI feedback contribute to this thesis:

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

[Article 6] S. Schwarz and M. Rupp, "Predictive quantization on the Stiefel manifold," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 234–238, 2015

[Article 7] S. Schwarz, "Robust full-dimension MIMO transmission based on limited feedback angular-domain CSIT," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 1–20, Mar 2018

3.5.2. Summary of Scientific Contributions

The main scientific contributions of these publications are:

1. Development of efficient implicit CSI feedback algorithms that account for the realistic performance of practical wireless communication systems in *[Conf. 1]*.
2. Derivation of predictive CSI quantization methods for the Stiefel manifold, which provide improved performance over Grassmannian quantization in the range of low to moderate SNR in *[Article 6]*.
3. Design of efficient and robust explicit CSI feedback schemes for FD-MIMO and massive MIMO systems, based on a sparse angular channel decomposition, in *[Article 7]*.

[Conf. 1] Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE: In this paper, I develop efficient feedback algorithms for state-of-the-art implicit CSI feedback. The methods are based on MIESM to accurately estimate the performance of the considered BICM transceiver architecture, as e.g. implemented in LTE. The algorithms employ instantaneous CSI at the receiver for feedback calculation and are therefore sensitive to a delay in the feedback path in time-variant situations. I thus consider several modifications in *[Article 2]* to achieve robustness with respect to channel variations. In [126], we additionally propose modifications that improve the robustness of the feedback methods for channels with large delay- and Doppler-spread, by accounting for the corresponding ICI and ISI in the feedback calculation. Recently, our focused shifted towards feedback overhead reduction by exploiting spatial channel correlation in densely populated areas. In [127, 128], we propose a Gaussian-process regression based spatial CQI interpolation method, which facilitates such a feedback overhead reduction.

[Article 6] Predictive quantization on the Stiefel manifold: In this article, I extend my predictive quantization algorithm of [110] from the Grassmannian to the Stiefel manifold. This method enables support of SINR and interference leakage based multi-user and multi-cell transmission schemes with limited feedback. I show that RBD precoding with Stiefel manifold feedback outperforms BD precoding with Grassmannian feedback at low to moderate SNR. Moreover, the method enables a seamless transition between Grassmannian and Stiefel manifold feedback as a function of the SNR, simply by replacing manifold-specific formulas of the algorithm (e.g. for the calculation of geodesics and chordal distances). Predictive quantization provides significant performance improvements at low to moderate mobility; at high mobility, exhibiting little temporal correlation, the performance falls back to memoryless quantization.

[Article 7] Robust full-dimension MIMO transmission based on limited feedback angular-domain CSIT: In this journal paper, I propose an explicit algebraic CSI feedback method for FD-MIMO systems, which is based on a sparse angular channel decomposition. I develop the feedback method in unison with an interference leakage-based multi-user precoding method, which is suitable for high mobility scenarios, as described in more detail in Chapter 4. This combination enables robust multi-user operation of FD-MIMO systems at high mobility. Moreover, the CSI feedback exhibits robustness with respect to imperfect channel estimation at the UEs and provides a significant denoising gain, even at high channel estimation errors. The leakage-based precoding method accounts for CSIT imperfections due to quantization errors and movement of UEs, and thereby enhances the robustness of the multi-user transmission.

Calculation of the Spatial Preprocessing and Link Adaption Feedback for 3GPP UMTS/LTE

Stefan Schwarz, Christian Mehlführer and Markus Rupp
Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: {sschwarz, chmehl, mrupp} @nt.tuwien.ac.at

***Abstract*—This paper presents an efficient method for calculating the Precoding Matrix Indicator (PMI), Rank Indicator (RI) and Channel Quality Indicator (CQI) at a Long Term Evolution (LTE) User Equipment (UE). The indicators are required for spatial preprocessing and link adaption in the downlink of a 3GPP UMTS/LTE system. To reduce the computational burden for the UE, our method decomposes the problem into two separate steps, one of jointly evaluating the PMI and RI based on a mutual information metric and one of choosing the CQI value to achieve a given target Block Error Ratio (BLER) constraint. The performance of the method is evaluated utilizing an LTE downlink physical layer simulator. The influence of estimated channel knowledge on the feedback choice is investigated for Least Squares (LS) and Linear Minimum Mean Squared Error (LMMSE) channel estimators.**

***Index Terms*—LTE, MIMO, Precoding, Link-Adaption**

I. Introduction

In 3GPP's future mobile communication system UMTS Long Term Evolution (LTE) [1] the feedback for channel adaption comprises three values (Channel Quality Indicator (CQI), Rank Indicator (RI), Precoding Matrix Indicator (PMI)) in the so-called closed-loop spatial multiplexing transmission mode [2]. By the CQI the transmitter selects one of 15 modulation alphabet and code rate combinations for transmission. The RI informs the transmitter about the number of useful spatial transmission layers for the current MIMO channel (which is not more than four in the current standard), and the PMI signals the codebook index of the prefered precoding matrix [3]. Finding a jointly optimal solution for these three values will in many cases not be feasible due to feedback delay constraints and limited signal processing hardware. It is therefore necessary to reduce computational complexity, which we achieve by separating the overall optimization process into several steps of finding local independent optima for the three values, thereby sacrificing overall optimality.

LTE is an Orthogonal Frequency Division Multiple Access (OFDMA) system with a system bandwidth of up to 20 MHz (1200 subcarriers). The time schedule for User Equipment (UE) feedback is given by the duration of one subframe (1 ms). The time-frequency grid spanned by the subcarriers and temporal samples within such a subframe is divided into several Resource Blocks (RBs) consisting of 12 subcarriers and 6 temporal samples each. Depending on the mode of operation the feedback granularity ranges from one PMI and CQI value per Resource Block (RB) up to just one value for the full system bandwidth [2]. The optimal feedback values will depend on the subcarrier and time instant, which necessitates some kind of majority decision.

In [4] we already introduced a PMI feedback scheme that is based on maximizing the sum mutual information over subcarriers. This method will be specialized and optimized here for linear receivers and will be used to jointly evaluate the optimal RI and PMI value in Section III.

The CQI value is chosen to achieve a given Block Error Ratio (BLER) target (BLER ≤ 0.1, a typical operating point for mobile communication systems). This choice is based on a mapping between post equalization Singal to Interference and Noise Ratio (SINR) and CQI for a Single Input Single Output (SISO) AWGN channel, which is evaluated by simulations. It is therefore necessary to map the SINR experienced on every fading subcarrier to an equivalent AWGN channel Signal to Noise Ratio (SNR), a method that is already well known from link level abstraction e.g. [5], [6], [7]. For this purpose we apply Mutual Information Effective SINR Mapping (MIESM) as well as Exponential Effective SINR Mapping (EESM) in Section IV. The system model is introduced in Section II and simulation results for the full feedback scheme and different antenna configurations are presented in Section V. In Section VI we analyze the algorithm and show where it gains in complexity compared to jointly optimizing RI, PMI and CQI. We suggest some additional modifications to further reduce the computational effort.

II. System Model

LTE converts a frequency selective channel into a number of narrowband frequency flat channels, by adopting OFDM. The input-output relation on subcarrier k, assuming M_R receive and N_T transmit antennas, at sampling time instant n is given by

$$\mathbf{y}_{k,n} = \mathbf{H}_{k,n}\mathbf{W}_i\mathbf{x}_{k,n} + \mathbf{n}_{k,n}, k \in 1,...K, n \in 1,...N. \quad (1)$$

$\mathbf{y}_{k,n} \in \mathbb{C}^{M_\mathrm{R}\times 1}$ is the received symbol vector, $\mathbf{H}_{k,n} \in \mathbb{C}^{M_\mathrm{R}\times N_\mathrm{T}}$ is the channel matrix experienced on subcarrier k at time instant n, $\mathbf{x}_{k,n} \in \mathcal{A}^{L\times 1}$ is the transmit symbol vector with $\mathcal{A}$ being the utilized symbol alphabet and $\mathbf{n}_{k,n} \sim \mathcal{CN}(0, \sigma_n^2 \cdot \mathbf{I})$ is white, complex-valued Gaussian noise with variance σ_n^2. The channel matrix and noise variance are assumed to be known by the receiver. The dimension of the transmit symbol vector

978-1-4244-7071-6/10/$26.00

depends on the number of useful spatial transmission layers L.

Spatial preprocessing is carried out with the precoding matrix $\mathbf{W}_i \in \mathcal{W}$. Here i denotes the index within the codebook of precoding matrices $\mathcal{W}$, defined in [3]. Depending on the feedback granularity, the precoder $\mathbf{W}_i$ will be either constant over only one RB or over the total system bandwidth and subframe duration.

The received symbol vector $\mathbf{y}_{k,n}$ is filtered by a linear equalizer given by a matrix $\mathbf{F}_{k,n} \in \mathbb{C}^{L \times M_\mathrm{R}}$. The output of this filter is the post-equalization symbol vector $\mathbf{r}_{k,n}$

$$\mathbf{r}_{k,n} = \mathbf{F}_{k,n}\mathbf{y}_{k,n} = \underbrace{\mathbf{F}_{k,n}\mathbf{H}_{k,n}\mathbf{W}_i}_{\mathbf{K}_{k,n} \in \mathbb{C}^{L \times L}} \mathbf{x}_{k,n} + \mathbf{F}_{k,n}\mathbf{n}_{k,n}. \quad (2)$$

The linear receiver is typically chosen according to a zero forcing or minimum mean square error design criterion [8]. The input signal vector is normalized to unit power.

III. PMI and RI Feedback

We have already presented the calculation of the PMI in [4], but we will specialize results here for linear receivers and also extend the idea to allow for the evaluation of the RI as well.

The basic idea is to choose the precoder that maximizes the mutual information for a specific subcarrier- $(1 \dots K)$ and temporal-range $(1 \dots N)$ of interest (which is at least a single RB and can be up to the full system bandwidth and subframe duration). Denoting $I_{k,n}$ the mutual information of the resource element (k, n) we obtain

$$\mathbf{W}_j = \arg \max_{\mathbf{W}_i \in \mathcal{W}} \sum_{k=1}^{K} \sum_{n=1}^{N} I_{k,n}(\mathbf{W}_i). \quad (3)$$

In [4] we have considered the pre-equalization mutual information for this choice, which achieves optimal results for maximum likelihood receivers, but not for linear ones in combination with MIMO systems (for interference free MISO systems linear receivers are optimal as well). Therefore we will now use the post-equalization mutual information which is given in terms of the post-equalization $\mathrm{SINR}_{k,n,l}$ as

$$I_{k,n} = \sum_{l=1}^{L} \log_2(1 + \mathrm{SINR}_{k,n,l}) \quad (4)$$

in bits per channel use, with L denoting the number of spatial transmission layers.

The post-equalization SINR on layer l equals

$$\mathrm{SINR}_{k,n,l} = \frac{|\mathbf{K}_{k,n}(l,l)|^2}{\sum_{i \neq l} |\mathbf{K}_{k,n}(l,i)|^2 + \sigma_n^2 \sum_i \mathbf{F}_{k,n}(l,i)}, \quad (5)$$

where $\mathbf{K}_{k,n}(l,i)$ refers to the element in the lth row and ith column of matrix $\mathbf{K}_{k,n}$ (see Equation (2)). The first term in the denominator corresponds to inter stream interference and the second term to noise enhancement. This expression assumes perfect channel knowledge and no inter cell interference. Of course such impairments can also be considered in the expression. For example, including a channel estimator in the system will just increase the effective noise variance by the mean square error of the channel estimator $\mathrm{MSE}_{k,n}$

$$\mathrm{SINR}_{k,n,l} = \frac{|\mathbf{K}_{k,n}(l,l)|^2}{\sum_{i \neq l} |\mathbf{K}_{k,n}(l,i)|^2 + \tilde{\sigma}_n^2 \sum_i \mathbf{F}_{k,n}(l,i)} \quad (6)$$

$$\tilde{\sigma}_n^2 = \sigma_n^2 + \mathrm{MSE}_{k,n}. \quad (7)$$

Simulation results in [4] have shown that it suffices to calculate the mutual information (4) just for a single channel matrix value per RB to come up with the optimal PMI. This approach is also adopted here. The single channel value can be obtained by averaging the channel over the corresponding Resource Elements (REs). This considerably reduces the computational complexity of the feedback calculation, but as Section V shows, it also entails a rate loss.

The feedback strategy for the PMI and RI values involves two steps (assuming different PMIs on every RB):

1) Calculate the post-equalization mutual information (4) for all possible precoders from $\mathcal{W}_L$ and spatial layer numbers $L = 1 \dots \min(M_R, N_T)$ for all resource blocks.
2) Find the combination of layer number and precoders that maximizes the sum mutual information over all resource blocks. The RI is given by this layer number and the PMIs by the codebook indices of the precoders. Of course a layer number L can only be combined with the corresponding precoders from $\mathcal{W}_L$.

IV. CQI Feedback

LTE uses the same modulation order and code rate (corresponding to a CQI value) for all resources allocated to a UE. Nevertheless, RB dependent CQI feedback is supported to give the scheduler the opportunity to schedule users on favourable resources. There are transmission modes defined in [3] that allow for independent codewords per spatial layer, but also for a single codeword for several layers. All these possibilities must be captured by a reasonable feedback strategy.

Our feedback strategy is based on averaging the post-equalization SINR over all resources of interest. This can include SINRs corresponding to single or multiple RBs per layer but also to RBs of different layers. Effective SINR Mapping (ESM) methods map several SINR values to an equivalent SNR value of a SISO AWGN channel (see [5], [6], [7] for details). This equivalent AWGN channel has similar BLER performance as the original OFDM system. In our work we have considered the EESM and MIESM methods. Mathematically the mapping is given by

$$\mathrm{SNR}_{\mathrm{eff}} = \beta f^{-1}\left(\frac{1}{R}\sum_{r=1}^{R} f\left(\frac{\mathrm{SINR}_r}{\beta}\right)\right), \quad (8)$$

where R corresponds to the number of resources of interest. For EESM the function f corresponds to an exponential, for MIESM it is given by the Bit Interleaved Coded Modulation (BICM) capacity [9]. Both methods require the calibration of the CQI dependent β value that adjusts the mapping to the different code rates and modulation alphabets. The goal of

the calibration is to obtain a close match between the BLE of the equivalent AWGN channel and the BLER of the rea fading channel. This calibration was carried out according t a relatively low complex procedure explained in [10]. The values obtained for EESM and MIESM can be found in ou LTE physical layer simulator that can be downloaded at [11

The mapping from SNR_{eff} to a corresponding CQI valu is carried out such that a BLER lower than 0.1 is achieved For this purpose SISO AWGN simulations have been carrie out for each CQI value that delivered this mapping, whic turned out to be a linear function (see [12] for details on th procedure).

The CQI feedback value is the highest possible value (whic delivers the highest throughput) with BLER $\leq$ 0.1 for th equivalent SISO AWGN channel.

The described method was also compared to another poss bility that jointly chooses RI and CQI to maximize the number of transmitted bits. Both methods have shown equivalent performance in all test cases investigated, but the described method has lower complexity as the ESM SINR averaging only has to be performed once.

V. Simulation Results

This section presents simulation results obtained with a standard compliant LTE physical layer simulator [13]. Simulations are carried out for a 2×1, 2×2 and 4×2 antenna system. A block fading channel model is assumed; that is, the channel is constant during one subframe duration and is fading independently between subframes. The feedback is sent to the transmitter with a delay of 0, meaning that the feedback values are calculated before the actual transmission. Antennas are assumed to be spatially uncorrelated. Simulations are carried out with a single UE occupying the full system bandwidth. As mentioned in Section III the channel is averaged over one RB before calculating the SINRs according to (5). The main simulation parameters are summarized in Table I.

TABLE I
Simulation parameters

Parameter	Value
System bandwidth	1.4 MHz
Number of subcarriers K	72
Feedback delay	0 TTI
Channel Model	ITU-T VehA [14]
Antenna configurations	2 transmit, 1 receive (2×1)
	2 transmit, 2 receive (2×2)
	4 transmit, 2 receive (4×2)
Receiver	Zero Forcing ZF
Feedback granularity	full bandwidth
Channel estimator	perfect channel knowledge

A. Antenna Configuration: 2×1

We first consider the 2×1 antenna configuration. As there is no source of interference in a single cell 2×1 system, the zero forcing receiver is equivalent to the Maximum Likelihood (ML) receiver. The feedback values, calculated according to Sections III and IV, are directly applied at the transmitter.

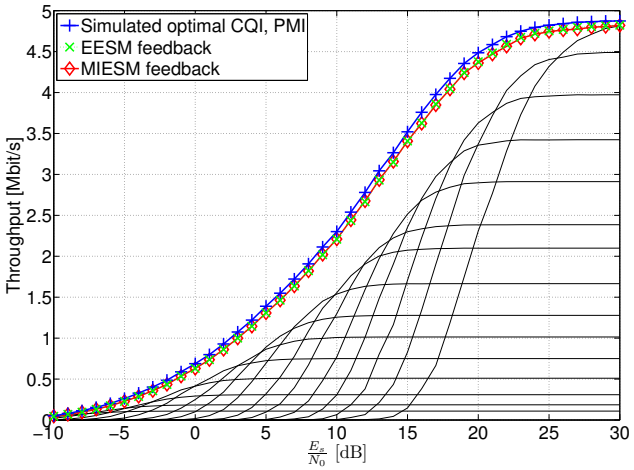


Fig. 1. Throughput over symbol energy to noise power spectral density for a 2×1 VehA channel.

Just a single PMI and CQI value is used for the full system bandwidth. The RI is equal to one.

Figure 1 shows simulated throughput versus transmit energy to noise power spectral density (SNR) obtained for this setup. For every SNR value 5000 channel realizations were simulated. The blue line with plus markers corresponds to an ideal choice of PMI and CQI that maximizes throughput. This choice is obtained by simulating every channel and noise realization with all possible combinations of PMI and CQI values and storing the result of the best combination. The red diamond marked and green cross marked lines correspond to the proposed feedback scheme, when applying MIESM or EESM for SINR averaging. There is virtually no difference between the two methods if they are well calibrated. The black lines show the throughput for fixed CQI values using just PMI feedback. The line with the smallest throughput at 30 dB corresponds to CQI 1 (4 QAM, code rate ~ 0.076), and the one with largest throughput to CQI 15 (64 QAM, code rate ~ 0.925). Adapting the CQI value to the current channel conditions brings a large gain of about $4 - 5$ dB. Our proposed feedback method looses about $0.5 - 1$ dB compared to the optimal choice.

Figure 2 shows the BLER obtained during the same simulation. As can be seen, the BLER target is achieved if the SNR is larger than approximately -5 dB. Below that value even a CQI value of 1 delivers a higher BLER. From 0 dB SNR upwards, the BLER fluctuates around 0.01. This is because in every subframe the CQI is adjusted to achieve BLER ≤ 0.1. Because the SNR range that is mapped to a certain CQI value has a width of approximately 2 dB (see [12]), in most cases the BLER is well below 0.1. The figure also shows the 95% confidence intervals for the MIESM simulation (similar for EESM). The slight tedency to increasing BLER with increasing SNR is caused by the calibration of MIESM or by the choice of the SNR-CQI mapping intervals.

In the next step a channel estimator is included in the

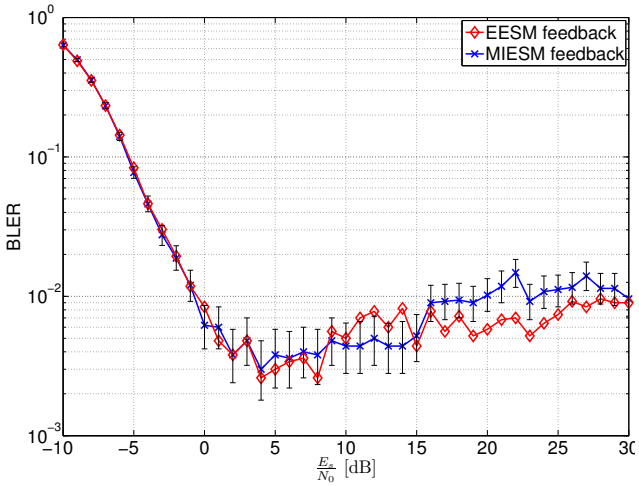


Fig. 2. Block error ratio over symbol energy to noise power spectral density for a 2×1 VehA channel.

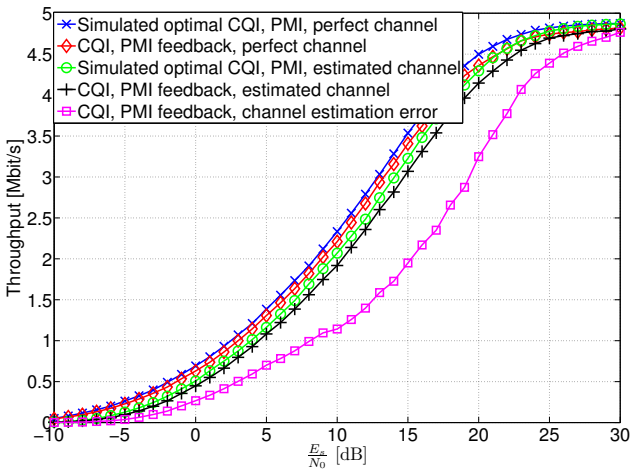


Fig. 3. Throughput over symbol energy to noise power spectral density for a 2×1 VehA channel with perfect and estimated channel knowledge.

system. For this purpose Least Squares (LS) [15] and Linear Minimum Mean Squared Error (LMMSE) [16] channel estimators are employed. Figure 3 compares the throughput curves obtained in this case with the ones with perfect channel knowledge for the LS channel estimator. The channel estimation error is incorporated into the post-equalization SINR expression as in (6). Due to the noise enhancement caused by the channel estimator there is a performance loss of about $1 - 1.5\,\text{dB}$ for the optimal choice of the feedback values as well as our proposed feedback strategy (compare the lines with cross and diamond markers and the ones with circle and plus markers). The magenta circle marked line shows the behaviour if the channel estimation MSE is not considered in the feedback calculation. The performance drops because the effective channel and therefore the CQI value is overestimated. Also the BLER can not be kept below 0.1 in this case.

When employing the LMMSE channel estimator, the

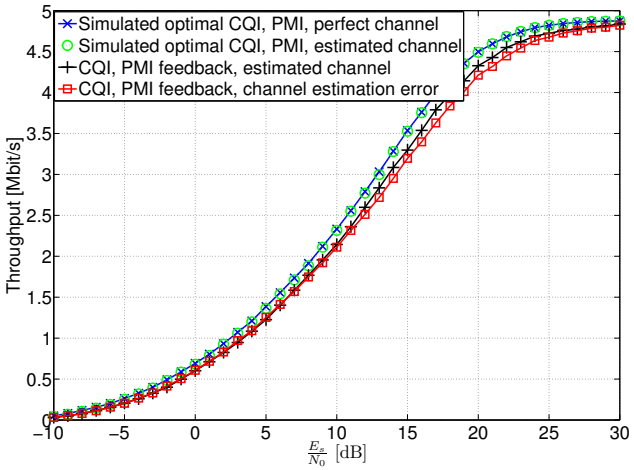


Fig. 4. Throughput over symbol energy to noise power spectral density for a 2×1 VehA channel with perfect and estimated channel knowledge.

throughput degradation is much less severe as Figure 4 shows. There is almost no difference between the optimal choice lines employing estimated or perfect channel knowledge. The feedback method looses about $0.5 - 1\,\text{dB}$ with knowlegdge of the channel estimation error and $0.5 - 2\,\text{dB}$ without.

B. Antenna Configuration: 2×2

Next a 2×2 system is investigated. In this case also RI feedback is supported and the number of spatial layers is adapted according to the feedack. Figure 5 shows the simulation results obtained for this case. The green line with plus markers uses all feedback capabilities (PMI, RI and CQI). For this result the channel is averaged over an RB before calculating SINR values. The red line with circle markers shows the performance of the proposed full feedback scheme if the channel is not averaged over an RB, but individual SINRs are calculated for every Resource Element (RE). At high SNR values $\sim 30\,\text{dB}$ this method delivers a performance gain of almost $2\,\text{dB}$ and should therefore be considered whenever complexity is not an issue (in the 2×1 case no gain was observed). The optimal performance, obtained by exhausitive search, is shown in blue with cross markers. Our feedback method looses about $0.5 - 1\,\text{dB}$ in SNR if channel averaging is not applied. The black line with diamond markers employs PMI and CQI feedback but fixes the number of spatial streams to $\text{RI} = 2$ while the magenta square marked line fixed it to $\text{RI} = 1$. A spatial stream number of two results in poor performance at low SNR. At $1\,\text{Mbit/s}$ throughput the dual stream mode looses about $4\,\text{dB}$ compared to the stream adaptive mode and also the single stream mode. In this regime the performance is therefore dominated by single stream operation. This shows that it is beneficial to exploit diversity and array gain instead of multiplexing gain at low SNR.

In Figure 6 the BLERs corresponding to the throughput curves in Figure 5 are depicted. If all feedback values are adapted or if RI is fixed to one the BLER meets the target

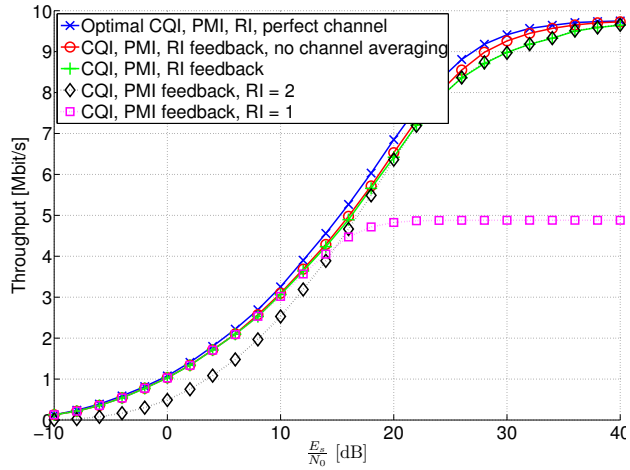


Fig. 5. Throughput over symbol energy to noise power spectral density for a 2×2 VehA channel.

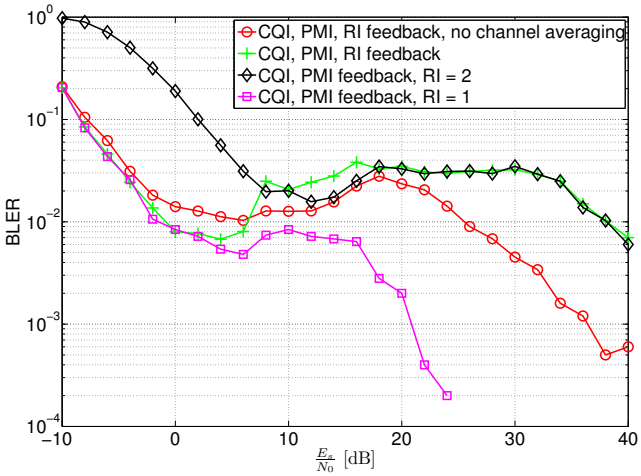


Fig. 6. BLER over symbol energy to noise power spectral density for a 2×2 VehA channel.

(BLER ≤ 0.1) already for SNR ≥ -8 dB, while it requires more than 2 dB of SNR if the stream number is fixed to two. As soon as dual stream transmission becomes more dominant (at approximately 7 dB SNR) the BLER increases, but still meets the target (green plus marked line). The figure shows that the BLER when employing no channel averaging (red circle marked line) starts decreasing at approximately 25 dB while this happens only at 35 dB (green plus marked line) when the channel is being averaged. It is also this regime in which the throughput degradation of channel averaging occurs.

C. Antenna Configuration: 4×2

In the previous configurations with two transmit antennas, the number of possible precoders according to the standard [3] is small. For two transmit antennas there are just four precoders for single layer transmission and two precoders for dual layer transmission. With four transmit antennas, the amount of precoders grows to sixteen for every layer number,

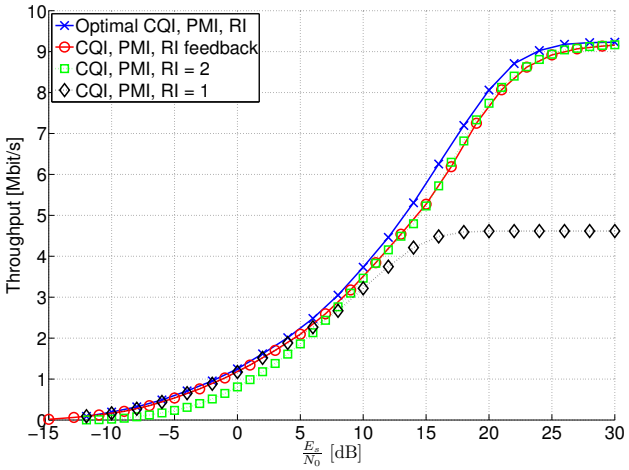


Fig. 7. Throughput over symbol energy to noise power spectral density for a 4×2 VehA channel.

which makes the choice more complex. Nonetheless, our feedback method works well as Figure 7 shows. Channel averaging is applied in the feedback calculation procedure. The loss in SNR compared to the optimal choice is similar to the 2×2 case and is approximately $0 - 1.5$ dB depending on the throughput. Again, at low SNR the transmission is dominated by single stream operation and at ~ 7 dB, dual stream operation outperforms single stream operation. The BLER performance is similar as in the 2×2 configuration. Comparing the throughput performance of the 2×2 and 4×2 configurations shows that the 4×2 system gains approximately 3 dB SNR at 3 Mbit/s throughput. In saturation at 30 dB SNR the throughput of the 4×2 system is less because there are more reference symbols due to the larger amount of transmit antennas.

VI. Computational Complexity Gains

The complete feedback algorithm consists of the following steps:

1) Computation of the post-equalization SINRs and mutual informations for all rank and precoding matrix combinations and all RBs using (4), (5) (up to 32 combinations for 4×2).
2) Choice of the rank and precoding matrix combination that maximizes the sum mutual information (3).
3) Calculation of the effective SNR using ESM (8) and mapping to a corresponding CQI value.

The first step has considerable computational complexity, as it requires computing the receive filter for all precoders and ranks. Complexity is reduced here, by not considering every subcarrier and temporal sample on its own, but just a single value per RB. This reduces the amount of computations by a factor of 72. For channels with low delay spread, this amount can be even further reduced without degrading the performance (cf. [4]). The complexity of the second step is negligible. The third step requires SNR averaging for all

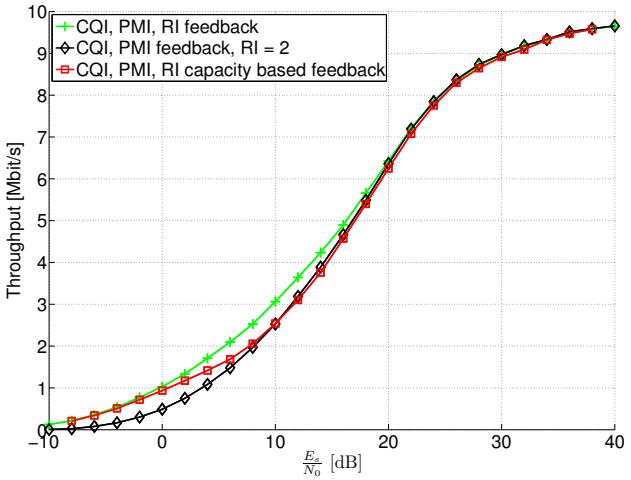


Fig. 8. Throughput over symbol energy to noise power spectral density for a 2×2 VehA channel.

possible CQI values (15 in our case), which is also complex. A feedback method that jointly optimizes RI, PMI and CQI would need to repeat this task for all rank and precoder combinations. Therefore, by choosing RI and PMI separately from CQI, we gain here up to a factor of 32 in complexity, depending on the antenna configuration.

A further reduction of computational complexity is possible by choosing PMI and RI from the theoretical capacity given by

$$I_{k,n} = \log_2 \det \left(\mathbf{I}_L + \frac{1}{\sigma_n^2} \mathbf{W}_i^H \mathbf{H}_{k,n}^H \mathbf{H}_{k,n} \mathbf{W}_i \right). \qquad (9)$$

In [4] we show that this entails an SNR loss of $0 - 1.5\,\text{dB}$ for the PMI choice depending on the antenna configuration. The receive filter then only needs to be calculated for the chosen combination of PMI and RI to find the appropriate CQI. Figure 8 compares the performance of this method with the previous one for a 2×2 channel. At low SNR ($\sim -5\,\text{dB}$) both methods transmit in single stream mode and choose the same precoder. But at around 5 dB the theoretical capacity is a too optimistic estimate for the performance of the zero forcing receiver. The feedback method switches to rank 2, which would deliver better performance with an ML receiver, but not with the linear receiver. Nevertheless the choice of the precoder is almost perfect ($\sim 0.1\,\text{dB}$ loss due to wrong precoder choice). In a 2×1 system both methods perform similar, because the linear receiver obtains ML performance.

VII. Conclusion

In this paper we present a suboptimal, reduced complexity PMI, RI and CQI feedback method for 3GPP UMTS/LTE. These feedback values are used for spatial preprocessing and link adaption at the transmitter (eNodeB). We show that our method performs close to optimal (in terms of throughput) for different antenna configurations and that the imposed BLER target is met. We also investigate the influence of channel estimation errors on our method and see that the performance is similar if the estimation error is included in the feedback calculation. Neglecting the channel estimation error deterioriates the performance of the method considerably.

Acknowledgement

This work has been funded by A1 Telekom Austria AG and the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology.

References

[1] 3GPP, "Technical Specification Group Radio Access Network; (E-UTRA) and (E-UTRAN); Overall description; Stage 2," September 2008. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36300.htm.

[2] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 8)," March 2009. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36213.htm.

[3] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)," September 2009. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36211.htm.

[4] S. Schwarz, M. Wrulich, and M. Rupp, "Mutual Information based Calculation of the Precoding Matrix Indicator for 3GPP UMTS/LTE," in Proc. IEEE Workshop on Smart Antennas 2010, (Bremen, Germany), February 2010.

[5] L. Wan, S. Tsai, and M. Almgren, "A Fading-Insensitve Performance Metric for a Unified Link Quality Model," in Proc. IEEE Wireless Communications & Networking Conference WCNC, 2006.

[6] X. He, K. Niu, Z.He, and J. Lin, "Link Layer Abstraction in MIMO-OFDM System," in Proc. International Workshop on Cross Layer Design, 2007.

[7] R. Sandanalakshmi, T. Palanivelu, and K. Manivannan, "Effective SNR Mapping for Link Error Prediction in OFDM based Systems," in Proc. IET-UK International Conference on Information and Communication Technology in Electrical Sciences ICTES, 2007.

[8] D. Tse and P. Viswanath, Fundamentals of Wireless Communications. Cambridge University Press, 2008.

[9] G. Caire, G. Taricco, and E. Biglieri, "Capacity of bit-interleaved channels," Electron. Lett., vol. 32, issue 12, pp. 1060–1061, June 1996.

[10] A. M. Cipriano, R. Visoz, and T. Sälzer, "Calibration Issues of PHY Layer Abstractions for Wireless Broadband Systems," in Proc. IEEE Vehicular Technology Conference Fall VTC2008-Fall, (Calgary, Canada), September 2008.

[11] [Online]. Available: http://www.nt.tuwien.ac.at/ltesimulator/.

[12] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in Proc. 71st Vehicular Technology Conference VTC2010-Spring, 2010.

[13] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the Long Term Evolution Physical Layer," in Proc. 17th European Signal Processing Conference EUSIPCO 2009, (Glasgow, Scotland), August 2009. [Online]. Available: http://publik.tuwien.ac.at/files/PubDat_175708.pdf.

[14] ITU, "Recommendation ITU-R M.1225: Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000," tech. rep., ITU, 1997.

[15] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in Proc. IEEE 45th Vehicular Technology Conference VTC1995, vol. 2, pp. 815–819, 1995.

[16] S. Omar, A. Ancora, and D. T. M. Slock, "Performance analysis of general pilot-aided linear channel estimation in LTE OFDMA systems with application to simplified MMSE schemes," in Proc. IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–6, September 2008.

Predictive Quantization on the Stiefel Manifold

Stefan Schwarz, *Member, IEEE*, and Markus Rupp, *Senior Member, IEEE*

***Abstract*—In this letter, we consider time-varying complex-valued $n \times m$ matrices $\mathbf{H}[k]$ ($m \leq n$) and propose a predictive quantizer for the eigenvectors of the Gramian $\mathbf{H}[k]\mathbf{H}[k]^{\mathrm{H}}$, which operates on the associated compact Stiefel manifold. The proposed quantizer exploits the temporal correlation of the source signal to provide high-fidelity representations with significantly reduced quantization codebook size compared to memoryless schemes. We apply the quantizer to channel state information quantization for limited feedback based multi-user MIMO, employing regularized block-diagonalization precoding. We demonstrate significant rate gains compared to block-diagonalization precoding using Grassmannian predictive feedback.**

***Index Terms*—Adaptive quantization, Grassmann manifold, limited feedback, multi-user MIMO, Stiefel manifold.**

I. Introduction

In multiple-input multiple-output (MIMO) wireless communications, the Grassmann manifold plays a central role in providing channel state information (CSI) to the transmitter over limited capacity feedback links [1]. The importance of the Grassmannian is justified by the fact that subspace information at the transmitter is sufficient for attaining the maximum degrees of freedom (DoF) of single-user MIMO systems with less receive antennas than transmit antennas [2], as well as that of multi-user MIMO broadcast channels [3]–[5] and MIMO interference channels [6]–[8]. DoF optimal strategies that are based on subspace information only, such as, block-diagonalization (BD) precoding [9] and interference-alignment (IA) [10], [11], however, often suffer a substantial signal to noise ratio (SNR) loss with respect to channel capacity, especially in case of spatially correlated channels [12]–[15]. In the low to intermediate SNR regime, significant throughput gains are possible using precoding methods that account for the SNR experienced on the individual channel eigenmodes, e.g., singular value decomposition (SVD) based precoding along with water-filling power allocation for single-user MIMO [16], regularized block-diagonalization (RBD) precoding for multi-user MIMO [17] and the max-SINR algorithm for IA [18]. To apply such methods, however, subspace information is not sufficient anymore; the transmitter rather requires knowledge of the directions and the magnitudes of the channel eigenmodes, i.e., the eigenvectors and eigenvalues of the channel Gramian.

Manuscript received March 13, 2014; revised June 16, 2014; accepted August 26, 2014. Date of publication September 04, 2014; date of current version September 17, 2014. This work was supported by the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility, the A1 Telekom Austria AG, and by the KATHREIN-Werke KG. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zoltan Safar.

The authors are with the Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria (e-mail: sschwarz@nt.tuwien.ac.at; mrupp@nt.tuwien.ac.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2354258

There are basically two possibilities to provide this information to the transmitter by means of limited feedback: either quantizing the channel Gram/covariance matrix and applying an eigendecomposition at the transmitter to determine the eigenmodes [19]–[21], or directly quantizing the eigenmodes [22]. While the work in [19] considers independent quantization of the individual entries of the Gramian, the authors of [20] exploit the geometry of the convex cone of positive definite Gram matrices to enable differential quantization, thereby improving the quantization accuracy for temporally correlated channels.

In this letter, we propose to independently quantize the magnitudes and the directions of the eigenmodes, providing the possibility to vary the feedback bit-allocation between these two. This is an important feature for multi-user MIMO and IA, because direction information commonly must be provided with very high accuracy to avoid multi-user interference, whereas magnitude information only impacts the power allocation. In an attempt similar to [20], we exploit the geometry of the quantization problem, in our case the Stiefel manifold [23], [24], to enable efficient predictive quantization of the eigenvectors. The algorithm reuses the basic principles of adaptive quantization introduced with our predictive Grassmannian quantizer [25], [26]. For magnitude-quantization, we employ a vector quantizer that is optimized using Lloyd's well known algorithm (k-means clustering) [27]. In contrast to [20], our quantization scheme is not restricted to positive definite matrices, but can also operate with positive semi-definite Gramians, thus encompassing CSI quantization for the multi-user MIMO broadcast channel with less receive antennas per user than transmit antennas.

II. Stiefel Manifold Geometry

The compact Stiefel manifold $\mathcal{S}t(n,m)$ over the complex numbers $\mathbb{C}$ represents the set of all ordered orthonormal Parseval $m-$ frames in the vector space $\mathbb{C}^n$ ($m \leq n$):

$$\mathcal{S}t\left(n,m\right) = \left\{\mathbf{S} \in \mathbb{C}^{n\times m} | \mathbf{S}^{\mathrm{H}}\mathbf{S} = \mathbf{I}_m\right\}. \tag{1}$$

Here, $(\cdot)^{\mathrm{H}}$ denotes conjugate-transposition. To each point $\mathbf{S} \in \mathcal{S}t(n,m)$, a linear vector space $\mathcal{T}_{\mathcal{S}t}(\mathbf{S})$ of tangents is associated [28], which is defined by:

$$\mathcal{T}_{\mathcal{S}t}(\mathbf{S}) = \left\{\mathbf{T} \in \mathbb{C}^{n\times m} | \mathbf{S}^{\mathrm{H}}\mathbf{T} + \mathbf{T}^{\mathrm{H}}\mathbf{S} = \mathbf{0}_m\right\}. \tag{2}$$

For every $\mathbf{T} \in \mathcal{T}_{\mathcal{S}t}(\mathbf{S}_1)$ there exists a geodesic $\gamma(t,\mathbf{T})$ satisfying:

$$\mathbf{S}_1 = \gamma(0,\mathbf{T}), \mathbf{S}_2 = \gamma(1,\mathbf{T}), \mathbf{T} = \frac{\partial}{\partial t}\gamma(t,\mathbf{T})|_{t=0}. \tag{3}$$

Varying t from 0 to 1, the geodesic traces the shortest path on the manifold between $\mathbf{S}_1, \mathbf{S}_2 \in \mathcal{S}t(n,m)$; it extends the concept of a straight line from the Euclidean space to curved spaces. Given $\mathbf{S}_1$ and the tangent $\mathbf{T}$, the exponential map returns the corresponding geodesic endpoint $\mathbf{S}_2 = \gamma(1,\mathbf{T})$ [23]. Conversely, given two points $\mathbf{S}_1, \mathbf{S}_2 \in \mathcal{S}t(n,m)$, the logarithmic map returns the associated tangent $\mathbf{T} \in \mathcal{T}_{\mathcal{S}t}(\mathbf{S}_1)$ specifying the geodesic. The Grassmannian quantizers in [26], [29] employ the

geodesic for prediction. This approach, however, does not appear feasible for the Stiefel manifold, because the calculation of the required logarithmic map is computationally hard [30].

To circumvent this complexity issue, alternative retraction and lifting pairs between the manifold and its tangent space have been proposed in the literature in the context of arithmetic averaging over the compact Stiefel manifold [31]. We denote the lifting of $\mathbf{S}_2 \in \mathcal{S}t(n,m)$ to the tangent space $\mathcal{T}_{\mathcal{S}t}(\mathbf{S}_1)$ as:

$$\mathbf{T} = L\left(\mathbf{S}_1, \mathbf{S}_2\right) \in \mathcal{T}_{\mathcal{S}t}(\mathbf{S}_1). \tag{4}$$

The retraction of $\mathbf{T} \in \mathcal{T}_{\mathcal{S}t}(\mathbf{S}_1)$ onto the manifold we write as:

$$\mathbf{S}_2 = R\left(\mathbf{S}_1, \mathbf{T}\right) \in \mathcal{S}t\left(n,m\right). \tag{5}$$

While the curve on the manifold defined by $\gamma(t,\mathbf{T}) = R(\mathbf{S}_1, t\mathbf{T}), t \in [0,1]$ still satisfies Eq. (3), it is in general not a valid geodesic, i.e., it does not trace the shortest path between $\mathbf{S}_1$ and $\mathbf{S}_2$. A compatible retraction/lifting pair fulfills the following condition in a certain neighborhood $\mathcal{M}(\mathbf{S}_1) \subseteq \mathcal{S}t(n,m)$ of $\mathbf{S}_1 \in \mathcal{S}t(n,m)$:

$$\mathbf{S}_2 = R\left(\mathbf{S}_1, L(\mathbf{S}_1, \mathbf{S}_2)\right), \mathbf{S}_2 \in \mathcal{M}\left(\mathbf{S}_1\right). \tag{6}$$

Instead of a compatible retraction/lifting pair, however, we apply the mixed polar-decomposition retraction and orthographic lifting pair proposed in [32] in our simulations to reduce computational complexity, although this pair causes minor discrepancy when evaluating Eq. (6); see [32] for details.

III. Predictive Quantization

We consider a stationary stochastic process of matrices $\{\mathbf{H}[k] \in \mathbb{C}^{n\times m}\}$, with k being the time index. We denote the auto-correlation function of the stochastic process as:

$$\mathbf{\Gamma}(\ell) = \mathbb{E}\left(\mathrm{vec}\left(\mathbf{H}[k]\right)\mathrm{vec}(\mathbf{H}[k-\ell])^{\mathrm{H}}\right). \tag{7}$$

In our simulations in Section V, we generate the wireless channel matrices under the assumption that the auto-correlation at lag $\ell = 0$ follows a Kronecker correlation model [33]:

$$\mathbf{H}[k] = \mathbf{C}_n^{1/2}\bar{\mathbf{H}}[k]\mathbf{C}_m^{1/2}, \mathbf{\Gamma}(0) = \mathbf{C}_m \otimes \mathbf{C}_n. \tag{8}$$

The auto-correlation between matrices with lag $\ell > 0$ is determined by Clarke's model [34]:

$$\mathbf{\Gamma}(\ell) = J_0\left(2\pi\nu_d\ell\right)\mathbf{\Gamma}(0), \nu_d = f_d T_s, \tag{9}$$

where T_s denotes the temporal sampling rate, f_d is the maximum channel Doppler frequency, ν_d denotes the normalized Doppler frequency and $J_0(\cdot)$ is the zeroth order Bessel function of the first kind. We employ the wireless channel model proposed in [35], which is based on Jakes' sum-of-sinusoids approach [36], to obtain realistic channel realizations.

From $\{\mathbf{H}[k]\}$ we derive the stochastic process $\{\mathbf{U}[k]\}$ on the Stiefel manifold by applying a compact-size SVD:

$$\mathbf{H}[k] = \mathbf{U}[k]\mathbf{\Sigma}[k]\mathbf{V}[k]^{\mathrm{H}}, \mathbf{U}[k] \in \mathcal{S}t\left(n,m\right), \tag{10}$$

with $\mathbf{U}[k]$ denoting the matrix of left singular vectors, which is equal to the eigenvector matrix of the Gramian $\mathbf{H}[k]\mathbf{H}[k]^{\mathrm{H}}$.

In order to quantize $\mathbf{U}[k]$, the quantizer employs the adaptive quantization codebook $\mathcal{Q}[k]$:

$$\mathcal{Q}[k] = \left\{\mathbf{Q}_i \in \mathcal{S}t\left(n,m\right) | i \in \{1,\ldots,2^B\}\right\}, \tag{11}$$

with B denoting the number of bits used for representation. The quantized point on the Stiefel manifold is obtained by minimizing the metric $\mathrm{d}_{\mathrm{s}}^2(\mathbf{U}[k], \mathbf{Q}_i)$ defined in (13):

$$\hat{\mathbf{U}}[k] = \underset{\mathbf{Q}_i \in \mathcal{Q}[k]}{\arg\min} \quad \mathrm{d}_{\mathrm{s}}^2\left(\mathbf{U}[k], \mathbf{Q}_i\right). \tag{12}$$

To conserve the individual directions of the eigenvectors, we consider the sum of one-dimensional Grassmannian subspace chordal distances [23] as quantization metric:

$$\begin{aligned} \mathrm{d}_{\mathrm{s}}^2\left(\mathbf{U}, \mathbf{Q}\right) &= \sum_{j=1}^{m} \mathrm{d}_{\mathrm{g}}^2\left(\mathbf{u}_j, \mathbf{q}_j\right) \in [0, m], \\ \mathrm{d}_{\mathrm{g}}^2\left(\mathbf{A}, \mathbf{B}\right) &= m - \mathrm{tr}\left(\mathbf{A}^{\mathrm{H}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{A}\right), \\ \mathbf{A}^{\mathrm{H}}\mathbf{A} &= \mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{I}_m, \end{aligned} \tag{13}$$

with $\mathbf{u}_j, \mathbf{q}_j$ being the j th columns of $\mathbf{U}$ and $\mathbf{Q}$, respectively.

The basic idea of the proposed predictive quantizer is, in line with [25], to predict the point $\mathbf{U}[k]$ based on N_p past observations $\mathcal{U}[k] = \{\hat{\mathbf{U}}[k-1], \ldots, \hat{\mathbf{U}}[k-N_p]\}$. Then, a local codebook is generated around the prediction $\mathbf{U}_p[k]$ to cover a certain volume on the Stiefel manifold, whose size is adaptively adjusted to match the variance of the prediction error.

To obtain the prediction $\mathbf{U}_p[k]$, we modify the proposal in [37] to be applicable to the Stiefel manifold. The first step is to determine the center of $\mathcal{U}[k]$ on the Stiefel manifold. This can be accomplished with the iterative arithmetic averaging algorithm of [31]. Specifically, given an estimate $\mathbf{U}_c[n]$ of the center, with n denoting the iteration index, the averaging algorithm determines the N_p tangents $\mathbf{T}_i[n] = L(\mathbf{U}_c[n], \hat{\mathbf{U}}[k-i]), i \in \{1,\ldots,N_p\}$ and estimates the new center as:

$$\mathbf{U}_c[n+1] = R\left(\mathbf{U}_c[n], \bar{\mathbf{T}}[n]\right), \bar{\mathbf{T}}[n] = \frac{1}{N_p}\sum_{i=1}^{N_p}\mathbf{T}_i[n]. \tag{14}$$

The algorithm stops at iteration N when the variation of the center is sufficiently small. As initial estimate of the center we employ $\mathbf{U}_c[0] = \hat{\mathbf{U}}[k - \lceil N_p/2 \rceil]$. Notice that a proof of convergence of this averaging method has not been carried out in [31]. Extensive simulations in [31], however, indicate steady convergence even for relatively large-size problems. Still, we employ a maximum number of 50 iterations in our simulations after which we latest stop the fixed-point iteration.

We then apply a linear regression to the tangents $\mathbf{T}_i[N]$:

$$\hat{\mathbf{T}}_i[N] = \mathbf{T}^{(0)} + (N_p - i)\mathbf{T}^{(1)}, i \in \{1,\ldots,N_p\}, \tag{15}$$

$$\left\{\mathbf{T}^{(0)}, \mathbf{T}^{(1)}\right\} = \arg\min \sum_{i=1}^{N_p}\left|\mathbf{T}_i[N] - \hat{\mathbf{T}}_i[N]\right|_F^2. \tag{16}$$

Based on this regression, we calculate the prediction as:

$$\mathbf{U}_p[k] = R\left(\mathbf{U}_c[N], \mathbf{T}^{(0)} + N_p\mathbf{T}^{(1)}\right). \tag{17}$$

The construction of the quantization codebook $\mathcal{Q}[k]$ around $\mathbf{U}_p[k]$ follows the same adaptive principle as introduced in [25]. At each time instant k, the quantization codebook is made up of two partial codebooks $\mathcal{Q}[k] = \{\mathcal{Q}_+[k], \mathcal{Q}_-[k]\}$, with $\mathcal{Q}_+[k]$ covering a larger volume around $\mathbf{U}_p[k]$ than $\mathcal{Q}_-[k]$:

$$\mathcal{Q}_+[k] = \left\{\mathrm{expm}\left(\mathbf{B}_i s_+\right)\mathbf{U}_p[k] | \forall \mathbf{B}_i \in \mathcal{B}\right\}, \tag{18}$$

$$\mathcal{Q}_-[k] = \{\mathrm{expm}\left(\mathbf{B}_i s_-\right)\mathbf{U}_p[k] | \forall \mathbf{B}_i \in \mathcal{B}\}, \quad (19)$$

$$\mathcal{B} = \{\mathbf{B}_i = 1/2\left(\mathbf{A}_i^{\mathrm{H}} - \mathbf{A}_i\right) | \forall \mathbf{A}_i \in \mathcal{A}\}, \quad (20)$$

$$\mathcal{A} = \left\{\mathbf{A}_i \in \mathbb{C}^{n\times n} | i \in \{1,\ldots,2^{B-1}\}\right\}, \quad (21)$$

where $\mathrm{expm}(\cdot)$ denotes the matrix exponential. The base codebook $\mathcal{A}$ ideally consists of isotropically distributed matrices in $\mathbb{C}^{n\times n}$; in our simulations, we employ randomly generated matrices with independent and identically distributed (iid) Gaussian elements of unit variance. The parameters s_+, s_- control the spread of the codebook around the center $\mathbf{U}_p[k]$[31] and are adapted over time to track the optimal size of the volume covered by the quantization codebook:

$$s_+ = g^{\min(s[k-1]+1,0)}, s_- = g^{s[k-1]-1}, \quad (22)$$

with $g > 1$ denoting the volume growth rate. The update of the codebook scale parameter $s[k]$ follows the rule:

$$s[k] = \begin{cases} \min(s[k-1]+1, 0) & \text{if } \hat{\mathbf{U}}[k] \in \mathcal{Q}_+[k] \\ s[k-1]-1 & \text{otherwise} \end{cases}, \quad (23)$$

with $s[0] = 0$; see [25] for more details. Notice that all steps of the codebook construction are based on quantized values only and can hence be reproduced at the decoder.

IV. Limited Feedback Based Multi-user MIMO

We consider a single transmitter that is equipped with n antennas and transmits data to $U = \frac{n}{m}$ users, each having m receive antennas. The input-output relationship of user u is:

$$\mathbf{y}_u = \mathbf{H}_u^{\mathrm{H}}\mathbf{F}_u\mathbf{x}_u + \mathbf{H}_u^{\mathrm{H}} \sum_{s=1, s\neq u}^{U} \mathbf{F}_s\mathbf{x}_s + \mathbf{n}_u, \quad (24)$$

with $\mathbf{x}_u \in \mathbb{C}^{m\times 1}$ being user u's intended signal, $\mathbf{F}_u \in \mathbb{C}^{n\times m}$ denoting the precoding matrix and $\mathbf{H}_u^{\mathrm{H}} \in \mathbb{C}^{m\times n}$ representing the MIMO channel matrix. The additive receiver noise $\mathbf{n}_u$ is iid complex Gaussian distributed: $\mathbf{n}_u \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2\mathbf{I}_m)$. The transmit signals are normalized as $\mathbb{E}(\mathbf{x}_u\mathbf{x}_u^{\mathrm{H}}) = \mathbf{I}_m$. We assume the total transmit power to be equal to one, and employ the definition $\mathrm{SNR} = \frac{1}{\sigma_n^2}$. The power allocation among users and spatial streams is determined by the precoders. We denote the compact SVD of the channel matrix as $\mathbf{H}_u = \mathbf{U}_u\mathbf{\Sigma}_u\mathbf{V}_u^{\mathrm{H}}$.

We compare two different DoF optimal transmit strategies, that is, BD and RBD precoding [9], [17]. BD precoding calculates the precoders such that the multi-user interference is eliminated, without taking the received power of the intended signal into account. For this the transmitter requires knowledge of the m-dimensional subspaces spanned by $\mathbf{H}_u, \forall u$, which can be signaled efficiently using Grassmannian feedback [4]. With quantized CSI at the transmitter (CSIT), the residual multi-user interference depends on the quantization error $\mathrm{d}_{\mathrm{g}}^2(\mathbf{U}_u, \hat{\mathbf{H}}_u)$[4], [5], where $\hat{\mathbf{H}}_u$ denotes the quantized orthonormal basis representing the m-dimensional channel subspace.

RBD precoding, on the other hand, optimally trades-off multi-user interference for received signal power in a mean-squared error (MSE) sense and applies water-filling power allocation to maximize the achievable transmission rate

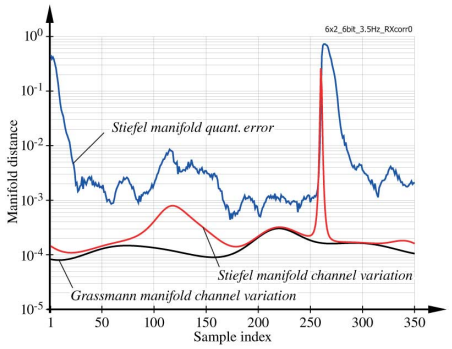


Fig. 1. Example of the temporal progress of the Grassmann and Stiefel manifold channel variation and the Stiefel manifold quantization error.

[17]. The calculation of the precoders in this case, however, requires knowledge of the directions and the magnitudes of the channel eigenmodes [38], i.e, the orthonormal matrix $\mathbf{U}_u$ and the m positive singular values $\boldsymbol{\sigma}_u = \mathrm{diag}(\mathbf{\Sigma}_u)$; here $\mathrm{diag}(\mathbf{A})$ stacks the diagonal elements of $\mathbf{A}$ in a column vector. We independently quantize this information employing the proposed Stiefel manifold quantizer and a Lloyd optimized vector quantizer for the vector $\boldsymbol{\sigma}_u \in \mathbb{R}_+^{m\times 1}$[27]. We evaluate the performance of the multi-user MIMO transmission in terms of the achievable sum rate:

$$\mathbb{R} = \sum_{u=1}^{U} \mathbb{E} log_2 \left| \mathbf{I}_m + \frac{1}{\sigma_n^2} \sum_{s=1}^{U} \mathbf{H}_u^{\mathrm{H}}\mathbf{F}_s\mathbf{F}_s^{\mathrm{H}}\mathbf{H}_u \right| \quad (25)$$

$$- \sum_{u=1}^{U} \mathbb{E} log_2 \left| \mathbf{I}_m + \frac{1}{\sigma_n^2} \sum_{\substack{s=1 \\ s\neq u}}^{U} \mathbf{H}_u^{\mathrm{H}}\mathbf{F}_s\mathbf{F}_s^{\mathrm{H}}\mathbf{H}_u \right|. \quad (26)$$

V. Simulations

In this section, we first evaluate the quantization MSE of the proposed predictive quantizer, and then apply the method for channel state information quantization to support limited feedback based multi-user MIMO. The Matlab code for reproduction of the presented results is available at [39].

We first investigate the convergence behavior of the algorithm for quantization on $\mathcal{St}(6,2)$ with a codebook of size $B = 6$ bits. The channel matrices are generated as described in Section III, assuming $\nu_d = 3.5\cdot 10^{-3}$ and $\mathbf{C}_n = \mathbf{I}_n, \mathbf{C}_m = \mathbf{I}_m$. Fig. 1 shows an example of the instantaneous quantization error of the predictive algorithm in comparison to the temporal channel variation on the Stiefel and Grassmann manifold, i.e., $\mathrm{d}_{\mathrm{s}}^2(\mathbf{U}[k-1], \mathbf{U}[k])$ and $\mathrm{d}_{\mathrm{g}}^2(\mathbf{U}[k-1], \mathbf{U}[k])$, respectively. Notice that by definition the channel variation on the Stiefel manifold is lower bounded by the variation on the Grassmannian; see Eq. (13). In the example, the algorithm converges within approximately 25 time instants. We observe that the variation of the eigenvectors over time, $\mathrm{d}_{\mathrm{s}}^2(\mathbf{U}[k-1], \mathbf{U}[k])$, is higher than the variation of the subspace spanned by these eigenvectors, $\mathrm{d}_{\mathrm{g}}^2(\mathbf{U}[k-1], \mathbf{U}[k])$. Specifically, at time instant 260, a significant change of the eigenvectors occurs although the subspace itself hardly varies. One may think that this is an artifact caused by a change of the order of the columns of $\mathbf{U}[k]$ as output by the SVD, e.g., the first column at time $[k-1]$ evolves to the second column at time $[k]$. This, however, is not the reason; we have corrected for changes in the SVD

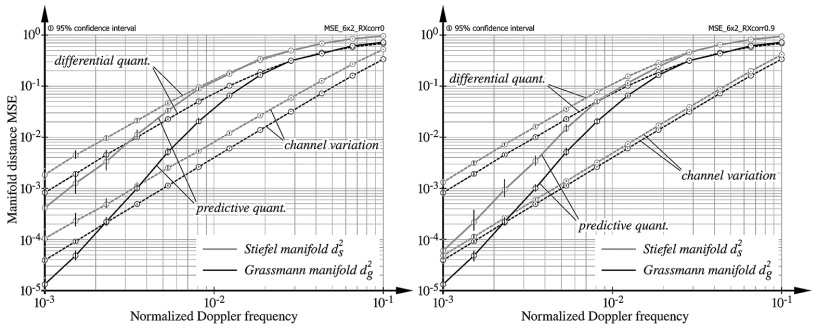


Fig. 2. MSE of the manifold distance achieved with differential and predictive quantization in dependence of the temporal channel variability and the spatial channel correlation (left: uncorrelated, right: strongly correlated). Also shown is the manifold temporal channel variation between consecutive time instants.

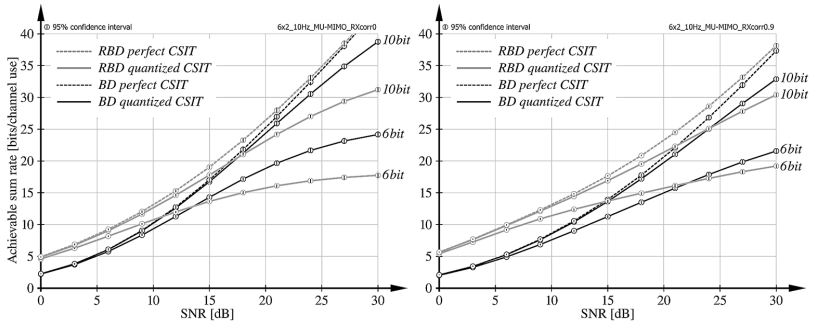


Fig. 3. Achievable sum rate obtained with BD and Grassmannian CSI quantization, as well as RBD and Stiefel manifold CSI quantization in dependence of the SNR. The left and right plots correspond to uncorrelated and strongly correlated receive antennas, respectively.

output-order. Such situations are detrimental to the performance, causing a fidelity degradation of predictive quantization on the Stiefel manifold compared to the Grassmannian.

This is visualized in Fig. 2, where we contrast the average quantization error of the proposed Stiefel manifold quantizer in terms of Eq. (13), with the Grassmannian error $\mathrm{d}_{\mathrm{g}}^2(\mathbf{U}_u, \hat{\mathbf{H}}_u)$ achieved by [26]. The figure also shows the performance of differential quantization, applying $\mathbf{U}_p[k] = \hat{\mathbf{U}}[k-1]$, as well as the variation of the channel itself. The left plot corresponds to the uncorrelated situation with $\mathbf{C}_m = \mathbf{I}_m$, whereas in the right plot we assume strongly correlated receive antennas, setting the off-diagonal elements of $\mathbf{C}_m$ equal to 0.9. We observe that the slope of the quantization error achieved with the Stiefel manifold quantizer is worse than the slope of the Grassmannian quantizer, due to such situations as the one observed in Fig. 1. With increasing receive-side correlation, however, the slopes become more and more similar, because the temporal variation of the individual directions of the eigenmodes within the subspace spanned by the channel matrix reduces. Also notice that Grassmannian quantization is not influenced by $\mathbf{C}_m$.

Finally, we evaluate the achievable transmission rate of BD and RBD precoding, employing limited feedback based on predictive Grassmannian and Stiefel manifold quantization, respectively. The corresponding results are shown in Fig. 3 for a normalized Doppler frequency of $\nu_d = 0.01$ and uncorrelated (left-hand side) as well as strongly correlated receive antennas (right-hand side), employing 6 bits and 10 bits of feedback per channel use. We assume perfect channel knowledge at the receiver; see [40] for an investigation on the impact of the channel estimation error for $\mathcal{S}t(n,1)$. For RBD precoding the singular values of the channel matrix must be quantized in addition to the singular vectors. At the considered Doppler frequency, however, this is achieved with a minor additional feedback overhead of 4 bits per 10 time instants without noticeable performance degradation. We observe in Fig. 3 that RBD precoding outperforms BD precoding at low to intermediate SNR, because the water-filling power allocation assigns power only to the strong eigenmodes of the users' channels. Conversely, at high SNR with quantized CSIT, BD achieves a higher rate than RBD, because the residual multi-user interference is smaller due to the improved performance of the Grassmannian quantizer compared to the Stiefel manifold quantizer. With increasing receive-side correlation, the situation changes in favor of RBD, due to the increased singular value disparity with growing correlation [41] and due to the improved performance of the Stiefel manifold quantizer.

VI. Conclusions

In this letter, we extend our previously published adaptive Grassmannian quantizer [25], [26] to the Stiefel manifold, utilizing recently proposed lifting and retraction pairs [31], [32] for the tangent space associated with points on the Stiefel manifold. We evaluate the performance of the proposed predictive quantizer and demonstrate substantial MSE gains when the temporal channel variation is sufficiently small. We apply the quantizer to channel state information quantization for limited feedback based multi-user MIMO transmission employing RBD precoding, and observe significant gains in achievable transmission rate at low to intermediate SNR compared to BD precoding using Grassmannian feedback.

Acknowledgment

The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged.

REFERENCES

[1] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, Oct. 2008.

[2] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, 2005.

[3] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, p. 5, Nov. 2006.

[4] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, Oct. 2008.

[5] S. Schwarz and M. Rupp, "Subspace quantization based combining for limited feedback block-diagonalization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5868–5879, 2013.

[6] M. Rezaee and M. Guillaud, "Limited feedback for interference alignment in the K-user MIMO interference channel," in *Proc. Inf. Theory Workshop*, Lausanne, Switzerland, Sep. 2012, pp. 1–5.

[7] O. El Ayach and R. Heath, "Grassmannian differential limited feedback for interference alignment," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6481–6494, Dec. 2012.

[8] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3908–3917, Aug. 2013.

[9] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[10] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[11] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[12] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[13] J. Lee and N. Jindal, "High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4787–4792, 2007.

[14] A. Tenenbaum and R. Adve, "Linear processing and sum throughput in the multiuser MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2652–2661, May 2009.

[15] S. Schwarz and M. Rupp, "Evaluation of distributed multi-user MIMO-OFDM with limited feedback," *IEEE Trans. Wireless Commun.*, Jun. 2014.

[16] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[17] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 953–961, 2008.

[18] K. Gomadam, V. Cadambe, and S. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *IEEE Global Telecommunications Conf.*, Nov. 2008, pp. 1–6.

[19] C.-B. Chae, D. Mazzarese, N. Jindal, and R. Heath, "Coordinated beamforming with limited feedback in the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1505–1515, Oct. 2008.

[20] D. Sacristan-Murga and A. Pascual-Iserte, "Differential feedback of MIMO channel Gram matrices based on geodesic curves," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3714–3727, Dec. 2010.

[21] R. Krishnamachari and M. Varanasi, "On the geometry and quantization of manifolds of positive semi-definite matrices," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4587–4599, Sep. 2013.

[22] Y. Li, S. Zhu, H. Tong, and M. Xu, "Enhanced limited rate implicit CSI feedback and its usage in covariance matrix based MU-MIMO," in *Wireless Commun. and Networking Conf.*, Apr. 2013, pp. 3067–3071.

[23] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Applicat.*, vol. 20, no. 2, pp. 303–353, 1998.

[24] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

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

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

[27] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[28] J. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.

[29] T. Inoue and R. Heath, "Grassmannian predictive coding for limited feedback multiuser MIMO systems," in *IEEE Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP)*, May 2011, pp. 3076–3079.

[30] Q. Rentmeesters, "Computing the Riemannian log map on the Stiefel manifold," in *20th Int. Symp. Mathematical Theory of Networks and Systems*, Melbourne, Australia, Jul. 2012.

[31] T. Kaneko, S. Fiori, and T. Tanaka, "Empirical arithmetic averaging over the compact stiefel manifold," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 883–894, Feb. 2013.

[32] S. Fiori, T. Kaneko, and T. Tanaka, "Mixed maps for learning a Kolmogoroff-Nagumo-type average element on the compact Stiefel manifold," in *2014 IEEE IntInt. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4518–4522.

[33] C. Oestges, "Validity of the Kronecker model for MIMO correlated channels," in *IEEE 63rd Vehicular Technology Conf., VTC 2006-Spring*, May 2006, vol. 6, pp. 2818–2822.

[34] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 957–1000, 1968.

[35] T. Zemen and C. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.

[36] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Piscataway, NJ, USA: Wiley-IEEE Press, 1994.

[37] Y. Zhang and M. Lei, "Robust Grassmannian prediction for limited feedback multiuser MIMO systems," in *IEEE Wireless Communications and Networking Conf. (WCNC)*, Apr. 2012, pp. 863–867.

[38] S. Schwarz and M. Rupp, "Subspace versus eigenmode quantization for limited feedback block-diagonalization," in *6th Int. Symp. Communications, Control and Signal Processing*, May 2014.

[39] "Predictive Quantization on the Stiefel Manifold," [Online]. Available: http://www.nt.tuwien.ac.at/about-us/staff/stefan-schwarz/reproducible-research/

[40] S. Schwarz and M. Rupp, "Adaptive channel direction quantization based on spherical prediction," in *IEEE Int. Conf. Communications*, Ottawa, ON, Canada, Jun. 2012.

[41] A. Zanella, M. Chiani, and M. Win, "On the marginal distribution of the eigenvalues of Wishart matrices," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1050–1060, Apr. 2009.

Schwarz *EURASIP Journal on Wireless Communications and Networking* (2018) 2018:58
https://doi.org/10.1186/s13638-018-1070-0

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Robust full-dimension MIMO transmission based on limited feedback angular-domain CSIT

Stefan Schwarz

Abstract

In this paper, we propose robust multi-user beamforming and precoding techniques for full-dimension MIMO transmission based on limited feedback. We propose to employ an over-complete basis decomposition in the angular domain to approximate the channel matrices as sums of few dominant specular components, facilitating efficient channel state information (CSI) quantization at the users. The selected expansion vectors of such a sparse approximation, parametrized by azimuth and elevation angles, are relatively robust with respect to channel estimation errors as well as channel variations over time. Based on this CSI feedback, we propose incoherent beamforming/precoding methods that make use only of the azimuth and elevation angles as well as the norm of the expansion coefficients and do not rely on coherent multipath interference to eliminate inter-user interference. Our optimization aims at maximizing the signal power or the achievable rate of a user, while limiting the amount of interference leakage caused to other users. To further improve robustness, we account for uncertainty in the angular channel decomposition in the proposed precoder optimization.

Keywords: Full-dimension MIMO, 3D beamforming, Robust precoding, Interference leakage, High mobility, Limited feedback

1 Introduction

Full-dimension multiple-input multiple-output (FD-MIMO) transmission refers to wireless transmission systems that support two-dimensional antenna arrays with a large number of antenna elements. This enables high-resolution beamforming in both, the elevation and the azimuth domain, in order to achieve space-division multiple access gains through spatial separation of users, as well as to enhance the energy efficiency of wireless data transmission by concentrating the radiated energy towards intended users [1–3]. The great potential of FD-MIMO systems has been recognized by standardization bodies, such as the third generation partnership project (3GPP); correspondingly, development and standardization efforts are ongoing within long-term evolution (LTE) Release 14 and 5G new radio (NR) to further enhance the initial realization of FD-MIMO in Release 13. Currently, most implementation proposals of FD-MIMO transceivers are based on hybrid architectures, where part of the signal processing is performed in base band and part in the analog radio frequency (RF) domain in order to limit the number of required RF chains. This approach is especially important in the millimeter wave band, where hardware complexity is a limiting factor [4–8]. When designing hybrid precoding systems, several challenges need to be addressed [9–11]: the design freedom of RF precoders is commonly restricted by hardware limitations, such as discrete angular resolution and constant modulus of the phase-shifting elements, as well as partial connectivity of the phase-shift network to limit insertion loss. An alternative to hybrid architectures, which potentially provides even higher spectral and energy efficiency in certain operating regimes, is to employ cheap low (even single bit) resolution analog to digital (AD)/digital to analog (DA) converters in FD-MIMO transceivers.

Correspondence: stefan.schwarz@nt.tuwien.ac.at
Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion, Institute of Telecommunications, Technische Universität (TU) Wien, Vienna, Austria

At the receive-side of most existing MIMO transceivers, high-resolution AD converters are employed to minimize the signal distortions introduced by the receiver. Such high-resolution AD converters are, however, among the most power hungry devices of the receiver chain. This motivates the use of low-resolution energy efficient AD converters in FD-MIMO systems to achieve energy-efficient operation [12–15]. At the transmit side, on the other hand, power expenditure is dominated by power amplifiers, which are usually required to operate within the high linearity regime to avoid signal distortion. Employing low-resolution DA converters relaxes these linearity requirements, allowing the amplifiers to operate closer to saturation thereby increasing their efficiency [16, 17].

In this work, our focus is on an enhancing the robustness of downlink three-dimensional FD-MIMO point to multi-point transmission with respect to imperfections of the channel state information at the transmitter (CSIT), especially due to high user mobility, without specifically addressing the restrictions of hybrid precoding or low-resolution AD/DA implementations. In MIMO transmission in general, and especially when considering multi-point communication, CSIT is an essential ingredient. CSIT for downlink transmission is commonly obtained either from uplink measurements using dedicated uplink pilot signals [18, 19] or from limited channel state information (CSI) feedback from the users [20–22]. We consider frequency division duplex (FDD) transmission in this paper, where CSIT is provided via feedback from the users over dedicated limited capacity feedback links. For this purpose, the users have to quantize the required CSI utilizing a proper quantization codebook. Two approaches are most common in current mobile communication systems [23]: (1) Implicit CSI feedback, where the users select preferred transmit beamformers/precoders from a predefined codebook given the current channel conditions. This has the advantage that the users can accurately estimate the expected performance during transmission, provided the channel conditions are quasi-static. (2) Explicit CSI feedback, where the users directly quantize the MIMO channel matrix. This approach is better suited for multi-user MIMO transmission because it facilitates calculation of efficient multi-user precoders at the transmitter on-the-fly. Either way, providing CSI via limited feedback is always prone to inaccuracies, on the one hand, caused by the quantization process, which is inevitable due to the limited capacity of the feedback links. On the other hand, CSI inaccuracies also result from the delay of the feedback path, leading to a mismatch of the CSI experienced during the quantization phase and the CSI valid during transmission. Several theoretical studies exist that evaluate the performance degradation due to imperfect and outdated CSIT for a variety of transmission scenarios [24–27].

Popular coherent multi-user beamforming/precoding methods, such as zero forcing (ZF) beamforming [28], block diagonalization (BD) precoding [29] and interference alignment (IA) [30, 31], utilize the phase difference between the multipath components contributing to the channel between the transmitter and the receiver to achieve constructive/destructive signal interference. The phases of multipath components, however, can change very quickly over time, especially for mobile users, and hence coherent beamforming/precoding methods react very sensitive to non-static scenarios [32, 33]. The corresponding performance degradation can partly be mitigated by robust coherent beamforming/precoding techniques that account for the CSIT imperfection. One prominent example of robust beamforming techniques is minimum variance robust adaptive beamforming [34], which accounts for imperfections in the estimation of signal covariance matrices. In [35], the authors achieve robustness w.r.t. CSI quantization by utilizing multiple receive antennas to improve system performance without the need for full channel feedback information; this idea is in essence similar to [36], where subspace-selection strategies for CSI feedback are discussed that enable mitigating the impact of CSI quantization. In [37], robust single-user MIMO precoding methods are proposed that maximize the worst-case received signal-to-noise ratio or minimize the worst-case error probability with imperfect CSIT. Furthermore, hybrid transceiver architectures that incorporate robustness against channel estimation errors are proposed in [38].

Contribution In this paper, we take a different approach to achieve robustness of multi-user beamforming/precoding w.r.t. CSIT imperfections due to limited feedback. Specifically, we employ an over-complete basis decomposition of the MIMO channel matrix at each receiver to obtain a sparse approximation of the CSI, which can efficiently be quantized and fed back to the transmitter. The over-complete basis decomposition is based on an angular-domain quantization codebook (also called dictionary following the nomenclature of sparse approximation literature), which is parametrized by azimuth and elevation signal departure angles. Under line of sight (LOS) conditions, the selected expansion vectors of the sparse approximation are robust w.r.t. channel estimation errors as well as channel variations over time. We demonstrate this robustness by means of numerical simulations employing state-of-the-art three-dimensional stochastic-geometric channel models. However, the complex-valued coefficients of the over-complete basis expansion are still significantly impacted by channel estimation errors and channel variations over time, which can impair the performance of coherent precoding techniques, such as, as ZF beamforming, BD precoding

and IA, that rely on destructive multipath interference to eliminate inter-user interference. We therefore propose to apply incoherent beamforming/precoding techniques that are inherently robust w.r.t. phase uncertainties of multipath components, similar to space-division multiple access (SDMA). Specifically, we propose leakage-based angular beamforming/precoding methods that make use only of the azimuth and elevation angles as well as the norm of the expansion coefficients and do not rely on the phase of the expansion coefficients to achieve destructive multipath interference. This approach also improves robustness of beamforming/precoding solutions at high mobility, where the relative phases of multipath components change very quickly over time. To further improve robustness we additionally account for uncertainty in the angular channel decomposition in the proposed precoder optimization. Naturally, this approach cannot compete with coherent precoding methods in case of highly accurate CSIT; however, our numerical simulations demonstrate significant gains in terms of achievable rate and symbol error ratio (SER) in case of imperfect CSIT under LOS conditions. The present work is an extension of our two conference publications [39, 40] to multi-user beamforming and multi-stream multi-user precoding. Specifically, in [39] we propose and approximately solve the FD-MIMO transmit beam pattern optimization problem of worst-case signal to leakage ratio (SLR) maximization, where we maximize the ratio of signal power radiated into an angular region of interest to leakage power radiated into an unintended angular region. A similar design and optimization principle guides us also in the present paper, yet extended to adaptive operation with limited feedback from the users and transmission of multiple data streams per user. In [40], we consider so-called double-sided FD-MIMO systems, where both transmitter and receiver are equipped with FD-MIMO antenna arrays, and we apply a similar over-complete basis decomposition as in the present paper to achieve efficient limited feedback operation. However, we do not consider robust multi-user beamforming/precoding in [40]. Hence, the present paper combines results of [39, 40] and extends them to adaptive limited-feedback based multi-user precoding with multi-stream transmission per user.

Organization In Section 2, we introduce our system model and state the underlying assumptions of this work. In Section 3, we present the proposed CSI feedback approach that is based on a sparse approximation of the channel utilizing an over-complete basis decomposition. Based on this CSI feedback, we then propose in Section 4 our robust leakage-bounded angular beamforming and precoding methods. We briefly summarize for the readers convenience in Section 5 the benchmark methods employed in our simulations, in order to clarify how these methods are applied in our considered scenario. Finally, in Section 6, we investigate the performance of the proposed beamformers and precoders by means of numerical simulations and we provide concluding remarks in Section 7.

Notation We denote scalars by upper and lower case italic letters (e.g., N, n), vectors by lower case bold letters (e.g., $\mathbf{x}$), and matrices by upper case bold letters (e.g., $\mathbf{H}$). To address the element in the i-th row and the j-th column of matrix $\mathbf{H}$, we use the notation $[\mathbf{H}]_{i,j}$. We write the trace of a matrix as $\mathrm{tr}\,(\mathbf{H})$, the transpose as $\mathbf{H}^{\mathrm{T}}$, the conjugate-transpose as $\mathbf{H}^{\mathrm{H}}$, the Frobenius norm as $\|\mathbf{H}\|$, the pseudo-inverse as $\mathrm{pinv}(\mathbf{H})$, and the vector obtained by stacking the columns of the matrix below each other as $\mathrm{vec}(\mathbf{H})$. We denote the logarithm of the determinant of a quadratic matrix $\mathbf{H}$ as $\log|\mathbf{H}|$. We denote sets by calligraphic letters (e.g., $\mathcal{A}$); the size of set $\mathcal{A}$ is $|\mathcal{A}|$. To calculate the expected value of a random variable r, we employ the notation $\mathbb{E}r$. The Dirac delta function is $\delta(t)$; the discrete Kronecker delta is δ_{ij}, i.e., $\delta(t) \to \infty \mathrm{iff} t = 0$ and $\delta_{ij} = 1 \mathrm{iff} i = j$. We denote the real part of a complex number z as $\Re(z)$, the imaginary part as $\Im(z)$, and the complex conjugate as z^*. We denote the uniform distribution on the interval $[a, b]$ as $\mathcal{U}(a, b)$ and the vector-valued complex-Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ as $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$.

2 System model

2.1 Channel model

In wireless communications, the channel between transmitter and receiver can be characterized by a double-directional time-variant channel impulse response [41]:

$$h(t, \tau, \boldsymbol{\Psi}_t, \boldsymbol{\Psi}_r) = \sum_{\ell=1}^{N_L} \alpha_\ell \delta(\tau - \tau_\ell) \delta(\boldsymbol{\Psi}_t - \boldsymbol{\Psi}_{t,\ell}) \delta(\boldsymbol{\Psi}_r - \boldsymbol{\Psi}_{r,\ell}), \tag{1}$$

with N_L denoting the number of multipath components and α_ℓ, τ_ℓ, respectively, representing the complex-valued amplitude and the propagation delay associated to path ℓ. The vectors $\boldsymbol{\Psi}_{t,\ell} = [\phi_{t,\ell}, \theta_{t,\ell}]^{\mathrm{T}}$ and $\boldsymbol{\Psi}_{r,\ell} = [\phi_{r,\ell}, \theta_{r,\ell}]^{\mathrm{T}}$ denote the azimuth and elevation angles of departure and arrival. Although not explicitly shown in (1), the parameters on the right-hand side may all implicitly depend on the absolute time t due to movement of users and scattering objects. Incorporating the responses of the transmit and receive antenna elements w.r.t. a plane wave from angular direction $\boldsymbol{\Psi}$, i.e., $a_t(\boldsymbol{\Psi})$ and $a_r(\boldsymbol{\Psi})$, we obtain the time-variant channel impulse response as:

$$h(t,\tau) = \int \cdots \int_{-\pi}^{\pi} a_t(\boldsymbol{\Psi}_t) a_r(\boldsymbol{\Psi}_r) h(t,\tau,\boldsymbol{\Psi}_t,\boldsymbol{\Psi}_r) d\boldsymbol{\Psi}_r d\boldsymbol{\Psi}_t = \sum_{\ell=1}^{N_L} \alpha_\ell \delta(\tau-\tau_\ell) a_r(\boldsymbol{\Psi}_{r,\ell}) a_t(\boldsymbol{\Psi}_{t,\ell}). \quad (2)$$

If we furthermore incorporate the overall impulse response of the transmit and receive filters $f(\tau)$, we obtain the transceiver-dependent time-variant channel impulse response as:

$$\tilde{h}(t,\tau) = \int_{\tau'} h(t,\tau') f(\tau-\tau') d\tau' = \sum_{\ell=1}^{N_L} \alpha_\ell f(\tau-\tau_\ell) a_r(\boldsymbol{\Psi}_{r,\ell}) a_t(\boldsymbol{\Psi}_{t,\ell}), \quad (3)$$

where again the time dependence of the parameters on the right-hand side is implicit. Under the well-known planar wave and narrowband array assumptions [42], we can obtain the time-variant impulse response of a MIMO system simply by concatenating the individual impulse responses of antenna element pairs:

$$\tilde{\mathbf{H}}(t,\tau) = \sum_{\ell=1}^{N_L} \alpha_\ell f(\tau-\tau_\ell) \mathbf{a}_r(\boldsymbol{\Psi}_{r,\ell}) \mathbf{a}_t(\boldsymbol{\Psi}_{t,\ell})^{\mathrm{T}} \quad (4)$$

with $\mathbf{a}_r(\boldsymbol{\Psi}) \in \mathbb{C}^{N_r \times 1}$, $\mathbf{a}_t(\boldsymbol{\Psi}) \in \mathbb{C}^{N_t \times 1}$ denoting the receive and transmit antenna array response vectors. For example, consider a uniform planar array (UPA) with N_v rows and N_h columns at the transmitter, with vertical antenna element spacing d_v and horizontal antenna element spacing d_h in multiples of the wavelength λ. Then, the response of the antenna element in the ℓ-th row and the k-th column with respect to a plane wave from direction $\boldsymbol{\Psi} = [\phi, \theta]^{\mathrm{T}}$ is [43]:

$$\left[\mathbf{a}_t^{\mathrm{UPA}}(\phi,\theta)\right]_{(\ell-1)N_h+k} = \frac{1}{\sqrt{N_t}} \exp\big(j2\pi (d_v(\ell-1)\cos\theta + d_h(k-1)\sin\phi\sin\theta)\big)\, g_e(\phi,\theta), \quad (5)$$

$$\ell \in \{1,\ldots,N_v\},\ k \in \{1,\ldots,N_h\},\ N_t = N_v N_h, \quad (6)$$

where angles are measured as illustrated in Fig. 1 in relation to the antenna array and the individual antenna element responses are arranged row-by-row in the vector $\mathbf{a}_t^{\mathrm{UPA}}(\phi,\theta) \in \mathbb{C}^{N_t \times 1}$. In (5), the function $g_e(\phi,\theta) \in \mathbb{C}$ denotes the complex-valued angle-dependent antenna element gain pattern, e.g., a Hertzian dipole pattern. We do not focus on a specific antenna array geometry in this work nor do we restrict the antenna element gain pattern $g_e(\phi,\theta)$. Notice, in (4), we do not account for coupling in-between antenna elements, which may be incorporated with additional coupling matrices [43].

2.2 Transmission model

We consider downlink transmission from a base station that is equipped with an antenna array of N_t antenna elements, e.g., a UPA of size $N_t = N_h \cdot N_v$ as illustrated in Fig. 1. The base station serves U users. The users are each equipped with N_r receive antennas; we assume $N_t \geq N_r$. We consider orthogonal frequency division multiplexing (OFDM) transmission; the MIMO-OFDM channel matrix $\mathbf{H}[n,k] \in \mathbb{C}^{N_r \times N_t}$ at OFDM symbol n and subcarrier k is obtained by sampling the Fourier transform of the time-variant impulse response $\tilde{\mathbf{H}}(t,\tau)$ of (4). We assume that time variations within OFDM symbols are small enough such that the inter-carrier interference due to Doppler shifts is negligible. In the following, we focus on a single subcarrier and omit the subcarrier index k for brevity. To

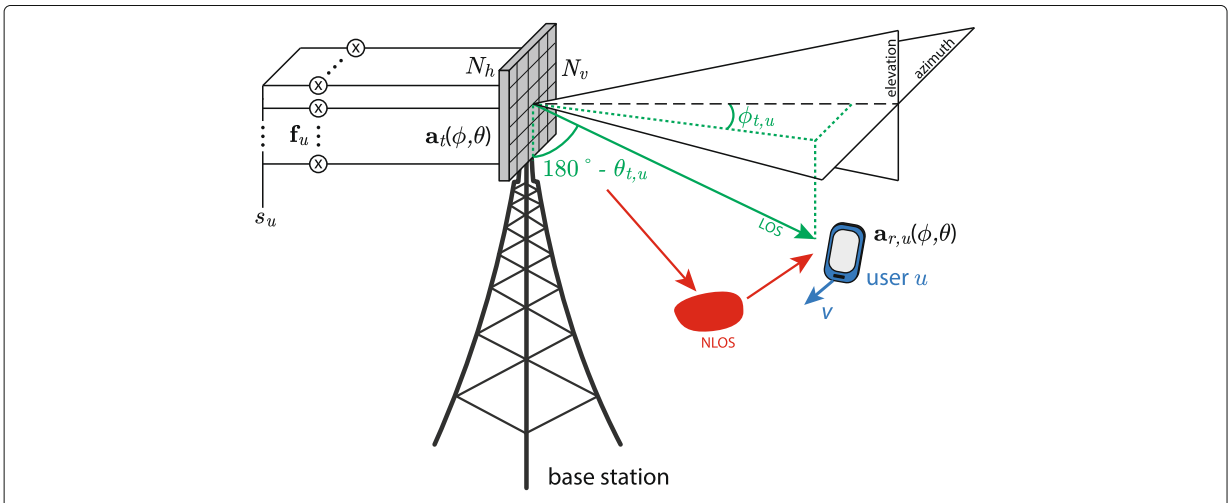


Fig. 1 Illustration of the considered FD-MIMO transmission system employing three-dimensional downlink beamforming

refer to the channel matrix associated to a specific user u, we employ a subscript, such as $\mathbf{H}_u[n]$.

The base station selects at each time instant a set $\mathcal{S}[n] \subseteq \{1,\ldots,U\}$ of users to be served in parallel, employing spatial multiplexing by means of linear beamforming/precoding. Assuming transmission of L_u streams to user $u \in \mathcal{S}[n]$, the base station applies the precoding matrix $\mathbf{F}_u[n] \in \mathbb{C}^{N_t \times L_u}$ to transmit the data symbol vector $\mathbf{s}_u[n] \in \mathbb{C}^{L_u \times 1}$ to user u. We assume $\mathbb{E}(\mathbf{s}_u[n]\,\mathbf{s}_u[n]^{\mathrm{H}}) = \mathbf{I}_{L_u}$ and account for the power allocation in $\mathbf{F}_u[n]$, such that $\mathrm{tr}\,(\mathbf{F}_u[n]\,\mathbf{F}_u[n]^{\mathrm{H}}) = 1/|\mathcal{S}[n]|$. The input-output relationship of user u hence is

$$\mathbf{y}_u[n] = \mathbf{H}_u[n] \sum_{j \in \mathcal{S}[n]} \mathbf{F}_j[n]\,\mathbf{s}_j[n] + \mathbf{n}_u[n] \in \mathbb{C}^{N_r \times 1}, \quad (7)$$

with $\mathbf{n}_u[n]$ denoting zero-mean complex Gaussian receiver noise of variance $\sigma^2_{n,u}$.

2.3 Targeted deployment scenarios

The beamforming and precoding methods proposed below are basically enhanced interference-aware SDMA techniques. As such, the methods require a certain degree of spatial separability of users in the angular domain. Such separability can mainly be guaranteed in limited scattering environments with a small number of dominant specular components per user and relatively weak diffuse background scattering, as common in LOS conditions. Even though this might sound restrictive for mobile communications, many scenarios envisioned for deployment of FD-MIMO systems are expected to exhibit dominant LOS transmission. Example scenarios where LOS propagation could dominate are the urban micro and macro scenarios illustrated in ([2], Fig. 2) and ([44], Fig. 2), where users on different floors of a building are served in parallel by means of spatial beamforming from base stations mounted on neighboring buildings. Another interesting deployment scenario is multi-user multiplexing in large indoor open spaces, e.g., within stadiums, airport terminals, shopping malls, where high-ceiling-mounted small cells and remote radio heads can provide LOS connectivity to users. Furthermore of interest are vehicular communication scenarios on highways and motorways, where FD-MIMO beamforming can be employed for multi-user transmission between road-side access points and vehicles as described in [45]. Notice also that channel fading models that contain only two dominant specular components in addition to weak diffuse background scattering, such as the two-wave diffuse power (TWDP), generalized two-ray (GRT), and fluctuating two-ray (FRT) fading models [46–50], are well suited to capture the behavior of channel measurements conducted in the millimeter wave band [49, 51], suggesting that millimeter wave transmissions are in many cases well characterized by few specular components. Importantly, in the millimeter wave band, even relatively low mobility of users (pedestrians) causes significant time selectivity of the channel due to the small wavelength; hence, CSIT becomes outdated even faster.

3 Channel state information feedback

We assume that the transmitter inserts pilot symbols in the OFDM time-frequency grid to support channel estimation for CSI feedback calculation at the receivers. Notice, these CSI pilots may be provided sparsely in the time-frequency grid to facilitate coarse channel estimation only for CSI calculation; additional more densely populated precoded pilot symbols can be employed to support accurate channel estimation for data detection, similar to LTE's CSI and UE-specific pilot symbols [52]. We consider channel quantization and feedback in the frequency domain; alternatively, it is also possible to quantize the channel impulse response [53]. We write the MIMO-OFDM channel matrix $\mathbf{H}_u[n]$ as a function of the estimated matrix $\hat{\mathbf{H}}_u[n]$ as:

$$\mathbf{H}_u[n] = \hat{\mathbf{H}}_u[n] + \mathbf{E}_u[n], \quad (8)$$

with $\mathbb{E}\left(\mathrm{vec}(\mathbf{E}_u[n])\,\mathrm{vec}(\mathbf{E}_u[n])^{\mathrm{H}}\right) = \sigma^2_{e,u}\mathbf{I}_{N_r N_t}$. If we assume that downlink CSI pilot signals are provided relatively sparse in the time-frequency domain, to keep the pilot signal overhead small, the channel estimation error variance $\sigma^2_{e,u}$ will be comparatively large.

In order to attain more robust behavior w.r.t. channel estimation errors and outdated CSIT, we propose to employ an over-complete basis decomposition of $\hat{\mathbf{H}}_u[n]$ to obtain a sparse approximation of $\mathbf{H}_u[n]$.[1] Choosing a proper quantization codebook/dictionary $\mathcal{D}$ for the decomposition, we can achieve a signal denoising gain that provides the wanted robustness. More specifically, we apply the following decomposition:

$$\hat{\mathbf{H}}_u[n] = \sum_{s=1}^{S} \mathbf{c}^s_u[n]\,(\mathbf{d}^s_u[n])^{\mathrm{T}} + \mathbf{R}_u[n], \quad (9)$$

with $\mathbf{d}^s_u[n] \in \mathcal{D}$ representing the s-th selected entry/atom from the dictionary $\mathcal{D}$ and $\mathbf{c}^s_u[n] \in \mathbb{C}^{N_r \times 1}$ being the corresponding expansion coefficient vector. Matrix $\mathbf{R}_u[n] \in \mathbb{C}^{N_r \times N_t}$ is the residual decomposition error. The dictionary $\mathcal{D} = \{\mathbf{d}_1,\ldots,\mathbf{d}_{|\mathcal{D}|}\}$, $\mathbf{d}_i \in \mathbb{C}^{N_t \times 1}$, $|\mathcal{D}| \geq N_t$ is chosen to span the entire N_t-dimensional complex Euclidean space, in order to be able to represent in principle any matrix error-free with sufficiently many expansion coefficients. Our goal is to obtain a good sparse approximation of $\hat{\mathbf{H}}_u[n]$ with a very small number S of expansion terms by means of a dictionary that supports easy parametrization and therefore efficient CSI quantization/feedback. This can be achieved if the channel matrix is compressible in the chosen dictionary [54], which basically means that the norm of the expansion coefficient vectors $\|\mathbf{c}^s_u[n]\|$

decreases exponentially with the index s (assuming sorting in decreasing order).

If we assume transmission under LOS conditions, the channel impulse response $\tilde{\mathbf{H}}(t,\tau)$ is dominated by the LOS specular component. This dominant component is then also present in the Fourier transform $\mathbf{H}_u[n]$. Correspondingly, we propose to utilize a dictionary that is matched to the transmit antenna array response vector $\mathbf{a}_t(\boldsymbol{\Psi})$, since this will result in a compressible decomposition of the channel matrix.[2] We therefore propose to set the dictionary as:

$$\mathcal{D} = \left\{ \mathbf{d}_i = \mathbf{a}_t(\boldsymbol{\Psi}_i) \middle| \boldsymbol{\Psi}_i \in \left\{\phi_1, \ldots, \phi_{N_\phi}\right\} \times \left\{\theta_1, \ldots, \theta_{N_\theta}\right\} \right\}. \tag{10}$$

With $|\mathcal{D}| = N_\phi N_\theta \geq N_t$, this dictionary spans the N_t-dimensional complex Euclidean space and can thus be used to decompose in principle any matrix; however, for a general unstructured matrix, such decomposition will not be compressible. Notice, as the codebook is matched to the transmit antenna array geometry, the users must be aware of this geometry which requires additional downlink signaling. However, since mainly UPAs are employed in practice, we expect this to be a minor issue. The size of the quantization codebook in the azimuth and elevation domains N_ϕ and N_θ should be chosen according to the spatial resolution of the antenna array, i.e., with a larger number of antenna elements in the horizontal/vertical dimension also the resolution of the quantization codebook in the corresponding angular domain should be increased. In order to be able to closely match the angular directions of the dominant specular components, we employ a dictionary with $|\mathcal{D}| \gg N_t$ in our simulations.

To obtain a sparse channel decomposition, we consider the following optimization problem:

$$\min_{\{\mathbf{c}_i \in \mathbb{C}^{N_r \times 1}\}_{i=1}^{S}, S \leq |\mathcal{D}|} \quad \sum_{i=1}^{S} \|\mathbf{c}_i\| \tag{11}$$

$$\text{subject to: } \left\| \hat{\mathbf{H}}_u[n] - \sum_{i=1}^{S} \mathbf{c}_i \mathbf{d}_i^{\mathrm{T}} \right\|^2 \leq \epsilon^2, \ \mathbf{d}_i \in \mathcal{D}$$

We determine an approximate solution of this optimization problem by applying a variation of the orthogonal matching pursuit (OMP) algorithm [55], as detailed in Algorithm 1. Notice, to approximately solve (11), we would actually have to adapt the number S of expansion terms in Algorithm 1 according to the desired approximation accuracy ϵ^2; however, in our simulations, we employ a fixed S. Notice the objective function in (11) can also be written as an $\ell 1$-norm $\|[\|\mathbf{c}_1\|, \ldots, \|\mathbf{c}_S\|]\|_1$ justifying its use for sparse decomposition. Given the sparse approximation, the CSI feedback is set equal to the corresponding codebook indices of the optimal azimuth and elevation angles $\left(\phi_u^s[n], \theta_u^s[n]\right)$ of the expansion vectors $\mathbf{d}_u^s[n]$. Additionally, we provide the norms of the expansion coefficient vectors $\left\|\mathbf{c}_u^s[n]\right\|$ as feedback information; the reason for this will become clear in Section 4. In our simulations in Section 6, we consider unquantized feedback of the expansion coefficient vector norms; we have investigated the impact of quantization of these norms for $N_r = 1$ in [39], showing that coarse quantization is possible without significant performance degradation.

Algorithm 1 Channel decomposition for CSI feedback calculation utilizing a variation of OMP.

1: Initialize the decomposed channel matrix $\mathbf{H}_d = \mathbf{0}$
2: Initialize the expansion matrix $\mathbf{D} = [\ ]$
3: **for** $s = 1$ to S **do**
4: Calculate the intermediate decomposition residual

$$\mathbf{R} = \hat{\mathbf{H}}_u[n] - \mathbf{H}_d \tag{12}$$

5: Calculate the decomposition metric

$$m_i = \left\|\mathbf{R}\mathbf{d}_i^*\right\|^2, \ \forall \mathbf{d}_i \in \mathcal{D} \tag{13}$$

6: Determine the s-th expansion vector

$$\mathbf{d}_u^s[n] = \arg\max_{\mathbf{d}_i \in \mathcal{D}} m_i \tag{14}$$

7: Update the expansion matrix

$$\mathbf{D} = \left[\mathbf{D}, \mathbf{d}_u^s[n]\right] \in \mathbb{C}^{N_t \times s} \tag{15}$$

8: Calculate the intermediate expansion coefficient matrix

$$\mathbf{C} = \hat{\mathbf{H}}_u[n] \ \mathrm{pinv}\,(\mathbf{D})^{\mathrm{T}} \in \mathbb{C}^{N_r \times s} \tag{16}$$

9: Update the decomposed channel matrix

$$\mathbf{H}_d = \mathbf{C}\mathbf{D}^{\mathrm{T}} \in \mathbb{C}^{N_r \times N_t} \tag{17}$$

10: **end for**
11: Set the expansion coefficient vectors

$$\left[\mathbf{c}_u^1[n], \ldots, \mathbf{c}_u^S[n]\right] = \mathbf{C} \tag{18}$$

4 Leakage-bounded angular beamforming and precoding

In this section, we propose multi-user beamforming and precoding methods that utilize the angular-domain CSI feedback described in the previous section. We consider first multi-user beamforming with a single stream per user in Section 4.1 and extend to multiple streams per user in Section 4.2.

4.1 Leakage-bounded angular beamforming

In this section, we assume that the base station transmits only a single stream $L_u = 1$ to each served user $u \in \mathcal{S}[n]$, i.e., the precoding matrices $\mathbf{F}_u[n]$ reduce to

beamforming vectors $\mathbf{f}_u[n]$. From the angular CSI feedback provided by the users, the base station has at time instant n delayed knowledge of the channel expansion vectors $\mathbf{d}_u^s[n-m]$ and the norm of the corresponding expansion coefficient vectors $\left\|\mathbf{c}_u^s[n-m]\right\|$, where m represent the delay of the CSI feedback path. To calculate the transmit beamformers, we interpret the expansion vectors $\mathbf{d}_u^s[n-m]$ as specular channel contributions and propose a beamformer optimization that maximizes the expected received signal power of user u over its specular components, while restricting the interference leakage caused to the other users over their respective specular components. To shorten notation, we omit the time indices $[n]$ and $[n-m]$ in the following derivations.

Consider the signal power received over the S specular components corresponding to the available CSIT:

$$\begin{aligned} P_u &= \mathbb{E}\left(\left\|\left(\sum_{s=1}^{S}\mathbf{c}_u^s\left(\mathbf{d}_u^s\right)^{\mathrm{T}}\right)\mathbf{f}_u s_u\right\|^2\right) \\ &= \mathrm{tr}\left(\sum_{s=1}^{S}\sum_{k=1}^{S}\mathbb{E}((\mathbf{c}_u^k)^{\mathrm{H}}\mathbf{c}_u^s)(\mathbf{d}_u^s)^{\mathrm{T}}\mathbf{f}_u\mathbf{f}_u^{\mathrm{H}}(\mathbf{d}_u^k)^*\right). \end{aligned} \tag{19}$$

The value of the inner product $(\mathbf{c}_u^k)^{\mathrm{H}}\mathbf{c}_u^s = r_u^{k,s}e^{j\varphi_u^{k,s}}$ depends on the relative phase-shift $\varphi_u^{k,s}$ between these expansion coefficient vectors. Since we do not provide feedback information about these relative phase-shifts, we assume $\varphi_u^{k,s} \sim \mathcal{U}(0, 2\pi)$ such that $\mathbb{E}(r_u^{k,s}e^{j\varphi_u^{k,s}}) = 0$. Notice, this assumption also makes sense for our targeted application scenario of high user mobility. In such scenarios, the angular CSI feedback corresponding to the expansion vectors $\mathbf{d}_u^s$ is relatively stable, whereas the phase of the expansion coefficient vectors $\mathbf{c}_u^s$ varies significantly over time. Hence, relying on such phase information at high mobility is not recommended. With this assumption of uncorrelated phases, the signal power is:

$$P_u = \sum_{s=1}^{S}\left\|\mathbf{c}_u^s\right\|^2(\mathbf{d}_u^s)^{\mathrm{T}}\mathbf{f}_u\mathbf{f}_u^{\mathrm{H}}(\mathbf{d}_u^s)^* \tag{20}$$

and the transmitter has sufficient CSIT to calculate this value. Similarly, we can determine the interference leakage power caused by the transmission to user u and received over specular component k of user j as:

$$L_{j,u}^k = \left\|\mathbf{c}_j^k\right\|^2(\mathbf{d}_j^k)^{\mathrm{T}}\mathbf{f}_u\mathbf{f}_u^{\mathrm{H}}(\mathbf{d}_j^k)^*. \tag{21}$$

With these definitions, we now proceed with the formulation of our optimization problem. We consider independent optimization of the transmit beamformers of the individual users according to the following optimization problem:

$$\begin{aligned} &\max_{\mathbf{f}_u\in\mathbb{C}^{N_t\times 1}} P_u \\ &\text{s. t.: } \|\mathbf{f}_u\|^2 \le 1/|\mathcal{S}|, \\ &\quad L_{j,u}^k \le L_{\max}, \quad \forall j\in\mathcal{S}, j\ne u, \forall k\in\{1,\ldots,S\}. \end{aligned} \tag{P1}$$

where the parameter $L_{\max}$ denotes the maximum tolerable interference level. Replacing $\max_{\mathbf{f}_u\in\mathbb{C}^{N_t\times 1}} P_u$ equivalently by $\min_{\mathbf{f}_u\in\mathbb{C}^{N_t\times 1}}(-P_u)$, this optimization problem can be formulated as a quadratically constrained quadratic program; however, since the matrix $-\sum_{s=1}^{S}\left\|\mathbf{c}_u^s\right\|^2(\mathbf{d}_u^s)^*(\mathbf{d}_u^s)^{\mathrm{T}}$, which determines the quadratic objective function, is negative semidefinite, the problem is in general non-convex and non-deterministic polynomial-time (NP)-hard. An approximate solution is possible by applying a semidefinite programming relaxation (SDR) to the optimization variable $\mathbf{f}_u$ and by recovering a feasible suboptimal rank-one beamforming solution through randomization [56]. The approximation performance of this approach in general deteriorates with increasing dimension N_t of the optimization variable as well as with growing number of constraints in (P1) [57, 58]. Hence, for our envisioned use case of FD-MIMO systems with large N_t and $|\mathcal{S}|$, the achieved approximation performance might not be satisfactory.

We therefore consider a modification of (P1) to avoid the SDR: Instead of maximizing the sum signal power over all S specular components of user u, we rather maximize the power received only over the strongest specular component $s=1$, while still considering the interference leakage caused over all specular components:

$$\begin{aligned} &\max_{\mathbf{f}_u\in\mathbb{C}^{N_t\times 1}} \left\|\mathbf{c}_u^1\right\|^2(\mathbf{d}_u^1)^{\mathrm{T}}\mathbf{f}_u\mathbf{f}_u^{\mathrm{H}}(\mathbf{d}_u^1)^* \\ &\text{s. t.: } \|\mathbf{f}_u\|^2 \le 1/|\mathcal{S}|, \\ &\quad L_{j,u}^k \le L_{\max}, \quad \forall j\in\mathcal{S}, j\ne u, \forall k\in\{1,\ldots,S\}. \end{aligned} \tag{P2}$$

This problem can be transformed to a convex problem by recognizing that (P2) is invariant w.r.t. the absolute phase of $\mathbf{f}_u$. Hence, we can apply the approach considered, e.g., in [59], and restrict $(\mathbf{d}_u^1)^{\mathrm{T}}\mathbf{f}_u$ to be real valued to obtain the following equivalent convex second-order cone program:

$$\begin{aligned} &\max_{\mathbf{f}_u\in\mathbb{C}^{N_t\times 1}, p_u\in\mathbb{R}} p_u \\ &\text{s. t.: } \|\mathbf{f}_u\|^2 \le 1/|\mathcal{S}|, \\ &\quad \Re\left((\mathbf{d}_u^1)^{\mathrm{T}}\mathbf{f}_u\right) \ge p_u, \\ &\quad \Im\left((\mathbf{d}_u^1)^{\mathrm{T}}\mathbf{f}_u\right) = 0, \\ &\quad L_{j,u}^k \le L_{\max}, \quad \forall j\in\mathcal{S}, j\ne u, \forall k\in\{1,\ldots,S\}. \end{aligned} \tag{P3}$$

We denote the optimal solution of this optimization problem as the *non-robust leakage-bounded angular beamformer*. As this is a second-order cone program, its

complexity scales with the square root of the number of cone constraints [60, 61].

The solution to problem (P3) allows to efficiently control the interference leakage caused to other users, provided the angular CSIT is accurate. However, since this CSIT is obtained by limited feedback, angular quantization is unavoidable which impairs the quality of the CSIT. Additionally, in mobile situations, the azimuth and elevation angles representing the specular components change over time, causing a mismatch between the angular CSI feedback and the actual angles, due to the processing delay of the feedback link. These effects increase the inter-user interference and thus deteriorate the performance of the system. To improve the robustness of our beamformer solution w.r.t. angular uncertainty, we consider a robust problem formulation in the following.

Given the angular feedback (ϕ_u^s, θ_u^s), we propose to incorporate angular uncertainty regions into the beamformer optimization problem. We denote the discrete angular uncertainty region corresponding to specular component s of user u as $T_u^s \subseteq \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [0, \pi]$, such that $(\phi_u^s, \theta_u^s) \in T_u^s$. In our investigations, we utilize symmetric rectangular regions around (ϕ_u^s, θ_u^s). The sizes of these regions can, e.g., be set according to the angular variation observed over the previous CSI feedback interval, i.e., $2\left|\phi_u^s[n-m] - \phi_u^s[n-2m]\right|$ and $2\left|\theta_u^s[n-m] - \theta_u^s[n-2m]\right|$, where the factor 2 accounts for the unknown direction of movement. The density of the lattice points considered within these regions, i.e., the angular resolution of T_u^s, must be chosen sufficiently large to ensure that no side-lobes of the radiation pattern are missed. As a guideline, consider a UPA with equal gain beamforming: here, zeros in the radiation pattern in azimuth occur at angles $\sin\phi = \pm\frac{n\lambda}{d_h N_h}$; hence, the angular resolution in azimuth must be chosen to assure that the peaks in-between such zeros are resolved with sufficient accuracy. Notice, though, that the complexity of the proposed optimization problems, more specifically the number of constraints, scales linearly with the size of each set T_u^s; hence, one should avoid unnecessary oversampling of T_u^s to reduce computational complexity.

Utilizing these angular uncertainty regions, we now formulate the optimization problem for the *robust leakage-bounded angular beamformer*:

$$\max_{\mathbf{f}_u \in \mathbb{C}^{N_t \times 1}} \left\|\mathbf{c}_u^1\right\|^2 (\mathbf{d}_u^1)^{\mathrm{T}} \mathbf{f}_u \mathbf{f}_u^{\mathrm{H}} (\mathbf{d}_u^1)^* \tag{P4}$$

$$\text{s. t.: } \|\mathbf{f}_u\|^2 \leq 1/|\mathcal{S}|,$$

$$\left\|\mathbf{c}_j^k\right\|^2 (\mathbf{a}_t(\boldsymbol{\Psi}_{i,j}^k))^{\mathrm{T}} \mathbf{f}_u \mathbf{f}_u^{\mathrm{H}} (\mathbf{a}_t(\boldsymbol{\Psi}_{i,j}^k))^* \leq L_{\max},$$

$$\forall j \in \mathcal{S}, j \neq u, \ \forall k \in \{1, \ldots, S\}, \ \forall \boldsymbol{\Psi}_{i,j}^k \in T_j^k.$$

In this problem, the vector $\mathbf{a}_t(\boldsymbol{\Psi}_{i,j}^k)$ denotes the transmit antenna array response evaluated at angle $\boldsymbol{\Psi}_{i,j}^k \in T_j^k$; here, the same antenna array response as in the CSI feedback dictionary (10) is used. Notice, we account for the angular uncertainty only in the interference terms, but not in the signal term. This is because the system reacts much more sensitive w.r.t uncertainty in the interference directions as compared to the signal direction, since nulls in the beamforming pattern are commonly very narrow whereas peaks are comparatively broad in the angular domain (see also the example in Fig. 2). Another reason for neglecting uncertainty in the angular direction of the intended signal is to keep the problem convex.[3] Problem (P4) can be brought into the form of a second-order cone program, similar to Problem (P3).

4.2 Leakage-bounded angular precoding

In this section, we extend the robust beamformer design to multi-user precoding with multiple streams per user.

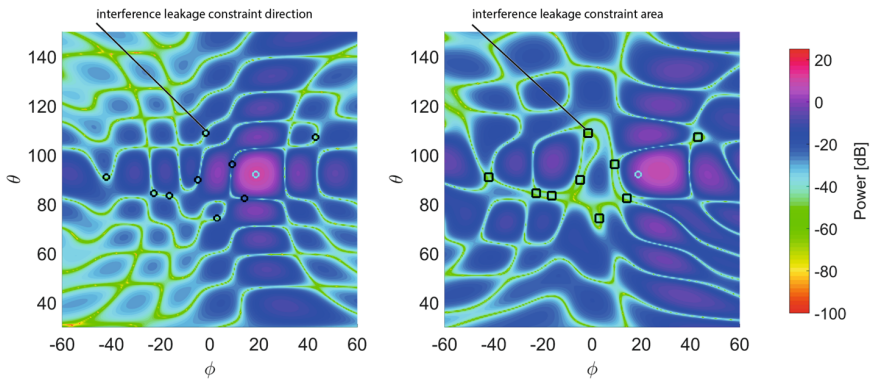


Fig. 2 Radiation beam pattern obtained with non-robust leakage-bounded angular beamforming (left) and with robust leakage-bounded angular beamforming (right). The nine angular uncertainty regions of robust leakage-bounded angular beamforming have a width of $\psi_{\max} = \pm 1.6°$ in this example

When considering multi-stream precoding, it is not sufficient to maximize the sum received power over all L_u streams of user u because this will generally lead to a rank one beamforming solution that steers all signal energy over the largest singular value of the channel. Hence, to obtain a multi-stream precoding solution of rank L_u, we have to consider the achievable rate of a user in terms of the mutual information. With the available CSIT, however, we can only obtain a coarse estimate of the achievable rate as detailed below.

Let us start by considering the achievable rate of user u in terms of the mutual information:

$$R_u = \mathbb{E}\log\left|\mathbf{I}_{N_r} + \left(\sigma_{n,u}^2\mathbf{I}_{N_r} + \sum_{\substack{j\in\mathcal{S}\\ j\neq u}}\mathbf{R}_{u,j}\right)^{-1}\mathbf{R}_u\right| \tag{22}$$

$$\mathbf{R}_u = \mathbf{H}_u\mathbf{F}_u(\mathbf{H}_u\mathbf{F}_u)^{\mathrm{H}},\ \ \mathbf{R}_{u,j} = \mathbf{H}_u\mathbf{F}_j(\mathbf{H}_u\mathbf{F}_j)^{\mathrm{H}}, \tag{23}$$

where $\mathbf{R}_u$ and $\mathbf{R}_{u,j}$ denote the signal and interference covariance matrices, respectively. We consider individual optimization of the achievable rate of each user w.r.t. its precoder, putting additional interference constraints on the signal leakage caused to the other users. This means that we do not attempt to jointly optimize the sum rate of all users; notice, though, that for single-stream beamforming, it has been shown in [62, 63] that such an approach can still achieve the maximal sum-rate, provided the leakage constraints are appropriately selected. When optimizing R_u w.r.t. $\mathbf{F}_u$, we can then focus on

$$\tilde{R}_u = \mathbb{E}\log\left|\sigma_{n,u}^2\mathbf{I}_{N_r} + \sum_{\substack{j\in\mathcal{S}\\ j\neq u}}\mathbf{R}_{u,j} + \mathbf{R}_u\right|, \tag{24}$$

since the remaining terms are independent of $\mathbf{F}_u$. Because the CSIT contains no information about the channel estimation error in (8) and the channel decomposition error in (9), the transmitter can only calculate approximations $\hat{\mathbf{R}}_u$ and $\hat{\mathbf{R}}_{u,j}$ of the covariance matrices utilizing the angular-domain CSI feedback; this gives a corresponding approximation $\hat{\tilde{R}}_u$ of $\tilde{R}_u$. In the Appendix, we derive the following upper bound on $\hat{\tilde{R}}_u$:

$$\hat{\tilde{R}}_u \leq \log\left|\sigma_{n,u}^2\mathbf{I}_S + (|\mathcal{S}|-1)\frac{SL_{\max}}{\mathrm{tr}\,(\mathbf{\Sigma}_u^2)}\mathbf{\Sigma}_u^2 + \mathbf{\Sigma}_u^2\mathbf{U}_u^{\mathrm{H}}\mathbf{C}_u\mathbf{U}_u\right|, \tag{25}$$

where the matrices $\mathbf{\Sigma}_u^2$ and $\mathbf{U}_u$, as defined in (36), can be calculated from the available CSIT and $\mathbf{C}_u = \mathbf{F}_u(\mathbf{F}_u)^{\mathrm{H}}$ denotes the transmit covariance matrix associated to user u.

We now utilize this upper bound to formulate the optimization problem for the *non-robust leakage-bounded angular precoder*:

$$\max_{\mathbf{C}_u\in\mathbb{C}^{N_t\times N_t}}\left|\sigma_{n,u}^2\mathbf{I}_S + (|\mathcal{S}|-1)\frac{SL_{\max}}{\mathrm{tr}\,(\mathbf{\Sigma}_u^2)}\mathbf{\Sigma}_u^2 + \mathbf{\Sigma}_u^2\mathbf{U}_u^{\mathrm{H}}\mathbf{C}_u\mathbf{U}_u\right| \tag{P5}$$

$$\text{s. t.:}\quad \mathrm{tr}\,(\mathbf{C}_u) \leq 1/|\mathcal{S}|,\ \mathbf{C}_u \succeq \mathbf{0},$$

$$L_{j,u}^k \leq L_{\max},\ \ \forall j\in\mathcal{S}, j\neq u,\ \forall k\in\{1,\ldots,S\},$$

with $L_{j,u}^k$ as defined in (38). Notice, this actually represents an SDR of the optimization w.r.t. the precoder $\mathbf{F}_u \in \mathbb{C}^{N_t\times L_u}$. In general, the optimal solution $\mathbf{C}_u^{(\mathrm{opt})}$ of (P5) is not guaranteed to be of rank L_u. Due to the definition of $\mathbf{\Sigma}_u^2$ and $\mathbf{U}_u$ in (36), rank $\mathbf{C}_u^{(\mathrm{opt})} \leq S$. This is because a solution of rank greater than S would steer part of the transmit energy into the null space of $\mathbf{U}_u$, which cannot maximize our objective function. The restriction rank $\mathbf{C}_u^{(\mathrm{opt})} \leq S$ also implies that we have to feed back at least as many specular components S as the number of data streams L_u to make sure that the solution $\mathbf{C}_u^{(\mathrm{opt})}$ can potentially support the transmission of L_u streams. Still, even with $S \geq L_u$, it can happen that the optimal solution $\mathbf{C}_u^{(\mathrm{opt})}$ has rank $\mathbf{C}_u^{(\mathrm{opt})} < L_u$; this is likely to be the case at low signal-to-noise ratio SNR where the transmission of less than L_u streams can provide an advantageous beamforming gain. If rank $\mathbf{C}_u^{(\mathrm{opt})} < L_u$, we simply transmit less than L_u streams over the eigenvectors corresponding to the non-zero eigenvalues of $\mathbf{C}_u^{(\mathrm{opt})}$, since this maximizes our objective function. If rank $\mathbf{C}_u^{(\mathrm{opt})} > L_u$, we apply Gaussian randomization to derive a feasible precoder of rank L_u, by appropriately modifying the randomization approaches described in more detail in [56, 64]. Notice, though, if we set $S = L_u$, i.e., we feed back as many specular components as the number of data streams, then the rank of $\mathbf{C}_u^{(\mathrm{opt})}$ is upper bounded by L_u and we therefore do not have to apply randomization at all. Thus, by setting $S = L_u$, we can guarantee that the SDR provides a globally optimal solution for the precoder $\mathbf{F}_u$.

Similar to (P4), we can also implement a *robust leakage-bounded angular precoder* by accounting for the angular uncertainty regions T_j^k:

$$\max_{\mathbf{C}_u\in\mathbb{C}^{N_t\times N_t}}\left|\sigma_{n,u}^2\mathbf{I}_S + (|\mathcal{S}|-1)\frac{SL_{\max}}{\mathrm{tr}\,(\mathbf{\Sigma}_u^2)}\mathbf{\Sigma}_u^2 + \mathbf{\Sigma}_u^2\mathbf{U}_u^{\mathrm{H}}\mathbf{C}_u\mathbf{U}_u\right| \tag{P6}$$

$$\text{s. t.:}\quad \mathrm{tr}\,(\mathbf{C}_u) \leq 1/|\mathcal{S}|,\ \mathbf{C}_u \succeq \mathbf{0},$$

$$\left\|\mathbf{c}_j^k\right\|^2 \mathrm{tr}\,(\mathbf{a}_t(\mathbf{\Psi}_{i,j}^k))^*(\mathbf{a}_t(\mathbf{\Psi}_{i,j}^k))^{\mathrm{T}}\mathbf{C}_u \leq L_{\max},$$

$$\forall j\in\mathcal{S}, j\neq u,\ \ \forall k\in\{1,\ldots,S\},\ \ \forall\mathbf{\Psi}_{i,j}^k \in T_j^k.$$

Applying a similar step as from (P2) to (P3), both problems (P5) and (P6) can be brought into convex form by adding an extra optimization variable $p_u \in \mathbb{R}$ and maximizing p_u subject to the logarithmic term in (P5), (P6)

being greater than or equal to p_u. However, such a determinant constraint is substantially more complex than the corresponding linear constraint in (P3). Hence, we consider in our simulations in Section 6.3 also as an alternative to determinant optimization a simple *per stream optimization*: we apply (P3) or (P4) L_u times to obtain beamformers that maximize the received power over the L_u strongest specular components and we concatenate these beamformers to obtain the L_u dimensional precoding matrix.

In our derivation, we assumed that the precoders of the other $|\mathcal{S}| - 1$ users are unknown when calculating the precoder of user u. Yet, in principle when optimizing the u-th user, the base station could already utilize the knowledge of the precoders calculated for those $u - 1$ users that have been optimized before. Similarly, an alternating optimization of precoders with a proper termination criterion could be applied to obtain an iterative approach that might provide better performance. For complexity reasons, however, we have not implemented such an approach in our simulations.

4.3 Implementation issues

4.3.1 Computational complexity

Our derivation in Section 4.1 shows that leakage-bounded transmit optimization as proposed in (P1) is in general an NP-hard quadratically constrained quadratic optimization problem, which cannot be solved efficiently. Approximate solution by means of an SDR is a convex optimization problem, which can be solved efficiently; yet, it is still computationally demanding, since the optimization variable of the SDR is of dimension $N_t \times N_t$ and hence grows quadratically with the size of the FD-MIMO antenna array. This complexity issue is relaxed by problem (P2) and its convex second-order cone programming reformulation (P3), in which the dimension of the optimization variable $\mathbf{f}_u$ grows only linearly with the size of the FD-MIMO antenna array. Since (P3) is a convex problem, it can be solved in polynomial time, e.g., by means of an interior point method; more details on solving second-order cone programs can for example be found in [65]. In problem (P4), we extend the number of constraints of our optimization problem by accounting for the angular uncertainty regions; this directly impacts the computational complexity of the problem. However, since the worst-case complexity of a second-order cone program scales with the square root of the number of constraints [60, 61], the increase in complexity is moderate. Considering the two multi-stream optimization problems (P5) and (P6), they both admit solution by means of an interior point method or a Newton conjugate-gradient approach as proposed in [66]. However, as our computer experiments have shown, this approach can be computationally demanding and slow. Therefore, the per-stream optimization approach proposed in Section 4.2 appears practically more relevant, since its complexity only grows linearly with the number of data streams per user.

4.3.2 Combination with hybrid architectures

As mentioned in the introduction of this paper, much research work on FD-MIMO systems is currently devoted to reducing system complexity through so-called hybrid architectures, which divide the precoding operation into an analog RF domain part and a digital base band processing part. The RF domain precoder thereby reduces the dimension of the effective base band channel, i.e., the product of channel matrix and RF domain precoder, which simplifies channel estimation, alleviates the CSI feedback overhead burden, and reduces the number of RF chains required. The RF domain precoder optimization is commonly required to provide a precoder solution with unit modulus matrix entries, to enable implementation with simple phase-shifter elements, and to be constant over the entire system bandwidth, to facilitate application in the analog domain. Both these constraints are not fulfilled by the leakage-bounded precoding optimization problem proposed in this paper. Nevertheless, we consider an extension of our leakage-bounded precoding optimization to hybrid architectures as promising future work.

Specifically, our leakage-bounded angular beamforming approach might be well suited for the design of the RF domain precoder, by applying the per stream optimization mentioned in Section 4.2 for each user more than L_u times in order to obtain a precoder of dimension equal to the number of available RF chains. However, the corresponding optimization problem (P3) or (P4) will in general not output a unit-modulus solution. It is well-known that the unit modulus precoder constraint of RF precoding is a non-convex constraint. A common solution to deal with this issue is to relax the constraint in the precoder optimization and to orthogonally project the resulting solution onto the set of unit modulus matrices [14]; this approach can also be applied to our problem at hand. Notice also, by spending twice the amount of phase-shifters it is actually possible to avoid the unit-modulus constraint all together [67].

A second issue that needs to be addressed is that the channel decomposition (11) is frequency selective due to the multipath delay spread introduced by the channel, which implies that the precoder also needs to be frequency selective and can thus not be implemented in the analog domain. However, in LOS situations, the frequency selectivity is not very distinct and can potentially be neglected for the cost of a slight frequency-dependent mismatch of the decomposition.

In principle, the so obtained analog RF domain precoder can then be combined with any base band precoding

approach that is designed for the effective base band channel, similar to the hybrid BD precoding scheme described in Section 5.2 below.

5 Benchmark methods

In this section, we summarize several well-known beamforming and precoding techniques that we employ as benchmarks for our proposed methods. We provide this section as a reference for the interested reader as well as to clarify how these benchmark methods are applied in our considered setup. Additionally, we propose in Section 5.2.4 a modification of the popular signal to leakage and noise ratio SLNR metric to enable its application with the angular-domain CSI feedback proposed in this paper.

5.1 Interference-unaware beamsteering

The following methods do not explicitly account for multi-user interference in the beamformer design. This methods are pure angle-of-departure-based beamsteering techniques, which can be viewed as typical SDMA implementations. As is well known, the performance of SDMA strongly depends on the selection of co-scheduled users [68], i.e., the transmitter should co-schedule users that experience close to orthogonal channels to avoid extensive multi-user interference. In our simulations in Section 6.2, we avoid the impact of imperfect user scheduling by employing an exhaustive search over all possibilities of co-scheduled users.

5.1.1 Max-gain beamsteering

The simplest beamforming/precoding technique is to steer the transmit signal along the L_u strongest angular directions provided by the CSI feedback:

$$\mathbf{F}_u = \frac{1}{\sqrt{|\mathcal{S}|}} \frac{\tilde{\mathbf{F}}_u}{\left\| \tilde{\mathbf{F}}_u \right\|}, \tag{26}$$

$$\tilde{\mathbf{F}}_u = \left[(\mathbf{d}_u^1)^*, \ldots, (\mathbf{d}_u^{L_u})^* \right]. \tag{27}$$

This technique does provide large signal gain; yet, its major weakness is that it does not account for interference among users. Hence, it only works properly if the base station serves users that are a priori well separated in the angular domain.

5.1.2 Chebyshev-tapered beamsteering

Max-gain beamsteering, as considered above, in general, produces large side lobe levels and thus potentially strong inter-user interference. Such large side lobe levels can be reduced by array weight tapering methods [69]. A popular weighting technique is Chebyshev weighting, which achieves the smallest main lobe width for a given side lobe level [70].

In order to steer the main lobe in an intended direction, array tapering is combined with the beamsteering approach of the previous section; i.e., the actual beamformer is obtained as the product of the array taper and the max-gain beamsteering vectors.

5.2 Interference-aware precoding

The following methods take multi-user interference into account and attempt to mitigate it.

5.2.1 Block-diagonalization precoding

BD precoding enables inter-user interference-free transmission of a total number of at most N_t data streams [29]. This is achieved by steering the transmit signal of a user into the null-space of the channel matrices of all other users. In our simulations in Section 6, we do not employ BD precoding in combination with our CSI feedback methods, but we rather assume perfect knowledge of $\mathbf{H}_u$ or its estimate $\hat{\mathbf{H}}_u$. We do this since BD precoding requires phase information which is not provided by our CSI feedback; see, e.g., [33], for an efficient CSI feedback method that can be applied with BD precoding.

5.2.2 Angle of departure block-diagonalization precoding

BD precoding is very sensitive to CSI imperfections since it relies on destructive multipath interference to eliminate multi-user interference [71–73]. It also has heavy CSI requirements as it requires knowledge of the entire channel matrix $\mathbf{H}_u$ of each user. To reduce the CSI requirements, the authors of [74] propose to calculate the precoder not for the entire channel matrix $\mathbf{H}_u$ but rather only for the dominant propagation paths between transmitter and receiver. Specifically, the authors of [74] consider single stream transmission per user and propose to calculate the corresponding ZF beamformer for the single strongest multipath component of each co-scheduled user. Rather than decomposing the frequency response channel matrix as we do in the present paper, the authors of [74] directly consider the time domain channel impulse response (4) and select the strongest multipath component according to the magnitude of α_ℓ; the multi-user ZF beamformer is then calculated from the corresponding transmit antenna array response vectors $\mathbf{a}_t(\mathbf{\Psi}_{t,\ell,u})$. This can be extended to multi-stream transmission per user by means of BD precoding in a straightforward way, by calculating the BD precoder from the L_u transmit antenna array response vectors corresponding to the angles of departure of the L_u strongest multipath components.

5.2.3 Hybrid block-diagonalization precoding

The angle of departure BD precoding approach described above does not achieve full multi-user interference cancelation, since it mitigates interference only over the strongest multipath component of each user. The residual multi-user interference leads to a significant performance

loss compared to regular BD precoding in case of perfect CSI. To reduce this performance gap, while still keeping CSI requirements at a reasonable level, the authors of [74] propose hybrid BD precoding. Here, the hybrid precoding principle is applied to first reduce the dimension of the effective base band channel matrix of each user, by employing a RF domain precoder. This RF domain precoder $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_t \times \sum_{u\in\mathcal{S}[n]} L_u}$ is obtained as the matched filter to the L_u strongest multipath components of each scheduled user. The base band BD precoder is then calculated for the product of channel matrix and RF domain precoder, which is called the effective base band channel. CSI requirements for estimating this effective base band channel are reduced compared to the full channel matrix, since the dimension of the effective channel matrix is smaller.

5.2.4 *Incoherent SLNR precoding*

SLNR beamforming is a popular coherent beamforming technique for multi-user broadcast and interference channels [75]. If full CSI is available at the base station, that is the full channel matrix $\mathbf{H}_u$ is known for each user u, the SLNR beamformer is obtained by maximizing a corresponding Rayleigh quotient. If we apply the CSI feedback of Section 3, however, the transmitter only has knowledge of the channel decomposition expansion vectors $\mathbf{d}_u^s$ and the norm of their corresponding expansion coefficients $\left\|\mathbf{c}_u^s\right\|$. Hence, the transmitter does not know whether the multipath transmissions over the different specular components $\mathbf{d}_u^s$ add up constructively or destructively, since it has no phase information available. We therefore modify the SLNR metric to consider the incoherent ratio of signal and leakage powers, utilizing the delayed angular CSI available at the transmitter:

$$\begin{aligned}
\hat{\mathrm{SLNR}}_u &= \frac{\mathrm{tr}\left(\mathbf{F}_u^{\mathrm{H}}\hat{\mathbf{P}}_u\mathbf{F}_u\right)}{\mathrm{tr}\left(\sigma_{n,u}^2\frac{N_r}{L_u}\mathbf{I}_{L_u}+\mathbf{F}_u^{\mathrm{H}}\hat{\mathbf{L}}_u\mathbf{F}_u\right)},\\
\hat{\mathbf{P}}_u &= \sum_{s=1}^{S}\left\|\mathbf{c}_u^s\right\|^2(\mathbf{d}_u^s)^*(\mathbf{d}_u^s)^{\mathrm{T}},\\
\hat{\mathbf{L}}_u &= \sum_{\substack{j\in\mathcal{S}\\ j\neq u}}\sum_{s=1}^{S}\left\|\mathbf{c}_j^s\right\|^2(\mathbf{d}_j^s)^*(\mathbf{d}_j^s)^{\mathrm{T}}.
\end{aligned} \tag{28}$$

If we restrict to semi-unitary precoding with equal power allocation, such that $\mathbf{F}_u^{\mathrm{H}}\mathbf{F}_u = \frac{1}{L_u|\mathcal{S}|}\mathbf{I}_{L_u}$, we can pull the precoder in the denominator of (28) left and right out of the sum, taking into account that $\mathrm{tr}\left(\mathbf{F}_u^{\mathrm{H}}\mathbf{F}_u\right) = 1/|\mathcal{S}|$. This leads to a generalized Rayleigh quotient that is maximized by the following semi-unitary precoder:

$$\mathbf{F}_u = \frac{1}{\sqrt{L_u|\mathcal{S}|}}\tilde{\mathbf{F}}_u, \tag{29}$$

$$\tilde{\mathbf{F}}_u = \max\mathrm{eig}\,\mathrm{vec}_{L_u}\left(\sigma_{n,u}^2\frac{N_r}{L_u}|\mathcal{S}|\,\mathbf{I}_{N_t}+\hat{\mathbf{L}}_u\right)^{-1}\hat{\mathbf{P}}_u, \tag{30}$$

where the operator $\max\mathrm{eig}\,\mathrm{vec}_{L_u}$ calculates the L_u orthonormal eigenvectors corresponding to the largest eigenvalues.

6 Simulations

In this section, we evaluate the performance of the proposed beamforming and precoding methods by means of numerical simulations. We first provide in Section 6.1 an example of the radiation beam pattern obtained by leakage-bounded angular beamforming with and without robustness constraints. This facilitates the development of a basic understanding of the behavior of the beamforming solutions produced by the proposed optimization problem, specifically with respect to the influence of the robustness constraints. In Section 6.2, we compare the achievable transmission rate of the proposed leakage-bounded angular beamforming methods to the classical interference-unaware beamsteering methods summarized in Section 5.1. In these simulations we consider a simple Rician-fading channel model, which allows to easily control the compressibility of the CSI in the angular domain via the Rician K-factor. In Section 6.3, we then utilize the more realistic QuaDRiGa channel model [76] to compare the proposed leakage-bounded angular precoding methods to the interference-aware precoding methods summarized in Section 5.2[4]. We first of all evaluate the compressibility of the channel realizations obtained from QuaDRiGa in the angular domain. Based on these results, we then investigate the performance of the different precoding methods in terms of the achievable transmission rate. We furthermore simulate the SER of uncoded transmission as a function of the variance of the channel estimation error σ_e^2 in (8), demonstrating the robustness of the proposed leakage-bounded beamformer. To find solutions to the proposed optimization problems, we apply CVX [77].

6.1 Radiation beam pattern

In our first example, we illustrate the radiation beam pattern obtained by leakage-bounded angular beamforming with and without robustness constraints. We consider single-stream beamforming and assume a UPA of isotropic antenna elements at the transmitter of size $N_t = N_h \cdot N_v = 11 \cdot 11 = 121$. We assume that $U = 10$ users are randomly placed within an angular area stretching in azimuth ϕ from $-60°$ to $60°$ and in elevation θ from $70°$ to $110°$[5]. In this example, we consider only a single specular component $S = 1$ per user; in the robust leakage-bounded optimization problem, we additionally assume

an angular uncertainty region of width $\psi_{\max} = \pm 1.6^\circ$ in azimuth and elevation around each specular component. This uncertainty represents, for example, LOS transmission to users that move on a radius of distance $d = 50\,\text{m}$ from the base station with $v = 100\,\text{km/h}$ when high-resolution CSI feedback is provided every $\tau = 50\,\text{ms}$ ($\psi_{\max} = \frac{v\tau}{d}$).

In Fig. 2, we compare the radiation beam patterns of robust and non-robust leakage-bounded angular beamforming for a user that is located at approximately $(\phi, \theta) = (22^\circ, 95^\circ)$ (indicated with the green circle) and nine leakage areas corresponding to the other $U-1$ users (indicated in black). In this example, we set the maximum tolerable interference level to $L_{\max} = -40\,\text{dB}$ relative to the unit transmit power. In Fig. 2 (left), we observe that the radiation pattern of non-robust leakage-bounded angular beamforming shows very narrow valleys wherever interference leakage constraints are active. Therefore, the transmission to the intended user does not cause substantial interference to the other users, provided the angular CSIT is accurate. However, due to the narrow width of the valleys at the interference leakage constraints, the residual inter-user inference grows significantly if the angular CSIT is not accurate. To compensate for such angular uncertainty, we consider interference leakage uncertainty regions of width $\psi_{\max}$ in the robust leakage-bounded angular beamformer. As shown in Fig. 2 (right), the interference leakage is then low over the entire uncertainty regions, thus improving the robustness with respect to imperfect angular CSIT. However, the maximum achievable gain of the intended user is slightly reduced, since it is not possible any more to perfectly steer the signal power towards the intended user while still satisfying all leakage constraints. Therefore, there exists a trade-off between the achievable signal gain of the intended user and the robustness of the interference leakage constraints.

6.2 Comparison to interference-unaware beamsteering

In our second example, we investigate multi-user beamforming with a single stream per user, comparing the interference-unaware angle of departure-based SDMA beamsteering methods of Section 5.1 and our robust/non-robust leakage-bounded angular beamformers. We consider users with a single receive antenna $N_r = 1$ and a transmitter with a UPA of size $N_t = N_h \cdot N_v = 11 \cdot 11$. We employ a Rician-fading channel model to generate the channel vector $\mathbf{h}_u \in \mathbb{C}^{1\times N_t}$ according to:

$$\mathbf{h}_u = \sqrt{\frac{K}{K+1}} e^{j\beta_u} \mathbf{a}_t^{\text{UPA}}(\phi_u, \theta_u)^{\text{T}} + \sqrt{\frac{1}{K+1}} \mathbf{h}_{u,d}, \quad (31)$$

with $\beta_u \sim \mathcal{U}(-\pi, \pi)$ denoting a random phase-shift, $\phi_u \sim \mathcal{U}(-60^\circ, 60^\circ)$ and $\theta_u \sim \mathcal{U}(70^\circ, 110^\circ)$ representing the LOS directions, and $\mathbf{h}_{u,d} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ being a diffuse background scattering component. The Rician K-factor controls the relative strength of the LOS component. We do not utilize OMP in this example to decompose the channel vector, but we rather assume that the receivers employ a dictionary that is matched to the UPA and that the receivers are able to estimate the LOS angles with an accuracy of $\pm 2^\circ$. Hence, the CSIT utilized for beamforming is $\mathbf{a}_t^{\text{UPA}}(\hat{\phi}_u, \hat{\theta}_u)$ with $\hat{\phi}_u \sim \mathcal{U}(\phi_u - 2^\circ, \phi_u + 2^\circ)$ and $\hat{\theta}_u \sim \mathcal{U}(\theta_u - 2^\circ, \theta_u + 2^\circ)$ corresponding to a dictionary with 4° angular resolution. In our simulations, we consider a total number of $U = 10$ users; out of these ten users we select the best subset $\mathcal{S} \subseteq \{1, \ldots, U\}$ of users through exhaustive search with varying size $|\mathcal{S}| \in [1, U]$ to avoid an impact of imperfect user scheduling. We set the noise variance equal to $\sigma_{n,u}^2 = 10^{-3}$. For Chebyshev-tapered beamsteering, we set the side-lobe level equal to the maximum of noise power and residual interference caused by the diffuse component:

$$\text{SLL}_u = \frac{1}{|\mathcal{S}| - 1} \max\left(\sigma_{n,u}^2, \frac{1}{K+1}\right), \quad (32)$$

such that the multi-user interference caused by the side-lobes is in the order of the maximum of diffuse background interference and noise power.

In Fig. 3, we show the achievable rate performance of the interference-unaware beamsteering methods and our robust/non-robust leakage-bounded angular beamformers for varying K-factors. The performance of the interference-unaware beamsteering methods (max. gain and Cheb. tapered beamsteering) is plotted for the case of perfect angular CSIT; however, their performance with imperfect CSIT of accuracy $\pm 2^\circ$ is virtually the same. The performance of leakage-bounded beamforming is shown for perfect angular CSIT as well as for imperfect angular CSIT with robust/non-robust interference leakage constraints. We observe that our leakage-bounded beamforming method can achieve better performance and can serve more users in parallel than the interference-unaware beamsteering schemes (max. gain and Cheb. tapered beamsteering). With decreasing Rician K-factor (left to right) the performance deteriorates since the residual interference over the diffuse background scattering component is not mitigated by our beamformer optimization. Still, even with the relatively low value of $K = 10$ our proposed method outperforms the interference-unaware beamsteering schemes. We furthermore observe in Fig. 3 that the performance loss due to imperfect angular CSIT can partly be compensated with our robust beamforming method as compared to the non-robust approach.

6.3 Comparison to interference-aware precoding

For the remaining simulations, we utilize the QUADRIGA channel model to test our proposed methods on more realistic channel realizations. We summarize the most important simulation parameters in Table 1, which have

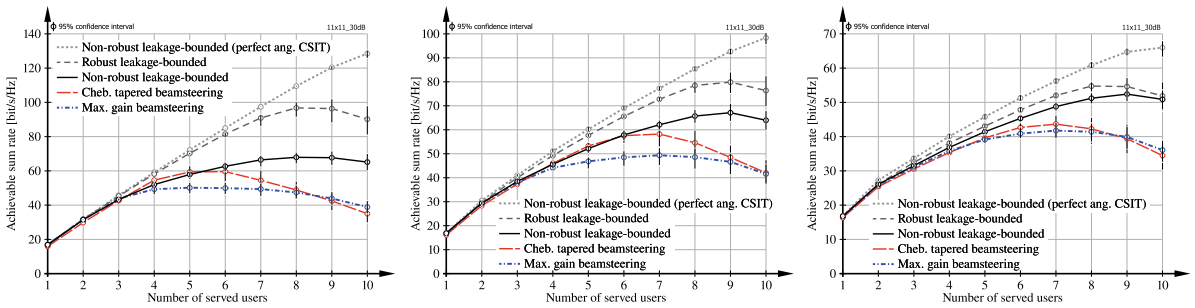


Fig. 3 Comparison of the achievable sum rate of different single-stream beamforming methods with varying K factor. **a** Achievable sum rate with $K = 10^3$. **b** Achievable sum rate with $K = 10^2$. **c** Achievable sum rate with $K = 10$

been selected to be LTE standard compliant. We consider users that are uniformly distributed within a sector of 120° with a maximum distance of 500 m from the transmitter and we assume that the users move perpendicular to the transmit antenna array bore-sight direction with 150 km/h. For such user placement, the variation of the elevation angle is in most cases very small and we therefore consider beamforming only in the azimuth domain utilizing a high-resolution horizontal uniform linear array (ULA) of 50 antenna elements. In our simulations, we do not consider macroscopic pathloss even though it is considered in the output of the QuaDRiGa channel model. We therefore normalize the channel matrices obtained from the QuaDRiGa channel model to satisfy $\mathbb{E}\left(\|\mathbf{H}_u[n]\|^2\right) = N_r N_t$. We apply this normalization since we do not consider multi-user scheduling in this section, which would be necessary in case of strong received signal power disparities of different users. We employ a dictionary/quantization codebook that is matched to the horizontal ULA and we consider an azimuthal angular resolution for CSI quantization of 1°[6]. We consider a block-fading channel model with constant channel during each transmission time interval (TTI).

Table 1 Simulation parameters employed to setup the QuaDRiGa channel model

Parameter	Value
System bandwidth	10 MHz (600 OFDM subcarriers)
Carrier frequency	$f_c = 2.1$ GHz
Transmission time interval (TTI)	1 ms
Number of OFDM symbols per TTI	14
Transmitter height	25 m
Receiver height	1.5 m
Scenario	3GPP UMa LOS/NLOS
	3GPP UMi LOS/NLOS
	BERLIN UMa LOS/NLOS
Number of receive antennas	$N_r \in \{1, 2\}$
Receive antenna model	Dipole ($N_r = 1$), x-pol ($N_r = 2$)
Number of transmit antennas	$N_t = N_h \cdot N_v = 50 \cdot 1$
Transmit antenna model	3GPP-3D
Antenna element spacing	Half-wavelength
User placement	Uniform within 120° sector
	with maximum distance 500 m
User movement	Linear track perpendicular to
	antenna boresight direction
User speed	150 km/h

In our first simulation of this section, we investigate the compressibility of the per-subcarrier channel matrices $\mathbf{H}_u[n]$ in the angular domain. That is, we apply the OMP algorithm to decompose $\mathbf{H}_u[n]$ in the angular domain using a dictionary that is matched to the employed ULA and we determine the average norm of the expansion coefficient vectors $\left\|\mathbf{c}_u^s[n]\right\|^2$ of the decomposition. The corresponding results are shown in Fig. 4 for different simulation scenarios as specified in the QuaDRiGa channel model. We observe that in the LOS situations, the first expansion coefficient contains the majority of the channel energy, especially in the 3GPP compliant scenarios; this expansion coefficient corresponds to the LOS direction. The second largest coefficient lies already approximately an order of magnitude below the first coefficient. In the non-line of sight (NLOS) scenarios, however, the channel energy distributes more equally over many expansion coefficients. Correspondingly, compressibility with the employed dictionary is worse. In the remaining simulations we therefore consider LOS transmission and employ the 3GPP compliant UMa scenario. We also conducted simulations in NLOS situations, where our proposed method however does not achieve a significant gain over the interference-aware benchmark precoders.

In the next simulation, we compare the achievable transmission rate of the interference-aware precoding schemes summarized in Section 5.2 to our robust leakage-bounded

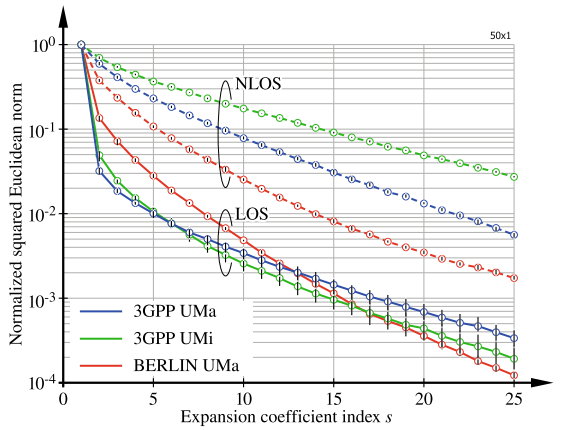


Fig. 4 Angular-domain compressibility of the channel matrices obtained with the QuaDRiGa channel model. Blue: 3GPP UMa; green: 3GPP UMi; red: BERLIN UMa

precoding method. We consider transmission to $U = 5$ users in parallel with $L_u = 2$ streams per user; hence, the total number of spatially multiplexed data streams is ten. We have seen above that in the LOS scenario most channel energy is contained in a single specular component. To efficiently support transmission of two spatial streams per user, we therefore employ a distributed antenna system (DAS) composed of two remote radio heads (RRHs) that are separated by 25 m. Each RRH radiates the same transmit signal; that is, the radio heads are fed with the same RF signal, e.g., via radio-over-fiber connections. Correspondingly, the two RRHs do not appear as independently steerable antenna arrays since they are fed with the same transmit signal; hence, the effective number of actively steerable transmit antennas (and thus the size of the channel matrix that must be estimated) is still $N_t = 50$. In that way, we guarantee that each user observes two strong specular components of approximately equal strength corresponding to the two antenna arrays of the RRHs.

We assume that the receivers estimate and feed back the CSI only at the first TTI, to enable calculation of the precoders; these precoder are then utilized for the entire transmission duration of 150 TTIs. For BD precoding and hybrid BD precoding, we assume perfect unquantized CSIT of the (effective) base band channel at the first TTI. For angle of departure BD precoding, we assume perfect knowledge of the LOS angles w.r.t. the two RRHs and hence perfect knowledge of the corresponding transmit antenna array response vectors. For the other schemes, we utilize our OMP-based channel decomposition to determine the $S = 2$ strongest specular components, in order to facilitate transmission of two spatial streams. We have observed that the performance does not improve if we consider $S > L_u$ expansion coefficients; when S is too large, the performance rather deteriorates since too many interference leakage constraints are active. For BD precoding and its hybrid variant, we assume perfect base band CSIT on each of the 600 subcarriers within the first TTI. For angular BD precoding, only broadband feedback of the angles of departure of the strongest multipath components needs to be provided, since this CSI is obtained from the time domain channel impulse response; notice, though, that this either requires estimation of the time domain channel impulse response or direction estimation of the dominant angles from uplink training signals [74]. For the remaining schemes, we consider a CSI feedback resolution of 60 subcarriers, i.e., the $S = 2$ strongest specular components are provided for every 60th subcarrier. Therefore, to estimate the full channel matrix at the receiver for CSI feedback calculation, we require a total of $600/60 \cdot N_t = 500$ orthogonal pilot-symbols within the first TTI, amounting to an overhead of $500/(14 \cdot 600) \approx 6\,\%$ within the first TTI. However, since we provide CSI feedback only every 150th TTI the pilot overhead for CSI feedback calculation is negligible.

For our robust leakage-bounded precoders, we determine the size of the angular uncertainty regions to match the uncertainty of the LOS direction caused by the user movement. To avoid unnecessarily large uncertainty regions, we recalculate the robust leakage-bounded precoders every 25 TTIs with correspondingly increasing angular uncertainty regions. Notice, in this recalculation, only the size of the uncertainty regions is updated and not the angular CSIT itself; it does therefore not require renewed CSI feedback from the users. In Fig. 5, we show the achievable rate performance of the considered precoding schemes. We observe that regular BD precoding as well as hybrid BD precoding achieve very high data

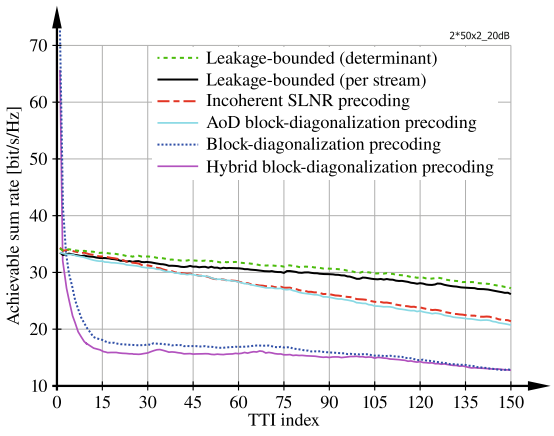


Fig. 5 Time-evolution of the achievable rate of different precoding schemes assuming constant precoders over time and time-variant channels

rate in the first TTI since at this time the CSIT is perfect and thus interference-free transmission is achieved; notice, we assume a noise variance of $\sigma_{n,u}^2 = 10^{-2}$ in this simulation. However, the performance of BD precoding and its hybrid variant deteriorates very quickly over time due to increasing residual interference caused by growing precoder mismatch, since the base band channel varies quickly over time. Our proposed leakage-bounded precoding scheme cannot compete with BD precoding in the first TTI. This is because it does not achieve interference-free transmission, since there is always residual interference caused by the channel decomposition error. Even with S very large and therefore a small channel decomposition error, our method does not achieve the multiplexing capabilities of BD precoding because it is an incoherent scheme that does not make use of coherent multipath interference, as BD precoding does to eliminate multi-user interference. Yet, our method is much less sensitive to the temporal variation of the channel matrices, since the employed angular channel decomposition is comparatively stable over time. This means that the angular directions of the largest channel expansion coefficients change only slowly over time and therefore the leakage-based precoding solution performs well even after many TTIs. Notice, the incoherent SLNR precoder as well as the angle of departure BD precoder achieve virtually the same performance as our scheme at the beginning of the transmission; however, their performance deteriorates faster because they do not consider angular uncertainty regions. Nevertheless, both methods provide a reasonably good trade-off between robustness and computational complexity. We did attempt to incorporate angular uncertainty regions in the SLNR metric, but we were not successful. In Fig. 5, we also compare the performance of leakage-bounded precoding utilizing the determinant optimization approach and the per stream optimization approach described in Section 4.2. We observe that per stream optimization exhibits only a small rate degradation compared to determinant optimization and therefore might be practically more relevant due to lower computational complexity.

In our final example, we investigate the sensitivity of leakage-bounded beamforming and BD precoding with respect to channel estimation errors. For this purpose, we distort the output $\mathbf{H}_u[n]$ of the QUADRIGA channel model to emulate a channel estimation error:

$$\hat{\mathbf{H}}_u[n] = \sqrt{1-\sigma_e^2}\,\mathbf{H}_u[n] + \sigma_e^2\,\mathbf{E}_u[n], \tag{33}$$

with $[\mathbf{E}_u[n]]_{i,j} \sim \mathcal{CN}(0,1)$ denoting the normalized channel estimation error and σ_e^2 controlling its relative strength, such that $\mathbb{E}\left(\left\|\hat{\mathbf{H}}_u[n]\right\|^2\right) = N_r N_t$[7]. We consider uncoded 16 QAM transmission to a varying number of users $U \in \{3,5,7\}$ with $L_u = 1$ stream per user. We use high-resolution angular quantization of a single $S = 1$ specular component, such that the CSI uncertainty of this single expansion vector is only due to the channel estimation error.

In Fig. 6, we show the obtained SER as a function of the channel estimation error variance σ_e^2. We observe that leakage-bounded beamforming exhibits an SER floor due to residual interference caused by the channel decomposition error, which grows with the number of users served in parallel. Yet, the performance of the method is hardly impacted by the channel estimation error; this robustness is achieved by the signal denoising properties of the employed angular basis decomposition. Only for very large channel estimation errors the decomposition fails due to the growing impact of $\mathbf{E}_u[n]$. BD precoding, on the other hand, is much more sensitive w.r.t. channel estimation errors, since such errors directly cause residual inter-user interference; yet, for small error variance it outperforms our proposed method. We want to stress that the performance shown in (8) is for uncoded transmission; with channel coding, the SER can of course be substantially reduced.

7 Conclusions

Multi-user beamforming and precoding approaches are in general very sensitive to CSI imperfections at the transmitter. Most existing approaches are not efficiently applicable in high-mobility situations, since channel matrices change so quickly over time that already small processing delays can lead to substantial inter-user interference. In this paper, we provide a remedy to this problem by considering non-coherent beamforming/precoding in the

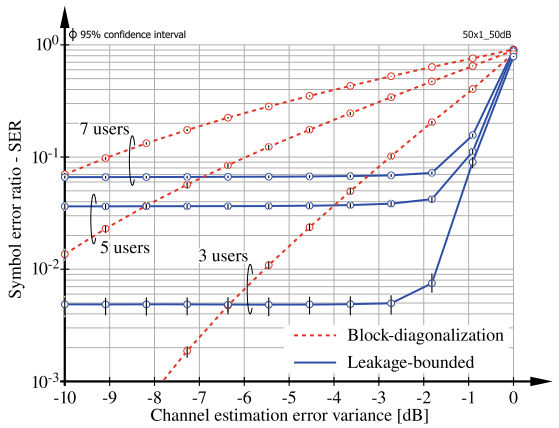


Fig. 6 Sensitivity of the SER achieved with BD precoding and leakage-bounded beamforming with respect to the channel estimation error for uncoded 16 QAM transmission. Red: Block diagonalization; blue: leakage-bounded

angular domain with robustness w.r.t. angular uncertainty. In addition to improving the robustness of the beamformer/precoder, this has the further advantage that only a minimal amount of CSIT is required, namely the angles of departure of the most significant specular components together with their respective gains as obtained from a low-rank basis decomposition. We show that the relevant CSI can efficiently be quantized by applying a variant of the OMP algorithm. Naturally, under ideal circumstances of perfect CSIT, such a non-coherent approach cannot compete with coherent methods in terms of the achievable multiplexing gain. Also, it is not well suited for strong scattering environments with a large number of significant specular components because the low-rank channel decomposition has a large residual error in such cases. Nevertheless, the proposed technique can provide substantial improvements in situations with few dominant specular components, as for example expected in the millimeter wave regime, and it enables robust multi-user beamforming/precoding with low requirements on CSIT accuracy.

Endnotes

[1] A basis of a vector space is said to be over-complete if it is complete even after the removal of a vector from the basis.

[2] A related low-rank basis expansion model has recently also been investigated in [78]; however, in [78] the authors consider an orthogonal basis expansion model, whereas we consider an over-complete basis set containing far more entries $|\mathcal{D}|$ than the dimension N_t of the vector space.

[3] We investigate a related max-min problem that accounts for uncertainty in the signal direction in [39]; it can only be approximately solved via SDR.

[4] The simulation results were generated using QuaDRiGa Version 1.4.8-571.

[5] Notice, the antenna boresight direction is $(\phi,\theta) = (0°, 90°)$.

[6] Notice, for multi-user beamforming with a single stream per user we investigate the impact of the angular CSI resolution in [40] and demonstrate that an angular resolution of 4° achieves close to optimal performance for a UPA of size $10 \cdot 10$. Since the spatial resolution of the 50 element ULA considered here is better, we increase the angular CSI resolution to 1°.

[7] The definition of σ_e^2 here is slightly different as compared to (8), in order to keep the energy of the estimated channel constant irrespective of σ_e^2.

Appendix

Upper bound on the approximation $\hat{\tilde{R}}_u$

In this appendix we derive an upper bound on the approximation $\hat{\tilde{R}}_u$ of (24) that can be calculated from the provided CSI feedback. We start by applying Sylvester's determinant theorem and Jensen's inequality to (24):

$$\tilde{R}_u \leq \log \left| \sigma_{n,u}^2 \mathbf{I}_{N_t} + \mathbb{E}\left(\mathbf{H}_u^{\mathrm{H}} \mathbf{H}_u \sum_{j \in \mathcal{S}} \mathbf{F}_j \mathbf{F}_j^{\mathrm{H}} \right) \right| \tag{34}$$

Let us next consider the approximation of a single interference term in detail:

$$\begin{aligned} \mathbb{E}(\mathbf{H}_u^{\mathrm{H}} \mathbf{H}_u \mathbf{F}_j \mathbf{F}_j^{\mathrm{H}}) &\approx \mathbb{E}\left(\sum_{s=1}^{S} \sum_{k=1}^{S} \left(\mathbf{d}_u^k\right)^* \left(\mathbf{c}_u^k\right)^{\mathrm{H}} \mathbf{c}_u^s \left(\mathbf{d}_u^s\right)^{\mathrm{T}} \mathbf{C}_j \right) \\ &= \mathbb{E}\left(\sum_{s=1}^{S} \left\| \mathbf{c}_u^s \right\|^2 \left(\mathbf{d}_u^s\right)^* \left(\mathbf{d}_u^s\right)^{\mathrm{T}} \mathbf{C}_j \right), \end{aligned} \tag{35}$$

with $\mathbf{C}_j = \mathbf{F}_j \mathbf{F}_j^{\mathrm{H}}$ being the transmit covariance matrix. In this equation, we again utilized the phase independence of different specular components as in the steps from (19) to (20). We next apply an eigen-decomposition to the sum of the specular components:

$$\sum_{s=1}^{S} \left\| \mathbf{c}_u^s \right\|^2 (\mathbf{d}_u^s)^* (\mathbf{d}_u^s)^{\mathrm{T}} = \mathbf{U}_u \mathbf{\Sigma}_u^2 \mathbf{U}_u^{\mathrm{H}}, \tag{36}$$

$$\mathbf{U}_u \in \mathbb{C}^{N_t \times S}, \mathbf{U}_u^{\mathrm{H}} \mathbf{U}_u = \mathbf{I}_S, \ \mathbf{\Sigma}_u^2 \in \mathbb{R}^{S \times S}, \tag{37}$$

where we assume that $S \leq N_t$ as our focus is on large-scale antenna arrays and low-rank channel decomposition.

Let us next consider the interference leakage power caused by the transmission to user j w.r.t. all specular components of user u:

$$\begin{aligned} L_{u,j} &= \sum_{s=1}^{S} L_{u,j}^s = \sum_{s=1}^{S} \mathrm{tr}\left(\mathbf{c}_u^s (\mathbf{d}_u^s)^{\mathrm{T}} \mathbf{C}_j (\mathbf{d}_u^s)^* (\mathbf{c}_u^s)^{\mathrm{H}} \right) \\ &= \sum_{s=1}^{S} \left\| \mathbf{c}_u^s \right\|^2 (\mathbf{d}_u^s)^{\mathrm{T}} \mathbf{C}_j (\mathbf{d}_u^s)^* = \mathrm{tr}\left(\mathbf{U}_u \mathbf{\Sigma}_u^2 \mathbf{U}_u^{\mathrm{H}} \mathbf{C}_j \right). \end{aligned} \tag{38}$$

In the precoder optimization problem (P5), we restrict this interference leakage to satisfy $L_{u,j}^s \leq L_{\max}$ and hence $L_{u,j} \leq S L_{\max}$. From this upper bound on the interference leakage power, we can determine an upper bound on the interference leakage covariance (35). For that purpose, we decompose the precoding matrix in terms of the S dimensional basis $\mathbf{U}_u$ and a basis $\mathbf{U}_u^{\perp}$ of its $N_t - S$ dimensional orthogonal complement:

$$\mathbf{F}_j = \mathbf{U}_u \mathbf{F}_j^{\parallel} + \mathbf{U}_u^{\perp} \mathbf{F}_j^{\perp}, \tag{39}$$

$$\mathbf{F}_j^{\parallel} = \mathbf{U}_u^{\mathrm{H}} \mathbf{F}_j,\ \mathbf{F}_j^{\perp} = \left(\mathbf{U}_u^{\perp}\right)^{\mathrm{H}} \mathbf{F}_j. \tag{40}$$

With this, we can reformulate the leakage power as:

$$L_{u,j} = \mathrm{tr}\left(\mathbf{\Sigma}_u^2 \mathbf{F}_j^{\parallel} \left(\mathbf{F}_j^{\parallel}\right)^{\mathrm{H}}\right) \leq SL_{\max}. \tag{41}$$

We next assume $\mathbf{F}_j^{\parallel} \in \mathbb{C}^{S \times L_j}$ to be isotropically distributed, such that in combination with (41) we get

$$\mathbb{E}\left(\mathbf{F}_j^{\parallel} \left(\mathbf{F}_j^{\parallel}\right)^{\mathrm{H}}\right) = k_j \mathbf{I}_S, \tag{42}$$

$$k_j \leq \frac{SL_{\max}}{\mathrm{tr}\left(\mathbf{\Sigma}_u^2\right)}. \tag{43}$$

Utilizing these results in (35), we obtain:

$$\mathbb{E}\left(\mathbf{U}_u \mathbf{\Sigma}_u^2 \mathbf{U}_u^{\mathrm{H}} \mathbf{C}_j\right) \preceq \frac{SL_{\max}}{tr(\mathbf{\Sigma}_u^2)} \mathbf{U}_u \mathbf{\Sigma}_u^2 \mathbf{U}_u^{\mathrm{H}}, \tag{44}$$

where $\preceq$ means that the difference between the right-hand side and left-hand side is positive semidefinite. This brings us to the desired upper bound on the approximation $\hat{\tilde{R}}_u$:

$$\tilde{R}_u \approx \hat{\tilde{R}}_u \leq \log\left|\sigma_{n,u}^2 \mathbf{I}_S + (|\mathcal{S}| - 1)\frac{SL_{\max}}{\mathrm{tr}\left(\mathbf{\Sigma}_u^2\right)} \mathbf{\Sigma}_u^2 + \mathbf{\Sigma}_u^2 \mathbf{U}_u^{\mathrm{H}} \mathbf{C}_u \mathbf{U}_u\right|, \tag{45}$$

where we again applied Sylvester's determinant theorem.

Abbreviations

3GPP: Third generation partnership project; AD: Analog to digital; BD: Block diagonalization; BS: Base station; CoMP: Coordinated multipoint transmission; CSI: Channel state information; CSIT: Channel state information at the transmitter; DAS: Distributed antenna system; DA: Digital to analog; DoF: Degrees of freedom; FD: Full-dimension; FDD: Frequency division duplex; FD-MIMO: Full-dimension multiple-input multiple-output; FRT: Fluctuating two-ray; GRT: Generalized two-ray; IA: Interference alignment; i.i.d.: Independent and identically distributed; ITS: Intelligent transport systems; LOS: Line of sight; LTE: Long term evolution; MBMS: Multimedia broadcast multicast service; eMBMS: Evolved MBMS; MC-IC: Multicast interference channel; MIMO: Multiple-input multiple-output; MISO: Multiple-input single-output; NLOS: Non-line of sight; NP: Non-deterministic polynomial-time; NR: New radio; OFDM: Orthogonal frequency division multiplexing; OMP: Orthogonal matching pursuit; QAM: Quadrature amplitude modulation; RF: Radio frequency; RRH: Remote radio head; SDMA: Space-division multiple access; SDR: Semidefinite programming relaxation; SER: Symbol error ratio; SINR: Signal to interference and noise ratio; SLNR: Signal to leakage and noise ratio; SLR: Signal to leakage ratio; SNR: Signal to noise ratio; SVD: Singular value decomposition; TDD: Time division duplex; TTI: Transmission time interval; TWDP: Two-wave diffuse power; UC-IC: Unicast interference channel; UE: User equipment; UMTS: Universal mobile telecommunications system; ULA: Uniform linear array; UPA: Uniform planar array; ZF: Zero forcing

Acknowledgements

The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged. The author acknowledges the TU Wien University Library for the financial support through its Open Access Funding Program.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Authors' contributions

All contributions are from the sole author SS. The author read and approved the final manuscript.

Authors' information

Not applicable.

Competing interests

The author declares that he has no competing interests.

Publisher's Note

Received: 9 August 2017 Accepted: 23 February 2018
Published online: 14 March 2018

References

1. H Yang, T Marzetta, Performance of conjugate and zero-forcing beamforming in large-scale antenna systems. IEEE J. Sel. Areas Commun. **31**(2), 172–179 (2013)
2. Y Kim, H Ji, J Lee, YH Nam, BL Ng, I Tzanidis, Y Li, J Zhang, Full dimension MIMO (FD-MIMO): The next evolution of MIMO in LTE systems. IEEE Wirel. Commun. **21**(3), 92–100 (2014)
3. YH Nam, MS Rahman, Y Li, G Xu, E Onggosanusi, J Zhang, JY Seol, in *Information Theory and Applications Workshop (ITA)*. Full dimension MIMO for LTE-Advanced and 5G (IEEE, San Diego, 2015), pp. 143–148
4. M Kim, YH Lee, MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels. IEEE Trans. Veh. Technol. **64**(6), 2714–2720 (2015)
5. F Sohrabi, W Yu, Hybrid digital and analog beamforming design for large-scale antenna arrays. IEEE J. Sel. Topics Signal Process. **10**(3), 501–513 (2016)
6. R Rajashekar, L Hanzo, Hybrid beamforming in mm-Wave MIMO systems having a finite input alphabet. IEEE Trans. Commun. **64**(8), 3337–3349 (2016)
7. A Liu, VKN Lau, Impact of CSI knowledge on the codebook-based hybrid beamforming in massive MIMO. IEEE Trans. Signal Process. **64**(24), 6545–6556 (2016)
8. H Wang, W Wang, VKN Lau, Z Zhang, Hybrid limited feedback in 5G cellular systems with massive MIMO. IEEE Syst. J. **11**(1), 50–61 (2017)
9. C Rusu, R Méndez-Rial, N González-Prelcic, RW Heath, Low complexity hybrid precoding strategies for millimeter wave communication systems. IEEE Trans. Wirel. Commun. **15**(12), 8380–8393 (2016)
10. QUA Nadeem, A Kammoun, M Debbah, MS Alouini, Design of 5G full dimension massive MIMO systems. IEEE Trans. Commun. **PP**(99), 1–1 (2017)
11. N Li, Z Wei, H Yang, X Zhang, D Yang, Hybrid precoding for mmWave massive MIMO systems with partially connected structure. IEEE Access. **5**, 2–15151 (1514)
12. O Orhan, E Erkip, S Rangan, in *2015 Information Theory and Applications Workshop (ITA)*. Low power analog-to-digital conversion in millimeter wave systems: impact of resolution and bandwidth on performance (IEEE, San Diego, 2015), pp. 191–198
13. S Jacobsson, G Durisi, M Coldrey, U Gustavsson, C Studer, Throughput analysis of massive MIMO uplink with low-resolution ADCs. IEEE Trans. Wirel. Commun. **16**(6), 4038–4051 (2017)
14. WB Abbas, F Gomez-Cuba, M Zorzi, Millimeter wave receiver efficiency: a comprehensive comparison of beamforming schemes with low resolution ADCs. IEEE Trans. Wirel. Commun. **16**(12), 8131–8146 (2017)
15. K Roth, JA Nossek, Achievable rate and energy efficiency of hybrid and digital beamforming receivers with low resolution ADC. IEEE J. Sel. Areas Commun. **35**(9), 2056–2068 (2017)
16. TE Bogale, LB Le, Massive MIMO and mmWave for 5G wireless HetNet: potential benefits and challenges. IEEE Veh. Technol. Mag. **11**(1), 64–75 (2016)

17. LN Ribeiro, S Schwarz, M Rupp, ALF de Almeida, Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs. CoRR. **abs/1709.05139** (2017)
18. J Zuo, J Zhang, C Yuen, W Jiang, W Luo, Multicell multiuser massive MIMO transmission with downlink training and pilot contamination precoding. IEEE Trans. Veh. Technol. **65**(8), 6301–6314 (2016)
19. Y Han, J Lee, Uplink pilot design for multi-cell massive MIMO networks. IEEE Commun. Lett. **20**(8), 1619–1622 (2016)
20. S Schwarz, R Heath, M Jr. Rupp, Single-user MIMO versus multi-user MIMO in distributed antenna systems with limited feedback. EURASIP J. Adv. Sig. Proc. **2013**(1), 54 (2013)
21. W Shen, L Dai, Y Shi, B Shim, Z Wang, Joint channel training and feedback for FDD massive MIMO systems. IEEE Trans. Veh. Technol. **65**(10), 8762–8767 (2016)
22. MS Sim, J Park, CB Chae, RW Heath, Compressed channel feedback for correlated massive MIMO systems. J. Commun. Netw. **18**(1), 95–104 (2016)
23. S Schwarz, T Philosof, M Rupp, Signal processing challenges in cellular assisted vehicular communications. IEEE Signal Process. Mag. **34**(2), 47–59 (2017)
24. MA Maddah-Ali, D Tse, Completely stale transmitter channel state information is still very useful. IEEE Trans. Inf. Theory. **58**(7), 4418–4431 (2012)
25. S Yang, M Kobayashi, D Gesbert, X Yi, Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT. IEEE Trans. Inf. Theory. **59**(1), 315–328 (2013)
26. J Chen, P Elia, Toward the performance versus feedback tradeoff for the two-user MISO broadcast channel. IEEE Trans. Inf. Theory. **59**(12), 8336–8356 (2013)
27. H Joudeh, B Clerckx, Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach. IEEE Trans. Commun. **64**(11), 4847–4861 (2016)
28. C Peel, B Hochwald, A Swindlehurst, A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization. IEEE Trans. Commun. **53**(1), 195–202 (2005)
29. Q Spencer, A Swindlehurst, M Haardt, Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. IEEE Trans. Signal Process. **52**(2), 461–471 (2004)
30. V Cadambe, S Jafar, Interference alignment and degrees of freedom of the K-user interference channel. IEEE Trans. Inf. Theory. **54**(8), 3425–3441 (2008)
31. M Maddah-Ali, A Motahari, A Khandani, Communication over MIMO X channels: interference alignment, decomposition, and performance analysis. IEEE Trans. Inf. Theory. **54**(8), 3457–3470 (2008)
32. J Zhang, RW Heath Kountouris, JG Andrews, Mode switching for the multi-antenna broadcast channel based on delay and channel quantization. EURASIP J. Adv. Signal Process. **2009**(1), 802548 (2009)
33. S Schwarz, R Heath, Rupp Jr M, in *38th International Conference on Acoustics, Speech and Signal Processing*. Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems (IEEE, Vancouver, 2013), pp. 5021–5025
34. SA Vorobyov, Principles of minimum variance robust adaptive beamforming design.Signal Process. **93**(12), 3264–3277 (2013). special Issue on Advances in Sensor Array Processing in Memory of Alex B. Gershman
35. WS Chua, C Yuen, YL Guan, F Chin, Robust multi-antenna multi-user precoding based on generalized multi-unitary decomposition with partial CSI feedback. IEEE Trans. Veh. Technol. **62**(2), 596–605 (2013)
36. S Schwarz, M Rupp, Evaluation of distributed multi-user MIMO-OFDM with limited feedback. IEEE Trans. Wirel. Commun. **13**(11), 6081–6094 (2014)
37. J Wang, M Bengtsson, B Ottersten, DP Palomar, Robust MIMO precoding for several classes of channel uncertainty. IEEE Trans. Signal Process. **61**(12), 3056–3070 (2013)
38. F Khalid Speidel, Robust hybrid precoding for multiuser MIMO wireless communication systems. IEEE Trans. Wirel. Commun. **13**(6), 3353–3363 (2014)
39. S Schwarz, T Philosof, M Rupp, in *50th Asilomar Conference on Signals Systems and Computers*. Convex-optimization based geometric beamforming for FD-MIMO arrays (IEEE, Asilomar, 2016)
40. S Schwarz, M Rupp, in *50th Asilomar Conference on Signals Systems and Computers*. Limited feedback based double-sided full-dimension MIMO for mobile backhauling (IEEE, Asilomar, 2016)
41. P Almers, E Bonek, A Burr, N Czink, M Debbah, V Degli-Esposti, H Hofstetter, P Kyösti, D Laurenson, G Matz, A Molisch, C Oestges, H Özcelik, Survey of channel and radio propagation models for wireless MIMO systems. EURASIP J. Wirel. Commun. Netw. **2007**(019070), 19 (2007)
42. H Özcelik, *Indoor MIMO channel models, 2004*. (Wien, Technische Universität, Dissertation, 2005). http://resolver.obvsg.at/urn:nbn:at:at-ubtuw:1-17999
43. CA Balanis, *Antenna theory: analysis and design*, 3rd. (Wiley, New Jersey, 2005)
44. H Ji, Y Kim, J Lee, E Onggosanusi, Y Nam, J Zhang, B Lee, B Shim, Overview of full-dimension MIMO in LTE-advanced pro. IEEE Commun. Mag. **55**(2), 176–184 (2017)
45. J Choi, V Va, N Gonzalez-Prelcic, R Daniels, CR Bhat, RW Heath, Millimeter-wave vehicular communication to support massive automotive sensing. IEEE Commun. Mag. **54**(12), 160–167 (2016)
46. GD Durgin, TS Rappaport, DA de Wolf, New analytical models and probability density functions for fading in wireless communications. IEEE Trans. Commun. **50**(6), 1005–1015 (2002)
47. M Rao, FJ Lopez-Martinez, A Goldsmith, in *2014 48th Annual Conference on Information Sciences and Systems*. Statistics and system performance metrics for the two wave with diffuse power fading model (IEEE, Princeton, 2014), pp. 1–6
48. M Rao, FJ Lopez-Martinez, MS Alouini, A Goldsmith, MGF approach to the analysis of generalized two-ray fading models. IEEE Trans. Wirel. Commun. **14**(5), 2548–2561 (2015)
49. JM Romero-Jerez, FJ Lopez-Martinez, JF Paris, A Goldsmith, in *2016 IEEE Globecom Workshops*. The fluctuating two-ray fading model for mmWave communications (IEEE, Washington, 2016), pp. 1–6
50. S Schwarz, Outage investigation of beamforming over random-phase finite-scatterer MISO channels. IEEE Signal Process. **24**(7), 1029–1033 (2017)
51. J Zhang, W Zeng, X Li, Q Sun, KP Peppas, New results on the fluctuating two-ray model with arbitrary fading parameters and its applications. IEEE Trans. Veh. Technol. **PP**(99), 1–1 (2017)
52. E Dahlman, S Parkvall, J Sköld, *4G LTE/LTE-advanced for mobile broadband*. (Elsevier Academic Press, Oxford, 2011)
53. O El Ayach, R Heath Jr., Grassmannian differential limited feedback for interference alignment. IEEE Trans. Signal Process. **60**(12), 6481–6494 (2012)
54. YC Eldar, G Kutyniok, *Compressed sensing: theory and applications*. (Cambridge Univ. Press, Cambridge, 2012)
55. YC Pati, R Rezaiifar, PS Krishnaprasad, in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, (1993), pp. 40–44
56. Z-Q Luo, W-K Ma, A-C So, Y Ye, S Zhang, Semidefinite relaxation of quadratic optimization problems. IEEE Signal Process. Mag. **27**(3), 20–34 (2010)
57. Z Luo, ND Sidiropoulos, P Tseng, S Zhang, Approximation bounds for quadratic optimization with homogeneous quadratic constraints. SIAM J. Optim. **18**(1), 1–28 (2007)
58. T-H Chang, Z-Q Luo, C-Y Chi, Approximation bounds for semidefinite relaxation of max-min-fair multicast transmit beamforming problem. IEEE Trans. Signal Process. **56**(8), 3932–3943 (2008)
59. R Zhang, YC Liang, Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks. IEEE J. Sel. Topics Signal Process. **2**(1), 88–102 (2008)
60. Y Nesterov, A Nemirovski, *Interior-point polynomial algorithms in convex programming: theory and applications*. (Philadelphia: Society for Industrial and Applied Mathematics, SIAM, 1994)
61. MS Lobo, L Vandenberghe, S Boyd, H Lebret, Applications of second-order cone programming. Linear Algebra Appl. **284**(1), 193–228 (1998). international Linear Algebra Society (ILAS) Symposium on Fast Algorithms for Control, Signals and Image Processing
62. R Zhang, S Cui, Cooperative interference management with MISO beamforming. IEEE Trans. Signal Process. **58**(10), 5450–5458 (2010)
63. S Schwarz, M Rupp, Transmit optimization for the MISO multicast interference channel. IEEE Trans. Commun. **63**(12), 4936–4949 (2015)
64. Z-Q Luo, T-H Chang, in *Convex Optimization in Signal Processing and Communications*, ed. by D Palomar, Y Eldar. SDP relaxation of

homogeneous quadratic optimization: approximation and applications (Cambridge University Press, 2010)
65. F Alizadeh, D Goldfarb, Second-order cone programming. Math Program. **95**, 3–51 (2001)
66. C Wang, D Sun, K-C Toh, Solving log-determinant optimization problems by a newton-cg primal proximal point algorithm. SIAM J. Optim. **20**(6), 2994–3013 (2010)
67. YP Lin, SH Tsai, in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (Thic structures for rf beamforming, 2017), pp. 1–5
68. Y Lee, Y Nam, J So, in *2012 18th Asia-Pacific Conference on Communications (APCC)*. Optimal user selection algorithm for opportunistic space division multiple access systems (IEEE, Jeju Island, 2012), pp. 922–923
69. HL Van Trees, *Detection, estimation, and modulation theory. Part IV., Optimum array processing.* (Wiley-Interscience, New York, 2002)
70. F-I Tseng, DK Cheng, Optimum scannable planar arrays with an invariant sidelobe level. Proc IEEE. **56**(11), 1771–1778 (1968)
71. N Jindal, MIMO broadcast channels with finite-rate feedback. IEEE Trans. Inf. Theory. **52**(11), 5 (2006)
72. N Ravindran, N Jindal, Limited feedback-based block diagonalization for the MIMO broadcast channel. IEEE J. Sel. Areas Commun. **26**(8), 1473–1482 (2008)
73. S Schwarz, M Rupp, Subspace quantization based combining for limited feedback block-diagonalization. IEEE Trans. Wirel. Commun. **12**(11), 5868–5879 (2013)
74. X Zhang, J Tadrous, E Everett, F Xue, A Sabharwal, in *2015 49th Asilomar Conference on Signals, Systems and Computers*. Angle-of-arrival based beamforming for FDD massive MIMO, (2015), pp. 704–708
75. M Sadek, A Tarighat, A Sayed, A leakage-based precoding scheme for downlink multi-user MIMO channels. IEEE Trans. Wirel. Commun. **6**(5), 1711–1721 (2007)
76. S Jaeckel, L Raschkowski, K Börner, L Thiele, Quadriga: A 3-D multi-cell channel model with time evolution for enabling virtual field trials. IEEE Trans. Antennas Propag. **62**(6), 3242–3256 (2014)
77. M Grant, S Boyd, CVX: Matlab software for disciplined convex programming, version 2.1. (2014). http://cvxr.com/cvx
78. H Xie, F Gao, S Zhang, S Jin, A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model. IEEE Trans. Veh. Technol. **66**(4), 3170–3184 (2017)

4. Multi-User Resource Allocation and Robust Transmission Optimization

CSIT allows to optimize the wireless transmissions to accommodate the demands and requirements of mobile network users. From a network operator standpoint, it is important to employ the available resources as efficiently as possible, in order to maximize the network utilization and to fully capitalize on the operational network infrastructure. This implies optimizing the multi-user resource allocation, with respect to efficiency, fairness and user satisfaction, as well as, the spatial signature of the transmit signals through MIMO beamforming and precoding.

CSIT in reality though is never perfect. Especially in high mobility scenarios, CSIT imperfections are the consequence of increased channel estimation errors [61], degraded performance of differential/predictive CSI feedback algorithms [100], as well as, significant channel aging during the CSIT acquisition phase [70]. These effects impair the spectral efficiency in particular of MIMO transmissions and increase the outage probability of the communication system. To compensate for such performance degradation, robust transceiver designs are necessary that balance the strive for efficiency with the risk of signal outages.

Chapter Outline

In this chapter, I discuss methods for utilizing CSIT to optimize the efficiency, robustness and QoS experience of multi-user transmissions. I first focus on multi-user scheduling algorithms, i.e. the space-time-frequency allocation of the available resources amongst UEs, to exploit multi-user diversity, achieve fairness amongst UEs and to satisfy QoS requirements. I then discuss MIMO transmission with imperfect CSIT, putting the scope on robust worst-case optimization and outage-based beamformer/precoder design.

4.1. Multi-User Scheduling and Admission Control

Multi-user scheduling is a central means for exploiting the multi-user channel diversity in order to enhance the spectral efficiency and the reliability of the data transmissions, as well as, for satisfying QoS requirements of the users. It is well-known that under iid Gaussian channel fading conditions of multiple users, the multi-user diversity gain can provide a double-logarithmic improvement of the sum rate with the number of users served by the multi-user system, by scheduling users on favorable time-frequency resources [129]. However, practical mobile communication systems cannot only be designed for maximizing the system sum rate; they rather have to strike a balance between maximizing the system efficiency and satisfying the QoS requirements of their users.

Key issues of scheduling and admission control: The following key issues have to be addressed by the multi-user scheduling and admission control algorithms of the communication system, to enable efficient, satisfactory and dependable operation:

- Trading-off spectral efficiency for fairness of the resource allocation.
- Satisfying strict QoS requirements of inelastic traffic.
- Controlling admission of applications with strict dependability requirements.

These issues are specifically relevant for novel 5G mobile networks, which are expected to support a variety of service requirements beyond best effort.

Consideration of the realistic throughput performance: Multi-user scheduling in realistic systems is complicated by the fact that the throughput of a UE does in general not grow strictly linear with the amount of allocated resources in time-frequency selective channels. We already encountered this issue in Section 3.3.1, when discussing CQI feedback calculation; there, I have shown that the non-linear MIESM function is suitable to estimate the achievable rate of BICM architectures, when transmitting over time-frequency selective channels. The non-linearity of MIESM, however, implies that scheduling optimization problems also become non-linear and, hence, difficult to solve. I provide a remedy to this issue in *[Conf. 2]*, by applying a local linearization of the BICM functions involved in MIESM; details are given in Section 4.3.

4.1.1. Spectral Efficiency versus Fairness

In mobile communications, multiple users commonly experience significant channel quality differences, due to the location-dependency of the SINR in cellular networks. This implies that scheduling strategies that purely focus on maximizing the achievable rate of the system, such as [130], mostly assign resources only to a small subset of users, which experience good channel conditions, whereas the other UEs are not served by the system. Contrarily, schedulers that attempt to balance the average throughput of all UEs, such as, $\max. \min.$ throughput scheduling [131], have to assign excessive resources to UEs with bad channel conditions, providing just little pay-off for UEs with good channel conditions. Hence, the notion of the fairness of a resource allocation has attracted significant interest by researchers.

Quantifying fairness: There exist a number of approaches to quantify the fairness of a resource allocation, the most prominent being Jain's fairness index [132]:

$$J = \frac{\left(\sum_{u=1}^{U} \bar{T}_u\right)^2}{U \sum_{u=1}^{U} \bar{T}_u^2}, \tag{4.1}$$

with $\bar{T}_u$ denoting the average throughput of UE u. This index ranges from $\frac{1}{U}$, when only one user is served, to one, when all users achieve equal throughput. In practice, often proportional fair (PF) scheduling is applied [133], where the amount of resources allocated to a UE is proportional to the average channel quality of this UE. PF scheduling is a special case of the more general Pareto-optimal class of so-called α-fair utility functions [134], for which α is a free parameter that determines the extent of fairness of the resource allocation ($\alpha = 0$ corresponds to maximum throughput scheduling, $\alpha = 1$ to proportional fair, $\alpha \to \infty$ to $\max. \min.$ scheduling). However, fixing an α does not correspond to a specific value of Jain's fairness index; this relationship depends on the channel quality of the UEs. This circumstance makes it difficult for network operators to decide for a preferred value of α.

Multi-user scheduling with adjustable fairness: In [135], I provide an alternative approach, by including Jain's fairness index explicitly in the scheduling optimization problem. Specifically, I consider specifying a target lower bound on Jain's fairness index in the scheduling problem, while at the same time attempting to maximize the sum throughput of the system. As illustrated in Figure 4.1, a lower bound on Jain's fairness index corresponds to a second-order cone constraint in the achievable rate region of the multi-user system. To efficiently determine the sum-rate optimal point within the intersection of the second-order cone fairness constraint and the achievable rate region of the users, I propose in [135] to employ α-fair utility maximization in combination with a line-search over α. A similar approach is proposed in [136], where the authors moreover consider constraints imposed by

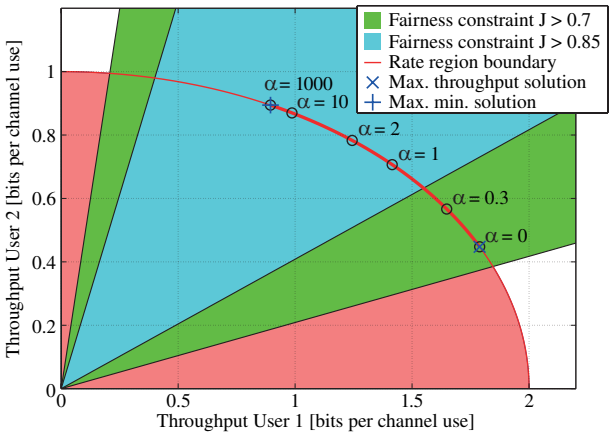


Figure 4.1.: Illustration of second-order cone fairness constraints within the hypothetical rate region of two users. Varying the parameter α of the α-fair utility functions allows to Pareto-optimally trade-off fairness for efficiency. The illustration is taken from my publication [135] © 2011 IEEE.

real-time traffic. A comprehensive survey on opportunistic scheduling strategies for wireless communications is provided in [137].

4.1.2. Handling QoS Requirements

Providing fairness to users of a mobile network is an important concern for network operators; yet, even the most fair resource allocation can still just support best effort services. Handling traffic that imposes strict QoS requirements in mobile communications is difficult, due to the often unpredictable micro- and macroscopic channel fading conditions. QoS requirements include average and minimum throughput constraints, maximum packet delay constraints for real-time traffic, minimum SINR constraints to guarantee coverage, as well as, acceptable packet loss rates. 3GPP currently employs 21 so-called QoS class identifiers (QCIs) to specify the QoS requirements of applications, ranging from voice, over video and vehicle to everything (V2X) messages, to augmented reality applications [138].

QoS-aware scheduling: To provide packet latency guarantees, reduced packet loss rate, and good fairness, the modified-largest weighted delay first and the exponential/PF algorithms have been introduced in [139, 140]. They exploit knowledge of the acceptable packet loss probability and the head of line delay of the packet buffer to achieve their targets. In [141], the authors propose a two-level scheduling algorithm, combining a frame-level scheduler, which determines the amount of data that each real-time service should transmit within a frame, with a proportional fair RE scheduler that exploits channel diversity. We propose a similar two-level approach in [142], where we initially, in the first level, schedule

real-time traffic using an exponentially weighted PF metric; afterwards, in the second level, we assign the remaining resources to non real-time traffic. The exponential weighting of the PF metric thereby accounts for the latency constraint of the respective traffic class and prioritizes data packets that are close to their deadline.

4.1.3. Admission Control under Strict Dependability Requirements

For some applications that impose strict dependability requirements, it is preferable to deny access to the communication system if these requirements cannot be satisfied, rather than granting access to an insufficiently reliable system. This is, e.g., the case in safety-critical information exchange for vehicular communications, where an unreliable connectivity cannot be tolerated. In such situations, it is better to adopt a fall-back solution (e.g., a more conservative driving-mode of vehicles), rather than facing serious problems if the undependable connection fails. It is therefore important to be able to estimate, whether or not the QoS requirements of a UE can be satisfied, given the channel conditions and the load caused by the already connected UEs.

Outage Probability of Wireless Transmissions

The issue of granting access to the communication system relates to the calculation of the outage probability of wireless transmissions, in order to gauge whether the rate requirements of the UEs can be satisfied with a sufficiently high probability. The outage probability of multi-antenna wireless transmissions is strongly impacted by the quality of the CSIT. This has been analyzed in the scientific literature under a host of assumptions:

- Single-cell single-user multiple-input single-output (MISO): In [143], the authors show that CSIT imperfections, which are treated as a Gaussian error with respect to the actual channel state, lead to a reduction of the SNR and a loss of diversity.
- Single-cell multi-user MISO: In [144], the authors calculate the outage probability under the assumption that only the channel correlation matrices are available as CSIT. The calculations in the multi-user situation are a lot more involved compared to the single-user situation, since the intended signal and the residual multi-user interference are received over the same channel and are therefore correlated.
- Multi-cell single-user MISO: In [145], the authors characterize the entire outage-region of MISO interference networks with imperfect CSIT, where again the CSIT error is treated as an additive Gaussian distortion of the actual channel state. The

authors of [145] show that the outage-region shrinks significantly with increasing CSIT error.

Outage probability of finite scatterer directional channels: In *[Article 9]*, I revisit the single-cell single-user MISO channel and calculate the outage probability under the assumption of a finite scatterer directional channel model with imperfect CSIT, similar to (3.7). The considered finite scatter directional MISO channel model is:

$$\mathbf{h}_{u,i}[n,k] = \sum_{s=1}^{S} \mathbf{h}_{u,i}^{(s)}[n,k] + \mathbf{h}_{u,i}^{(d)}[n,k], \tag{4.2}$$

where $\mathbf{h}_{u,i}^{(s)}[n,k]$ represents the channel response of the s-th specular scattering component and $\mathbf{h}_{u,i}^{(d)}[n,k]$ corresponds to diffuse background scattering, which is assumed to be iid Gaussian distributed. The channel response of the specular components is, up to the path-loss $\gamma_{u,i}^{(s)}$ and a phase shift $\varphi_{u,i}^{(s)}[n,k]$, determined by the transmit antenna array response $\mathbf{a}_i(\boldsymbol{\Psi})$ with respect to a plane wave in angular direction $\boldsymbol{\Psi}_{\boldsymbol{u,i}}^{\boldsymbol{(s)}}[k]$:

$$\mathbf{h}_{u,i}^{(s)}[n,k] = \sqrt{\gamma_{u,i}^{(s)}}\, \mathbf{a}_i\Big(\boldsymbol{\Psi}_{\boldsymbol{u,i}}^{\boldsymbol{(s)}}[n,k]\Big)\; e^{j\varphi_{u,i}^{(s)}[n,k]}. \tag{4.3}$$

I investigate the impact of imperfect knowledge of the angular directions $\boldsymbol{\Psi}_{\boldsymbol{u,i}}^{\boldsymbol{(s)}}[n,k]$ and the phases $\varphi_{u,i}^{(s)}[n,k]$ at the transmitter. I consider these two terms individually, since in a time-variant scenario, the angular directions can be assumed relatively stable, as they vary slowly with the macroscopic movement of UEs, whereas the phases vary significantly with movements in the order of the wavelength. The model also captures the CSI feedback quantization errors of the limited feedback algorithm I propose in *[Article 7]*. It moreover reduces to the simple Gaussian error model, if perfect knowledge of the specular components and no knowledge of the diffuse component is considered; the diffuse component would then correspond to the Gaussian error. The calculation of the outage probability for this channel model in closed-form is in general not feasible; one has to resort to numerical integration.

In [146], I generalize these results further, to the case of multi-cell multi-user MISO transmission. The evaluation of the outage probability here relates to the distribution of a non-central definite quadratic form in complex Gaussian random variables, whose study has a long standing history [147–151]. Again, calculations in closed-form are restricted to relatively simple special cases.

4.2. MIMO Transmission with Imperfect CSIT

MIMO techniques are a fundamental component of state-of-the-art wireless communication systems. They provide substantial improvements in spectral efficiency through spatial

multiplexing and interference mitigation, in reliability by means of diversity transmission, and in network coverage through beamforming. Novel multi-antenna transmission concepts, such as, massive MIMO, FD-MIMO and distributed/cell-free massive MIMO [65, 152, 153], continue to attract significant research interest to push the performance of wireless networks.

Yet, the promised performance of MIMO techniques is often not achieved in practice, due to CSIT imperfections, whether because of estimation errors, feedback quantization or due to channel aging. As discussed in Section 4.1.3, the CSIT imperfections impair the outage probability of wireless communications; even more, as mentioned in Chapter 3, they fundamentally affect the achievable DoF of multi-user transmissions. Below, I summarize existing theoretical results on the performance of multi-antenna transmissions with CSIT imperfections, discuss transmission schemes that provide robustness against imperfect CSIT, and tie in my own contributions to these fields.

4.2.1. DoF with Imperfect CSIT

CSIT imperfections have a significant impact on the performance of multi-user transmissions, since they are the cause of multi-user interference. Ultimately, if the variance of the residual multi-user interference σ^2_{MUI} caused by imperfect CSIT is constant, this leads to an interference limitation of the achievable transmission rate at high SNR, effectively reducing the DoF of the system to zero. To avoid this interference limitation, the multi-user interference σ^2_{MUI} has to diminish with growing SNR to achieve non-zero DoF. Specifically, the following results have been established in the scientific literature:

- Linear transceiver architectures: To achieve the optimal DoF D with linear beamforming/precoding schemes, such as, BD precoding and IA, the variance of the residual multi-user interference must scale inversely proportional to the SNR: $\sigma^2_{\text{MUI}} \propto \left(P/\sigma^2_{z,u}\right)^{-1}$ [98, 154]. If the variance scales as $\sigma^2_{\text{MUI}} \propto \left(P/\sigma^2_{z,u}\right)^{-\alpha}, \alpha \in [0,1]$, the DoF are reduced to αD [83, 155].

- Non-linear transceiver architectures: With a rate-splitting approach, which requires non-linear SIC receivers, it is possible to improve the DoF of the MISO broadcast channel (BC) with imperfect CSIT to $1 + \alpha(D-1) \geq \alpha D, \alpha \in [0,1]$ [156].

- Outdated CSIT: If the feedback delay is so large that the current channel state is not correlated to the CSIT feedback, it is still possible to achieve non-trivial DoF with the retrospective interference alignment space-time coding scheme proposed in [157]. Specifically, the DoF of the U user MISO BC with sufficiently many transmit antennas and outdated CSIT approach $U/\log(U)$ for $U \to \infty$. The DoF of the most general case of imperfect and outdated CSIT are derived in [158].

Relationship to CSIT Acquisition

The DoF results above relate to the CSIT acquisition schemes discussed in Chapter 3. Specifically, achieving the optimal DoF determines the overhead of these schemes:

- Reverse link channel estimation: The residual multi-user interference σ^2_{MUI} is directly proportional to the reverse link channel estimation error. Since the channel estimation error is determined by the density of pilot symbols in the time-frequency RE grid, as well as, their transmit powers, these parameters need to be adapted according to the required estimation accuracy. This can, e.g., be achieved by repurposing the pilot-pattern optimization method proposed in [61].
- Analog CSI feedback: This is analyzed for multi-user MIMO in [155] and for IA in [154]. The authors show that the optimal DoF can be achieved, as long as the SNR of the reverse feedback link is comparable to the SNR of the forward link.
- Digital CSI feedback: For this type of feedback, the residual multi-user interference is determined by the quantization codebook size, which grows exponentially with the number of feedback bits B. It thus follows that B must grow in proportion to the logarithm of the SNR. Specifically, assuming that the number of transmit antennas N_i at the BS is larger than the number of receive antennas M_u at the UEs:
  - BD precoding with number of streams per user $\ell_u[n,k] = M_u$ requires $B \propto (N_i - M_u)M_u \log\left(\frac{P}{\sigma^2_{z,u}}\right)$ as shown in [96].
  - BD precoding with $\ell_u[n,k] \leq M_u$ requires $B \propto (N_i - M_u)\ell_u[n,k] \log\left(\frac{P}{\sigma^2_{z,u}}\right)$ as I show in [100], which requires application of my proposed SQBC receiver.

If a dynamic reallocation of time-frequency resources between the up- and downlink directions is possible, the investigation of the optimal feedback/training overhead becomes even more involved. One then has to find the optimal resource partitioning as a function of the achievable rate in the downlink. Spending more resources on CSIT acquisition reduces the multi-user interference and thus improves the SINR; yet, it leaves less resources for the downlink transmission. Optimal resource partitioning is considered, e.g., in [154, 159].

4.2.2. Robust Transmission Schemes

Robust transmission schemes attempt to cope with the CSIT imperfections, by explicitly accounting for them in the transceiver design. Such schemes cannot beat the DoF results mentioned above at very high SNR; yet, they can provide a significant performance gain within practically relevant SNR ranges, by reducing the susceptibility of the transmission

with respect to imperfect CSIT. The discussed methods are especially relevant in high mobility scenarios, since these are prone to CSIT imperfections; see *[Article 1]*, *[Article 2]*.

Robust Beamforming and Precoding

Robust beamforming and precoding schemes have been proposed in the scientific literature under a host of assumptions on the CSIT imperfection and for various transmission scenarios:

- Single-user MIMO transmissions: The problem of worst-case optimization for single-user MIMO transmissions has been treated by many authors [160–165]. Commonly, the goal is to optimize the worst-case of a system utility function over the channel state with respect to the transmit precoder, assuming that the channel matrix lies within an uncertainty set in a neighborhood of the available CSIT. The problem is generally treated in [165] for several utility functions and classes of CSIT uncertainty sets, and a convex formulation of the worst-case optimization is derived.
- Multi-user MIMO systems: Robust optimization for multi-user MISO systems is considered in [166]. The authors consider QoS constraints in terms of SINR lower bounds or MSE upper bounds, and develop a convex optimization framework based on semi-definite programming (SDP), where the channel uncertainty is taken into account with a Frobenius norm constraint. Similar robust optimizations are investigated in [167, 168]. The multi-user MIMO case is treated in [169], where the authors consider joint optimization of the transmit precoders and the receiver equalizers under probabilistic MSE upper bounds. The derived optimization problem is non-convex in general and the authors thus resort to a sub-optimal alternating optimization.
- MIMO interference channel (IC): Robust transceivers for the MIMO IC are developed in [170] assuming a Frobenius norm upper bound on the channel uncertainty. The authors propose a worst-case SINR optimization, which leads to a non-deterministic polynomial-time (NP) hard quadratically constrained quadratic program (QCQP). The authors propose an alternating optimization based on a semi-definite relaxation (SDR) of the QCQP.

Angular-domain leakage bounded precoding: In *[Article 7]*, I develop robust transceivers for the multi-user MIMO downlink under the finite scatterer directional channel model introduced in Section 4.1.3 for MISO transmissions. The underlying assumption of my robust transceiver optimization is that the BS has imperfect knowledge of the signal angles $\mathbf{\Psi}_{u,i}^{(s)}[n,k]$ and no knowledge about the respective signal phases $\varphi_{u,i}^{(s)}[n,k]$; this assumption lies outside the classes of CSIT uncertainty sets previously considered in the literature. The CSIT uncertainty model implies that the transmitter cannot rely on destructive

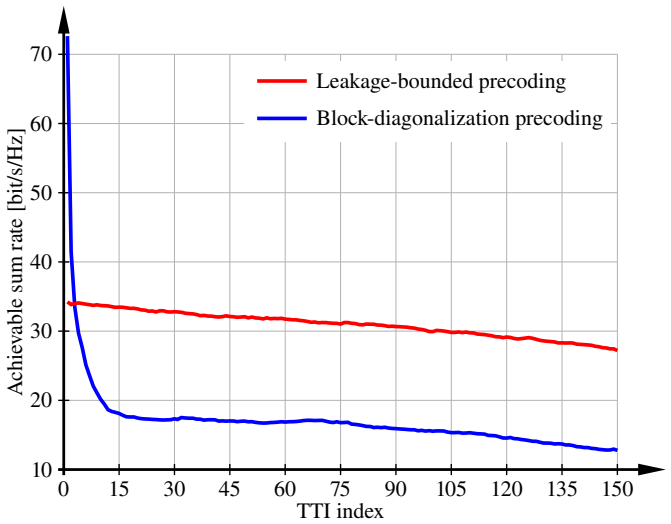


Figure 4.2.: Comparison of the geometric leakage-bounded multi-user precoding scheme with the limited CSI feedback method of *[Article 7]*, and BD precoding with perfect CSIT in the first TTI. Details of the simulated setup are given in *[Article 7]*.

multi-path interference to cancel multi-user interference; it rather mitigates interference by avoiding significant signal leakage into certain angular directions. Such a space division multiple access (SDMA) like scheme provides improved robustness with respect to CSIT imperfections over schemes that rely on coherent multi-path interference. Yet, it can achieve the optimal DoF only under pure LOS transmission; otherwise, the scheme exhibits some residual multi-user interference.

To emphasize the excellent robustness of the angular-domain leakage-bounded precoding method, I include here a result from *[Article 7]* in Figure 4.2, comparing this incoherent transmission scheme to the coherent BD precoding. These simulation results have been obtained with the Quadriga channel model [122], assuming LOS conditions[1] and a high user mobility of 150 km/h. CSIT feedback is provided only every 150 TTIs $= 150$ ms. The performance of the coherent BD precoding schemes is very good over the first few TTIs, where the perfect CSIT provided in the first TTI is highly correlated to the current channel state, but it falls off very quickly afterwards. The leakage-bounded precoding method operates with imperfect angular-domain CSIT from the first TTI on. Since these angles change slowly over time, the performance degradation over time is much more benign.

FD-MIMO and mmWave systems: FD-MIMO and mmWave systems frequently apply hybrid transceiver architectures to reduce the hardware complexity. This implies that the precoding is partitioned into a base band and an RF processing part [38]. To achieve robustness in such systems, both, the digital base band processing and the analog RF pro-

[1] Notice, though, that even under LOS conditions there are additional NLOS components in the channel.

cessing should be designed for imperfect CSIT. My method of *[Article 7]* lends itself to an implementation in the RF domain, since it is a pure beam-steering approach, which can in principle be realized with a PSN. Additionally, it can be combined with robust base band processing, which operates on the effective beamformed channel as discussed in more detail in *[Article 7]*.

Outage Optimized Transmission

As discussed in Section 4.1.3, the outage probability is a critical parameter for services that require dependability of the transmission link. Outage optimal beamforming and precoding schemes under CSIT imperfections have thus been developed for several transmission scenarios. The general goal of these schemes is to find beamformers/precoders that ensure that the outage probability falls below a prescribed threshold:

- Point-to-point MISO transmission: The authors of [171] develop outage optimal precoding schemes for channel covariance feedback and channel mean feedback.
- Multi-user MISO downlink transmission: In [172], the authors consider the multi-user MISO downlink under a Gaussian distributed CSIT error. The authors apply the method of convex restriction to obtain a so-called safe tractable approximation [173], which provides a feasible but possibly not optimal solution of the original problem.
- MIMO IC: This scenario is treated in [174], where the authors propose an iterative alternating optimization to determine outage-optimal beamformers. The method is guaranteed to converge; yet, not necessarily to a global optimum.

Outage-optimal beamforming for MISO TWDP fading channels: In [175] and *[Article 9]*, I consider point-to-point MISO transmission under the finite scatterer directional channel model of Equation (4.2) restricted to $S = 2$ dominant specular components. Assuming unknown phases $\varphi_{u,i}^{(s)}[n,k]$, the effective channel including beamforming follows the two wave with diffuse power (TWDP) fading channel model. This channel model causes fading that can be even worse than Rayleigh fading in terms of the outage probability [176], depending on the relative strength of the two dominant specular components. This relative strength of the specular components can partly be controlled by the applied beamformer. I develop a local gradient optimization approach for outage optimized beamforming, which operates on the associated Grassmann manifold of one-dimensional unit-norm beamformers. I moreover demonstrate that the method outperforms the classical robust beamforming approach of [177]. As shown by our recent measurement results [80], the considered situation with $S = 2$ dominant specular components is highly relevant for indoor mmWave

communications; moreover, since the wavelength of mmWaves is so small, it is sensible to assume unknown/random phases $\varphi_{u,i}^{(s)}[n,k]$ under a robust design perspective.

Robust Scheduling and Rate Adaptation

In addition to adopting robust MIMO beamforming/precoding schemes, it is important to account for the CSIT uncertainty also in the channel quality estimates that are utilized for multi-user scheduling and transmission rate adaptation. In *[Article 8]*, I derive an SINR lower bound for BD precoding with imperfect (quantized) CSIT, which enables robust transmission rate adaptation and multi-user scheduling. Similarly, in [121] I develop a multi-user scheduler that is suitable for the precoding method developed in *[Article 7]*. In general, transmission rate adaptation and multi-user scheduling should be developed to specifically address the impact of CSIT imperfections under the respective applied precoding scheme, to enable robust transmissions.

4.3. Scientific Contributions and Publications

My research work on multi-user scheduling and resource allocation started during my time as a PhD student, in the context of scheduling based on LTE standard compliant CSI feedback, as described in *[Conf. 2]* and [135]. In my later publications, such as, *[Article 8]* and [121], multi-user scheduling was not the major focus of my research, but rather a necessary byproduct of the proposed CSI feedback and transmission schemes, to demonstrate their performance under realistic circumstances. Robust transceiver design, as described, e.g., in *[Article 7]*, *[Article 9]*, [146, 175], came into the focus of my work as a post-doctoral researcher, especially within my Christian Doppler Research Laboratory on Dependable Wireless Connectivity for the Society in Motion, since robustness, reliability and outage investigations are essential for dependable transceiver designs.

4.3.1. Selected Publications

The following publications within the fields of multi-user scheduling, robust transceiver design and outage investigations contribute to this thesis:

[Conf. 2] 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

[Article 7] S. Schwarz, "Robust full-dimension MIMO transmission based on limited feedback angular-domain CSIT," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 1–20, Mar 2018

[Article 8] S. Schwarz and M. Rupp, "Evaluation of distributed multi-user MIMO-OFDM with limited feedback," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6081–6094, Aug. 2014

[Article 9] S. Schwarz, "Outage investigation of beamforming over random-phase finite-scatterer MISO channels," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1029–1033, July 2017

4.3.2. Summary of Scientific Contributions

The main scientific contributions of these publications are:

1. Derivation of an efficient linear optimization framework for multi-user scheduling in BICM-based transceiver architectures, such as, 3GPP LTE, and formulation and solution of standard scheduling problems as linear programs (LPs) in *[Conf. 2]*.
2. Development of robust FD-MIMO precoding schemes for finite scatterer directional channels, based on imperfect angular-domain CSIT, in *[Article 7]*.
3. Design of an entire limited feedback based transceiver architecture for BD precoding multi-user MIMO transmission in OFDMA systems, including transmission rate adaptation and multi-user scheduling based on quantized CQI feedback, in *[Article 8]*.
4. Investigation of the outage probability of beamforming over finite scatterer directional MISO channels with imperfect CSIT in *[Article 9]*.

[Conf. 2] Low complexity approximate maximum throughput scheduling for LTE: In this conference publication, I derive an efficient framework for multi-user scheduling in BICM MIMO-OFDMA transceiver architectures, with application to 3GPP LTE. As mentioned in Section 4.1, a critical issue of such BICM architectures is that the achievable rate of a user does not scale linearly with the amount of allocated resources, but is rather determined by the non-linear MIESM, which complicates the solution of multi-user scheduling optimization problems. To circumvent this issue, I propose in *[Conf. 2]* a local linearization of the BICM functions involved in MIESM, which leads to a computationally efficient linear optimization framework for multi-user scheduling. The local linearization is justified by the fact that the SNR variations due to the microscopic fading are effectively limited to a few [dB] around the mean. This framework enables an implementation of

popular scheduling strategies as LPs. I moreover utilize this linearization in [135] to derive a throughput maximizing scheduling algorithms that supports adjustable fairness in terms of Jain's index.

[Article 7] Robust full-dimension MIMO transmission based on limited feedback angular-domain CSIT: I discussed a part of the scientific contribution of this publication already in Section 3.5, within the context of CSI quantization and feedback. I consider the article here again, because it also contains significant scientific contributions within the field of multi-antenna precoding optimization. Specifically, I derive in this article robust beamforming and precoding methods for efficient multi-user transmission in FD-MIMO systems with angular-domain CSI feedback. The algorithm thereby relies on feedback of the azimuth and elevation angles of the dominant scattering components of each user, to calculate multi-user precoders that limit the amount of multi-user interference of the transmission. This approach is successful in limited scattering environments, when the channels of the UEs contain few dominant scattering components. The algorithm additionally allows for integration of an angular CSIT uncertainty, to compensate for the imperfections caused by the CSI quantization and by the movement of UEs. It therefore achieves significant improvements over BD precoding in high mobility situations, when the CSIT is partially outdated, e.g. due to a delay in the feedback path.

[Article 8] Evaluation of distributed multi-user MIMO-OFDM with limited feedback: In this article, I consider the downlink of an OFDM-based LTE compliant multi-carrier system, in which remote radio units are employed to enable distributed multi-user MIMO transmission. The UEs provide CSI feedback to the BS via limited capacity feedback links, causing residual multi-user interference due to CSIT imperfections. Building on Grassmannian CSI quantization concepts, I propose an effective CSI feedback clustering approach to exploit the correlation of the wireless channel in the frequency domain. I moreover derive the corresponding channel quality feedback for transmission rate adaptation and multi-user scheduling, and propose a practical greedy multi-user scheduler that achieves a performance close to the optimal exhaustive search. These methods enable robust BD-precoding based multi-user MIMO transmission in OFDM systems.

[Article 9] Outage investigation of beamforming over random-phase finite scatterer MISO channels: The finite scatterer directional channel model is a popular model for representation of FD-MIMO systems, especially in the context of mmWave transmissions, which commonly exhibit limited scattering. In this article, I consider beamforming over such channels under the assumption of imperfect CSIT about the angular directions in azimuth and elevation of the scattering components, as well as, uncertainty in the relative

phases of signals arriving at the receivers via these scatterers. In practice, acquiring relatively accurate CSIT about the angular directions is realistic, since these change only slowly over time, whereas the signal phases are much more sensitive to movement of UEs and/or scatterers. The calculation of the outage probability under these assumptions is generally not feasible in closed-form, but requires numerical integration. I further utilize the developed outage expressions for outage optimized beamforming in [175] and generalize them to the case of multi-cell transmissions in [146], where I moreover apply them for multi-user scheduling and outage-constrained admission control to support dependable services.

Low Complexity Approximate Maximum Throughput Scheduling for LTE

Stefan Schwarz, Christian Mehlführer and Markus Rupp
Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: {sschwarz, chmehl, mrupp} @nt.tuwien.ac.at

***Abstract*—In this paper we address the challenge of multiuser scheduling in the downlink of 3GPP UMTS/LTE. Long Term Evolution (LTE) imposes the constraint of using the same code rate, modulation order and transmit power for all resources a User Equipment (UE) is scheduled onto. This, in addition to the lack of channel knowledge, prohibits theoretical concepts such as capacity maximization to be applied for resource allocation. Based on the Channel Quality Indicator (CQI) feedback we derive a linearized model for multiuser scheduling. In contrast to other proposals we use Mutual Information Effective SNR Mapping (MIESM) to calculate an average CQI value for all UE resources. This enables a rate increase while still guaranteeing an imposed Block Error Ratio (BLER) constraint. The proposed framework can also be applied to implement other scheduling strategies. This is demonstrated by comparing different standard schedulers in terms of achieved throughput and fairness.**

***Index Terms*—LTE, OFDMA, multiuser scheduling, adaptive resource allocation, linear programming**

I. Introduction

Orthogonal Frequency Division Multiple Access has been adopted in emerging broadband wireless access networks such as 3GPP UMTS/LTE [1] and IEEE 802.16x (WiMAX) [2] due to its inherent immunity to intersymbol interference and scheduling flexibility in resource allocation. This flexibility allows the exploitation of frequency, temporal and multiuser diversity offered by the wireless broadcast channel. By employing sophisticated multiuser scheduling algorithms, which transmit data to different users on favourable resources, a high system capacity can be achieved.

In our work we focus on the downlink of Long Term Evolution (LTE). In this system the time-frequency grid (resource grid [3]) spanned by OFDM is divided into several Resource Blocks (RBs). User Equipment (UE) resource allocation is carried out on a Resource Block (RB) or a subband basis (each subband consists of several contiguous RBs). Sophisticated scheduling requires the UEs to feed back for each RB (or subband) the supported Channel Quality Indicator (CQI) value, corresponding to a specific code rate - modulation order combination [4]. The standard specifies different feedback granularity possibilities, ranging from wideband (just a single CQI value for all considered RBs) to best M feedback (a distinct CQI value for the M best RBs). For our simulations we consider distinct CQI values for every RB. In our work in [5] and [6] we proposed a spatial preprocessing and link adaption (CQI) feedback scheme based on Mutual Information Effective SNR Mapping (MIESM). MIESM is a well know technique from link level abstraction [7], [8]. It allows to map the SINR experienced on several resources to an equivalent AWGN channel SNR. In [5] and [6] we have shown that MIESM-based wideband feedback achieves close to optimal performance in terms of throughput while fulfilling an imposed constraint on a maximum allowed Block Error Ratio (BLER). In this paper we use the same feedback strategy, but on an RB basis. The required MIESM averaging is carried out at the scheduler after resource allocation.

In this paper we first formulate the sum rate maximization resource allocation problem in Section II based on MIESM to calculate the supported CQI value. This leads to a nonlinear binary integer program. We then use a linear approximation in Section III to come up with a linearized model, which allows to solve the resource allocation problem very efficiently with a Linear Program (LP). We show that this LP is actually equivalent to a best CQI scheduler. Next we give simulation results in Section IV that compare our proposed sum rate maximizing scheduler with the one proposed in [9]. Afterwards in Section V we show that the linearized model is directly applicable to other scheduling strategies as well. In Section VI we compare the performance of different scheduling strategies in terms of throughput and fairness. Concluding remarks will be provided in Section VII.

II. Sum Rate Maximizing Scheduler

Consider the downlink of an OFDM single antenna (SISO) multiuser LTE system. Let N be the number of available RBs and K the number of users. Every user k feeds back a CQI vector $\mathbf{CQI}_k \in \{1, \ldots, \mathrm{CQI}^{(\max)}\}^{N\times 1}$ containing supported CQI values for the N RBs (in LTE $\mathrm{CQI}^{(\max)} = 15$).

These CQIs correspond to supported modulation order - code rate combinations (given in Table 7.2.3-1 in [4]) for the individual RBs. If a UE k is served on several RBs it is necessary to find an average supported CQI value $\overline{\mathrm{CQI}}_k$. We achieve this by first mapping the CQI values of the considered RBs to corresponding SNR values. Next we compute an equivalent AWGN channel SNR value $\mathrm{SNR}_{\mathrm{eq},k}$ by applying MIESM and from this value we arrive at $\overline{\mathrm{CQI}}_k$.

In [10] the author shows that, assuming a BLER target of ≤ 0.1, the mapping function from SNR to CQI is linear. Therefore $\mathbf{CQI}_k$ is linearly related to a corresponding quantized

SNR vector $\mathbf{SNR}_k^{[\mathrm{dB}]} \in \{\mathrm{SNR}^{(1)}, \ldots, \mathrm{SNR}^{(\max)}\}^{N\times 1}$

$$\mathbf{SNR}_k^{[\mathrm{dB}]} = s_1 \mathbf{CQI}_k + s_2 \mathbf{1}, \tag{1}$$

where s_1 and s_2 are coefficients obtained from the linear mapping function and $\mathrm{SNR}^{(i)}$ is the quantized SNR value corresponding to $\mathrm{CQI}^{(i)}$. The notation $(.)^{[\mathrm{dB}]}$ indicates that the value is given in dB.

The goal of a sum rate maximizing scheduler is to allocate resources such that the sum of the user throughputs is maximized. Let $\mathbf{b}_k \in \{0,1\}^{N\times 1}$ be a binary vector indicating which RBs are allocated to user k

$$\mathbf{b}_k(n) = 1 \Longleftrightarrow \text{RB } n \text{ is allocated to user } k, \tag{2}$$

with $\mathbf{b}_k(n)$ corresponding to the nth value of the vector $\mathbf{b}_k$. In a SISO system an RB can only be used by a single UE, therefore RB allocations must not overlap

$$\mathbf{b}_j^T \cdot \mathbf{b}_i = 0 \quad \forall i \neq j. \tag{3}$$

Note that this assumption need not be true in a multiuser MIMO system, because different spatial layers may be allocated to different users on overlapping resources. The supported CQI value of user k, $\overline{\mathrm{CQI}}_k$, is computed by averaging $\mathbf{SNR}_k$ with the help of MIESM to obtain an equivalent AWGN channel SNR, $\mathrm{SNR}_{\mathrm{eq},k}$. This SNR is then mapped back to the CQI domain via the inverse of the linear mapping function. Because the CQI value must be an integer in the range $1, \ldots, \mathrm{CQI}^{(\max)}$ the result has to be rounded down:

$$\mathrm{SNR}_{\mathrm{eq},k} = \beta f^{-1}\left(\frac{1}{||\mathbf{b}_k||_1}\sum_{n=1}^{N} f\left(\frac{\mathbf{SNR}_k(n)\mathbf{b}_k(n)}{\beta}\right)\right) \tag{4}$$

$$\overline{\mathrm{CQI}}_k = \left\lfloor \frac{\mathrm{SNR}_{\mathrm{eq},k}^{[\mathrm{dB}]} - s_2}{s_1} \right\rfloor \tag{5}$$

The function f is given by the Bit Interleaved Coded Modulation (BICM) capacity [11]. The variable β is a calibration factor used to adjust the mapping to the different code rates and modulation alphabets. Therefore it depends on the CQI value. Theoretically it would be necessary to repeat the averaging for all different β values, but the calibration has shown that β is always close to one and therefore set equal to one for simplicity. The whole nonlinear averaging and mapping procedure of eqs. (1), (4) and (5) is condensed in the function R, which yields the spectral efficiency (in bits per channel use) by mapping the average CQI value $\overline{\mathrm{CQI}}_k$ to its corresponding spectral efficiency. The throughput of user k in bits/s, T_k, therefore equals

$$\mathrm{T}_k = c \cdot R(\mathbf{CQI}_k, \mathbf{b}_k) \cdot ||\mathbf{b}_k||_1, \tag{6}$$

where c is a constant that transforms from bits/channel use to bits/s.

Finally, the sum rate maximization problem can be formulated:

$$\{\mathbf{b}_1^\star, \ldots, \mathbf{b}_K^\star\} = \underset{\{\mathbf{b}_1,\ldots,\mathbf{b}_K\}}{\operatorname{argmax}} \sum_{k=1}^{K} \mathrm{T}_k \tag{7}$$

$$\text{subject to:}$$

$$\mathbf{b}_j^T \cdot \mathbf{b}_i = 0 \quad \forall i \neq j$$

$$\mathbf{b}_k(n) \in \{0,1\} \quad \forall n, k.$$

This is a highly nonlinear binary integer program, for which no efficient solution exists. It cannot be implemented in realtime because scheduling decisions must be carried out in every subframe, that is, every 1 ms. Therefore it is necessary to further simplify the model which is achieved in the following section.

III. Linearized Model

In LTE, the CQI feedback from a single UE can only span up to four values, as only 2 bits of feedback are allowed per resource [4]. These 2 bits per RB signal an offset value to an average CQI value, which is also fed back. This constraint can be utilized to simplify the nonlinear optimization problem by linearly approximating the BICM function f necessary for MIESM Eq. (4). Figure 1 shows linear MMSE fits of the envelope of the BICM curves. Normally the BICM functions are given over SNR, but due to the linear mapping between SNR and CQI, they are here directly given over CQI. The

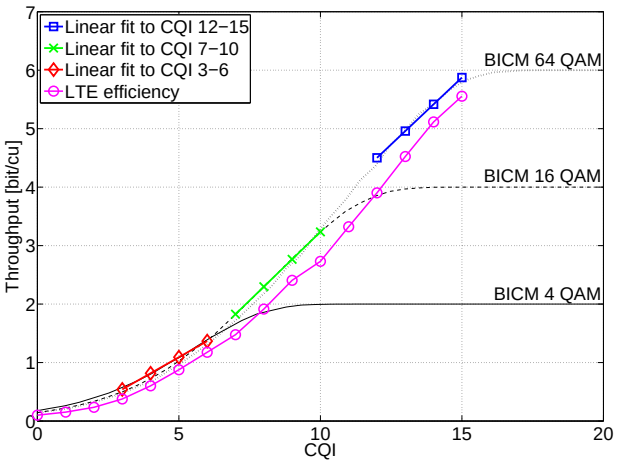


Fig. 1. Linear MMSE fits to the envelope of the BICM functions.

figure shows that for a range of four CQI values linear approximations are reasonable. By applying this approximation, SNR averaging boils down to computing the arithmetic mean. Still, to calculate the supported CQI value $\overline{\mathrm{CQI}}_k$ according to (5), a nonlinear operation (rounding) is necessary. In order to achieve linearity, this operation is ignored for the resource allocation process and noninteger CQI values are allowed. Note that these approximations are just applied during resource allocation. As soon as this process is completed, MIESM is used for every

UE and its RBs to come up with an integer supported CQI value.

One possibility to compute the efficiency corresponding to a noninteger CQI value is to linearly interpolate the efficiencies corresponding to integer CQI values defined in the LTE standard [4] (cf. the magenta circle marked line in Fig. 1). Another possibility is using the theoretical BICM functions. We consider BICM for this purpose (as both are almost parallel, simulation results differed only marginally using either of the two). In order to avoid nonlinearities it is necessary to make use of the linear fits again. Which linear fit has to be applied depends on the actual value of $\overline{\mathrm{CQI}}_k$. Since $\overline{\mathrm{CQI}}_k$ has to be in the range $[\min(\mathbf{CQI}_k), \ldots, \max(\mathbf{CQI}_k)]$ the appropriate fit can be chosen in advance.

Using the above assumptions, the rate of user k, $\overline{\mathrm{R}}_k$, in bits per channel use becomes

$$\overline{\mathrm{R}}_k = d_k \frac{\mathbf{CQI}_k^T \mathbf{b}_k}{||\mathbf{b}_k||_1} + e_k, \tag{8}$$

where d_k and e_k are the coefficients from the linear fit. The throughput equals

$$\begin{aligned} \mathrm{T}_k &= c \cdot \overline{\mathrm{R}}_k ||\mathbf{b}_k||_1 = \\ &= c \cdot d_k \mathbf{CQI}_k^T \mathbf{b}_k + c \underbrace{e_k ||\mathbf{b}_k||_1}_{\mathbf{e}_k^T \mathbf{b}_k} = \\ &= (\underbrace{c \cdot d_k \mathbf{CQI}_k^T + c \cdot \mathbf{e}_k^T}_{\mathbf{c}_k^T}) \mathbf{b}_k. \end{aligned} \tag{9}$$

The step from the second to the third line is possible as $\mathbf{b}_k$ is binary and $\mathbf{e}_k = e_k \mathbf{1}$. In (9) the user throughput finally is linear in the RB allocation $\mathbf{b}_k$.

For convenience let us introduce the following vector notation

$$\mathbf{b} = \mathrm{vec}\left(\begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_K^T \end{bmatrix} \right) \in \{0,1\}^{N \cdot K \times 1} \tag{10}$$

$$\mathbf{c} = \mathrm{vec}\left(\begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_K^T \end{bmatrix} \right) \in \mathbb{R}^{N \cdot K \times 1}. \tag{11}$$

$\mathbf{b}$ contains the RB allocation for all UEs in the form that the first K rows correspond to the first RB of users $1, \ldots, K$, the next K rows to the second RB of all users and so on. Similarly the vector $\mathbf{c}$ contains the corresponding rates.

With this notation the sum rate maximization problem can be written as following binary linear program

$$\begin{aligned} \mathbf{b}^\star = \underset{\mathbf{b}}{\operatorname{argmax}}\, & \mathbf{c}^T \mathbf{b} \\ \text{subject to:} & \\ & \mathbf{A} \cdot \mathbf{b} \leq \mathbf{1}_N \\ & \mathbf{b}(n) \in \{0,1\} \ \ \forall n. \end{aligned} \tag{12}$$

The matrix $\mathbf{A} \in \{0,1\}^{N \times KN}$ ensures that every RB is used at most by a single UE:

$$\mathbf{A} = \begin{bmatrix} \overbrace{11\ldots1}^{K} & \overbrace{0\ldots}^{K\cdot N} & & 0 \\ 0\ldots & 11\ldots1 & 0\ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0\ldots & & & 11\ldots1 \end{bmatrix} \tag{13}$$

Problem (12) can be solved efficiently by a binary solving method such as branch and bound [12].

Investigating the matrix $\mathbf{A}$ shows that it is not even necessary to solve a binary program. Under certain conditions it is possible to apply integer relaxation to a binary linear program without loss of optimality and ensurance that the solution is integer valued. This is possible whenever the constraint matrix $\mathbf{A}$ is totally unimodular[1] and the right hand side of the constraints is integer valued [12]. It is easily verified that these conditions are fulfilled for the given constraint matrix. Therefore the problem can be solved as an LP (e.g. with the simplex method) with additional constraints $0 \leq \mathbf{b}(n) \leq 1 \ \ \forall n$.

The solution can also be obtained in a different way by examining the structure of the problem. Each RB can at most be assigned to a single UE. Due to the linearity of the problem each RB will be assigned. No RB will be left out, although this might be the case if the problem was nonlinear; that is, the CQIs were allowed to vary over a larger range. As a consequence the UE with the largest corresponding value in $\mathbf{c}_k$ will obtain the RB revealing the scheduler identical to a best CQI scheduler that schedules the UE with the highest CQI value. This intuition is proven analytically in Appendix A. We conclude that the best CQI scheduler is sum rate maximizing for LTE under the given approximations.

The presented approach may also be applied in multiuser scenarios with many UEs even if the CQIs are not guaranteed to lie in a range of four values. In these cases multiuser diversity will enforce that UEs are scheduled on their best RBs, automatically leading to low CQI variation (see Section VI).

IV. Simulation Results for two Users

In this section we compare our proposed scheduler (Approximate Max. Throughput (AMT)), based on the LP relaxation, with the Best CQI (BCQI) scheduler and a sum rate maximizing scheduler proposed by R. Kwan et.al. in [9] (Kwan Max. Throughput (KMT)). In [9] the authors are tackling the challenge of resource allocation under the conditions given by the LTE standard (limited channel knowledge, single CQI value per UE). In order to solve the optimization problem, they assume that a UE can only support the lowest CQI value of all

[1] every square non-singular submatrix is unimodular; that is, it has determinant ± 1 and integer entries

RBs it is assigned to. This will be shown to entail a rate loss compared to our solution although the imposed constraint on a maximum allowed BLER is fulfilled by both approaches. In [13] the authors also deal with the problem of scheduling under the constraint of a single adaptive modulation and coding scheme per user, but in a more abstract way. This fact in addition to the complex structure of the solution prevents it from direct application in a realtime system.

In order to compare different schedulers we use a standard compliant LTE physical layer simulator [14] that is publicly available [15]. We implement the slightly suboptimal ($\sim 5\%$ rate loss) scheduling strategy proposed in [9] due to the high complexity of the optimal integer linear program solution.

In the first simulation we consider a scenario with two UEs and a difference in the mean SNR of the two UEs of $\Delta\text{SNR} = 3\,\text{dB}$. The parameters of the simulation are summarized in Table I. We assume a blockfading channel

TABLE I
SIMULATION PARAMETERS

Parameter	Value
System bandwidth	1.4 MHz
Number of subcarriers	72
Number of RBs N	12
Number of users K	2
Channel Model	ITU-T VehA [16]
Antenna configuration	1 transmit, 1 receive (1×1)
Receiver	Zero Forcing ZF
Schedulers	Best CQI (BCQI)
	Approx. Max. Throughput (AMT)
	Kwan. Max. Throughput (KMT)

model with a constant channel during one subframe duration (1 ms) and channel realizations independent between subframes. The CQI feedback from the UEs arrives with zero delay, meaning that the scheduler knows the feedback before the actual transmission. We assume distinct feedback values for all RBs, but we do not explicitly enforce the constraint that the CQI feedback must lie in a range of four values (in most of the cases this is anyway fulfilled). We use a full transmit buffer assumption for our simulations; that is, users fully utilize all resources they get.

Figure 2 shows the sum throughput of both UEs for the three different schedulers over the SNR. BCQI and AMT perform similar. Our proposed scheduler gains about $1.8-2\,\text{dB}$ compared to the proposal of Kwan et.al.

Figure 3 shows a comparison of the BLERs of the two UEs when different schedulers are employed. There is a slight difference in the BLER performance of BCQI and AMT, because the RB assignment is not unique if both users feed back the same CQI value for a resource. The user with the higher SNR (UE1) needs about $-2\,\text{dB}$ SNR to achieve the imposed target BLER for all schedulers, while the worse user (UE2) requires about $-1\,\text{dB}$ SNR. The reason for this is that UE2 in general gets less resources than UE1 and therefore the codeblocksize is smaller which impairs the code performance. The BLER of KMT is lower than that of our proposed method because it underestimates the channel and uses a too conservative CQI value.

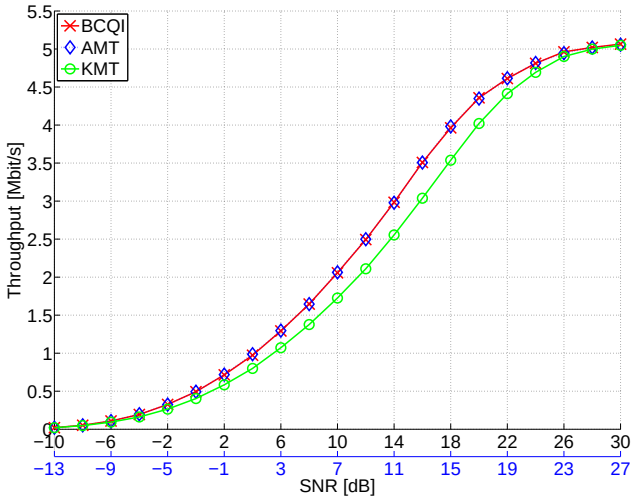


Fig. 2. Sum throughput obtained with different schedulers plotted over SNR for UE 1 and UE 2.

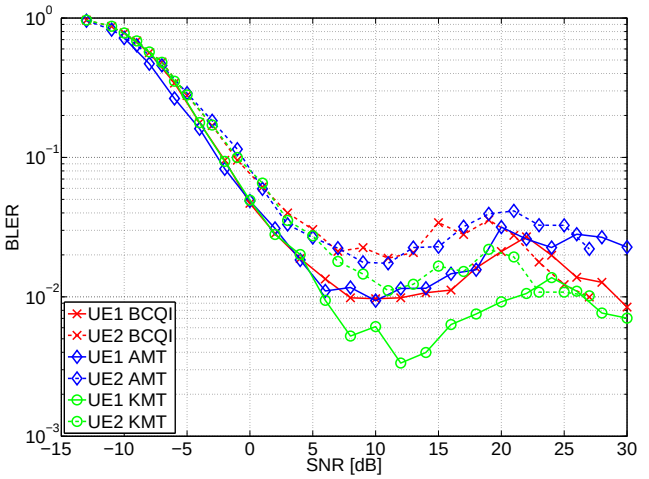


Fig. 3. Block error ratio obtained with different schedulers plotted over SNR for UE 1 and UE 2.

V. Application to Other Scheduling Strategies

In this section we show how the proposed framework can be used to implement other scheduling strategies. We consider some fair schedulers namely the Resource Fair (RF) scheduler, the MaxMin. scheduler and the Proportional Fair (PF) scheduler.

A. Resource Fair Scheduler

The RF scheduler tries to maximize the sum rate of all UEs while guaranteeing fairness with respect to the number of RBs a UE gets. This can be easily achieved by imposing the additional constraint

$$||\mathbf{b}_k||_1 = \frac{N}{K} \quad \forall k, \tag{14}$$

if $\frac{N}{K}$ is integer, otherwise some UEs will get $\lfloor \frac{N}{K} \rfloor$ while others get $\lceil \frac{N}{K} \rceil$ to make up for the total number of RBs (to guarantee

fairness one should randomize this decision). This can be easily achieved by including an additional row for each UE in the matrix $\mathbf{A}$ (13) that sums up all RBs of a UE. This does not harm the unimodularity of the matrix, so the problem can be solved as an LP.

B. MaxMin. Scheduler

The task of a MaxMin. scheduler is to maximize the minimum of the user throughputs. This scheduler is Pareto optimal, meaning that the rate of one UE cannot be increased without decreasing the rate of another UE that has a lower rate than the one considered [17]. The optimization problem can be formulated as

$$\begin{aligned} \{\mathbf{b}_1^\star, \ldots, \mathbf{b}_K^\star\} = \underset{\{\mathbf{b}_1,\ldots,\mathbf{b}_K\}}{\operatorname{argmax}} \min_k \mathbf{c}_k^T \mathbf{b}_k \qquad (15) \\ \text{subject to:} \\ \mathbf{b}_j^T \cdot \mathbf{b}_i = 0 \ \ \forall i \neq j \\ \mathbf{b}_k(n) \in \{0,1\} \ \ \forall n,k. \end{aligned}$$

Introducing the variable ϵ allows to recast the problem into a linear integer program

$$\begin{aligned} \{\mathbf{b}_1^\star, \ldots, \mathbf{b}_K^\star, \epsilon^\star\} = \underset{\{\mathbf{b}_1,\ldots,\mathbf{b}_K,\epsilon\}}{\operatorname{argmax}} \ \epsilon \qquad (16) \\ \text{subject to:} \\ \epsilon \leq \mathbf{c}_k^T \mathbf{b}_k \ \ \forall k \\ \mathbf{b}_j^T \cdot \mathbf{b}_i = 0 \ \ \forall i \neq j \\ \mathbf{b}_k(n) \in \{0,1\} \ \ \forall n,k. \end{aligned}$$

Next we combine the different RB vectors $\mathbf{b}_k$ into a single vector $\mathbf{b}$ as in (10), append the variable ϵ at the end of the vector to get $\widetilde{\mathbf{b}} \in \mathbb{R}^{(KN+1)\times 1}$ and put the $\mathbf{c}_k$ into a matrix $\mathbf{C} \in \mathbb{R}^{K \times KN+1}$

$$\widetilde{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ \epsilon \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \overbrace{\overbrace{c_{1,1}}^{K} \ 0 \ 0 \ \ldots}^{K\cdot N+1} & 0 & \ldots & c_{1,2} & 0 & 0 & \ldots & 1 \\ 0 & c_{2,1} & 0 & \ldots & 0 & c_{2,2} & 0 & \ldots & 1 \\ 0 & 0 & c_{3,1} & \ldots & 0 & 0 & c_{3,2} & \ldots & 1 \\ \vdots & & & \ddots & & & & \ddots & \vdots \end{bmatrix}$$

where $c_{i,j} = -\mathbf{c}_i(j)$. Using this notation the optimization problem can be written as

$$\begin{aligned} \widetilde{\mathbf{b}}^\star = \underset{\widetilde{\mathbf{b}}}{\operatorname{argmax}} \ [\mathbf{0}_{KN}, 1] \cdot \widetilde{\mathbf{b}} \qquad (17) \\ \text{subject to:} \\ \mathbf{C} \cdot \widetilde{\mathbf{b}} \leq \mathbf{0}_K \\ \mathbf{A} \cdot \mathbf{b} \leq \mathbf{1}_N \\ \mathbf{b}(n) \in \{0,1\} \ \ \forall n. \end{aligned}$$

This linear binary integer program cannot be relaxed to an LP without sacrificing optimality, as the constraint matrices are not totally unimodular. Nevertheless, we apply the relaxation here and round the solution simply to the nearest integer. Simulations have shown that the relaxation only entails a minimal rate loss.

C. Proportional Fair Scheduler

A scheduling P is proportionally fair if and only if, for any feasible scheduling S, it satisfies:

$$\sum_k \frac{\overline{\mathrm{T}}_k^{(S)} - \overline{\mathrm{T}}_k^{(P)}}{\overline{\mathrm{T}}_k^{(P)}} \leq 0 \qquad (18)$$

where $\overline{\mathrm{T}}_k^{(S)}$ is the temporal average rate of user k by scheduler S [17]. In [18] necessary and sufficient conditions for a multicarrier scheduler to be proportionally fair are derived. Based on these conditions a slightly suboptimal reduced-complexity algorithm is derived in [19] which can directly be applied with the proposed framework. We use this algorithm with a window size of 10 subframes for the exponential window that is used for averaging the user throughput.

VI. Simulation Results for Multiple Users

In this section we compare different schedulers in terms of their achieved throughput and fairness. Additionally to the schedulers presented in previous sections we use a round robin (RR) scheduler, that schedules users with a fixed pattern, such that every UE gets the same number of contiguous resources. We quantify fairness using Jain's fairness index [20]

$$J(\mathbf{T}) = \frac{\left(\sum_{k=1}^K \mathbf{T}(k)\right)^2}{K \sum_{k=1}^K \mathbf{T}(k)^2}, \qquad (19)$$

where $\mathbf{T}$ is a vector of measured (simulated) user throughputs. Jain's fairness index equals one if all throughputs are the same and perfect fairness is achieved. With decreasing fairness, Jain's fairness index approaches zero. We consider here absolute fairness, meaning that we don't take into account the SNR differences in our fairness measure.

The simulation setup consists of a single cell SISO scenario with 25 UEs having average SNRs ranging from 1 to 25 dB in 1 dB steps. Our simulation parameters are summarized in Table II. Figure 4 shows the throughput achieved by the different UEs when applying several resource allocation schemes. The max. throughput schedulers (AMT, KMT, BCQI) achieve high throughputs for UEs with good SNR but users with low SNR are never severed. The figure also shows, that the scheduler proposed by Kwan et.al. performs as good as our proposal in this case. This is due to multiuser diversity, which causes UEs to be scheduled only on their best RBs, where the CQI values are hardly varying. Figure 5 shows the sum throughput achieved in the cell. The rate maximizing schedulers behave similar and outperform the others in terms of throughput, because they only serve UEs with good channel conditions. The round robin scheduler performs worst, as it does not take into account the channel state for resource allocation. In terms of fairness, the situation more or less reverses, as Figure 6

TABLE II
SIMULATION PARAMETERS

Parameter	Value
System bandwidth	10 MHz
Number of subcarriers	600
Number of RBs N	100
Number of users K	25
Channel Model	3GPP TU [21]
Antenna configuration	1 transmit, 1 receive (1×1)
Receiver	Zero Forcing ZF
Schedulers	Best CQI (BCQI)
	Approx. Max. Throughput (AMT)
	Kwan. Max. Throughput (KMT)
	Round Robin (RR)
	Proportional Fair (PF)
	Resource Fair (RF)
	MaxMin.

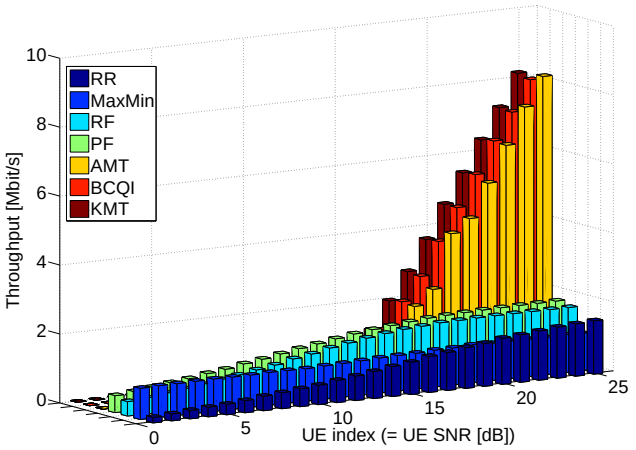


Fig. 4. Throughputs simultaneously achieved by UEs with different average SNRs for several schedulers.

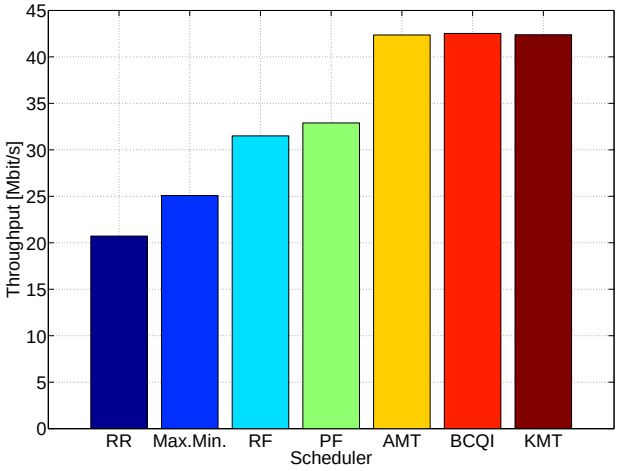


Fig. 5. Sum of user throughputs achieved with different schedulers.

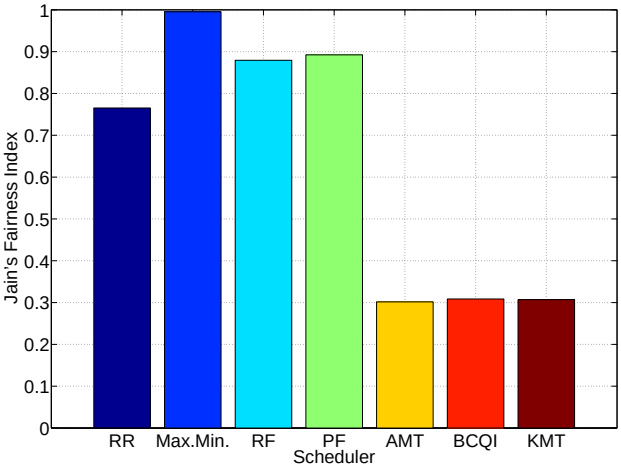


Fig. 6. Fairness achieved with different schedulers.

shows. In accordance to Figure 4 pure rate maximization is not compatible with fairness. The best fairness is achieved with the maxmin. scheduler, which conforms to it's design goal. Good fairness is also achieved with the PF and RF schedulers. These two also deliver high sum throughput and therefore seem to be a good compromise. Not taking into account the channel conditions for resource allocation, as the RR scheduler does, clearly is a bad choice, because neither high fairness nor high throughput can be achieved.

All simulation results as well as the corresponding MATLAB code will be made available online in the next release (v.1.5) of our physical layer LTE simulator [15].

VII. CONCLUSION

In this paper we formulate the sum rate maximization resource allocation problem in the framework of Long Term Evolution. We develop a linearized model to simplify the nonlinear combinatorial optimization problem to a simple linear program. We show that solving this linear program is equivalent to allocating resources to the users with the best channel conditions (best CQI). A comparison to a rate maximizing scheduling strategy proposed by Kwan et.al. in [9] is carried out by simulations. Our scheduler achieves better results for small user numbers, while for large user numbers the performance is similar. Next we show how the proposed framework can be used to implement other schedulers. We compare several schedulers in terms of achieved throughput and fairness by simulations. These show that proportional fair and resource fair schedulers deliver a good compromise between fairness and throughput.

ACKNOWLEDGEMENT

This work has been funded by A1 Telekom Austria AG and the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology.

APPENDIX A

We show analytically that the AMT scheduler is equivalent to the BCQI scheduler by considering the Karush-Kuhn-Tucker (KKT) optimality conditions [22]. For this purpose we

reformulate the linear relaxation of the optimization problem (12) in scalar notation. We set the first constraint equal to one, because every resource is used due to the linearity of the problem.

$$\{b^\star_{1,1},\dots,b^\star_{N,K}\} = \underset{\{b_{1,1},\dots,b_{N,K}\}}{\operatorname{argmax}} \sum_{i=1}^{N}\sum_{j=1}^{K} c_{i,j}b_{i,j} \qquad (20)$$

subject to:

$$\sum_{j=1}^{K} b_{i,j} = 1 \;\; \forall i \in \{1,\dots,N\}$$

$$0 \le b_{i,j} \le 1 \;\; \forall i \in \{1,\dots,N\},\; j \in \{1,\dots,K\}$$

The Lagrangian $L(\mathbf{b},\lambda,\nu)$ is given by

$$L(\mathbf{b},\lambda,\mu) = \sum_{i=1}^{N}\sum_{j=1}^{K} c_{i,j}b_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{K}\lambda_{i,j}(b_{i,j}-1) + \dots$$
$$+ \sum_{i=1}^{N}\sum_{j=1}^{K}\lambda_{i+N,j}(-b_{i,j}) + \sum_{i=1}^{N}\nu_i\left(\sum_{j=1}^{K} b_{i,j} - 1\right) \qquad (21)$$
$$\mathbf{b} = [b_{1,1}, b_{1,2}, \dots, b_{N,K}]^T \in \mathbb{R}^{NK\times 1}$$
$$\lambda = [\lambda_{1,1}, \lambda_{1,2}, \dots, \lambda_{2N,K}]^T \in \mathbb{R}^{2NK\times 1}$$
$$\nu = [\nu_1, \nu_2, \dots, \nu_N]^T \in \mathbb{R}^{N\times 1}.$$

The KKT conditions result in the following system of equations:

$$(i)\quad b_{i,j} \le 1 \;\; \forall i,j \qquad (ii)\quad b_{i,j} \ge 0 \;\; \forall i,j$$
$$(iii)\quad \sum_{j=1}^{K} b_{i,j} = 1 \;\; \forall i \qquad (iv)\quad \lambda_{i,j} \ge 0 \;\; \forall i,j$$
$$(v)\quad \lambda_{i,j}\cdot(b_{i,j}-1) = 0 \;\; \forall i,j \qquad (vi)\quad \lambda_{i+N,j}\cdot b_{i,j} = 0 \;\; \forall i,j$$
$$(vii)\; \lambda_{i,j} - \lambda_{i+N,j} = c_{i,j} - \nu_i \;\; \forall i,j$$

Consider next the case $c_{i,j} - \nu_i > 0$.

$$c_{i,j} - \nu_i > 0 \overset{(vii)}{\Longrightarrow} \lambda_{i,j} > \lambda_{i+N,j} \overset{(iv)}{\Longrightarrow} \lambda_{i,j} > 0 \overset{(v)}{\Longrightarrow} b_{i,j} = 1$$
$$\overset{(vi)}{\Longrightarrow} \lambda_{i+N,j} = 0 \overset{(iii)}{\Longrightarrow} b_{i,n} = 0 \;\; \forall n \ne j \overset{(iv,vi)}{\Longrightarrow} \lambda_{i+N,n} \ge 0$$
$$\overset{(v)}{\Longrightarrow} \lambda_{i,n} = 0 \overset{(vii)}{\Longrightarrow} c_{i,n} - \nu_i \le 0 \Longrightarrow c_{i,j} > c_{i,n} \;\; \forall n.$$

In this case the user j with the largest CQI value on resource i is served. Considering the case $c_{i,j} - \nu_i < 0$ leads to $b_{i,j} = 0$ following a similar argumentation as above. This means that the user j is not served on resource i. The last possibility to be investigated is $c_{i,j} - \nu_i = 0$.

$$c_{i,j} - \nu_i = 0 \overset{(vii)}{\Longrightarrow} \lambda_{i,j} = \lambda_{i+N,j} \overset{(v,vi)}{\Longrightarrow} \lambda_{i,j} = \lambda_{i+N,j} = 0$$

If the user j is the only one for which $c_{i,j} - \nu_i = 0$ holds and there is no user k with $c_{i,k} - \nu_i > 0$ than (iii) necessitates $b_{i,j} = 1$ and the user gets the resource i. If there are more users $\{j,k,l,\dots\}$ that have $c_{i,n} - \nu_i = 0 \;\; n \in \{j,k,l,\dots\}$ than time sharing of the resource i between the users, such that $\sum_{n\in\{j,k,l,\dots\}} b_{i,n} = 1$ is fulfilled, is an optimal solution. Employing the simplex method for solving the LP, a time sharing solution will not be produced, as this does not correspond to a corner point of the feasible set, but rather to a point on a connecting surface between corner points [23]. Therefore, in the case of equal CQI values for several users, a single user will get the resource.

REFERENCES

[1] 3GPP, "Techn. Spec. Group Radio Access Network; (E-UTRA) and (E-UTRAN); Overall description; Stage 2 (release 8)," Sept. 2008. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36300.htm.

[2] IEEE, "IEEE Std 802.16-2009," May 2009. [Online]. Available: http://standards.ieee.org/getieee802/download/802.16-2009.pdf.

[3] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)," September 2009. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36211.htm.

[4] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 8)," March 2009. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36213.htm.

[5] S. Schwarz, M. Wrulich, and M. Rupp, "Mutual Information based Calculation of the Precoding Matrix Indicator for 3GPP UMTS/LTE," in Proc. IEEE Workshop on Smart Antennas 2010, (Bremen, Germany), February 2010.

[6] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in Wireless Advanced (WiAD), 2010 6th Conference on, (London, UK), June 2010.

[7] L. Wan, S. Tsai, and M. Almgren, "A Fading-Insensitve Performance Metric for a Unified Link Quality Model," in Proc. IEEE Wireless Communications & Networking Conference WCNC, 2006.

[8] X. He, K. Niu, Z.He, and J. Lin, "Link Layer Abstraction in MIMO-OFDM System," in Proc. International Workshop on Cross Layer Design, 2007.

[9] R. Kwan, C. Leung, and J. Zhang, "Multiuser Scheduling on the Downlink of an LTE Cellular System," Research Letters in Communications, 2008. Hindawi Publishing Corporation.

[10] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in Proc. 71st Vehicular Technology Conference VTC2010-Spring, 2010.

[11] G. Caire, G. Taricco, and E. Biglieri, "Capacity of bit-interleaved channels," Electron. Lett., vol. 32, issue 12, pp. 1060–1061, June 1996.

[12] C. H. Papadimitriou and K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity. Dover Pubn Inc, 2000.

[13] G. Gotsis, D. Komnakos, and P. Constantinou, "Linear Modeling and Performance Evaluation of Resource Allocation and User Scheduling for LTE-like OFDMA networks," in Proc. IEEE International Symposium on Wireless Communication Systems ISWCS, 2009.

[14] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the Long Term Evolution Physical Layer," in Proc. 17th European Signal Processing Conference EUSIPCO 2009, (Glasgow, Scotland), August 2009. [Online]. Available: http://publik.tuwien.ac.at/files/PubDat_175708.pdf.

[15] [Online]. Available: http://www.nt.tuwien.ac.at/ltesimulator/.

[16] ITU, "Recommendation ITU-R M.1225: Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000," tech. rep., ITU, 1997.

[17] F. Kelly, "Charging and rate control for elastic traffic," European Transactions on Telecommunications, vol. 8, 1997.

[18] H. Kim and Y. Han, "A Proportional Fair Scheduling for Multicarrier Transmission Systems," IEEE Communications Letters, vol. 9, no. 3, 2005.

[19] Z. Sun, C. Yin, and G. Yue, "Reduced-Complexity Proportional Fair Scheduling for OFDMA Systems," in Proc. IEEE International Conference on Communications, Circuits and Systems, 2006. vol. 2.

[20] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," Tech. Rep. TR-301, DEC, September 1984.

[21] 3GPP, "Technical Specification Group Radio Access Networks; Deployment aspects (Release 8)," Dezember 2008. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25943.htm.

[22] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.

[23] J. Chimneck, "Practical Optimization: a Gentle Introduction," 2000. [Online]. Available: http://www.sce.carleton.ca/faculty/chinneck/po.html.

Evaluation of Distributed Multi-User MIMO-OFDM With Limited Feedback

Stefan Schwarz, *Member, IEEE*, and Markus Rupp, *Senior Member, IEEE*

***Abstract*—In this article, we investigate the performance of cellular networks employing remote radio units to enable distributed multi-user MIMO transmission using block-diagonalization precoding. We consider the downlink of an OFDM-based LTE compliant multi-carrier system in which the users provide channel state information (CSI) to the base station via limited capacity feedback links. With limited feedback, residual interference between the spatially multiplexed users cannot be avoided, causing an interference-limitation of the achievable downlink throughput. Efficient CSI quantization is therefore central in such systems to achieve a performance gain over single-user MIMO. Building on Grassmannian CSI quantization concepts, we propose an effective CSI feedback clustering approach in this work to exploit the correlation of the wireless channel in the frequency domain. We furthermore derive the corresponding channel quality feedback and propose a practical greedy multi-user scheduler.**

***Index Terms*—Limited feedback, multi-user MIMO, block-diagonalization, OFDM, distributed antenna systems.**

I. Introduction

In this paper, we investigate multi-user multiple-input multiple-output (MU-MIMO) transmission based on block-diagonalization (BD) precoding [1] with limited feedback in the downlink of orthogonal frequency division multiplexing (OFDM) based cellular networks. Although strictly suboptimal in terms of the achievable transmission rate, many authors consider BD precoding as practically important for its complexity advantage compared to non-linear techniques, such as vector perturbation [2], and because it achieves the sum degrees of freedom of the MIMO broadcast channel. MU-MIMO, however, is sensitive to the accuracy of the channel state information (CSI) at the transmitter (CSIT). With imperfect CSIT, e.g., due to limited capacity feedback links, residual multi-user interference causes an interference limitation of the achievable throughput that can only be avoided by scaling the feedback overhead linearly with the logarithmic signal-to-noise ratio (SNR) [3]–[5]. To reduce the feedback burden, the channel correlation over time, frequency and space must be exploited during CSI quantization. Differential and predictive limited feedback algorithms that utilize the temporal correlation of the wireless channel have been proposed in [6]–[8] for single stream transmission per user (zero-forcing beamforming) and have been extended in [9], [10] to support multiple streams per user. Additionally, the frequency correlation of the channel can be exploited for feedback overhead reduction. Clustering consecutive OFDM subcarriers and providing the best representation for each cluster as CSI has been proposed in [11] for beamforming systems. Feeding back CSI on a subset of OFDM subcarriers and employing interpolation at the transmitter to determine the CSI on the intermediate subcarriers has been considered in [12]–[14].

The efficiency of MU-MIMO can be improved with remote radio units (RRU) that are distributed over the cell area to form a distributed antenna system (DAS) [15], [16]. DASs improve the macro-diversity of cellular networks and reduce the average access distance to the base station [17]. They have hence been successfully employed for coverage improvement [18], to reduce the outage probability of cellular networks [19] and to increase the network capacity [17]. Distributing antenna arrays over the cell area can cause significant macroscopic pathloss variations with respect to the individual arrays. This can be exploited during CSI quantization by providing more accurate CSI for the strong antenna arrays, e.g., by means of feedback bit allocation [20] or by matching the quantization codebook to the spatial statistics of the channel [16], [21].

Another important aspect of MU-MIMO is the selection of users that are served in parallel. For zero-forcing beamforming, it has been shown that a selection of users with orthogonal channels implies asymptotic optimality with a growing number of users [22]. Based on this observation practical greedy scheduling algorithms have been proposed that achieve close-to-optimal performance, most famously, semi-orthogonal user selection (SUS) [22]. The SUS idea has been extended to operate with limited feedback in [23] and similar greedy schedulers for BD precoding have been proposed in [24].

Summary of Contributions: In this paper, we propose a practically viable BD-based MU-MIMO OFDM transceiver architecture in Section II that operates with limited feedback. The major focus of the paper is on the exploitation of the frequency correlation of the wireless channel to achieve efficient CSI quantization in multi-carrier OFDM. In Section III, we briefly review geodesic CSI interpolation and propose an alternative efficient CSI clustering approach that generalizes subspace quantization based combining (SQBC) [5] from single-carrier to multi-carrier transmission. With this approach, we achieve a significant mean squared error (MSE) reduction and throughput improvement compared to interpolation-based methods, especially when cluster sizes are large compared to the channel coherence bandwidth. We combine CSI clustering and interpolation with our previously proposed CSI quantizers to achieve efficient limited feedback operation. We furthermore propose a

Manuscript received October 30, 2013; revised April 11, 2014 and June 3, 2014; accepted July 29, 2014. Date of publication August 7, 2014; date of current version November 7, 2014. The associate editor coordinating the review of this paper and approving it for publication was M. Vu.

The authors are with the Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria (e-mail: sschwarz@nt.tuwien.ac.at; mrupp@nt.tuwien.ac.at).

Digital Object Identifier 10.1109/TWC.2014.2346191

TABLE I
TABLE OF FREQUENTLY USED VARIABLES AND PARAMETERS

Notation	Support	Meaning
P_t	$\mathbb{R}$	transmit power
P_u	$\mathbb{R}$	transmit power of user u
N_t	$\mathbb{N}$	number of transmit antennas
M_u	$\mathbb{N}$	number of receive antennas of user u
ℓ	$\mathbb{N}$	total number of data streams
ℓ_u	$\mathbb{N}$	number of data streams of user u
$\bar{\ell}_u$	$\mathbb{N}$	total number of streams excluding user u
U	$\mathbb{N}$	total number of users
S	$\mathbb{N}$	number of spatially multiplexed users
$\mathbf{H}_u$	$\mathbb{C}^{N_t \times M_u}$	channel matrix of user u
$\mathbf{F}_u$	$\mathbb{C}^{N_t \times \ell_u}$	precoding matrix of user u
$\mathbf{G}_u$	$\mathbb{C}^{M_u \times \ell_u}$	antenna combiner of user u
$\mathbf{H}_u^{\text{eff}}$	$\mathbb{C}^{N_t \times \ell_u}$	effective channel matrix of user u
$\bar{\mathbf{H}}_u$	$\mathbb{C}^{N_t \times \ell_u}$	orthonormal basis for $\text{span}\left(\mathbf{H}_u^{\text{eff}}\right)$
$\hat{\mathbf{H}}_u$	$\mathbb{C}^{N_t \times \ell_u}$	orthonormal basis obtained from quantization

lower bound on the expected per-stream signal-to-interference-and-noise ratio (SINR) as channel quality feedback, enabling effective multi-user scheduling with low complexity when combined with SUS. We investigate the performance of the proposed methods and algorithms by means of Monte-Carlo simulations in Section IV and apply the proposed transceiver architecture in a 3GPP LTE compliant heterogeneous cellular network in Section V, employing the open-source *MATLAB*-based Vienna LTE-A link level simulator [25].

Notation: $\mathbf{a} \in \mathcal{N}_{\mathbb{C}}(\mathbf{m}, \mathbf{R})$ defines a complex-valued circularly symmetric Gaussian random vector $\mathbf{a}$, with mean $\mathbf{m}$ and covariance matrix $\mathbf{R}$. The Grassmann manifold of m-dimensional subspaces in the n-dimensional complex Euclidean space ($m \leq n$) is written as $\mathcal{G}(n, m)$. The conjugate-transpose of a matrix $\mathbf{A}$ is $\mathbf{A}^{\text{H}}$. $\mathbb{E}(\cdot)$ denotes the expected value. The operator $[\mathbf{A}]_{m:n,i:j}$ selects rows m to n and columns i to j of $\mathbf{A}$; $\text{span}\,(\mathbf{A})$ denotes the space spanned by the columns of $\mathbf{A}$, $\text{null}\,(\mathbf{A})$ defines the null space of $\mathbf{A}$ and $\text{tr}\,(\mathbf{A})$ is the trace. Symbols for frequently used variables and parameters are listed in Table I.

II. SYSTEM MODEL

We consider MU-MIMO transmission in the downlink of a cellular network. The transmitter employs OFDM modulation to convert the frequency selective channel into a set of non-interfering frequency-flat subcarriers indexed by n. The input-output relationship of user u at subcarrier n is

$$\mathbf{y}_u[n] = (\mathbf{H}_u[n]\mathbf{G}_u[n])^{\text{H}}\,\mathbf{F}_u[n]\mathbf{x}_u[n] + (\mathbf{H}_u[n]\mathbf{G}_u[n])^{\text{H}} \sum_{s=1, s\neq u}^{S} \mathbf{F}_s[n]\mathbf{x}_s[n] + \mathbf{G}_u[n]^{\text{H}}\mathbf{z}_u[n]. \quad (1)$$

Here, $\mathbf{H}_u[n] \in \mathbb{C}^{N_t \times M_u}$ denotes the complex-valued channel matrix between the N_t transmit antennas and the M_u receive antennas.[1] The matrices $\mathbf{F}_s[n] \in \mathbb{C}^{N_t \times \ell_s}, s \in \mathcal{S} = \{1, \ldots, S\}$ denote the precoders of the S spatially multiplexed users. To simplify notations, we omit a subcarrier dependency of the schedule $\mathcal{S}$ and of the number of spatial streams $\ell_s \leq M_s$ of user s. The vector $\mathbf{x}_s[n] \in \mathbb{C}^{\ell_s \times 1}$ contains the spatially multiplexed transmit symbols of user s; it is normalized as $\mathbb{E}(\mathbf{x}_s[n]\mathbf{x}_s[n]^{\text{H}}) = \mathbf{I}_{\ell_s}$. $\mathbf{G}_u[n] \in \mathbb{C}^{M_u \times \ell_u}$ is the antenna combiner applied by user u to separate the intended signal from the multi-user interference. We are interested in the practically relevant situation $M_u < N_t$, in which the users cannot in general completely eliminate the multi-user interference without the aid of the transmitter. We refer to the product $\mathbf{H}_u^{\text{eff}}[n] = \mathbf{H}_u[n]\mathbf{G}_u[n]$ as the effective channel matrix. The vector $\mathbf{z}_u[n]$ denotes the sum of Gaussian receiver noise and other-cell interference. We suppose that the users treat the other-cell interference as additional noise and assume $\mathbf{z}_u[n] \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_z^2\mathbf{I}_{M_u})$. Notice that Gaussianity of the other-cell interference may not be fulfilled in all cases, e.g., if there are few dominant interferers. Then, receivers that estimate the interference statistics can provide better performance [26].

The considered system model recently gained interest in the context of joint transmission coordinated multi-point transmission (CoMP) [27], [28], where the N_t transmit antennas are distributed among several cooperating transmission points. In our simulations in Section IV, we apply the proposed algorithms for joint transmission in a DAS, in which RRUs are connected to the macro base station over low-latency high-bandwidth dedicated connections (e.g., radio over fiber, microwave radio links) to enable coherent transmission from spatially separated antenna arrays. We account for these effects by decomposing the channel as

$$\mathbf{H}_u[n] = \mathbf{\Gamma}_u^{1/2}\bar{\mathbf{H}}_u[n], \quad \mathbf{\Gamma}_u = \text{diag}\left(\gamma_u^{(1)}, , \ldots, \gamma_u^{(N_t)}\right), \quad (2)$$

where the channel gain $\gamma_u^{(j)}$ characterizes the large-scale fading with respect to transmit antenna j, and the matrix $\bar{\mathbf{H}}_u[n]$ captures the frequency selective small-scale fading.

A. Block-Diagonalization Precoding

To optimize the achievable throughput of the system in (1), a joint optimization of the multi-user schedule $\mathcal{S}$, the precoders $\mathbf{F}_s[n]$ and the antenna combiners $\mathbf{G}_s[n]$ is required. Iterative algorithms have been proposed to find locally optimal solutions to this non-convex joint-optimization problem, e.g., [29], [30]. These approaches, however, involve a significant signaling overhead between the transmitter and the receivers to determine the solution; also, it is not immediately clear how to enable efficient limited feedback operation. In this paper we hence pursue a different practically feasible approach, in which we restrict the precoders to BD and we determine the antenna combiners independently at the users based on purely selfish arguments; thus, we do not consider joint transceiver optimization. With our approach, the users select the antenna combiners to either maximize the gain of the intended signal, ignoring the residual multi-user interference due to limited feedback; or they focus on reducing the residual multi-user interference by minimizing the CSI quantization error.

With BD precoding [1], the transmitter calculates the precoders $\mathbf{F}_s[n], \forall\, s \in \mathcal{S}$ such that the interference at each user due to all other users is eliminated. In the original proposal of BD precoding [1], the authors construct the solution by invoking the conditions $\mathbf{H}_u[n]^{\text{H}}\mathbf{F}_s[n] = \mathbf{0},\ \forall\, s \neq u$. In the considered

[1] We apply the conjugate transpose of $\mathbf{H}_u[n]$ and $\mathbf{G}_u[n]$ in Eq. (1) to simplify later notations.

scenario with $\ell_u \leq M_u$ we are satisfied if the interference is zero in the ℓ_u-dimensional subspace selected by the antenna combiner $\mathbf{G}_u[n]$, leading to the conditions

$$\mathbf{H}_u^{\text{eff}}[n]^{\text{H}}\mathbf{F}_s[n] = \mathbf{0}, \quad \forall\, u, s \in \mathcal{S} \text{ and } s \neq u, \quad (3)$$

$$\text{rank}\left(\mathbf{H}_u^{\text{eff}}[n]^{\text{H}}\mathbf{F}_u[n]\right) = \ell_u, \quad \forall\, u \in \mathcal{S}. \quad (4)$$

Assuming full-rank channel matrices, a solution exists provided the feasibility conditions $\ell_u \leq M_u$ and $\sum_{s\in\mathcal{S}} \ell_s \leq N_t$ are satisfied [1]. The precoders are obtained from

$$\mathbf{F}_u[n] \in \text{null}\left(\overline{\mathbf{H}}_u[n]\right), \ \text{rank}\left(\mathbf{F}_u[n]\right) = \ell_u, \ \forall\, u \in \mathcal{S}, \quad (5)$$

$$\overline{\mathbf{H}}_u[n] = \left[\tilde{\mathbf{H}}_1[n], \ldots, \tilde{\mathbf{H}}_{u-1}[n], \tilde{\mathbf{H}}_{u+1}[n], \ldots, \tilde{\mathbf{H}}_S[n]\right]^{\text{H}}, \quad (6)$$

$$\text{span}\left(\tilde{\mathbf{H}}_s[n]\right) = \text{span}\left(\mathbf{H}_s^{\text{eff}}[n]\right), \quad (7)$$

$$\tilde{\mathbf{H}}_s[n]^{\text{H}}\tilde{\mathbf{H}}_s[n] = \mathbf{I}_{\ell_s}, \ \tilde{\mathbf{H}}_s[n] \in \mathbb{C}^{N_t \times \ell_s}, \ \forall\, s \in \mathcal{S}. \quad (8)$$

In this equation, the matrices $\tilde{\mathbf{H}}_s[n]$ form orthonormal bases for $\text{span}(\mathbf{H}_s^{\text{eff}}[n])$ and can, e.g., be obtained from a singular value decomposition (SVD) of $\mathbf{H}_s^{\text{eff}}[n]$. 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 $u \in \mathcal{S}$, we obtain interference-free transmission to all users, i.e.,

$$\mathbf{y}_u[n] = \mathbf{H}_u^{\text{eff}}[n]^{\text{H}}\mathbf{F}_u[n]\mathbf{x}_u[n] + \mathbf{G}_u[n]^{\text{H}}\mathbf{z}_u[n]. \quad (9)$$

The precoder $\mathbf{F}_u[n]$ obtained from (5) is unique up to right-multiplication with any full-rank $\ell_u \times \ell_u$ matrix. In [1], the authors remove this ambiguity by treating the interference-free input-output relationship (9) as a single-user MIMO system, and additionally performing SVD-based precoding over this single-user channel. This is feasible if the channel matrices $\mathbf{H}_s^{\text{eff}}[n], \forall\, s$ are available at the transmitter. In this paper, however, we target the application of well studied Grassmannian quantization algorithms for limited feedback operation. The Grassmannian conveys only subspace information, i.e., neither information about the magnitudes of the singular values of $\mathbf{H}_s^{\text{eff}}[n]$ nor about the orientation of the individual singular vectors. Hence, power allocation is not an option with Grassmannian feedback and we thus consider equal power allocation across users, spatial streams and subcarriers

$$\mathbf{F}_u[n]^{\text{H}}\mathbf{F}_u[n] = \frac{P_t}{S\ell_u}\tilde{\mathbf{F}}_u[n]^{\text{H}}\tilde{\mathbf{F}}_u[n] = P_u\mathbf{I}_{\ell_u}, \quad (10)$$

with P_t being the per-subcarrier transmit power. This approach is suitable for high SNR and has the advantage that the base station only requires knowledge of the ℓ_s-dimensional subspaces $\text{span}(\mathbf{H}_s^{\text{eff}}[n]), \forall\, s$, to determine the precoders, which can be conveyed using Grassmannian feedback [10]. In the low to intermediate SNR regime, however, performance improvements are possible, e.g., with regularized BD and water-filling power allocation, which can be enabled with CSIT on the Stiefel manifold rather than on the Grassmannian; see [31].

B. Limited Feedback Model

The subspace $\text{span}(\mathbf{H}_s^{\text{eff}}[n])$ is a point on the Grassmann manifold of ℓ_s-dimensional subspaces in the N_t-dimensional complex Euclidean space $\mathcal{G}(N_t, \ell_s)$ [32], [33]. To represent this point, we employ an orthonormal basis $\tilde{\mathbf{H}}_s[n]$ as in (7) and (8). Perfect knowledge of $\tilde{\mathbf{H}}_s[n], \ \forall\, s \in \mathcal{S}$ is required at the base station to achieve zero multi-user interference. With limited feedback, user u feeds back a quantized version $\hat{\mathbf{H}}_u[n]$ obtained by minimizing a quantization metric

$$\hat{\mathbf{H}}_u[n] = \underset{\mathbf{Q}_j \in \mathcal{Q}_u}{\arg\min}\ \Omega\left(\mathbf{H}_u[n], \mathbf{Q}_j\right), \quad (11)$$

$$\mathcal{Q}_u = \left\{\mathbf{Q}_j \in \mathbb{C}^{N_t \times \ell_u} | \mathbf{Q}_j^{\text{H}}\mathbf{Q}_j = \mathbf{I}_{\ell_u}, j \in \{1, \ldots, 2^b\}\right\}, \quad (12)$$

with $\mathcal{Q}_u$ denoting the Grassmannian quantization codebook and b being the number of feedback bits. The transmitter treats the quantized subspace $\hat{\mathbf{H}}_u[n]$ as the actual subspace $\tilde{\mathbf{H}}_u[n]$ and calculates the precoders from (5), replacing $\overline{\mathbf{H}}_u[n]$ with

$$\hat{\overline{\mathbf{H}}}_u[n] = \left[\hat{\mathbf{H}}_1[n], \ldots, \hat{\mathbf{H}}_{u-1}[n], \hat{\mathbf{H}}_{u+1}[n], \ldots, \hat{\mathbf{H}}_S[n]\right]^{\text{H}}. \quad (13)$$

We decompose the effective channel $\mathbf{H}_u^{\text{eff}}[n]$ into its range space component $\mathbf{H}_u^{\text{eff,r}}[n]$ and its left null space component $\mathbf{H}_u^{\text{eff,n}}[n]$ with respect to the orthonormal basis $\hat{\mathbf{H}}_u[n]$

$$\mathbf{H}_u^{\text{eff}}[n] = \underbrace{\hat{\mathbf{H}}_u[n]\hat{\mathbf{H}}_u[n]^{\text{H}}\mathbf{H}_u^{\text{eff}}[n]}_{\mathbf{H}_u^{\text{eff,r}}[n]} + \underbrace{\left(\mathbf{I}_{N_t} - \hat{\mathbf{H}}_u[n]\hat{\mathbf{H}}_u[n]^{\text{H}}\right)\mathbf{H}_u^{\text{eff}}[n]}_{\mathbf{H}_u^{\text{eff,n}}[n]}. \quad (14)$$

Using this notation and recalling that $\hat{\mathbf{H}}_u[n]^{\text{H}}\mathbf{F}_s[n] = 0, \ \forall\, s \neq u$, due to the BD construction, the input-output relationship with quantized CSIT is

$$\mathbf{y}_u = \mathbf{H}_u^{\text{eff}}[n]^{\text{H}}\mathbf{F}_u[n]\mathbf{x}_u[n] + \mathbf{H}_u^{\text{eff,n}}[n]^{\text{H}}\sum_{s=1, s\neq u}^{S}\mathbf{F}_s[n]\mathbf{x}_s[n] + \mathbf{G}_u[n]^{\text{H}}\mathbf{z}_u[n]. \quad (15)$$

III. CSI Feedback Strategies

CSIT imperfections cause residual interference in multi-user precoding systems. The rate loss with respect to perfect CSIT is determined by the chordal distance quantization error [3]–[5]

$$\text{d}_{\text{c}}^2\left(\tilde{\mathbf{H}}_u[n], \hat{\mathbf{H}}_u[n]\right) = \ell_u - \text{tr}\left(\tilde{\mathbf{H}}_u[n]^{\text{H}}\hat{\mathbf{H}}_u[n]\hat{\mathbf{H}}_u[n]^{\text{H}}\tilde{\mathbf{H}}_u[n]\right), \quad (16)$$

which is thus often employed as CSI quantization metric, not only for BD precoding [6], [8], but also for interference alignment [34] and signal to leakage and noise ratio beamforming [35]. In this section, we propose CSI feedback strategies for multi-carrier OFDM, extending the single-carrier quantization

metric (16), such as to determine the best subspace representation for a cluster of subcarriers. We compare this Grassmannian clustering approach to interpolation-based feedback, employing linear geodesic interpolation. As demonstrated in Section IV, clustering provides an advantage over interpolation in case the distance between subcarriers for which CSI is fed back, so called CSI pilots, is large compared to the coherence bandwidth of the channel.

To perform multi-user scheduling, the base station additionally requires information about the achievable transmission rates of the users. For this purpose, we derive a lower bound on the expected SINR of a user with imperfect CSIT in Section III-B, exploiting the BD construction.

A. Channel Subspace Feedback

a) Channel subsampling and geodesic interpolation: With channel subsampling, the receiver provides CSI feedback only for a subset of the OFDM subcarriers, i.e., the CSI pilots. After receiving the feedback, the transmitter determines the CSI on the remaining subcarriers by drawing equidistant samples of the geodesic between neighboring CSI pilots. We briefly summarize the approach below, as it has already been proposed in [14]; however, we employ different formulations of the tangent and geodesic, following [32], [33], [36].

Considering a specific CSI pilot n_1, we assume that user u employs the selfish maximum eigenmode transmission (MET) antenna ccombiner investigated in [5] to determine the effective channel matrix. The corresponding CSI quantization metric is

$$\Omega_{\mathrm{MET}}\left(\mathbf{H}_u[n_1], \mathbf{Q}_j\right) = \mathrm{d}_{\mathrm{c}}^2\left(\left[\mathbf{U}_u[n_1]\right]_{:,1:\ell_u}, \mathbf{Q}_j\right), \quad (17)$$

$$\mathbf{H}_u[n_1] = \mathbf{U}_u[n_1]\mathbf{\Sigma}_u[n_1]\mathbf{V}_u[n_1]^{\mathrm{H}}, \quad (18)$$

where (18) denotes an SVD of the channel matrix with singular values in decreasing order. With MET, the user selects the ℓ_u largest eigenmodes as the effective channel, providing a potentially large channel gain and corresponding high SNR. Given the quantized CSI at two consecutive pilot positions n_1 and n_2, the transmitter determines the tangent $\mathbf{T}_u[n_1]$ defining the geodesic between $\mathrm{span}(\hat{\mathbf{H}}_u[n_1])$ and $\mathrm{span}(\hat{\mathbf{H}}_u[n_2])$ according to [36]

$$\mathbf{U}\tan(\mathbf{\Phi})\mathbf{V}^{\mathrm{H}} = \left(\mathbf{I}_{N_t} - \hat{\mathbf{H}}_u[n_1]\hat{\mathbf{H}}_u[n_1]^{\mathrm{H}}\right) \hat{\mathbf{H}}_u[n_2]\left(\hat{\mathbf{H}}_u[n_1]^{\mathrm{H}}\hat{\mathbf{H}}_u[n_2]\right)^{-1}, \quad (19)$$

$$\mathbf{T}_u[n_1] = \mathbf{U}\mathbf{\Phi}\mathbf{V}^{\mathrm{H}}, \quad (20)$$

where $\mathbf{U}\tan(\mathbf{\Phi})\mathbf{V}^{\mathrm{H}}$ represents an SVD of the right-hand side of (19) and $\mathbf{\Phi}$ is calculated as the arc tangent of the diagonal matrix $\tan(\mathbf{\Phi})$. With the tangent from (20), the base station then obtains the linearly interpolated subspace at distance $\Delta n = (n_2 - n_1)\, p,\ p \in [0,1]$ from $\mathrm{span}(\hat{\mathbf{H}}_u[n_1])$, using the geodesic defined in [32]

$$\hat{\mathbf{H}}_u[n_1 + \Delta n] = \hat{\mathbf{H}}_u[n_1]\mathbf{V}\cos(\mathbf{\Phi}\, p)\mathbf{V}^{\mathrm{H}} + \mathbf{U}\sin(\mathbf{\Phi}\, p)\mathbf{V}^{\mathrm{H}}. \quad (21)$$

b) SQBC-based clustering: As an alternative to CSI interpolation we propose a novel CSI clustering scheme, which minimizes the average subspace chordal distance per cluster, implying minimal residual interference due to the mismatch between the channel and the clustered and quantized CSI. This approach is suitable for large CSI pilot distances compared to the channel coherence bandwidth, entailing unsatisfactory performance of linear interpolation due to significant channel variations in between CSI pilots. Below we consider a specific CSI feedback cluster and assume for simplicity that the corresponding N subcarriers are indexed by $n \in \mathcal{N} = \{1, \ldots, N\}$.

We define the best *unquantized Grassmannian subspace representation* $\tilde{\mathbf{H}}_u$ of the cluster by minimizing the average chordal distance with respect to the channel matrices $\mathbf{H}_u[n]$, $\forall\, n \in \mathcal{N}$

$$\tilde{\mathbf{H}}_u = \arg\min_{\tilde{\mathbf{H}}} \frac{1}{N}\sum_{n=1}^{N} \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{U}_u[n], \tilde{\mathbf{H}}\right), \quad \text{subject to : } \tilde{\mathbf{H}} \in \mathbb{C}^{N_t \times \ell_u},\ \tilde{\mathbf{H}}^{\mathrm{H}}\tilde{\mathbf{H}} = \mathbf{I}_{\ell_u}, \quad (22)$$

with $\mathbf{U}_u[n]$ being obtained from an SVD of the channel matrix as in (18). We can extend this approach to account for CSI quantization as well and define the best *quantized Grassmannian subspace representation* as

$$\hat{\mathbf{H}}_u = \arg\min_{\mathbf{Q}_j} \frac{1}{N}\sum_{n=1}^{N} \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{U}_u[n], \mathbf{Q}_j\right), \quad \text{subject to : } \mathbf{Q}_j \in \mathcal{Q}_u = \left\{\mathbf{Q}_j \in \mathbb{C}^{N_t \times \ell_u} | \mathbf{Q}_j^{\mathrm{H}}\mathbf{Q}_j = \mathbf{I}_{\ell_u}\right\}. \quad (23)$$

We obtain the solution to problem (22) according to

$$\tilde{\mathbf{H}}_u = \left[\bar{\mathbf{U}}_u\right]_{:,1:\ell_u} = \bar{\mathbf{U}}_u^{(\ell_u)}, \quad \bar{\mathbf{U}}_u\bar{\mathbf{\Lambda}}_u\bar{\mathbf{U}}_u^{\mathrm{H}} = \bar{\mathbf{R}}_u,\ \bar{\mathbf{R}}_u = \frac{1}{N}\sum_{n=1}^{N}\mathbf{U}_u[n]\mathbf{U}_u[n]^{\mathrm{H}}, \quad (24)$$

with $\bar{\mathbf{U}}_u\bar{\mathbf{\Lambda}}_u\bar{\mathbf{U}}_u^{\mathrm{H}}$ denoting an eigen-decomposition of $\bar{\mathbf{R}}_u$. Matrix $\bar{\mathbf{R}}_u$ can be interpreted as a *subspace correlation matrix*, with $\bar{\mathbf{\Lambda}}_u$ containing its eigenvalues in decreasing order. Similarly, the solution of (23) leads to the following quantization metric

$$\begin{aligned}\hat{\mathbf{H}}_u^{(\mathrm{SQBC})} &= \arg\min_{\mathbf{Q}_j \in \mathcal{Q}_u} \ell_u - \mathrm{tr}\left(\bar{\mathbf{\Lambda}}_u\left(\bar{\mathbf{U}}_u^{\mathrm{H}}\mathbf{Q}_j\mathbf{Q}_j^{\mathrm{H}}\bar{\mathbf{U}}_u\right)\right) \\ &\triangleq \arg\min_{\mathbf{Q}_j \in \mathcal{Q}_u} \mathrm{d}_{\mathrm{c,w}}^2(\bar{\mathbf{U}}_u, \mathbf{Q}_j, \bar{\mathbf{\Lambda}}_u) \\ &= \Omega_{\mathrm{SQBC}}\left(\{\mathbf{H}_u[1], \ldots, \mathbf{H}_u[N]\}, \mathbf{Q}_j\right). \end{aligned} \quad (25)$$

In this equation, we define the *weighted chordal distance* $\mathrm{d}_{\mathrm{c,w}}^2(\bar{\mathbf{U}}_u, \mathbf{Q}_j, \bar{\mathbf{\Lambda}}_u)$ with weighting matrix $\bar{\mathbf{\Lambda}}_u$. This diagonal weighting matrix specifies the importance of the individual eigenmodes of the subspace correlation matrix in the quantization problem (25). Notice that this quantization problem is a generalization of SQBC [5] from single-carrier systems to multi-carrier feedback. The corresponding antenna combiners $\mathbf{G}_u[n], \forall\, n \in \mathcal{N}$ that must be applied by user u to achieve this minimal quantization and representation error can be obtained similar to [5].

c) Dimensionality adaptation: SQBC considers all M_u eigenmodes of the channel matrix $\mathbf{H}_u[n]$ as equally important during quantization, which potentially causes a weak channel gain of the effective channel matrix [5]. This is also the case with SQBC-based clustering as proposed in (25). This can be problematic in case the multi-user transmission is noise limited, rather than interference limited, because it is then more important to achieve a large effective channel gain, instead of a minimal CSI quantization error. To account for such situations, we propose to trade-off between MET and SQBC by considering only the subset of d largest eigenmodes of the channel matrices during quantization, with $\ell_u \leq d \leq M_u$. Hence, instead of the weighted chordal distance quantization metric (25) we employ

$$\begin{aligned}\hat{\mathbf{H}}_u^{(d)} &= \underset{\mathbf{Q}_j \in \mathcal{Q}_u}{\arg\min}\ \mathrm{d}_{\mathrm{c,w}}^2\left(\bar{\mathbf{U}}_u^{(d)}, \mathbf{Q}_j, \bar{\mathbf{\Lambda}}_u^{(d)}\right), \\ &\quad \bar{\mathbf{U}}_u^{(d)} \bar{\mathbf{\Lambda}}_u^{(d)} \left(\bar{\mathbf{U}}_u^{(d)}\right)^{\mathrm{H}} = \bar{\mathbf{R}}_u^{(d)}, \\ \bar{\mathbf{R}}_u^{(d)} &= \frac{1}{N}\sum_{n=1}^{N} \mathbf{U}_u^{(d)}[n]\mathbf{U}_u^{(d)}[n]^{\mathrm{H}},\ \mathbf{U}_u^{(d)}[n] = [\mathbf{U}_u[n]]_{:,1:d}\,.\end{aligned} \tag{26}$$

With this metric, we determine the best ℓ_u-dimensional Grassmannian subspace representation in the quantization codebook $\mathcal{Q}_u$ in terms of the average chordal distance to the d largest eigenmodes of the channel matrices within the cluster. Notice that this does not imply changing the number of streams ℓ_u, as is commonly considered in MIMO spatial multiplexing; we only select the dimension d of the *search space*, in which we find the best ℓ_u-dimensional subspace.

The natural question that arises is how to select the dimension d. To answer this question, we propose to employ an estimate of the pre-equalization achievable transmission rate with BD precoding and quantized CSIT. According to [37], the instantaneous mutual information between the channel input and output on subcarrier n, determining the achievable transmission rate, is

$$\begin{aligned}R_u^{(d)}[n] = \log_2 \det\Bigg(&\mathbf{I}_{M_u} + \mathbf{H}_u[n]^{\mathrm{H}}\mathbf{S}_u[n]\mathbf{H}_u[n] \\ &\left(\sigma_z^2\mathbf{I}_{M_u} + \mathbf{C}_u^{(d)}[n]\right)^{-1}\Bigg),\end{aligned} \tag{27}$$

with $\mathbf{S}_u[n]$ denoting the covariance matrix of the channel input and $\mathbf{C}_u^{(d)}[n]$ being the interference covariance matrix. At the time of calculation of the CSI feedback, user u does not have knowledge of the precoders because the multi-user schedule is not yet fixed and the CSI of the other users is not available to user u. To determine an estimate of the mutual information, we hence consider the precoders as random and take them into account in the covariance matrices[2]

$$\mathbf{S}_u[n] = \mathbb{E}\Big(\mathbf{F}_u[n]\mathbf{x}_u[n](\mathbf{F}_u[n]\mathbf{x}_u[n])^{\mathrm{H}}\Big) = \mathbb{E}\big(\mathbf{F}_u[n]\mathbf{F}_u[n]^{\mathrm{H}}\big), \tag{28}$$

[2]Notice that this implies regarding the precoders as part of the channel input that should be estimated by the receiver.

$$\mathbf{C}_u^{(d)}[n] = \sum_{s=1, s\neq u}^{S} \mathbf{H}_u[n]^{\mathrm{H}}\mathbb{E}\left(\mathbf{F}_s[n]\mathbf{F}_s[n]^{\mathrm{H}}\right)\mathbf{H}_u[n], \tag{29}$$

where the expectation is with respect to the transmit signals and precoders. The precoder $\mathbf{F}_u[n]$ depends only on the channels of the other users. As these channels are unknown to user u, we assume $\mathbf{F}_u[n]$ to be isotropically distributed, implying

$$\mathbf{S}_u[n] = P_u\mathbb{E}\left(\tilde{\mathbf{F}}_u[n]\tilde{\mathbf{F}}_u[n]^{\mathrm{H}}\right) = \frac{P_t}{N_t S}\mathbf{I}_{\ell_u}, \tag{30}$$

with $\tilde{\mathbf{F}}_u[n]$ as defined in (10). Due to the BD construction, the precoders $\mathbf{F}_s[n]$ of the other users are restricted to the left null space of $\hat{\mathbf{H}}_u^{(d)}$. Considering this in the calculation of $\mathbf{C}_u^{(d)}[n]$, and assuming that $S = N_t/\ell_u$ users are served in parallel each over ℓ_u streams, we obtain

$$\mathbf{C}_u^{(d)}[n] = \frac{P_t}{N_t}\mathbf{H}_u[n]^{\mathrm{H}}\left(\mathbf{I}_{M_u} - \hat{\mathbf{H}}_u^{(d)}\left(\hat{\mathbf{H}}_u^{(d)}\right)^{\mathrm{H}}\right)\mathbf{H}_u[n]. \tag{31}$$

With these estimates, we select the *preferred dimensionality* d_u by maximizing the sum rate

$$d_u = \underset{\ell_u \leq d \leq M_u}{\arg\max} \sum_{n=1}^{N} R_u^{(d)}[n]. \tag{32}$$

d) Quantization codebooks: In the simulation-based study conducted in Section V, we employ memoryless and predictive quantization codebooks to quantize the subspaces selected with the methods proposed above to provide CSIT. We present the constructions of the applied codebooks in [8], [10], [16], respectively. In case of memoryless quantization, the users quantize the CSI at each TTI independently utilizing predefined quantization codebooks. We consider memoryless quantization based on random vector quantization (RVQ) and correlated RVQ. Correlated RVQ exploits the pathloss difference experienced in DASs to provide more accurate CSIT by matching the statistics of the codebook to the channel gain matrix (2) [16]. This is achieved by determining the subspace codebook from random matrices that are distributed according to (2), assuming $\bar{\mathbf{H}}_u[n]$ as i.i.d. Gaussian:

$$\begin{aligned}\mathcal{Q}_u^{(\mathrm{corr})} = \Big\{\mathbf{Q}_j^{(\mathrm{corr})}\ \Big|&\mathbf{Q}_j^{(\mathrm{corr})}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}} = \mathbf{\Gamma}_u^{1/2}\bar{\mathbf{H}}_j, \\ &\bar{\mathbf{H}}_j \in \mathbb{C}^{N_t \times M_u}, [\bar{\mathbf{H}}_j]_{m,n} \sim \mathcal{N}_{\mathbb{C}}(0,1)\Big\},\end{aligned} \tag{33}$$

with $\mathbf{Q}_j^{(\mathrm{corr})}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ representing a compact SVD of $\mathbf{\Gamma}_u^{1/2}\bar{\mathbf{H}}_j$. Simulation results are averaged over random realizations of the matrices $\bar{\mathbf{H}}_j$. If the number of streams ℓ_u is less than M_u, we further tailor $\mathcal{Q}_u^{(\mathrm{corr})}$ to SQBC and MET, respectively. In case of MET, the ℓ_u largest eigenmodes of $\mathbf{\Gamma}_u^{1/2}\bar{\mathbf{H}}_j$ are quantized; hence, we include only the first ℓ_u columns of the matrices $\mathbf{Q}_j^{(\mathrm{corr})}$, corresponding to the ℓ_u largest eigenmodes, in our quantization codebook:

$$\mathcal{Q}_u^{(\mathrm{MET})} = \left\{\mathbf{Q}_j^{(\mathrm{MET})} = \left[\mathbf{Q}_j^{(\mathrm{corr})}\right]_{:,1:\ell_u}\ \middle|\ \mathbf{Q}_j^{(\mathrm{corr})} \in \mathcal{Q}_u^{(\mathrm{corr})}\right\}. \tag{34}$$

With SQBC, on the other hand, there exists no preferred subspace orientation within $\mathrm{span}(\mathbf{\Gamma}_u^{1/2}\bar{\mathbf{H}}_j)$; hence, we include randomly drawn ℓ_u dimensional subspaces of $\mathrm{span}(\mathbf{Q}_j^{(\mathrm{corr})})$ in the codebook:

$$\mathcal{Q}_u^{(\mathrm{SQBC})} = \left\{ \mathbf{Q}_j^{(\mathrm{SQBC})} = \mathbf{Q}_j^{(\mathrm{corr})}\mathbf{U}_j \,\middle|\, \mathbf{Q}_j^{(\mathrm{corr})} \in \mathcal{Q}_u^{(\mathrm{corr})} \right. , \quad \left. \mathbf{U}_j \in \mathbb{C}^{M_u \times \ell_u}, \mathbf{U}_j^{\mathrm{H}}\mathbf{U}_j = \mathbf{I}_{\ell_u} \right\}. \quad (35)$$

Here, multiplication of $\mathbf{Q}_j^{(\mathrm{corr})}$ with the isotropically distributed matrix $\mathbf{U}_j$ generates an orthonormal basis $\mathbf{Q}_j^{(\mathrm{SQBC})}$ that spans a uniformly distributed subspace within $\mathrm{span}(\mathbf{Q}_j^{(\mathrm{corr})})$. Similarly, in case we consider dimensionality adaptation, the SQBC codebook is determined from the d columns of $\mathbf{Q}_j^{(\mathrm{corr})}$ corresponding to the d largest eigenmodes. In case of predictive quantization, the quantization codebook adapts to the temporal evolution of the channel to achieve high-fidelity quantization in low to moderate mobility scenarios [8], [10].

We finally mention that Grassmannian feedback is not only relevant for BD-based MU-MIMO, but also for single-user MIMO [38] and for MIMO interference alignment [34]. Grassmannian clustering is an appropriate approach whenever the achievable transmission rate is determined by the residual multi-user interference due to CSIT imperfections, i.e., for BD and interference alignment in the high SNR regime [3]–[5], [34]. At low to intermediate SNR alternative quantization and clustering metrics that explicitly account for the channel gain may obtain better results.

e) Comments on complexity: To determine the CSI feedback for MET with interpolation, the users have to apply a single SVD per CSI pilot to obtain the largest eigenmodes of the channel matrix. This is advantageous compared to SQBC-based clustering, where the users have to apply an SVD on each subcarrier to calculate the subspace correlation matrix. On the other hand, from the perspective of the transmitter, SQBC-based clustering does not require any further processing of the CSI feedback, whereas MET feedback necessitates geodesic interpolation to derive the CSI in between pilot positions for each user. It thus depends on the capabilities of the transmitter and the users which of the two strategies is favorable in terms of complexity.

B. Channel Quality Feedback

To determine the multi-user resource allocation, the base station estimates the achievable data rate of a given schedule. For this purpose, we propose BD specific channel quality feedback in this section. The achievable rate depends on the set of users that is served in parallel; therefore, it must be possible to update the channel quality feedback at the base station depending on the considered user set. In practical communication systems, the post-equalization SINR is commonly employed for rate adaptation and thus determines the achievable throughput [39]. We hence consider an estimate of the post-equalization SINR as CQI. As the multi-user schedule and the precoders are unknown at the time the feedback is calculated, we propose to apply a lower bound on the expected value of the SINR similar to the bound derived in [23] for zero-forcing beamforming. We provide the details of the derivation in Appendix. The lower bound on the expected SINR of user u on stream ν according to (61) is

$$\begin{aligned}
\tilde{\beta}_{\nu,u}[n] &= \frac{c_S[n]\sigma_{\nu,u}^{\mathrm{eff}}[n]^2}{\sigma_z^2 + c_I[n]\sigma_{\nu,u}^{\mathrm{eff}}[n]^2}, \\
c_S[n] &= \frac{P_t}{S(N_t - \bar{\ell}_u)} \left(1 - \frac{\mathrm{d_c^2}(\hat{\mathbf{H}}_u[n], \mathbf{B}_u[n])}{\ell_u}\right) \\
&\quad \left(1 - \frac{\mathrm{d_c^2}(\hat{\mathbf{H}}_u[n], \tilde{\mathbf{H}}_u[n])}{\ell_u}\right), \\
c_I[n] &= \frac{P_t}{N_t \ell_u}\, \mathrm{d_c^2}(\hat{\mathbf{H}}_u[n], \tilde{\mathbf{H}}_u[n]), \\
\mathbf{\Sigma}_u^{\mathrm{eff}}[n] &= \mathrm{diag}\left(\sigma_{1,u}^{\mathrm{eff}}[n], \ldots, \sigma_{\ell_u,u}^{\mathrm{eff}}[n]\right). \quad (36)
\end{aligned}$$

with $\mathbf{\Sigma}_u^{\mathrm{eff}}[n]$ denoting the matrix of singular values of the effective user channel $\mathbf{H}_u^{\mathrm{eff}}[n]$, $\tilde{\mathbf{H}}_u[n]$ as defined in (7), $\hat{\mathbf{H}}_u[n]$ being the quantized channel subspace, $\mathbf{B}_u[n]$ denoting an orthonormal basis for the orthogonal complement of the other served users' quantized effective channels (see (48)) and $\bar{\ell}_u = \sum_{s \in \mathcal{S}\setminus u} \ell_s$. We observe that the residual multi-user interference is determined by the chordal distance quantization error in $c_I[n]$.

Considering the constant $c_S[n]$, we see that the first two factors cannot be determined by the user u, because S, $\bar{\ell}_u$ and the other users' effective channels are unknown to user u. However, the user can calculate the last term, which depends on the quantization error. Also, the user knows the constant $c_I[n]$ and the singular values in $\mathbf{\Sigma}_u^{\mathrm{eff}}[n]$, because they depend on local CSI only. We therefore propose the following value as per-stream CQI feedback

$$\mathrm{CQI}_{\nu,u}[n] = \frac{\sigma_{\nu,u}^{\mathrm{eff}}[n]^2 \left(1 - \frac{\mathrm{d_c^2}(\hat{\mathbf{H}}_u[n], \tilde{\mathbf{H}}_u[n])}{\ell_u}\right)}{\sigma_z^2 + \frac{P_t}{N_t \ell_u}\, \mathrm{d_c^2}(\hat{\mathbf{H}}_u[n], \tilde{\mathbf{H}}_u[n]) \sigma_{\nu,u}^{\mathrm{eff}}[n]^2}. \quad (37)$$

With this feedback, the base station can estimate the achievable user rate $R_u[n]$ for a given schedule $\mathcal{S}$, because the remaining two factors of $c_S[n]$ are available at the transmitter

$$\tilde{\beta}_{\nu,u}[n] = \frac{P_t}{S(N_t - \bar{\ell}_u)} \left(1 - \frac{\mathrm{d_c^2}(\hat{\mathbf{H}}_u[n], \mathbf{B}_u[n])}{\ell_u}\right) \mathrm{CQI}_{\nu,u}[n], \quad (38)$$

$$R_u[n] \approx \sum_{\nu=1}^{\ell_u} \log_2\left(1 + \tilde{\beta}_{\nu,u}[n]\right). \quad (39)$$

With feedback clustering or interpolation, the users provide CQI feedback only once per cluster. Then, we propose to calculate an average CQI for each cluster, by linearly averaging the quantization error and the squared singular values.

Utilizing the proposed CQI feedback for multi-user scheduling, we obtain reasonably close to optimal results as demonstrated in Section IV. Still, for transmission rate adaptation the SINR estimate is not sufficiently accurate. Instead, we consider

TABLE II
SUS-BASED MULTI-USER SCHEDULING ALGORITHM

Initialize:	Weighted sum rate $R = 0$
	Set of scheduled users $\mathcal{S} = \{\}$
	Set of potential users $\mathcal{P} = \{1, \ldots, U\}$
	Number of scheduled users $S = 0$
Repeat the following steps until the maximum number of $S = \frac{N_t}{L}$ users is served:	
1	Calculate an orthonormal basis $\mathbf{B}$ for the space spanned by the channels of the users in $\mathcal{S}$ ($\mathbf{B} = \mathbf{0}$ if $\mathcal{S} = \{\}$)
2	Find the semi-orthogonal user set $\tilde{\mathcal{S}}$, by determining all users $s \in \mathcal{P}$ satisfying the SUS condition: $\mathrm{tr}\left(\hat{\mathbf{H}}_s^{\mathrm{H}} \mathbf{B}\mathbf{B}^{\mathrm{H}} \hat{\mathbf{H}}_s\right) \leq \alpha_{\mathrm{SUS}} L$
3	If the SUS user set $\tilde{\mathcal{S}}$ is empty, stop the algorithm and serve the users in $\mathcal{S}$
4	Otherwise, calculate the estimated achievable rate R_s in Eq. (39) for all users $s \in \tilde{\mathcal{S}}$ under the assumption that user s is served in parallel with the users in $\mathcal{S}$
5	Determine the user $s \in \tilde{\mathcal{S}}$ that achieves the largest weighted rate: $\hat{s} = \mathrm{argmax}_{s \in \tilde{\mathcal{S}}} \frac{R_s}{T_s}$ with T_s denoting the average throughput of user s achieved over the past
6	Update the estimated achievable rates $R_s, \forall s \in \mathcal{S}$, assuming that $\hat{s}$ is served in addition to the users in $\mathcal{S}$
7	Calculate the weighted sum rate of the schedule $\{\mathcal{S}, \hat{s}\}$: $\bar{R} = \sum_{s \in \{\mathcal{S}, \hat{s}\}} \frac{R_s}{T_s}$
8	If $\bar{R} \geq R$, add user $\hat{s}$ to $\mathcal{S}$ and update S; set $R = \bar{R}$ and remove $\hat{s}$ from $\mathcal{P}$
	Otherwise, stop the algorithm and serve the users in $\mathcal{S}$

(37) only as an initial CQI for multi-user scheduling. As soon as the schedule is fixed, the users can estimate the instantaneous SINR accurately and feed back this value for rate adaptation. Naturally this approach is only useful if the schedule is fixed for several TTIs, implying a loss of temporal multi-user diversity. However, accounting for the downlink signaling overhead involved in changing the resource allocation, the performance loss is reasonable in the considered low to moderate mobility scenarios.

C. Multi-User Scheduling

The throughput achieved with BD precoding depends strongly on the set of users $\mathcal{S}$ that is served in parallel. To avoid the complexity of an exhaustive search, we propose a greedy proportional fair scheduler that is based on a preselection of users utilizing the SUS algorithm [22] and exploits the CQI feedback as well the transmission rate estimate derived in the previous section to determine the final user set. We summarize the scheduling procedure in Table II for $\ell_s = L, \forall\, s$ and N_t/L is integer-valued; otherwise, the scheduler must perform an explicit validation of the feasibility conditions of BD precoding. A total number of U users is attached to the base station. In step 2, the scheduler performs a preselection of users, based on their subspace distance to the already served users, invoking the SUS principle. Only if a user is close to orthogonal to the served users, he is considered as a potential additional user. The SUS parameter α_{SUS} specifies the meaning of "close to orthogonal" [22]. In the presented simulations we set $\alpha_{\mathrm{SUS}} = 0.35$; we observed a dependency of the optimal value of α_{SUS} on the number of users and the antenna configuration. The value $\alpha_{\mathrm{SUS}} = 0.35$ was selected as a good compromise for all considered scenarios; slight improvements may be achieved by optimizing α_{SUS} for each specific situation. The proposed scheduler, however, employs the semi-orthogonal user set $\tilde{\mathcal{S}}$ only as a preselection of users for complexity reduction. Given $\tilde{\mathcal{S}}$, it determines in step 5 the user $\hat{s} \in \tilde{\mathcal{S}}$ that achieves the highest rate when scheduled in parallel with the already selected users in $\mathcal{S}$. Provided adding this user improves the weighted sum rate, it is included in the schedule $\mathcal{S}$; otherwise the algorithm is terminated. In contrast to pure SUS scheduling, where the number of users served in parallel is basically determined by the parameter α_{SUS} only, we hence employ a greedy scheduler utilizing the proposed rate estimate to decide on the number of served users.

IV. PERFORMANCE INVESTIGATION

In this section, we investigate the performance of the proposed CSI feedback strategies and the efficiency of the multi-user scheduler by means of Monte-Carlo simulations. To quantify the frequency-selectivity of the wireless channel, we employ the normalized sampling bandwidth

$$B_s = \Delta f_s N \tau_{\mathrm{RMS}}, \tag{40}$$

with Δf_s [Hz] being the OFDM subcarrier spacing, N denoting the cluster size and τ_{RMS} specifying the root mean square delay spread of the channel. We also specify the feedback reduction ratio (FRR) achieved by feedback clustering, i.e., the number of subcarriers of the system divided by the number of clusters used for CSI feedback. Notice, however, that the performance achieved at a specific FRR depends on the coherence bandwidth of the wireless channel, which is explicitly accounted for in the normalized sampling bandwidth. The minimal FRR considered is equal to 12, corresponding to one CSI feedback value per LTE resource block. We employ the spatial Kronecker correlation model defined in [40] to determine the correlation between the entries of the channel matrix assuming receive-side correlation only, specified with the single parameter $\alpha_{\mathrm{corr}} \in [0, 1]$ ($\alpha_{\mathrm{corr}} = 0$ means no correlation). We conduct standard-compliant link-level throughput simulations of a single cell utilizing the Vienna LTE-A link-level simulator [25]. We do not consider distributed antennas in this section, i.e., $\mathbf{\Gamma}_u = \mathbf{I}_{N_t}$.

In the first simulation, we cross-compare the chordal distance MSE obtained by MET with geodesic interpolation and SQBC-based clustering. We consider a channel of size $N_t \times M_u = 8 \times 4$ and select an $\ell_u = 1$-dimensional subspace as CSI feedback. The system bandwidth is 5 MHz (300 subcarriers) and

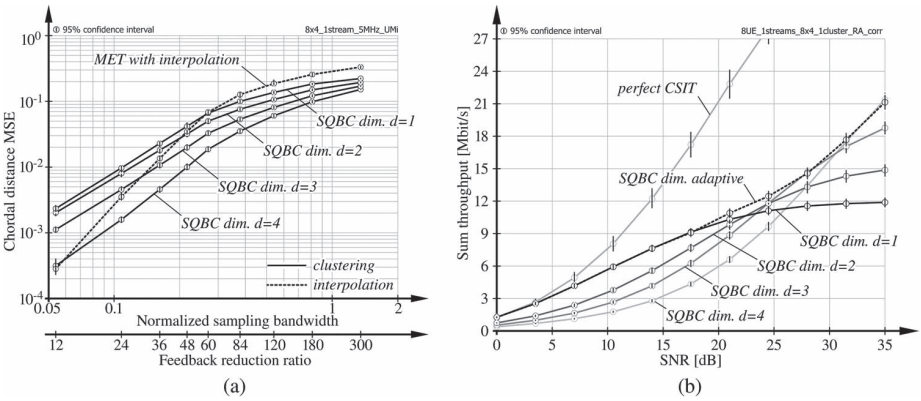


Fig. 1. Chordal distance MSE of feedback clustering and interpolation and throughput investigation of SQBC-based clustering. (a) MSE of feedback clustering and interpolation. (b) Impact of the SQBC dimensionality (FRR $= 72$).

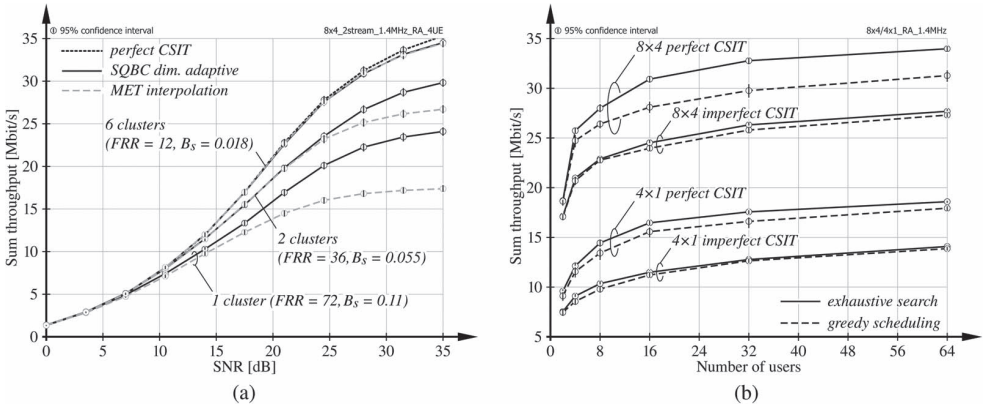


Fig. 2. Comparison of SQBC-based clustering and MET with interpolation and investigation of SUS-based greedy scheduling. (a) Throughput in dependence of the cluster size. (b) Greedy scheduling versus exhaustive search.

the channel follows the SCME *urban micro* model [41] having a coherence bandwidth [42] of 680 kHz. The users provide unquantized feedback; the chordal distance error is therefore only due to the clustering and interpolation, respectively.

Fig. 1(a) shows the results of the simulation. We observe that SQBC clustering achieves a substantial MSE improvement over geodesic interpolation at large CSI pilot distance (respectively cluster size) N. The performance of SQBC clustering depends on the dimensionality d considered in (26). The value of d determines the dimension of the *search space* in the optimization problem (26); the larger d, the better the subspace representation. With SQBC clustering we thus exploit the degrees of freedom provided by the $M_u \geq \ell_u$ excess receive antennas to improve the performance of clustering. This, however, implies a channel gain reduction, which is significant in case of spatially correlated channels as observed below.

In the next simulation, we investigate the throughput reduction with respect to perfect CSIT due to SQBC-based feedback clustering. In this simulation, we employ the *rural area* channel model [43] with a coherence bandwidth of 2 MHz and we assume strongly correlated receive antennas ($\alpha_{\mathrm{corr}} = 0.9$). The users employ a single CSI feedback cluster to represent the 1.4 MHz system bandwidth ($\mathrm{FRR} = 72$, $B_s = 0.11$) and apply the interference-averaged MMSE equalizer proposed in [44], which exploits the BD construction to estimate the residual multi-user interference and reduce its impact on the intended signal.

As shown in Fig. 1(b), at low SNR a dimensionality of $d = 1$ achieves the best results. With $d = 1$, the subspace representation is obtained from the maximum eigenmodes of the channels. The BD precoder ensures that the selected subspace is free of interference, thus, enabling a large channel gain over the interference-free subspace. Due to clustering, however, the single subspace representation is imperfect for a given subcarrier, implying residual multi-user interference after precoding. Therefore, the throughput is interference-limited at high SNR. With a larger dimensionality $d > 1$, the chordal distance error is decreased, causing a reduction of the residual multi-user interference. Hence, we can trade-off the channel gain for the residual interference by varying d. The optimal dimensionality depends on the SNR. As demonstrated in Fig. 1(b), our proposed dimensionality adaptation is able to identify the optimal dimensionality. Notice that varying d has no impact on the number of transmitted data streams, i.e., $\ell = 8$ in all cases. Due to the strong receive antenna correlation, causing significant singular value disparity of the channel [45], the gap between $d = 1$ and $d > 1$ is large at low SNR.

In Fig. 2(a) we consider the same scenario as above, but with only four users getting $\ell_u = 2$ streams each, assuming uncorrelated receive antennas $\alpha_{\mathrm{corr}} = 0$. In this investigation,

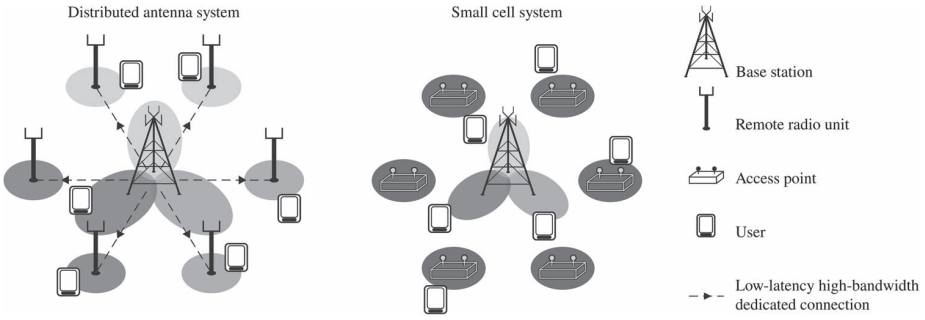


Fig. 3. Investigated heterogeneous cellular networking architectures.

the users apply an increasing number of feedback clusters to reduce the throughput gap with respect to perfect CSIT. We compare the performance achieved by MET feedback with geodesic interpolation to SQBC-based clustering with dimensionality adaptation. We observe that SQBC feedback outperforms MET at high SNR and with large cluster sizes, due to the reduced subspace error. With dimensionality adaptation clustering performs never worse than interpolation.

Finally, we investigate the efficiency of the proposed scheduler in Fig. 2(b), by comparing to the throughput achieved with the optimal exhaustive search schedule. We consider two different configurations, i.e., $N_t \times M_u = 4 \times 1$ with $\ell_u = 1$ and $N_t \times M_u = 8 \times 4$ with $\ell_u = 2$. Hence, the base station serves at most four users in parallel in both cases. We apply the *rural area* channel model with $\alpha_{\text{corr}} = 0$ and use a single feedback cluster to represent the 1.4 MHz system bandwidth (FRR $= 72, B_s = 0.11$). All users experience an average SNR of 20 dB. We observe that the proposed greedy scheduler obtains a similar multi-user diversity as the exhaustive search scheduler, albeit an approximately constant rate loss that is dependent on the configuration of the system. Especially in the realistic situation of imperfect CSIT, the greedy algorithm achieves a throughput that is close to the rate obtained with the optimal schedule at a significantly reduced complexity.

V. Network Architecture Comparison

In this chapter, we apply the proposed methods and algorithms in a simulation-based study to obtain a realistic performance comparison of two heterogeneous cellular networking architectures, that is, a DAS and a system comprising uncoordinated small cells as visualized in Fig. 3.

In the DAS, each sector of the 120° sectorized network contains, in addition to the four antennas at the macro base station, two RRU with two omni-directional antennas each, forming a distributed MIMO system. We place the RRU uniformly on a ring of radius $(2/3)r_c$ around the macro base station, with $r_c = 500$ m being the nominal cell radius. The base station has a transmit power of $P_t = 43$ dBm, which can be arbitrarily divided among the antenna arrays. We employ the 2D antenna gain pattern defined in [46] for the macro base stations, having a 3 dB beamwidth of 65°, a maximum gain of 15 dB and a maximum attenuation of 20 dB.

In the small cell system, we replace the RRU with uncoordinated micro base stations that are equipped with two transmit antennas and have a transmit power of $P_t = 37$ dBm each. These micro base stations operate autonomously in the same 1.4 MHz wide frequency band as the macro base station. We reduce the transmit power of the macro base stations to $P_t = 40$ dBm to have the same total transmit power of $(2 \cdot 37 + 40)$ dBm $= 43$ dBm in both scenarios. Notice that the total number of transmit antennas is the same as well.

TABLE III
Simulation Parameters Applied for the Networking Architecture Comparison

Parameter	Value
System bandwidth	1.4 MHz (72 subcarriers)
Carrier frequency	$f_c = 2$ GHz
Noise power spectral density	-174 dBm/Hz
Channel model	pedestrian A [47]
Number of receive antennas	$M_u = 2$
Transmit antennas per sector	$N_t = 8$
Total transmit power per sector	$P_t = 43$ dBm
Number of users per sector	$U \in [2, 32]$
Number of data streams per user	$\ell_u = 1$
Number of feedback clusters	1 (FRR $= 72$, $B_s = 0.05$)
Number of feedback bits	$b \in [1, 10]$
Spatial correlation parameter	$\alpha_{\text{corr}} = 0$
Maximum Doppler frequency	$f_d = 10$ Hz
Multi-user scheduling	proportional fair + SUS
MIMO receiver	MMSE (interference averaged) [44]

We consider $U \in [2, 32]$ users per sector that are uniformly distributed over the cell area and are equipped with $M_u = 2$ antennas each. The users are served over a single stream, $\ell_u = 1$, by the strongest base station; hence, we do not apply load balancing [48]. The received signal strength and thus the channel gain matrix $\mathbf{\Gamma}_u$ depend on the pathloss, according to the model specified in [46, Section 4.5.2]. We do not take shadow fading into account. This simplifies the simulations, because the region of interest is clearly delimited by the geometry of the cell. Also, with shadow fading a uniform placement of RRU is not reasonable. The *PedA* channel model [47] with a coherence bandwidth of 4.4 MHz determines the frequency selectivity of the channel, while the temporal evolution follows Clarke's model. We assume a low mobility scenario with a normalized Doppler frequency of $\nu_d = T_s f_d = 1$ ms $\cdot$ 10 Hz $= 0.01$, with T_s being the LTE subframe duration and f_d denoting the Doppler frequency. We conduct extensive link-level simulations of the three sectors shown in Fig. 3. We also consider other cells that are placed on a hexagonal grid, but the other-cell interference caused by these cells is only modeled, utilizing the Gamma distribution model proposed in [15], [16]. This model determines the first- and second-order moments of the

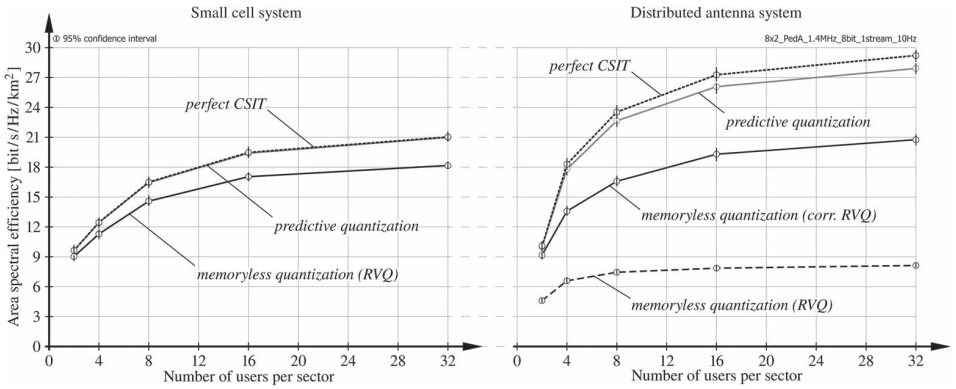


Fig. 4. Area spectral efficiency achieved in the DAS and the small cell system in dependence of the number of users per sector. An $N_t \times M_u = 8 \times 2$ antenna configuration is considered, transmitting $\ell_u = 1$ stream per user.

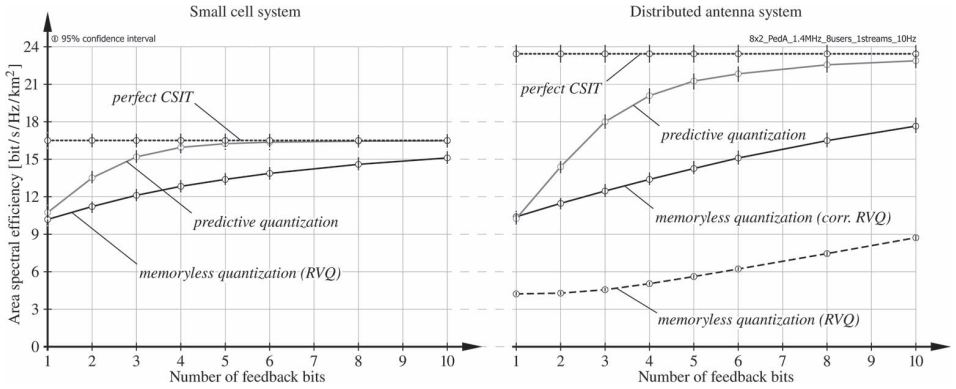


Fig. 5. Area spectral efficiency achieved in the DAS and the small cell system in dependence of the feedback overhead. An $N_t \times M_u = 8 \times 2$ antenna configuration is considered, serving $U = 8$ user over $\ell_u = 1$ stream per user.

interference power observed at a given user position and matches a Gamma distribution to these moments, which is then applied to randomly generate interference. We evaluate the results in terms of the ASE, employing the discrete approximation presented in [15]. Important simulation parameters are summarized in Table III.

In Fig. 4, we plot the ASE in dependence of the number of users, assuming a single feedback cluster $(\text{FRR} = 72)$ that is quantized with $b = 8$ bits per subframe. Comparing the two networking architectures with perfect CSIT, we observe a gain of approximately 40–45% with the DAS, because it enables coherent data transmission from all distributed antennas. With memoryless quantization, using RVQ and correlated RVQ respectively (see Section III-A), this gain reduces to 15–25%, due to the larger residual multi-user interference incurred in the DAS with imperfect CSIT. Notice that the DAS with RVQ does not achieve a multiplexing gain, because the CSIT is not accurate enough; hence, the scheduler only selects one user per TTI. This is because the RVQ codebook does not account for the pathloss differences experienced with respect to different RRU and hence the transmitter does not know which RRU are favorable for the user. Correlated RVQ accounts for the pathloss differences in the codebook construction, avoiding such issues. With predictive quantization we obtain close to optimal performance in the considered low mobility scenario. The gap between perfect CSIT and quantized CSIT is smaller in the small cell system than in the DAS. This is because in the small cell system users only provide CSI to the macro base station or the micro base station they are attached to, i.e., either a 4×2 or a 2×2 channel matrix is quantized, whereas in the DAS the channel matrix is of size 8×2.

In the second simulation, we evaluate the impact of the feedback overhead, to assess the robustness of the observed gains with respect to CSIT imperfections. Fig. 5 presents the results assuming $U = 8$ users per sector. The small cell system achieves almost perfect performance with memoryless quantization and 10 bit of feedback per TTI; with predictive quantization already 4 bit of feedback provide quasi-perfect CSIT in this low mobility scenario. In the DAS, the difference between quantized and perfect CSIT is larger, thus requiring a larger feedback overhead to get close-to-optimal performance. Importantly, the DAS always outperforms the uncoordinated small cell system in our investigations, provided the pathloss differences of the channel gain matrix $\mathbf{\Gamma}_u$ are exploited during quantization, i.e., with correlated RVQ and with predictive quantization. If the quantizer ignores the pathloss differences (RVQ in Fig. 5) the efficiency is strongly reduced.

VI. Conclusion

In this paper, we propose an efficient CSI feedback clustering method for BD-based multi-user MIMO-OFDM with limited

feedback. Based on the observation that channel subspace feedback is sufficient for BD precoding with equal power allocation, we derive the best Grassmannian subspace representation for a cluster of OFDM subcarriers in terms of minimizing the average chordal distance error, which determines the residual multi-user interference due to imperfect CSIT. Especially with excess receive antennas and a low density of CSI pilots, we obtain a significant performance improvement over interpolation-based methods. We propose an effective greedy multi-user scheduler that is based on a lower bound of the expected per-stream SINR of the users and achieves close to optimal performance with quantized CSIT. We apply the proposed methods to compare the performance of small cell systems and DASs with limited feedback and observe spectral efficiency gains of DASs over uncoordinated small cells. We can interpret this evaluation as an indicator for the performance gain that can be expected by enabling joint transmission CoMP in small cell networks compared to uncoordinated transmission. More detailed investigations are, however, required considering shadow fading, more realistic positioning of the RRU, varying node density, etc., to confirm these trends.

APPENDIX
SINR LOWER BOUND FOR BLOCK-DIAGONALIZATION PRECODING

In this appendix, we provide the derivation of the SINR lower bound for limited feedback-based BD precoding, extending the zero-forcing beamforming bound of [23] to multiple streams per user. We omit the index $[n]$ for brevity. The expected value of the post-equalization SINR of user u, assuming that the user applies a semi-unitary antenna combiner, is

$$\mathbb{E}(\beta_u) = \mathbb{E}\left(\frac{\mathrm{tr}(\mathbf{S}_u)}{\mathrm{tr}\left(\sigma_z^2\mathbf{I}_{\ell_u}+\mathbf{R}_u\right)}\right) \geq \frac{\mathbb{E}(\mathrm{tr}\left(\mathbf{S}_u\right))}{\sigma_z^2\ell_u+\mathbb{E}\left(\mathrm{tr}(\mathbf{R}_u)\right)}, \tag{41}$$

$$\mathbf{S}_u = \left(\mathbf{H}_u^{\mathrm{eff}}\right)^{\mathrm{H}}\mathbf{F}_u\mathbf{F}_u^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff}} = P_u\left(\mathbf{H}_u^{\mathrm{eff}}\right)^{\mathrm{H}}\tilde{\mathbf{F}}_u\tilde{\mathbf{F}}_u^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff}}, \tag{42}$$

$$\mathbf{R}_u = \sum_{s\in\mathcal{S},s\neq u} P_s\left(\mathbf{H}_u^{\mathrm{eff}}\right)^{\mathrm{H}}\tilde{\mathbf{F}}_s\tilde{\mathbf{F}}_s^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff}} = \sum_{s\in\mathcal{S},s\neq u}\mathbf{R}_u^{(s)}, \tag{43}$$

where we utilized Jensen's inequality $\mathbb{E}(f(x)) \geq f(\mathbb{E}(x))$ to obtain the lower bound in (41), with $f(x) = a/(b+x), x = \mathrm{tr}(\mathbf{R}_u), a = \mathbb{E}(\mathrm{tr}(\mathbf{S_u})), \mathrm{b} = \sigma_{\mathrm{z}}^2\ell_{\mathrm{u}}$. We next derive the expected values of the covariance matrices of the signal term and the interference terms.

a) Signal term: Applying the channel decomposition of (14), we write the covariance matrix of the intended signal of user u as

$$\mathbb{E}(\mathbf{S}_u) = P_u\mathbb{E}\left(\left(\mathbf{H}_u^{\mathrm{eff,r}}+\mathbf{H}_u^{\mathrm{eff,n}}\right)^{\mathrm{H}}\tilde{\mathbf{F}}_u\tilde{\mathbf{F}}_u^{\mathrm{H}}\left(\mathbf{H}_u^{\mathrm{eff,r}}+\mathbf{H}_u^{\mathrm{eff,n}}\right)\right), \tag{44}$$

At the time the CQI feedback is calculated, the precoder $\tilde{\mathbf{F}}_u$ is unknown. Because the precoder is determined solely by the channels of the other served users according to (5), which are unknown to user u, we assume the precoder to be isotropically distributed. It follows that the expected values of the mixed terms between $\mathbf{H}_{u,i}^{\mathrm{eff,r}}$ and $\mathbf{H}_{u,i}^{\mathrm{eff,n}}$ are equal to zero, e.g.,

$$\left(\mathbf{H}_u^{\mathrm{eff,r}}\right)^{\mathrm{H}}\mathbb{E}\left(\tilde{\mathbf{F}}_u\tilde{\mathbf{F}}_u^{\mathrm{H}}\right)\mathbf{H}_u^{\mathrm{eff,n}} = \frac{\ell_u}{N_t}\left(\mathrm{H}_u^{\mathrm{eff,r}}\right)^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff,n}} = \mathbf{0}, \tag{45}$$

$$\Rightarrow \mathbb{E}(\mathbf{S}_u) = P_u\mathbb{E}\left(\left(\mathbf{H}_u^{\mathrm{eff,r}}\right)^{\mathrm{H}}\tilde{\mathbf{F}}_u\tilde{\mathbf{F}}_u^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff,r}}\right) + P_u\mathbb{E}\left((\mathbf{H}_u^{\mathrm{eff,n}})^{\mathrm{H}}\tilde{\mathbf{F}}_u\tilde{\mathbf{F}}_u^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff,n}}\right). \tag{46}$$

Neglecting the signal power received over $\mathbf{H}_u^{\mathrm{eff,n}}$, the covariance matrix is lower bounded as

$$\mathbb{E}(\mathbf{S}_u) \succeq \mathbb{E}(\tilde{\mathbf{S}}_u) = P_u\mathbb{E}\left(\left(\mathbf{H}_u^{\mathrm{eff,r}}\right)^{\mathrm{H}}\tilde{\mathbf{F}}_u\tilde{\mathbf{F}}_u^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff,r}}\right), \tag{47}$$

in the sense that $\mathbb{E}(\mathbf{S}_u - \tilde{\mathbf{S}}_u)$ is positive semidefinite. Notice that this bound is tight in case of unquantized feedback or if orthogonal users are scheduled.

According to the BD construction (5), the precoder of user u lies in the left null space of the other users' quantized channels. Considering an orthonormal basis $\mathbf{B}_u$ for this space,

$$(\hat{\bar{\mathbf{H}}}_u)^{\mathrm{H}}\mathbf{B}_u = \mathbf{0}, \quad \mathbf{B}_u \in \mathbb{C}^{N_t\times N_t-\bar{\ell}_u}, \quad \mathbf{B}_u^{\mathrm{H}}\mathbf{B}_u = \mathbf{I}_{N_t-\bar{\ell}_u}, \tag{48}$$

with $\hat{\bar{\mathbf{H}}}_u$ from (13) and $\bar{\ell}_u = \sum_{s\in\mathcal{S}\setminus u}\ell_s$, the precoder can be expressed as

$$\tilde{\mathbf{F}}_u = \mathbf{B}_u\mathbf{Q}_u, \quad \mathbf{Q}_u \in \mathbb{C}^{N_t-\bar{\ell}_u\times\ell_u}, \quad \mathbf{Q}_u^{\mathrm{H}}\mathbf{Q}_u = \mathbf{I}_{\ell_u}. \tag{49}$$

The assumption that $\tilde{\mathbf{F}}_u$ is isotropically distributed implies that $\mathbf{Q}_u$ is isotropically distributed and statistically independent of $\mathbf{B}_u$, leading to

$$\mathbb{E}(\tilde{\mathbf{S}}_u) = P_u\frac{\ell_u}{N_t-\bar{\ell}_u}\mathbb{E}\left(\left(\mathbf{H}_u^{\mathrm{eff,r}}\right)^{\mathrm{H}}\mathbf{B}_u\mathbf{B}_u^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff,r}}\right)$$
$$= P_u\frac{\ell_u}{N_t-\bar{\ell}_u}\mathbb{E}\left(\left(\mathbf{H}_u^{\mathrm{eff}}\right)^{\mathrm{H}}\hat{\mathbf{H}}_u\hat{\mathbf{H}}_u^{\mathrm{H}}\mathbf{B}_u\mathbf{B}_u^{\mathrm{H}}\hat{\mathbf{H}}_u\hat{\mathbf{H}}_u^{\mathrm{H}}\mathbf{H}_u^{\mathrm{eff}}\right), \tag{50}$$

where $\mathbf{H}_u^{\mathrm{eff,r}}$ has been substituted from (14). Next, we decompose $\mathbf{B}_u$ with respect to $\hat{\mathbf{H}}_u$

$$\mathbf{B}_u = \hat{\mathbf{H}}_u\hat{\mathbf{H}}_u^{\mathrm{H}}\mathbf{B}_u + \left(\mathbf{I}_{N_t} - \hat{\mathbf{H}}_u\hat{\mathbf{H}}_u^{\mathrm{H}}\right)\mathbf{B}_u$$
$$= \hat{\mathbf{H}}_u\mathbf{U}_u\cos(\mathbf{\Phi}_u)\mathbf{V}_u^{\mathrm{H}} + \left(\mathbf{I}_{N_t} - \hat{\mathbf{H}}_u\hat{\mathbf{H}}_u^{\mathrm{H}}\right)\mathbf{B}_u, \tag{51}$$

where the second equality follows from a compact SVD of $\hat{\mathbf{H}}_u^{\mathrm{H}}\mathbf{B}_u$. User u does neither know the CSI feedback of the other users nor the schedule $\mathcal{S}$ during feedback calculation; hence, the user cannot determine $\mathbf{B}_u$ and we therefore assume it to be uniformly distributed on $\mathcal{G}(N_t, N_t-\bar{\ell}_u)$. Correspondingly, $\mathbf{U}_u$, $\cos(\mathbf{\Phi}_u)$ and $\mathbf{V}_u$ are statistically independent, and the $\ell_u\times\ell_u$ dimensional unitary matrix $\mathbf{U}_u$ is isotropically distributed [49, Theorem 1]. Diagonal matrix $\mathbf{\Phi}_u$ contains the principal angles between $\mathrm{span}(\hat{\mathbf{H}}_u)$ and $\mathrm{span}(\mathbf{B}_u)$; hence,

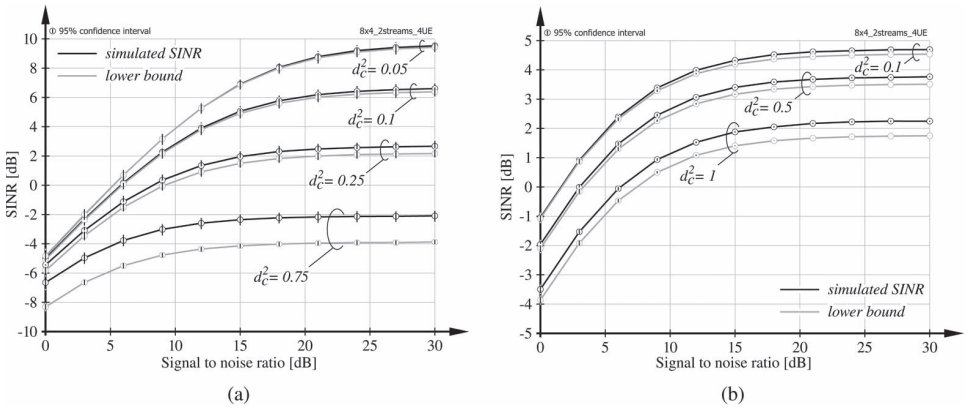


Fig. 6. Evaluation of the proposed lower bound on the expected SINR in dependence of the SNR P_t/σ_z^2. (a) Impact of the quantization error $\mathrm{d_c^2}(\mathbf{U}_u^{\mathrm{eff}}, \hat{\mathbf{H}}_u)$. (b) Impact of the subspace distance $\mathrm{d_c^2}(\hat{\mathbf{H}}_u, \mathbf{B}_u)$.

$\mathrm{tr}(\cos(\mathbf{\Phi}_u)^2) = \ell_u - \mathrm{d_c^2}(\hat{\mathbf{H}}_u, \mathbf{B}_u)$. Plugging (51) into (50) and considering the mentioned assumptions we obtain

$$\mathbb{E}(\tilde{\mathbf{S}}_u) = P_u \frac{\ell_u}{N_t - \bar{\ell}_u} \mathbb{E}\Bigg(\left(\mathbf{H}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \hat{\mathbf{H}}_u \left(\mathbf{U}_u \cos(\mathbf{\Phi}_u)^2 \mathbf{U}_u^{\mathrm{H}}\right) \hat{\mathbf{H}}_u^{\mathrm{H}} \mathbf{H}_u^{\mathrm{eff}} \Bigg) = P_u \frac{\ell_u}{N_t - \bar{\ell}_u} \mathbb{E}\left(\left(\mathbf{H}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \hat{\mathbf{H}}_u \hat{\mathbf{H}}_u^{\mathrm{H}} \mathbf{H}_u^{\mathrm{eff}} \right) \left(1 - \frac{\mathrm{d_c^2}(\hat{\mathbf{H}}_u, \mathbf{B}_u)}{\ell_u}\right), \quad (52)$$

where the second equality is due to the isotropy of $\mathbf{U}_u$. Finally, we apply an SVD to the effective channel $\mathbf{H}_u^{\mathrm{eff}} = \mathbf{U}_u^{\mathrm{eff}} \mathbf{\Sigma}_u^{\mathrm{eff}} (\mathbf{V}_u^{\mathrm{eff}})^{\mathrm{H}}$, and decompose $\hat{\mathbf{H}}_u$ with respect to $\mathbf{U}_u^{\mathrm{eff}} \in \mathbb{C}^{N_t \times \ell_u}$

$$\begin{aligned} \hat{\mathbf{H}}_u &= \mathbf{U}_u^{\mathrm{eff}} \left(\mathbf{U}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \hat{\mathbf{H}}_u + \left(\mathbf{I}_{N_t} - \mathbf{U}_u^{\mathrm{eff}} \left(\mathbf{U}_u^{\mathrm{eff}}\right)^{\mathrm{H}}\right) \hat{\mathbf{H}}_u \\ &= \mathbf{U}_u^{\mathrm{eff}} \mathbf{Y}_u \cos(\mathbf{\Theta}_u) \mathbf{W}_u^{\mathrm{H}} + \left(\mathbf{I}_{N_t} - \mathbf{U}_u^{\mathrm{eff}} \left(\mathbf{U}_u^{\mathrm{eff}}\right)^{\mathrm{H}}\right) \hat{\mathbf{H}}_u, \end{aligned} \quad (53)$$

with the second equality being obtained from a compact SVD of $(\mathbf{U}_u^{\mathrm{eff}})^{\mathrm{H}} \hat{\mathbf{H}}_u$. For a fixed quantized subspace $\hat{\mathbf{H}}_u$, we cannot further simplify this expression. However, considering the randomness of the applied quantization codebook, we can reasonably argue that for a fixed subspace quantization error $\mathrm{d_c^2}(\mathbf{U}_u^{\mathrm{eff}}, \hat{\mathbf{H}}_u) = \ell_u - \mathrm{tr}(\cos(\mathbf{\Theta}_u)^2)$, which is solely determined by the component of $\hat{\mathbf{H}}_u$ in the orthogonal complement of $\mathrm{span}(\mathbf{U}_u^{\mathrm{eff}})$, the range space component of $\hat{\mathbf{H}}_u$ within $\mathrm{span}(\mathbf{U}_u^{\mathrm{eff}})$ is random. We hence assume $(\mathbf{U}_u^{\mathrm{eff}})^{\mathrm{H}} \hat{\mathbf{H}}_u$ to be isotropically distributed, implying that $\mathbf{Y}_u$ is isotropic unitary. Using (53) in (52) and exploiting the isotropy of $\mathbf{Y}_u$, the covariance matrix is

$$\mathbb{E}(\tilde{\mathbf{S}}_u) = \mathbf{V}_u^{\mathrm{eff}} \left(\mathbf{\Sigma}_u^{\mathrm{eff}}\right)^2 \left(\mathbf{V}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \underbrace{P_u \frac{\ell_u}{N_t - \bar{\ell}_u} \left(1 - \frac{\mathrm{d_c^2}(\hat{\mathbf{H}}_u, \mathbf{B}_u)}{\ell_u}\right) \left(1 - \frac{\mathrm{d_c^2}\left(\mathbf{U}_u^{\mathrm{eff}}, \hat{\mathbf{H}}_u\right)}{\ell_u}\right)}_{c_S}. \quad (54)$$

The first term in brackets quantifies the overlap between the quantized user subspace $\hat{\mathbf{H}}_u$ and the other users' channels. If $\hat{\mathbf{H}}_u$ is in the left null space of the other users, i.e., $\mathrm{d_c^2}(\hat{\mathbf{H}}_u, \mathbf{B}_u) = 0$, we can fully steer the transmit signal of user u into the subspace $\mathrm{span}(\hat{\mathbf{H}}_u)$ without causing any interference. The term involving $\mathrm{d_c^2}(\mathbf{U}_u^{\mathrm{eff}}, \hat{\mathbf{H}}_u)$ is a consequence of (47).

b) Interference term: To estimate the inference covariance matrices, we impose additional assumptions: as user u does not know the number of streams of the other users, we assume that all users are served over $\ell_s = \ell_u$ streams, implying that $P_s = P_u, \forall\, s$. Also, we consider the worst case number of users $S = N_t/\ell_u$ in terms of residual multi-user interference.

With an orthonormal basis $\hat{\mathbf{H}}_u^{\perp} \in \mathbb{C}^{N_t \times N_t - \ell_u}$ of the orthogonal complement of $\hat{\mathbf{H}}_u$, we can rewrite the expected value of a single interference term as

$$\mathbb{E}(\mathbf{R}_u^{(s)}) = \mathbb{E}\left(\left(\mathbf{H}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \hat{\mathbf{H}}_u^{\perp} \left(\hat{\mathbf{H}}_u^{\perp}\right)^{\mathrm{H}} \left(\hat{\mathbf{H}}_u^{\perp} \mathbf{Q}_s \mathbf{Q}_s^{\mathrm{H}} \left(\hat{\mathbf{H}}_u^{\perp}\right)^{\mathrm{H}}\right) \hat{\mathbf{H}}_u^{\perp} \left(\hat{\mathbf{H}}_u^{\perp}\right)^{\mathrm{H}} \mathbf{H}_u^{\mathrm{eff}} \right), \quad (55)$$

where we substituted $\mathbf{H}_u^{\mathrm{eff,n}}$ from (14) and exploited the BD construction to express $\tilde{\mathbf{F}}_s$ in terms of the basis $\hat{\mathbf{H}}_u^{\perp}$ using the semi-unitary matrix $\mathbf{Q}_s \in \mathbb{C}^{N_t - \ell_u \times \ell_s}$. Because the user u does not know the precoders of the other users, we assume $\mathbf{Q}_s$ as uniformly distributed on $\mathcal{G}(N_t - \ell_u, \ell_s)$. With an SVD of $\mathbf{H}_u^{\mathrm{eff}}$, as in (53), we can write the covariance matrix as

$$\mathbb{E}\left(\mathbf{R}_u^{(s)}\right) = P_s \frac{\ell_s}{N_t - \ell_u} \mathbf{V}_u^{\mathrm{eff}} \mathbf{\Sigma}_u^{\mathrm{eff}} \mathbb{E}\left(\left(\mathbf{U}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \hat{\mathbf{H}}_u^{\perp} \left(\hat{\mathbf{H}}_u^{\perp}\right)^{\mathrm{H}} \mathbf{U}_u^{\mathrm{eff}} \right) \mathbf{\Sigma}_u^{\mathrm{eff}} \left(\mathbf{V}_u^{\mathrm{eff}}\right)^{\mathrm{H}}. \quad (56)$$

To determine the expected value in the center, we replace $\hat{\mathbf{H}}_u^{\perp} (\hat{\mathbf{H}}_u^{\perp})^{\mathrm{H}}$ with $(\mathbf{I}_{N_t} - \hat{\mathbf{H}}_u \hat{\mathbf{H}}_u^{\mathrm{H}})$

$$\mathbb{E}\left(\left(\mathbf{U}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \hat{\mathbf{H}}_u^{\perp} \left(\hat{\mathbf{H}}_u^{\perp}\right)^{\mathrm{H}} \mathbf{U}_u^{\mathrm{eff}} \right) = \mathbf{I}_{N_t} - \mathbb{E}\left(\left(\mathbf{U}_u^{\mathrm{eff}}\right)^{\mathrm{H}} \hat{\mathbf{H}}_u \hat{\mathbf{H}}_u^{\mathrm{H}} \mathbf{U}_u^{\mathrm{eff}} \right). \quad (57)$$

Applying the decomposition and argumentation from (53) and below, the expected value is

$$\mathbb{E}\left(\left(\mathbf{U}_u^{\mathrm{eff}}\right)^{\mathrm{H}}\hat{\mathbf{H}}_u^{\perp}(\hat{\mathbf{H}}_u^{\perp})^{\mathrm{H}}\mathbf{U}_u^{\mathrm{eff}}\right)=\mathbf{I}_{N_t}$$
$$-\mathbf{I}_{N_t}\left(1-\frac{\mathrm{d_c^2}\left(\mathbf{U}_u^{\mathrm{eff}},\hat{\mathbf{H}}_u\right)}{\ell_u}\right)=\frac{\mathrm{d_c^2}\left(\mathbf{U}_u^{\mathrm{eff}},\hat{\mathbf{H}}_u\right)}{\ell_u}\mathbf{I}_{N_t}. \quad (58)$$

With the assumptions $\ell_s=\ell_u$, $P_s=P_u$ and $S=N_t/\ell_u$, we obtain the sum interference as

$$\mathbb{E}(\mathbf{R}_u)=\underbrace{P_u\frac{\ell_u}{N_t-\ell_u}\left(\frac{N_t}{\ell_u}-1\right)\frac{\mathrm{d_c^2}\left(\mathbf{U}_u^{\mathrm{eff}},\hat{\mathbf{H}}_u\right)}{\ell_u}}_{c_I}$$
$$\mathbf{V}_u^{\mathrm{eff}}\left(\mathbf{\Sigma}_u^{\mathrm{eff}}\right)^2\left(\mathbf{V}_u^{\mathrm{eff}}\right)^{\mathrm{H}}. \quad (59)$$

c) SINR lower bound: With the results derived above we get the following lower bound on the expected SINR of user u with imperfect CSIT

$$\mathbb{E}(\beta_u)\geq\frac{c_S\mathrm{tr}\left(\left(\mathbf{\Sigma}_u^{\mathrm{eff}}\right)^2\right)}{\sigma_z^2\ell_u+c_I\mathrm{tr}\left(\left(\mathbf{\Sigma}_u^{\mathrm{eff}}\right)^2\right)}. \quad (60)$$

Observing that $\mathbf{\Sigma}_u^{\mathrm{eff}}$ is a diagonal matrix, we can also define the following per stream SINR lower bound

$$\mathbb{E}(\beta_{\nu,u})\geq\tilde{\beta}_{\nu,u}=\frac{c_S\left(\sigma_{\nu,u}^{\mathrm{eff}}\right)^2}{\sigma_z^2+c_I\left(\sigma_{\nu,u}^{\mathrm{eff}}\right)^2},\quad \nu\in\{1,\ldots,\ell_u\}$$
$$\mathbf{\Sigma}_u^{\mathrm{eff}}=\mathrm{diag}\left(\sigma_{1,u}^{\mathrm{eff}},\ldots,\sigma_{\ell_u,u}^{\mathrm{eff}}\right). \quad (61)$$

d) Evaluation of the bound: We evaluate the tightness of the proposed lower bound with simulations, assuming $N_t=8$, $M_u=4$, $\ell_u=2$ and $S=U=4$. The results are shown in Fig. 6. We first assume that the channels of all users are isotropically distributed and we investigate the impact of the quantization error $\mathrm{d_c^2}(\mathbf{U}_u^{\mathrm{eff}},\hat{\mathbf{H}}_u)$ on the tightness of the lower bound in Fig. 6(a). With decreasing quantization error the bound gets tighter, because the interference diminishes. This is in line with our derivation, because on the one hand, without interference, Jensen's inequality in (41) is tight and on the other hand, with diminishing quantization error, the null space component of the effective channel vanishes, implying tightness of (47). In Fig. 6(b), we investigate the impact of the overlap between the quantized subspace $\mathrm{span}(\hat{\mathbf{H}}_u)$ and the left null space of the other users' channels $\mathrm{span}(\mathbf{B}_u)$, assuming a fixed quantization error $\mathrm{d_c^2}(\mathbf{U}_u^{\mathrm{eff}},\hat{\mathbf{H}}_u)=0.5$. We observe that the bound gets tighter with decreasing subspace distance $\mathrm{d_c^2}(\hat{\mathbf{H}}_u,\mathbf{B}_u)$, due to (47). Also, the SINR improves because more signal energy is steered into the subspace spanned by the effective channel. The residual gap of the bound, however, does not vanish, due to Jensen's inequality in (41).

REFERENCES

[1] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[2] B. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[3] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[4] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, Oct. 2008.

[5] S. Schwarz and M. Rupp, "Subspace quantization based combining for limited feedback block-diagonalization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5868–5879, Nov. 2013.

[6] T. Inoue and R. Heath, Jr., "Grassmannian predictive coding for limited feedback multiuser MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. 3076–3079.

[7] Y. Zhang and M. Lei, "Robust Grassmannian prediction for limited feedback multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2012, pp. 863–867.

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

[9] D. Sacristan-Murga and A. Pascual-Iserte, "Differential feedback of MIMO channel Gram matrices based on geodesic curves," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3714–3727, Dec. 2010.

[10] S. Schwarz, R. Heath, Jr., and M. Rupp, "Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems," in *Proc. IEEE 38th Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 5021–5025.

[11] M. Wu, C. Shen, and Z. Qiu, "Feedback reduction based on clustering in MIMO-OFDM beamforming systems," in *Proc. 5th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, 2009, pp. 1–4.

[12] T. Li, L. Yang, and Z. He, "Interpolation-based multiuser precoding for MIMO-OFDM system with limited feedback," in *Proc. Int. Conf. Neural Netw. Signal Process.*, 2008, pp. 205–209.

[13] S. Schwarz and M. Rupp, "Adaptive channel direction quantization for frequency selective channels," in *Proc. 20th Eur. Signal Process. Conf.*, Bucarest, Romania, Aug. 2012, pp. 2536–2540.

[14] 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 Commun.*, vol. 6, no. 3, pp. 281–288, Feb. 2012.

[15] 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 Trans. Signal Process.*, vol. 59, no. 10, pp. 4885–4899, Oct. 2011.

[16] S. Schwarz, R. Heath, Jr., and M. Rupp, "Single-user MIMO versus multi-user MIMO in distributed antenna systems with limited feedback," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 54, Mar. 2013.

[17] W. Choi and J. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 69–73, Jan. 2007.

[18] P. Chow, A. Karim, V. Fung, and C. Dietrich, "Performance advantages of distributed antennas in indoor wireless communication systems," in *Proc. IEEE 44th Veh. Technol. Conf.*, Jun. 1994, vol. 3, pp. 1522–1526.

[19] K. Kerpez and S. Ariyavisitakul, "A radio access system with distributed antennas," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 1994, vol. 3, pp. 1696–1700.

[20] E. Park, H. Kim, H. Park, and I. Lee, "Feedback bit allocation schemes for multi-user distributed antenna systems," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 99–102, Jan. 2013.

[21] B. Clerckx, G. Kim, and S. Kim, "MU-MIMO with channel statistics-based codebooks in spatially correlated channels," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2008, pp. 1–5.

[22] T. Yoo and A. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *Proc. IEEE Int. Conf. Commun.*, May 2005, vol. 1, pp. 542–546.

[23] M. Trivellato, F. Boccardi, and F. Tosate, "User selection schemes for MIMO broadcast channels with limited feedback," in *Proc. IEEE 65th Veh. Technol. Conf.*, Dublin, Ireland, Apr. 2007, pp. 2089–2093.

[24] Z. Shen, R. Chen, J. Andrews, R. Heath, Jr., and B. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[25] S. Schwarz *et al.*, "Pushing the limits of LTE: A survey on research enhancing the standard," *IEEE Access*, vol. 1, pp. 51–62, May 2013.
[26] 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 Trans. Wireless Comm.*, vol. 9, no. 4, pp. 1472–1482, Apr. 2010.
[27] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.
[28] *LTE-Advanced (3GPP Rel.11) Technology Introduction*, Rohde & Schwarz, Munich, Germany, 2012, white Paper.
[29] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: MMSE balancing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3702–3712, Aug. 2008.
[30] J. Park, B. Lee, and B. Shim, "A MMSE vector precoding with block diagonalization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 569–577, Feb. 2012.
[31] S. Schwarz and M. Rupp, "Subspace versus eigenmode quantization for limited feedback block-diagonalization," in *Proc. 6th Int. Symp. Commun., Control Signal Process.*, Athens, Greece, May 2014.
[32] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, Apr. 1998.
[33] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2008.
[34] M. Rezaeekheirabadi and M. Guillaud, "Limited feedback for interference alignment in the k-user MIMO interference channel," in *Proc. Inf. Theory Workshop*, Lausanne, Switzerland, Sep. 2012, pp. 1–5.
[35] L. Qiang, Y. Yang, F. Shu, and W. Gang, "SLNR precoding based on QBC with limited feedback in downlink CoMP system," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, 2010, pp. 1–5.
[36] K. Schober, P. Janis, and R. Wichman, "Geodesical codebook design for precoded MIMO systems," *IEEE Commun. Lett.*, vol. 13, no. 10, pp. 773–775, Oct. 2009.
[37] R. Blum, "MIMO capacity with interference," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 793–801, Jun. 2003.
[38] D. Love and R. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
[39] S. Schwarz and M. Rupp, "Throughput maximizing feedback for MIMO OFDM based wireless communication systems," in *Proc. IEEE 12th Int. Workshop Signal Process. Adv. Wireless Commun.*, San Francisco, CA, USA, Jun. 2011, pp. 316–320.
[40] 3rd Generation Partnership Project (3GPP). (2010, Dec.). Technical specification group radio access network; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36101.htm
[41] D. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3G systems: Extending the 3GPP Spatial Channel Model (SCM)," in *Proc. IEEE 61st Veh. Technol. Conf.*, Stockholm, Sweden, 2005, vol. 5, pp. 3132–3136.
[42] T. Rappaport, *Wireless Communications: Principles and Practice*. London, U.K.: Dorling Kindersley, 2009, ser. Prentice Hall Communications Engineering and Emerging Technologies Series.
[43] "Technical specification group radio access networks; Deployment aspects (Release 8)," Sophia-Antipolis, France, Dec. 2008.
[44] S. Schwarz and M. Rupp, "Antenna combiners for block-diagonalization based multi-user MIMO with limited feedback," in *Proc. IEEE Int. Conf. Commun.*, Budapest, Hungary, Jun. 2013, pp. 127–132.
[45] A. Zanella, M. Chiani, and M. Win, "On the marginal distribution of the eigenvalues of Wishart matrices," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1050–1060, Apr. 2009.
[46] "Technical specification group radio access network; E-UTRA; Radio Frequency (RF) system scenarios (release 10)," Sophia-Antipolis, France, Oct. 2010.
[47] "Recommendation ITU-R M.1225: Guidelines for evaluation of radio transmission technologies for IMT-2000," ITU, Geneva, Switzerland, Tech. Rep., 1997.
[48] A. Sharma, A. Roy, S. Ghosal, R. Chaki, and U. Bhattacharya, "Load balancing in cellular network: A review," in *Proc. 3rd Int. Conf. Comput. Commun. Netw. Technol.*, 2012, pp. 1–5.
[49] B. Hassibi, "Random matrices, integrals and space-time systems," in *Proc. DIMACS Workshop Algebraic Coding Theory Inf. Theory*, Piscataway, NJ, USA, Dec. 2003.

Stefan Schwarz was born in Neunkirchen, Austria, in 1984. He received the B.Sc. in electrical engineering and the Dipl.-Ing. degree (M.Sc. equivalent) in telecommunications engineering with highest distinctions in 2007 and 2009, respectively, both at Vienna University of Technology. He also received the Dr.Techn. degree (Ph.D. equivalent) in telecommunications engineering with highest distinctions from Vienna University of Technology in 2013. In 2010 he received the honorary price of the Austrian Minister of Science and Research, for excellent graduates of scientific and artistic universities. Since 2008 he is working as a Project Assistant at the Mobile Communications group of the Institute of Telecommunications, Vienna University of Technology. His research interests are located in the broad fields of wireless communications and signal processing. In his dissertation he focused on limited feedback single- and multi-user MIMO-OFDM communication systems and multi-user scheduling algorithms. He actively serves as a reviewer for IEEE TRANSACTIONS ON COMMUNICATIONS, SIGNAL PROCESSING, WIRELESS COMMUNICATIONS and VEHICULAR TECHNOLOGY.

Markus Rupp received the Dipl.-Ing. degree from the University of Saarbruecken, Germany, in 1988 and the Dr.-Ing. degree from the Technische Universitaet Darmstadt, Germany, in 1993, where he worked with Eberhardt Haensler on designing new algorithms for acoustical and electrical echo compensation. From November 1993 until July 1995, he had a postdoctoral position at the University of Santa Barbara, California with Sanjit Mitra where he worked with Ali H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 until August 2001 he was a member of Technical Staff in the Wireless Technology Research Department of Bell-Labs at Crawford Hill, NJ, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11 and UMTS. Since October 2001 he is a Full Professor for Digital Signal Processing in Mobile Communications at the Technical University of Vienna where he served as Dean from 2005 to 2007. He was associate editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005, is currently Associate Editor of JASP EURASIP Journal of Advances in Signal Processing, JES EURASIP Journal on Embedded Systems, Research Letters in Signal Processing, and is elected AdCom member of EURASIP. He authored and co-authored more than 250 papers and patents on adaptive filtering, wireless communications, and rapid prototyping, as well as automatic design methods.

Outage Investigation of Beamforming Over Random-Phase Finite-Scatterer MISO Channels

Stefan Schwarz, *Member, IEEE*

***Abstract*—In this letter, we revisit single-user multiple-input single-output beamforming with imperfect channel state information (CSI) at the transmitter. We present a CSI estimation model that is suitable for finite-scatterer directional channel models with uncertainty in the relative phase shifts of the scattering components. We show that in this case the distribution of the effective beamformed channel does not follow one of the well-known channel fading distributions and leads to complicated outage calculations. We show that the presented model generalizes popular existing models and we relate our investigations for the special cases of one and two specular components to existing results. We furthermore consider signal outage optimal beamforming for the case of two specular components.**

***Index Terms*—Gradient optimization, Grassmann manifold, general two-ray (GTR) fading, signal outage, two-wave with diffuse power (TWDP) fading, manifold optimization.**

I. Introduction

BEAMFORMING is a well established and thoroughly investigated technique that is successfully employed in many commercial and standardized wireless communication systems, e.g., 3GPP long term evolution (LTE) and Wi-Fi. Beamforming methods have been proposed under a host of different design aspects, such as maximizing the signal power of users [1], [2], minimizing the interference amongst users [3]–[5], or managing the interference between multiple transmitters [6]–[11]. The theoretical performance of beamforming methods with perfect channel state information at the transmitter (CSIT) as well as imperfect CSIT has been established for many important design criteria such as achievable transmission rate [12]–[15] and outage probability [16]–[22]. Robust adaptive beamforming techniques have been developed [23]–[28] that are able to cope with various impairments encountered in practical realizations. Recently, the interest in beamforming methods for wireless communications in cellular networks is reigniting, specifically within the fields of full-dimension multiple-input multiple-output (MIMO) systems [29], three-dimensional (3-D) beamforming [30], massive MIMO communications [31], and transmission in the millimeter wave band [32].

Despite all of these existing results and contributions, we revisit in this letter multiple-input single-output (MISO) beamforming with imperfect CSIT for the seemingly simple case of single-user point-to-point transmission. Our contribution is to generalize the most commonly employed channel state information (CSI) estimation model, i.e., writing the actual channel as the sum of an imperfect estimate and an error that is complex-Gaussian distributed, to a more complex model that is suitable for finite-scatterer directional channels. Specifically, we consider in Section II, a channel that consists of multiple specular components that are relatively phase shifted with respect to each other with random unknown phase shifts. Beamforming over such a channel leads to fading behavior that can in general not be described with well-known fading distributions, such as Hoyt, Rayleigh, Rice, Nakagami-m, Weibull, or the recently proposed $\alpha-\eta-\kappa-\mu$ distribution [33]. In Section III, we summarize existing results that are valid for the case of a single specular component, or equivalently for multiple specular components with fixed known relative phase shifts, both cases amounting to Rician fading. In Section IV, we particularize our results to the case of two specular components with random relative phase shift. This case relates to the two-wave with diffuse power (TWDP) and the general two-ray (GTR) [34]–[36] fading distributions, which naturally generalize Rayleigh and Rician fading. The TWDP and GTR models can be suitable to represent transmission scenarios in the millimeter waveband where the number of significant scatterers is commonly very small [37] and where the phase of the signals varies quickly due to the small wavelength. Recently, in [38] and [39], the average symbol error probability and the capacity of single-input single-output TWDP channels have been derived. In Section IV-A, we propose a beamformer optimization that minimizes the signal outage probability for the TWDP case and find a generally suboptimal solution via a projected gradient optimization that operates on a Grassmann manifold; in Section V, we investigate its performance by Monte-Carlo simulations.

Notation: The transpose and conjugate-transpose of matrix $\mathbf{A}$ are $\mathbf{A}^{\mathcal{T}}$ and $\mathbf{A}^{\mathcal{H}}$, and the Frobenius norm is $\|\mathbf{A}\|$. The expected value of random variable x is $\mathbb{E}\{x\}$. The probability of random event $\mathcal{A}$ is $\mathbb{P}\{\mathcal{A}\}$. The vector-valued circularly-symmetric complex Gaussian distribution is $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathbf{C})$. The uniform distribution centered at a and having a width of b is $\mathcal{U}(a,b)$. The Rayleigh distribution with parameter σ is $\text{Rayl}(\sigma)$. The Rician distribution with shape K and scale Ω is $\text{Rice}(K,\Omega)$. The complex Grassmann manifold of 1-D subspaces in the N-dimensional complex Euclidean space is $\mathcal{G}_{\mathbb{C}}(N,1)$. The tangent space to $\mathcal{G}_{\mathbb{C}}(N,1)$ at point $\mathbf{x}$ is $\mathcal{T}(\mathbf{x})$. The generalized Marcum Q-function of order m is $Q_m(a,b)$.

Manuscript received March 26, 2017; revised May 3, 2017; accepted May 14, 2017. Date of publication May 18, 2017; date of current version May 26, 2017. The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Luca Sanguinetti.

The author is with the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion, Institute of Telecommunications, TU Wien, Wien 1040, Austria (e-mail: sschwarz@nt.tuwien.ac.at).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2705724

II. System Model

A. Channel and Transmission Model

We consider MISO transmission where the transmitter is equipped with N_t transmit antennas. We denote the frequency-flat channel between transmitter and receiver as $\mathbf{h} \in \mathbb{C}^{N_t \times 1}$. We consider linear beamforming and denote the beamformer applied for transmission of data symbols $s \in \mathbb{C}, \mathbb{E}\{|s|^2\} = 1$ as $\mathbf{f} \in \mathbb{C}^{N_t \times 1}$. The input–output relationship is

$$y = \mathbf{h}^{\mathcal{H}} \mathbf{f} s + n \tag{1}$$

with n being complex Gaussian receiver noise of variance σ_n^2; w.l.o.g., we assume a unit transmit power constraint $\|\mathbf{f}\|^2 \leq 1$.

We consider a finite-scatterer directional channel model [40], where the channel $\mathbf{h}$ is composed of S specular components $\mathbf{h}_s$ and a diffuse background scattering component $\mathbf{h}_d$

$$\mathbf{h} = \sum_{s=1}^{S} \mathbf{h}_s + \mathbf{h}_d \tag{2}$$

$$\mathbf{h}_s = e^{j\varphi_s} \left(\sqrt{\gamma_s}\, \bar{\mathbf{h}}_s\right) = e^{j\varphi_s}\, \tilde{\mathbf{h}}_s \tag{3}$$

with $\gamma_s \in \mathbb{R}_+$ denoting the macroscopic pathloss of specular component s and $\bar{\mathbf{h}}_s$ representing the normalized channel vector with $\|\bar{\mathbf{h}}_s\| = 1$. The variable $\varphi_s \in \mathbb{R}$ is the phase shift due to propagation delay, Doppler-effects and reflection/scattering properties of the environment. The elements of the diffuse component are assumed to be uncorrelated zero-mean complex-Gaussian distributed of variance σ_d^2.

The normalized channel vector $\bar{\mathbf{h}}_s$ is determined by the transmit antenna array response with respect to the signal azimuth/elevation angles of departure (ϕ_s, θ_s) of specular component s. For example, assuming a mutually uncoupled uniform planar array of isotropic antenna elements of size $N_t = N_v \times N_h$, with N_v, N_h, respectively, denoting the number of rows and columns of the antenna array that are spaced distances δ_v, δ_h (in multiples of the wavelength) apart, the normalized channel vector is [41]

$$\begin{aligned} &\left[\bar{\mathbf{h}}_s(\phi_s, \theta_s)\right]_{(\ell-1)N_h + k} \\ &= \frac{1}{\sqrt{N_t}}\, e^{j2\pi\left(\delta_v\,(\ell-1)\cos\theta_s + \delta_h\,(k-1)\sin\phi_s\,\sin\theta_s\right)} \\ &\ell \in \{1, \ldots, N_v\},\ k \in \{1, \ldots, N_h\} \end{aligned} \tag{4}$$

with $[\mathbf{h}]_i$ indexing the i-th element of vector $\mathbf{h}$.

B. CSI Estimation Model

The signal angles (ϕ_s, θ_s) and the pathloss γ_s are macroscopic channel properties that change relatively slowly over time. The relative phase shifts φ_s, however, can change significantly with movements in the order of only fractions of the wavelength, causing spatial microscopic fading due to varying multipath interference. We assume that the base station has estimates of the angles of departure, the pathloss, and the phase shifts, acquired either through uplink pilots or by means of limited feedback. To reflect the varying degree of uncertainty about macroscopic and microscopic channel properties, we assume a model of imperfect CSIT of the following form:

$$\hat{\mathbf{h}}_s = e^{j\hat{\varphi}_s} \left(\sqrt{\hat{\gamma}_s}\, \hat{\bar{\mathbf{h}}}_s\right) = e^{j\hat{\varphi}_s}\, \hat{\bar{\mathbf{h}}}_s \tag{5}$$

with $\hat{\mathbf{h}}_s$ denoting the CSIT, $\hat{\varphi}_s$ being the estimated phase shift and $\hat{\bar{\mathbf{h}}}_s$ denoting the estimated channel vector obtained from macroscopic angle and pathloss estimates. This estimated CSIT is related to the actual channel state according to

$$\begin{aligned} \mathbf{h}_s &= e^{j(\hat{\varphi}_s + \Delta\varphi_s)} \left(\sqrt{\alpha_s}\, \hat{\bar{\mathbf{h}}}_s + \sqrt{1-\alpha_s}\, \tilde{\mathbf{e}}_s\right) \\ &= e^{j(\hat{\varphi}_s + \Delta\varphi_s)} \sqrt{\alpha_s}\, \hat{\bar{\mathbf{h}}}_s + \sqrt{1-\alpha_s}\, \mathbf{e}_s \end{aligned} \tag{6}$$

with $\alpha_s \in [0, 1]$ representing the quality of the macroscopic CSIT estimate and $\mathbf{e}_s \sim \mathcal{CN}(\mathbf{0}, \frac{\gamma_s}{N_t}\mathbf{I})$ being the normalized macroscopic estimation error. We assume that the phase estimation $\hat{\varphi}_s$ is unbiased $\mathbb{E}\{\hat{\varphi}_s\} = \varphi_s$ and, hence, the phase estimation error follows a zero-mean probability density $\Delta\varphi_s \sim f_{\Delta\varphi_s}(\Delta\varphi_s), \mathbb{E}\{\Delta\varphi_s\} = 0$; we also assume that the phase estimation errors are statistically independent $\mathbb{E}\{\Delta\varphi_s \Delta\varphi_t\} = 0, \forall s \neq t$. The model (6) is consistent with the low-rank channel decompositions considered in [42]–[44]; these papers, however, do not make the phase estimation error explicit, but rather absorb it in an overall Gaussian estimation error. The model also reflects our limited feedback method proposed in [27], where only angular CSI is provided as feedback information and thus $\Delta\varphi_s \sim \mathcal{U}(0, 2\pi)$. The method in [27] can be extended to additionally provide phase CSI by uniformly quantizing φ_s with resolution $\Delta\phi_s$; then $\Delta\varphi_s \sim \mathcal{U}(0, \Delta\phi_s)$.

Since both, the estimation error $\mathbf{e}_s$ of the specular components as well as the diffuse channel $\mathbf{h}_d$ follow a zero-mean uncorrelated complex-Gaussian distribution, we can combine these contributions to form the effective diffuse channel

$$\begin{aligned} \tilde{\mathbf{h}}_d &= \mathbf{h}_d + \sum_{s=1}^{S} \sqrt{1-\alpha_s}\, \mathbf{e}_s \\ \tilde{\mathbf{h}}_d &\sim \mathcal{CN}\left(\mathbf{0}, \tilde{\sigma}_d^2\, \mathbf{I}\right),\ \tilde{\sigma}_d^2 = \sigma_d^2 + \sum_{s=1}^{S} (1-\alpha_s)\frac{\gamma_s}{N_t}. \end{aligned} \tag{7}$$

With $\tilde{\gamma}_s = \hat{\gamma}_s\, \alpha_s$, the channel vector can be compactly written

$$\mathbf{h} = \sum_{s=1}^{S} e^{j(\hat{\varphi}_s + \Delta\varphi_s)} \sqrt{\tilde{\gamma}_s} \hat{\bar{\mathbf{h}}}_s + \tilde{\mathbf{h}}_d. \tag{8}$$

The effective channel, including the beamformer, is then

$$\begin{aligned} z = \mathbf{h}^{\mathcal{H}}\mathbf{f} = re^{j\varphi} &= \sum_{s=1}^{S} r_s e^{j\vartheta_s} + r_d e^{j\vartheta_d} \\ r_s = \sqrt{\tilde{\gamma}_s}\left|\hat{\bar{\mathbf{h}}}_s^{\mathcal{H}}\mathbf{f}\right|,\ \vartheta_s &= \arg\left(\hat{\bar{\mathbf{h}}}_s^{\mathcal{H}}\mathbf{f}\right) - \hat{\varphi}_s - \Delta\varphi_s \\ f_{\vartheta_s}(\vartheta_s) = f_{\Delta\varphi_s}&\left(\arg\left(\hat{\bar{\mathbf{h}}}_s^{\mathcal{H}}\mathbf{f}\right) - \hat{\varphi}_s - \vartheta_s\right) \\ r_d \sim \mathrm{Rayl}&\left(\tilde{\sigma}_d\,\|\mathbf{f}\|\right),\ \vartheta_d \sim \mathcal{U}\left(0, 2\pi\right) \end{aligned} \tag{9}$$

with $f_{\vartheta_s}(\vartheta_s)$ denoting the probability density of ϑ_s. To the best of the author's knowledge, the probability distribution of z is not known for the general case. If the random phase shifts $\Delta\varphi_s$ are fixed $\forall s$ it can be seen that $z|\Delta\varphi_s \sim \mathcal{CN}(\sum_{s=1}^{S} r_s e^{j\vartheta_s}, \tilde{\sigma}_d^2\|\mathbf{f}\|^2)$. The random phase shifts $\Delta\varphi_s$ determine the mean of $z|\Delta\varphi_s$; hence, with random phase shifts, z follows a complex-Gaussian distribution with random mean.

C. Outage Probability

In order to calculate the outage probability of z, i.e., the probability that $|z| = r \leq r_{\min}$, we first assume that the random

phase shifts $\Delta\varphi_s$ and, thus, ϑ_s are fixed $\forall s$. We then have

$$r|\Delta\varphi_s \sim \text{Rice}\left(K_r, \Omega_r\right)$$
$$K_r = \frac{\left|\sum_{s=1}^{S} r_s e^{j\vartheta_s}\right|^2}{\tilde{\sigma}_d^2 \left\|\mathbf{f}\right\|^2},\ \Omega_r = \left|\sum_{s=1}^{S} r_s e^{j\vartheta_s}\right|^2 + \tilde{\sigma}_d^2 \left\|\mathbf{f}\right\|^2. \tag{10}$$

Notice, the beamformer $\mathbf{f}$ impacts the scale and shape parameter of this Rician distribution. The cumulative distribution function of the Rician fading model can be stated, e.g., in terms of the Marcum Q-function $Q_m(a,b)$ [45] and, thus, the outage probability with fixed phase shifts and $r_{\min} \geq 0$ is

$$P_{\text{out}|\Delta\varphi_s}(r_{\min}) = \mathbb{P}\left\{r \leq r_{\min}|\Delta\varphi_s\right\} = 1 - Q_1\left(\frac{\alpha}{\beta}, \frac{r_{\min}}{\beta}\right)$$
$$\alpha^2 = \frac{K_r}{1+K_r}\Omega_r,\ \beta^2 = \frac{\Omega_r}{2(1+K_r)}. \tag{11}$$

Calculation of the outage probability for random phase shifts requires removing the conditioning of $P_{\text{out}|\Delta\varphi_s}(r_{\min})$, involving multidimensional integration over the random phase shifts

$$P_{\text{out}}(r_{\min}) = \int \cdots \int_0^{2\pi} P_{\text{out}|\Delta\varphi_s}(r_{\min}) \times \prod_{s=1}^{S} f_{\Delta\varphi_s}(\Delta\varphi_s)\, d\left(\Delta\varphi_s\right)_{\forall s}. \tag{12}$$

The calculation of this integral is numerically only feasible for a small number S of scatterers. We, hence, consider below the cases $S = 1$ and $S = 2$ in more detail.

III. Single Specular Component

If we restrict to $S = 1$ (or equivalently assume $\Delta\varphi_s$ as fixed $\forall s$), we recover the CSI estimation model that is mostly considered in literature; that is, the actual channel is equal to an imperfect estimate plus an additive Gaussian distortion. In this case, the phase uncertainty does not impact the outage probability at all and the effective channel follows a Rician distribution. In [17], the structure of the outage optimal input covariance matrix is investigated, revealing that the optimal signaling directions are the CSIT estimate and its orthogonal complement; the optimal power allocation among these directions is, however, not provided in [17]. Yet, restricting to rank-one beamforming, the optimal solution is maximum ratio transmission (MRT) along the CSIT estimate.

IV. Two Specular Components

For the case $S = 2$ and $\Delta\varphi_s \sim \mathcal{U}(0, 2\pi)$, the probability distribution of z follows the TWDP fading model [35]; with a general distribution $f_{\Delta\varphi_s}(\Delta\varphi_s)$, z follows the GTR fading model [36]. Here, the probability density function (pdf) of the envelope r can be expressed in terms of an equivalent Rician pdf

$$f_{\text{GTR}}(r) = \int_0^{2\pi} f_{\text{Rice}}(r; K(1 + \boldsymbol{\Delta}\cos\delta))\, f_\delta(\delta)\, d\delta \tag{13}$$
$$K = \frac{r_1^2 + r_2^2}{\tilde{\sigma}_d^2 \left\|\mathbf{f}\right\|^2},\ \boldsymbol{\Delta} = \frac{2r_1 r_2}{r_1^2 + r_2^2} \tag{14}$$

with phase difference $\delta = \vartheta_1 - \vartheta_2$ between the two specular components and $f_\delta(\delta)$ denoting its pdf, obtained by convolving $f_{\vartheta_1}(\vartheta_1)$ with $f_{\vartheta_2}(-\vartheta_2)$ and wrapping onto $[0, 2\pi]$ [to avoid wrapping, the range of integration in (13) can be extended]. The outage probability $P_{\text{out}}(r_{\min})$ can be obtained by integrating (13) from 0 to $r_{\min}$. It is obvious that the beamformer $\mathbf{f}$ impacts the outage probability in a complicated manner. Nevertheless, we attempt outage-based beamformer optimization in the following.

A. Signal Outage Optimized Beamforming

In this section, we discuss beamforming strategies for the case of $S = 2$ specular components with the ultimate goal of minimizing the outage probability.

1) Benchmark Beamformers: The following beamforming strategies are employed as benchmarks in our numerical investigations. We assume w.l.o.g. $\tilde{\gamma}_1 \geq \tilde{\gamma}_2$.

a) Maximum ratio transmission—MRT [1]: $\mathbf{f}_{\text{MRT1}} = \hat{\tilde{\mathbf{h}}}_1$, maximizing the signal power received over the strongest specular component. Notice, we do not consider the sum channel here, since the random phase shift between the two specular components is unknown; it only makes senses to consider the sum channel for beamforming if the phase uncertainty is sufficiently small.

b) Zero-forcing beamforming—ZF [3]:

$$\mathbf{f}_{\text{ZF2}} = \tilde{\mathbf{f}}_{\text{ZF2}} / \left\|\tilde{\mathbf{f}}_{\text{ZF2}}\right\|,\ \tilde{\mathbf{f}}_{\text{ZF2}} = \left(\mathbf{I} - \hat{\tilde{\mathbf{h}}}_2 \hat{\tilde{\mathbf{h}}}_2^{\mathcal{H}}\right) \hat{\tilde{\mathbf{h}}}_1 \tag{15}$$

eliminating the signal variation due to multipath interference.

c) Maximum eigenmode transmission—MET: Maximizing the expected received signal power for entirely unknown phases $\Delta\varphi_s \sim \mathcal{U}(0, 2\pi)$, by aligning the beamformer with the left singular vector $\mathbf{f}_{\text{MET}} = \mathbf{u}_{\max}$ corresponding to the largest singular value of matrix $\mathbf{H} = [\sqrt{\tilde{\gamma}_1}\hat{\tilde{\mathbf{h}}}_1, \sqrt{\tilde{\gamma}_2}\hat{\tilde{\mathbf{h}}}_2]$.

d) Robust adaptive beamforming—RAB: Here, we determine the beamformer by applying the worst-case optimization problem proposed in [46, eq. (20)] to our scenario. Specifically, in our case the presumed steering vector of [46] is $\hat{\tilde{\mathbf{h}}}_1$, whereas the steering error vector $\boldsymbol{\delta}$ of [46] is determined by the relatively phase shifted second specular component as well as the effective diffuse component. Since our goal is to minimize the outage probability, we find the largest success probability p (one minus outage probability) in [46, eq. (20)] that provides a feasible beamformer solution satisfying the power constraint $\|\mathbf{f}\| \leq 1$. This can easily be achieved by applying a bisection line search. Notice, however, that the problem is not always feasible with the restriction $\|\mathbf{f}\| \leq 1$ even if $p = 0$, especially for large $r_{\min}$; in case of infeasibility we fall back to MRT.

2) Gradient-Based Optimization: The following optimization problem defines the signal outage optimal beamformer:

$$\min_{\mathbf{f},\|\mathbf{f}\|=1} P_{\text{out}}(\mathbf{f}, r_{\min}) = \min_{\mathbf{f},\|\mathbf{f}\|=1} \mathbb{P}\left\{r \leq r_{\min}\right\}. \tag{16}$$

As calculation of the outage probability requires numerical integration, a closed-form solution appears out of reach. We, therefore, resort to local optimization by using a projected gradient method that operates on the associated Grassmann manifold $\mathcal{G}_{\mathbb{C}}(N_t, 1)$ [47]. Notice, in (16), we can restrict to unit norm beamformers, ignoring beamformers with $\|\mathbf{f}\| < 1$, since, otherwise, we can always increase the larger signal component r_1 or r_2, while keeping the other smaller component unchanged (by steering in the null-space of the corresponding channel); this in-turn reduces the outage probability by decreasing the signal variation with varying phase.

As a first step, we need to calculate the gradient of the signal outage probability $\mathbb{P}\{r \leq r_{\min}\}$ w.r.t. to $\mathbf{f}$, which we denote

as $\nabla_{\mathbf{f}} P_{\text{out}}$. This calculation is tedious but straightforward and requires numerical integration. In a second step, we project this gradient onto the tangent space $\mathcal{T}(\mathbf{f})$ of the Grassmannian associated to the point $\mathbf{f} \in \mathcal{G}_{\mathbb{C}}(N_t, 1)$

$$\nabla_{\mathbf{f}}^{\mathcal{T}} P_{\text{out}} = \left(\mathbf{I} - \mathbf{f}\mathbf{f}^{\mathcal{H}}\right) \nabla_{\mathbf{f}} P_{\text{out}} \in \mathcal{T}(\mathbf{f}). \tag{17}$$

This projected gradient parametrizes a geodesic $\Gamma(t, \nabla_{\mathbf{f}}^{\mathcal{T}} P_{\text{out}}) \in \mathcal{G}_{\mathbb{C}}(N_t, 1)$ on the Grassmannian

$$\Gamma(t, \nabla_{\mathbf{f}}^{\mathcal{T}} P_{\text{out}}) = \mathbf{f} \cos\left(t\frac{\pi}{4}\right) - \frac{\nabla_{\mathbf{f}}^{\mathcal{T}} P_{\text{out}}}{\left\|\nabla_{\mathbf{f}}^{\mathcal{T}} P_{\text{out}}\right\|} \sin\left(t\frac{\pi}{4}\right). \tag{18}$$

We employ this geodesic as local search direction for a line search algorithm over the line parameter $t \in [0, 1]$. The most costly part is that the calculation of the outage probability and its gradient, since both require numerical integration.

3) Restricted Geodesic Optimization: The method presented in the previous section is not very practical, since it requires numerical calculation of a complicated double-integral in each step of the gradient line search. We, thus, propose in the following two simpler optimization problems that drastically restrict the beamformer search space. More specifically, we restrict the search for the optimal beamformer to two geodesics: one connecting the zero-forcing solution of scattering component to $\mathbf{f}_{\text{ZF2}}$ with the maximum ratio beamformer of scattering component one $\mathbf{f}_{\text{MRT1}}$; the other connecting the two maximum ratio beamformers $\mathbf{f}_{\text{MRT1}}$, $\mathbf{f}_{\text{MRT2}}$. Even though this appears very restrictive, it turns out that in our simulations the gradient search never achieves a better solution than the optimum restricted to these geodesics. Yet, we have not been able to formally prove that the optimum necessarily lies on these geodesics.

The geodesic on the manifold $\mathcal{G}_{\mathbb{C}}(N_t, 1)$ connecting two points $\mathbf{f}_1 \in \mathcal{G}_{\mathbb{C}}(N_t, 1)$ and $\mathbf{f}_2 \in \mathcal{G}_{\mathbb{C}}(N_t, 1)$ is

$$\Gamma(t, \mathbf{t}_{1,2}) = \mathbf{f}_1 \cos(\phi_{1,2} t) + \frac{\mathbf{t}_{1,2}}{\|\mathbf{t}_{1,2}\|} \sin(\phi_{1,2} t)$$

$$\mathbf{t}_{1,2} = \left(\mathbf{I} - \mathbf{f}_1 \mathbf{f}_1^{\mathcal{H}}\right) \frac{\mathbf{f}_2}{\mathbf{f}_1^{\mathcal{H}} \mathbf{f}_2}, \ \phi_{1,2} = \operatorname{atan} \|\mathbf{t}_{1,2}\|, \ t \in [0, 1]. \tag{19}$$

To find the optimal solutions on the geodesics $\mathbf{f}_{\text{ZF2}}$ to $\mathbf{f}_{\text{MRT1}}$ and $\mathbf{f}_{\text{MRT1}}$ to $\mathbf{f}_{\text{MRT2}}$, we apply a bisection w.r.t. $t \in [0, 1]$, employing $P_{\text{out}}(\Gamma(t, \mathbf{t}_{1,2}), r_{\min})$ to determine the nested intervals.

V. Simulation Results

In this section, we investigate the presented beamforming methods numerically. We assume $S = 2$, $N_t = 4$, $\tilde{\sigma}_d^2 = 0.2$, and $\Delta\varphi_s \sim \mathcal{U}(0, 2\pi), \forall s$, i.e., the worst case of totally unknown phase shifts, corresponding to the limited feedback approach of [27]. We compare the empirical outage probability with varying threshold $r_{\min}$ to the analytical calculation. For the gradient method, we generate 25 random beamformer realizations $\mathbf{f}_0$ that are used to initialize the gradient optimization and we plot the performance of the best solution, which is of course still not necessarily the global optimum. We plot results for two different cases: $\tilde{\gamma}_1 = \tilde{\gamma}_2 = 2$ and $\tilde{\gamma}_1 = 3, \tilde{\gamma}_2 = 1$.

In Fig. 1, we compare the performance of the presented benchmark schemes to the gradient and restricted geodesic optimization in dependence of the signal power threshold $r_{\min}^2$. In the figure, analytic (solid) as well as empirical (dashed) results perfectly agree. As expected the gradient approach outperforms the benchmark schemes, but shows exactly the same performance

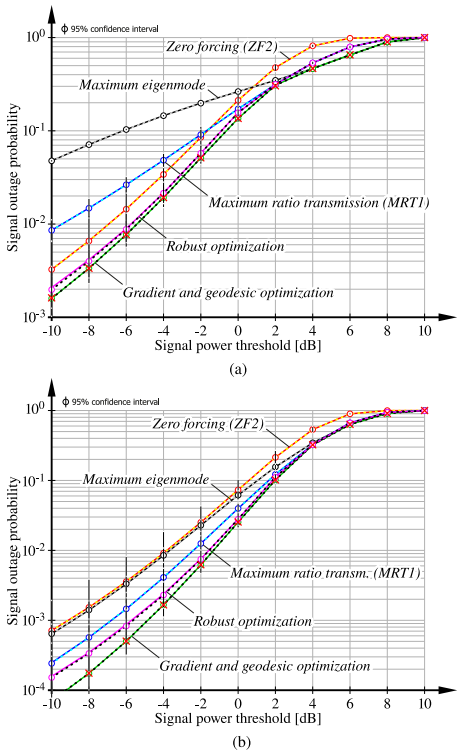


Fig. 1. Comparison of signal outage probabilities of different beamforming methods in dependence of the power threshold $r_{\min}^2$. (a) Signal outage probability with $\tilde{\gamma}_1 = \tilde{\gamma}_2 = 2$. (b) Signal outage probability with $\tilde{\gamma}_1 = 3$ and $\tilde{\gamma}_2 = 1$.

as the restricted geodesic optimization (cross markers). The robust optimization proposed in [46] comes close to the gradient optimization especially for equal-strength specular components. This method has the advantage that it can relatively easily be extended to $S > 2$ specular components, whereas the complexity of the gradient approach grows significantly, as each additional specular component implies an additional integral in the outage calculation. Notice though that the power-constrained robust adaptive beamforming (RAB) problem becomes infeasible in our simulations for $r_{\min}^2 \geq 2$ dB, in which case we heuristically fall back to MRT. We also observe in Fig. 1 that each of the three schemes MRT, zero-forcing beamforming (ZF), and maximum eigenmode transmission (MET) has a certain regime of $r_{\min}$ in which it outperforms the others. Comparing Fig. 1(a) and (b), we see that better performance is obtained if the total specular energy $\tilde{\gamma}_1 + \tilde{\gamma}_2 = 4$ distributes unequally over the two specular components as in Fig. 1(b).

VI. Conclusion

We investigate single-user MISO beamforming over finite-scatterer directional channels with imperfect CSIT, applying a CSI estimation model that accommodates varying degrees of uncertainty in microscopic and macroscopic CSI. The assumed model leads to channel fading behavior of the beamformed channel that has, to the best of the author's knowledge, not yet been investigated. We determine the outage probability for the general case in integral form and relate our results to the previously investigated cases of one and two randomly phase shifted specular components. We propose outage-optimized beamforming for the case of two scattering components and show that it outperforms existing schemes.

REFERENCES

[1] T. K. Y. Lo, "Maximum ratio transmission," in *Proc. IEEE Int. Conf. Commun.*, Jun. 1999, vol. 2, pp. 1310–1314.

[2] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[3] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[4] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[5] S. Schwarz, T. Philosof, and M. Rupp, "Leakage-based multicast transmit beamforming," in *Proc. Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 1–6.

[6] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[7] X. Shang, B. Chen, and H. Poor, "Multiuser MISO interference channels with single-user detection: Optimality of beamforming and the achievable rate region," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4255–4273, Jul. 2011.

[8] R. Mochaourab and E. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1128–1141, Mar. 2011.

[9] J. Park, G. Lee, Y. Sung, and M. Yukawa, "Coordinated beamforming with relaxed zero forcing: The sequential orthogonal projection combining method and rate control," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3100–3112, Jun. 2013.

[10] S. Schwarz and M. Rupp, "Exploring coordinated multipoint beamforming strategies for 5G cellular," *IEEE Access*, vol. 2, pp. 930–946, Aug. 2014.

[11] S. Schwarz and M. Rupp, "Transmit optimization for the MISO multicast interference channel," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4936–4949, Dec. 2015.

[12] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[13] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3834–3844, Oct. 2008.

[14] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, Oct. 2008.

[15] S. Schwarz and M. Rupp, "Subspace quantization based combining for limited feedback block-diagonalization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5868–5879, Nov. 2013.

[16] C. Chayawan and V. A. Aalo, "On the outage probability of optimum combining and maximal ratio combining schemes in an interference-limited rice fading channel," *IEEE Trans. Commun.*, vol. 50, no. 4, pp. 532–535, Apr. 2002.

[17] Y. Xie, C. N. Georghiades, and A. Arapostathis, "Minimum outage probability transmission with imperfect feedback for MISO fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1084–1091, May 2005.

[18] H. Li, Y. D. Yao, and J. Yu, "Outage probabilities of wireless systems with LCMV beamforming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3515–3523, Oct. 2007.

[19] T. Sriv, C. Chayawan, and W. Kumwilaisak, "Outage probability analysis of antenna array using maximum ratio combining in spherically invariant fading environment," *Wireless Commun. Mobile Comput.*, vol. 8, no. 2, pp. 231–244, 2008.

[20] V. S. Annapureddy, D. V. Marathe, T. R. Ramya, and S. Bhashyam, "Outage probability of multiple-input single-output (MISO) systems with delayed feedback," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 319–326, Feb. 2009.

[21] J. Park, Y. Sung, D. Kim, and H. V. Poor, "Outage probability and outage-based robust beamforming for MIMO interference channels with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3561–3573, Oct. 2012.

[22] E. Stathakis, J. Jaldn, L. K. Rasmussen, and M. Skoglund, "Outage region characterization for beamforming in MISO interference networks with imperfect CSI," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2378–2382, Dec. 2015.

[23] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[24] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.

[25] B. K. Chalise *et al.*, "Robust downlink beamforming based on outage probability specifications," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3498–3503, Oct. 2007.

[26] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Process.*, vol. 93, no. 12, pp. 3264–3277, 2013.

[27] S. Schwarz and M. Rupp, "Limited feedback based double-sided full-dimension MIMO for mobile backhauling," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Asilomar, CA, USA, 2016, pp. 27–33.

[28] S. Schwarz, T. Philosof, and M. Rupp, "Convex-optimization based geometric beamforming for FD-MIMO arrays," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Asilomar, CA, USA, 2016, pp. 679–683.

[29] H. Ji *et al.*, "Overview of full-dimension MIMO in LTE-advanced pro," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 176–184, Feb. 2017.

[30] S. M. Razavizadeh, M. Ahn, and I. Lee, "Three-dimensional beamforming: A new enabling technology for 5G wireless networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 94–101, Nov. 2014.

[31] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[32] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[33] M. D. Yacoub, "The α - η - κ - μ fading model," *IEEE Trans. Antennas Propag.*, vol. 64, no. 8, pp. 3597–3610, Aug. 2016.

[34] R. Esposito and L. Wilson, "Statistical properties of two sine waves in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 176–183, Mar. 1973.

[35] G. D. Durgin, T. S. Rappaport, and D. A. de Wolf, "New analytical models and probability density functions for fading in wireless communications," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1005–1015, Jun. 2002.

[36] M. Rao, F. J. Lopez-Martinez, M. S. Alouini, and A. Goldsmith, "MGF approach to the analysis of generalized two-ray fading models," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2548–2561, May 2015.

[37] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[38] M. Rao, F. J. Lopez-Martinez, and A. Goldsmith, "Statistics and system performance metrics for the two wave with diffuse power fading model," in *Proc. 48th Annu. Conf. Inf. Sci. Syst.*, Mar. 2014, pp. 1–6.

[39] N. Y. Ermolova, "Capacity analysis of two-wave with diffuse power fading channels using a mixture of gamma distributions," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2245–2248, Nov. 2016.

[40] P. Almers *et al.*, "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2007, no. 1, pp. 1–19, 2007.

[41] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Hoboken, NJ, USA: Wiley, 2005.

[42] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[43] E. Zöchmann, S. Schwarz, and M. Rupp, "Comparing antenna selection and hybrid precoding for millimeter wave wireless communications," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, 5 pages, 2016.

[44] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[45] J. I. Marcum, "Table of Q functions," RAND Corporation, Santa Monica, CA, USA, 1950.

[46] S. A. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5719–5724, Nov. 2008.

[47] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

5. Coordinated Transmission and Interference Management

Today's mobile cellular networks predominantly employ the entire available spectrum at all base stations simultaneously, in order to maximize the spatial spectrum reuse gains of the system. An observation from Martin Cooper in 2001 assessed that the wireless capacity doubles every 30 months [181]. This trend has been preserved to date and has improved the wireless capacity by a factor of one million during the last 45 years. The major part of this gain has been achieved by efficient spatial reuse of the spectrum [182], by means of network densification through small cells, distributed antenna systems (DASs) and relay nodes [183–185]. Network densification, however, leads to increasing interference between neighboring cells, which impairs both, the efficiency and the reliability of the wireless connectivity. In fact, several investigations have shown that in uncoordinated networks there can exist an optimal finite spatial density of BSs that maximizes the area spectral efficiency and the coverage of the network [186–188]. To circumvent these interference limitations, many multi-point coordination and joint transmission schemes have been developed over the last years, leading up to the proposal of cloud radio access networks (CRANs) and cell-free massive MIMO systems to achieve a large area coordination of mobile networks [189–191].

Chapter Outline

In this chapter, I provide a discussion of CoMP and joint transmission (JT) techniques for unicast and multicast scenarios, give an overview of my scientific contributions within these fields and tie in my corresponding publications that constitute this thesis. I first focus on the unicast scenario, which is a well investigated subject with many related scientific publications. Nevertheless, interference management for unicast scenarios has lost none of its topicality in the development of 5G mobile networks, since CRAN, distributed MIMO and cell-free massive MIMO promise significantly enhanced coordination capabilities [190–192].

The multicast scenario is less well understood, especially in the context of multi-cell coordination. I introduce in this chapter the underlying information-theoretic model of multi-cell multicasting, namely the multi-antenna MCIC, present coordination techniques for efficient multicasting in multi-cell networks and exhibit their potential to support cellular-assisted

vehicular communication scenarios. I do not address CSIT imperfections in this chapter in detail, even though they are as critical as in the multi-user MIMO transmission schemes discussed in Chapter 4. However, the basic insights on CSI feedback and CSIT imperfections of Chapters 3 and 4 carry over to multi-point coordination and joint transmission.

5.1. Interference Management in Unicast Scenarios

Interference management techniques in mobile cellular networks basically fall into one of two categories:

- CoMP techniques: These methods rely on coordination of the resource allocations and MIMO processing to mitigate the mutual interference of different cells. They require sharing of the CSI and of coordination information amongst cells for the optimization of the precoders and resource allocations. Mostly, such coordination techniques are implemented in a distributed fashion within the network.
- JT techniques: These methods apply cooperation amongst cells to jointly satisfy the QoS requirements of all UEs of the cooperating cells. They are commonly implemented by a central processor, which jointly optimizes the resource allocations and precoders of all cooperating cells and distributes the information accordingly. This requires gathering of all the CSI at the central processor and, in general, sharing of the data of all UEs amongst the cooperating cells. This implies that JT techniques require powerful backhaul connections in-between cooperating cells.

Both of these techniques have been standardized within 4G LTE. Yet, research for 5G CoMP and JT technologies is still ongoing, especially within the fields of CRAN antenna domain optimization [193–195], cell-free massive MIMO precoding and power control [153], and mmWave multi-point transmission coordination to enhance the macroscopic diversity and thereby reduce the outage probability [196, 197].

5.1.1. CoMP Techniques

Coordination techniques can be realized in the time-frequency domain, by dynamically allocating time-frequency resources amongst cells and UEs, and/or in the spatial domain, by coordinating beamforming and precoding strategies of neighboring cells.

Coordinated Resource Allocation

Coordinated scheduling: This method is basically a dynamic implementation of fractional frequency reuse (FFR) [198]. The underlying idea is to assign different time-frequency resources to UEs that are at the same cell edge but belong to different cells, thus, mitigating interference between them [199]. Such techniques can reduce the outage probability of cellular networks, require just little information exchange amongst BSs and are robust with respect to user mobility, since they do not rely on accurate CSIT. However, from a network capacity standpoint, coordinated scheduling provides hardly any performance gain [200].

Dynamic point selection (DPS): This method enables to receive data within consecutive TTIs from different cells, without requiring a hand-over between the cells [201]. This enhances the macroscopic diversity of the system and thus the reliability. Yet, DPS requires fast forwarding of data between cells, causing an increase of the backhaul network load.

Dual connectivity: Although not strictly a CoMP technique with respect to user data transmissions, dual connectivity nevertheless requires coordination of multiple transmission points to realize the control-plane user-plane splitting feature necessary to enable efficient and reliable high-mobility support in dense heterogeneous small cell networks [202]; see *[Article 2]* for a more detailed discussion.

Coordinated Beamforming and Precoding

Coordinated beamforming and precoding techniques exhibit a larger potential than coordinated resource allocation, because they allow to utilize all the available time-frequency resources by all cells simultaneously, while still being able to mitigate interference in the spatial domain through beamforming/precoding methods.

Coordinated beamforming for the MISO unicast interference channel: The MISO unicast interference channel (UCIC) is the information-theoretic model for unicast transmissions from multiple transmit antennas in multi-cell mobile networks. In general, neither the capacity region nor even the DoF region of such systems are known. Only for special cases, such as the two user SISO UCIC, the DoF region has been fully characterized [203], and it is known that rate-splitting and SIC are generally required to achieve the optimal DoF. In my further discussions, however, I restrict to linear beamforming/precoding (without rate-splitting) and single-user detection (no SIC), as such methods are practically

more relevant for reasons of complexity. The following main results have been established for this situation in the scientific literature:

- The optimal transmit covariance matrix has rank one; thus, beamforming is the optimal transmit strategy. The Pareto boundary of its achievable rate region is characterized in [204–207].
- Weighted sum-rate maximization can provide beamformers that achieve any point on the Pareto boundary of the rate region; yet, the optimization problem is NP hard [208].
- Iterative algorithms for local and even global optimization of the weighted sum-rate beamforming problem exist [209, 210]. In practice, though, mostly low-complexity (semi) closed-form beamformers are preferred, such as, SLNR beamforming [105] and multi-cell BD precoding [211].
- The outage region of coordinated beamforming with imperfect CSIT is characterized in [145], and robust beamforming methods are proposed, e.g., in [212].

In my article *[Article 10]*, I provide a comprehensive overview and performance comparison of coordinated transmit beamformers for the MISO UCIC with excess antennas at the receivers. Specifically, I consider MIMO systems, in which the receivers apply beamformers to convert the system to an effective MISO system from the perspective of transmit beamforming. I analyze the achievable rate region of a simplified but nonetheless instructive scenario with two transmitters and two receivers, for several combinations of transmit and receive beamformers, showing that SLNR and MMSE based schemes exhibit the best performance; see Section 5.3 for more details.

Coordinated beamforming for the MIMO unicast interference channel: Capacity optimal transmit strategies for this situation are not known in general. However, the DoF optimal linear transmit and receive beamforming/precoding strategy for the MIMO UCIC is known to be IA [213, 214]. This method has received significant attention, since it allows to achieve half of the DoF of the idealistic situation where cells a priori do not interfere with each other. The following main results have been established for IA:

- In general situations, with arbitrary number of transmit and receive antennas, IA requires so-called symbol-extensions, i.e. beamforming/precoding over space and time, to achieve the optimal DoF [213].
- When restricting to beamforming/precoding over space only by employing linear spatial transmit and receive filters, assessing the feasibility of an IA configuration, with a certain number of spatial streams for each UE, is a complicated problem on its own behalf. A polynomial-time feasibility test for this problem is developed in [215].

- Closed-form solutions for IA transmit and receive filters are available only for special cases; yet, many iterative optimization schemes exist, e.g., [104, 216, 217].
- In practically relevant SNR ranges it is often preferable to allow some residual interference between cells rather than canceling the interference entirely. This enables to achieve a favorable trade-off between the intended signal power and the residual interference, e.g., by maximizing the SINR [218].

I discuss the sensitivity of IA with respect to CSIT imperfections in Section 4.2. Moreover, I argue in Section 3.4 that subspace-based IA can be realized by explicit Grassmannian CSI feedback, while max-SINR IA requires explicit Gramian CSI feedback. Hence, my predictive Grassmannian and Stiefel manifold feedback algorithms, as explained in [110, 111], *[Article 6]*, are applicable for the CSIT acquisition of IA.

5.1.2. Joint Transmission Techniques

JT techniques in principle convert an information-theoretic interference channel to a broadcast channel, by centrally optimizing the base band processing of all involved transmission points. This implies that the transceiver architectures discussed in Chapter 4 are applicable for JT, with minor modifications, since the involved transmission points commonly impose individual transmit power constraints, rather than a sum-power constraint.

JT techniques have initially been adopted within DASs, because they require high-bandwidth low-latency connections between the cooperating transmission points. Since the backhaul of mobile networks becomes increasingly powerful, JT has lately gained interest in the context of CRAN, distributed MIMO and cell-free massive MIMO. JT has the potential to exploit the full DoF provided by the system, in contrast to CoMP, which can at most achieve half the DoF through IA. Additionally, JT can improve the reliability of the wireless connectivity, by enhancing the macroscopic diversity through multi-point transmissions.

Distributed Antenna Systems

DASs are wireless communication architectures that utilize multiple spatially (geographically) distributed transmission/reception points to serve users in wireless cellular networks [219]. DASs are static infrastructures, in which remote radio heads (RRHs) are connected to the BSs by high-bandwidth low-latency dedicated connections, such as, radio-over-fiber or dedicated micro-wave links [220]. In principle this structure enables JT from all RRHs that are connected to a BS, even though often simpler schemes, such as transmission point selection, are employed. DASs are successful in a number of aspects:

- Coverage improvement and reduction of the outage probability of wireless networks, especially indoors [221, 222].
- Network capacity enhancement and improvement of the ASE [223, 224].
- Improvement of the efficiency of simultaneous wireless information and power transfer (SWIPT) methods [225].

DASs versus collocated antennas: In my article [226], I analyze the ASE of DASs and of BSs with only collocated antennas, assuming the same total amount of antennas per cell. For both, single- and multi-user MIMO transmission with limited feedback, the performance gap with respect to perfect CSIT can be significantly reduced in DASs, provided the CSI quantization codebooks are designed to account for the path-loss differences between the RRHs. Additionally, in multi-user MIMO transmission, smart multi-user scheduling can reduce the residual inter-user interference in DASs, by co-scheduling UEs that are close to different geographically separated RRHs.

DASs versus small cells: In *[Article 8]*, I compare the ASE of DASs to heterogeneous networks that employ autonomous small cells instead of RRHs. I consider uncoordinated transmission in the small cell system and joint multi-user MIMO transmission in the DAS, in both cases with limited feedback and the same total feedback overhead budget. My results show that JT in the DAS provides higher ASE not only with perfect CSIT, but even with very low feedback rates. Yet, this performance gap might partly be alleviated by applying CoMP in the small cell system, which is not investigated in *[Article 8]*.

Dynamic distributed antenna systems: In *[Conf. 3]*, I investigate the performance of an enhanced version of DASs, called dynamic distributed antenna systems (dDASs), in which the association of RRHs to BSs is not fixed, but can rather be optimized to most efficiently satisfy the demand of the individual BSs. Such an architecture requires a reconfigurable front-haul network between the RRHs and the BSs. In principle, this can be realized wirelessly, through steerable antenna arrays at the BSs that operate in a dedicated band, e.g., in the mmWave regime as noted in *[Article 1]*. As I show in *[Conf. 3]*, dDASs have the potential to significantly reduce the outage probability of mobile networks. However, an optimal association of RRHs amongst BSs requires solving combinatorial optimization problems, which becomes computationally challenging with growing problem dimensions.

CRAN, Distributed MIMO and Cell-Free Massive MIMO

Realizing the full potential of network densification requires centralized coordination of the base band processing of larger spatial areas, to mitigate interference or to even exploit signals from many transmission points to reliably convey information. This is one of the motivations that underlies the proposal of CRAN architectures [189], where the base band processing is outsourced to so-called base band units (BBUs) on cloud computer servers. The actual radio transmissions are handled by RRHs, which are dynamically connected to the BBUs over a reconfigurable front-haul network. This is similar to a dDAS, however, in a dDAS the base band processing is not outsourced, but is rather performed by the existing BSs; dDASs can thus be thought of as an intermediate step towards a full-fledged CRAN. In CRAN, both CoMP and JT techniques are applicable; commonly a combination of both is considered: JT within so-called antenna domains and CoMP in-between them [194]. The second main motivation for CRAN is to enhance the scalability of the mobile network and to support better adaptability with respect to spatio-temporal capacity demand variations caused, e.g., by daily commuters. Specifically, the spatio-temporal distribution of the network capacity can be relatively easily varied and scaled by invoking a smaller or larger number of BBUs on the cloud, and by dynamically connecting them to the appropriate RRHs.

Cell-Free massive MIMO: The basic idea of coherent JT from many spatially distributed antennas to provide uniformly good service for all users in the network has been proposed under several different names, such as, virtual/distributed/network MIMO [227–229]. Its latest manifestation in the context of massive MIMO is known as cell-free massive MIMO [191], where a very large number of single-antenna access points simultaneously serve a much smaller number of users, employing computationally simple signal processing. Relying on the well known massive MIMO phenomena of favorable propagation and channel hardening, this enables the use of computationally efficient, globally optimal transceivers and simple pilot assignment schemes.

5.2. Interference Management in Multicast Scenarios

Multicasting implies the simultaneous transmission of the same information to multiple users. In mobile communications, this feature was initially implemented to support video broadcasting, e.g., venue casting at local sports events [230–233]. Other application scenarios include software/firmware/operating-system updates for mobile Internet of Things devices [234] and, most importantly, road traffic- and safety-relevant information broadcasting in cellular-assisted intelligent transportation systems (ITS) [235], *[Article 2]*.

Base band model of the multi-antenna MCIC: Multicasting can easily be accommodated in the base band system models described in Chapter 2. Rather than UE-specific transmit symbols and precoders as in Eq. (2.3), the same symbols are transmitted to all UEs:

$$\mathbf{x}_i[n,k] = \mathbf{F}_i[n,k]\mathbf{s}_i[n,k]. \tag{5.1}$$

The corresponding achievable rate of a user $u \in \mathcal{S}_i[n,k]$ served by BS i is:

$$\begin{aligned} R_u[n,k] &= \log_2 \det\left(\mathbf{I}_{M_u} + \mathbf{J}_u[n,k]^{-1}\mathbf{S}_u[n,k]\right), \\ \mathbf{S}_u[n,k] &= \mathbf{H}_{u,i}[n,k]\mathbf{F}_i[n,k]\ \left(\mathbf{H}_{u,i}[n,k]\mathbf{F}_i[n,k]\right)^{\mathrm{H}}, \\ \mathbf{J}_u[n,k] &= \sigma^2_{z,u}\mathbf{I}_{M_u} + \sum_{j\neq i}\mathbf{H}_{u,j}[n,k]\mathbf{F}_j[n,k]\ \left(\mathbf{H}_{u,j}[n,k]\mathbf{F}_j[n,k]\right)^{\mathrm{H}}. \end{aligned} \tag{5.2}$$

However, since the BS transmits the same information to all UEs, it has to adopt a multicast transmission rate $R^{(i)}[n,k]$ that enables satisfactory transmission to all served UEs:

$$R^{(i)}[n,k] = \min_{u\in\mathcal{S}_i[n,k]} R_u[n,k]. \tag{5.3}$$

5.2.1. Multicasting in Mobile Wireless Communications

In 3GPP mobile communication systems, multicasting is supported by three technologies:

- Multimedia broadcast multicast service: MBMS specifies the basic PHY processing and the logical channels required for multicasting in mobile networks.
- Multimedia broadcast single frequency network: MBSFN enables multicasting from multiple strictly time and frequency synchronized BSs simultaneously, forming a so-called MBSFN area. This provides a significant enhancement of the SINR within the MBSFN area, especially at the former cell edge between the involved BSs.
- Public warning system: PWS is mainly intended to broadcast geographically constrained warnings to the general public, e.g., storm, tsunami or earthquake warnings. The technology therefore supports geocasting, which makes it also suitable for cellular-assisted vehicular communication scenarios, as argued in *[Article 2]*.

In its current implementation, MBMS in LTE does neither support feedback based transmission rate or precoder adaptation, nor hybrid ARQ based retransmissions. This is because there exist no methods for reconciling the feedback of multiple UEs within the standard.

Multicasting for cellular-assisted vehicular communications: As I argue in *[Article 2]*, multicasting is an enabling technology for cellular-assisted exchange of road traffic- and safety-relevant information. In ITS, basically two types of messages are exchanged: 1) periodic messages to share vehicle status information, such as, location, direction and velocity; 2) event-driven sporadic messages, such as, emergency warnings. Since both types of messages are commonly intended for many vehicles in a given geographic area, they can most efficiently be distributed via geographically constrained multicasting, as supported by the interplay of MBMS/MBSFN/PWS.

MBMS/MBSFN based V2X in LTE: In [236], we investigate the potential of LTE to efficiently support the distribution of ITS related road traffic information by MBMS/MBSFN. Our focus thereby is on two important metrics: the latency of the packet delivery and the overhead caused by this information exchange on the mobile network. We thereby cross-compare the performance of unicasting and multicasting. Our system level simulation based results show that multicasting achieves approximately a factor ten reduction in transmission latency compared to unicasting, and at the same time requires much less time-frequency resources. Additionally, we demonstrate that transmission rate adaptation based on the minimum CQI of all multicast UEs provides a further improvement in latency and efficiency. Hence, we conclude that future releases of MBMS in LTE/5G should provide support for CQI feedback based rate adaptation for dependable and efficient ITS information exchange.

5.2.2. Optimization of Multi-Antenna Multicast Transmissions

Because in multicasting the same information symbol is transmitted to several UEs, the beamformer/precoder optimization of a multi-antenna multicast transmissions has to strike a balance that satisfies all UEs. Common optimization targets are therefore the maximization of the minimum rate/SINR of all multicast users, or the minimization of the transmit power subject to QoS (minimum rate/SINR) constraints.

Single-Cell Multicasting

In single-cell systems, no interference management/coordination is required. For the sake of completeness, I nevertheless summarize below the main results for such systems.

MISO optimization: The study of transmit optimization for single cell MISO physical-layer multicasting was initiated by the work of Sidiropoulos [237]. Since then a number of important results have been obtained:

- The problems of QoS constrained and $\max. \min.$ SINR multicast transmit beamforming are NP hard [237].
- An approximate solution of these problems is possible by an SDR [238, 239]. The approximation bound of this SDR is provided in [240], showing that beamforming is strictly suboptimal for multicasting in contrast to unicasting.
- To alleviate the approximation gap of the rank-one beamforming SDR, a rank-two approximation based on the Alamouti space-time coding scheme is derived in [241].
- For multicasting in massive MIMO systems, asymptotically optimal beamformers are studied in [242], showing that these beamformers are a linear combination of the users' channel vectors.
- Extensions of multicast transmit beamforming to multi-group multicasting, where several independent groups of multicast UEs are jointly served by a BS, are studied in [243]. In this situation, it is additionally necessary to control the interference between these multicast groups.

MIMO optimization: The above results have partly been extended to the MIMO case:

- Optimal single-stream multicast transmit beamforming for MIMO systems is considered in [244]. The corresponding optimization problem is generally NP hard; yet, for the special case of $N_i = 2$ transmit antennas, the authors of [244] succeed to provide a polynomial time algorithm.
- The extension of the above results to dual stream transmission is published in [245].
- A per-stream post-equalization $\max. \min.$ SINR precoder optimization for multi-stream multicasting is proposed in [246], assuming that the receivers apply decision feedback MMSE equalizers. The authors demonstrate that their method performs very close to the MIMO multicast channel capacity.

Multi-Cell Multicasting

In multi-cell systems, a coordination of the multicast beamforming/precoding strategies of multiple BSs is required to alleviate the inter-cell interference.

MISO optimization: Recently, a number of important contributions have been achieved for multicast transmission in the MISO MCIC:

- Multi-cell multicasting is studied under the assumption of statistical CSIT in [247]. The authors propose a beamforming problem for global $\max. \min.$ long-term SINR optimization, by applying a joint optimization over all transmitters. An efficient solution of this problem is provided in [248] by a quasi-convex formulation.
- The problem of global $\max. \min.$ SINR optimization with instantaneous CSIT is studied in [249]. The authors propose an iterative decentralized solution, by decomposing the problem into sub-problems that are approximately solved by an SDR.
- I study the problem of weighted sum-rate maximization for the MISO MCIC in *[Article 11]*. I discuss a few details of this optimization problem below.

Weighted sum-rate maximization for the MISO MCIC: The goal of weighted sum-rate maximization is to identify the Pareto-boundary of the achievable rate region of the multi-cell system. The corresponding optimization problem for multicast transmission is:

$$\max_{\mathbf{C}_1[n,k],\ldots,\mathbf{C}_J[n,k]} \quad \sum_{j=1}^{J} w_j R^{(j)}[n,k], \tag{5.4}$$
$$\text{subject to :} \quad \mathbf{C}_j[n,k] \succeq 0, \;\; \text{tr}\left(\mathbf{C}_j[n,k]\right) \leq P_j, \;\; \forall j \in \{1,\ldots,J\}.$$

Here, the weights $w_j \geq 0$ determine the point achieved on the Pareto-boundary of the rate region and $\mathbf{C}_j[n,k] = \mathbf{F}_j[n,k]\mathbf{F}_j[n,k]^{\text{H}}$ are the positive semi-definite input covariance matrices. As I show in *[Article 11]*, this problem is non-convex. Furthermore, a convex reformulation similar to the approach taken in the MISO UCIC [206] is not possible, due to the additional minimization over the user rates in (5.3). Moreover, not even a global optimization by adopting the monotonic optimization framework of [210] is possible, because the multicast rates $R^{(j)}[n,k]$ are not monotonic in the individual user rates.

In *[Article 11]*, I propose a local distributed optimization approach for Problem (5.4) as detailed in Section 5.3. This method can generally only be guaranteed to converge locally; yet, in my simulations the outcome turned out to be globally optimal, as proved by a vanishing duality gap. An exemplary achievable rate region of this leakage-based multicasting scheme is shown in Figure 5.1. In this figure, the performance of leakage-based multicasting is compared to the interference unaware $\max. \min.$ SINR multicasting scheme of [237] and to the generalization of BD precoding to multicast scenarios as proposed in *[Article 11]*.[1] Figure 5.1 shows that a coordination of the multicast transmissions of multiple cells can significantly enlarge the achievable rate region. However, these improvements diminish with

[1] A comparison to additional other schemes is included in *[Article 11]*.

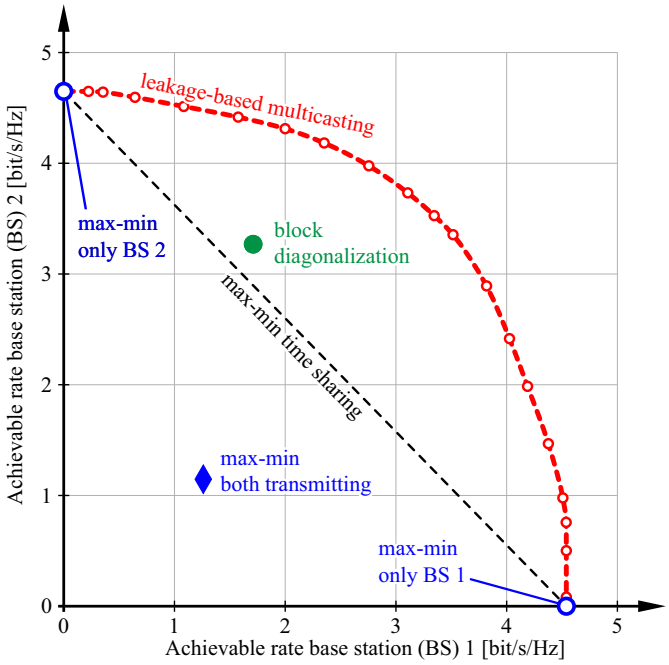


Figure 5.1.: Achievable rate region of the leakage-based multicasting scheme proposed in *[Article 11]* for a double-cell MISO multicasting scenario. The rate region in this case represents the global optimum as proved by a vanishing Lagrangian duality gap. The figure is a simplified version of *Fig. 2 (b) of [Article 11]*.

a growing number of multicast UEs per cell, if the number of transmit antennas does not scale accordingly.

MIMO optimization: The problem of coordinated precoding for the MIMO MCIC is largely unexplored. In [250], I restrict to single-stream transmit and receive beamforming and propose an iterative optimization of these beamformers. Specifically, I interpret the optimal rank-unconstrained input covariance matrix obtained from the optimization in *[Article 11]* as an SDR of the corresponding rank-one beamforming problem. Hence, I obtain a rank-one transmit beamforming solution by applying a Gaussian randomization to the optimal input covariance matrix [238], while the receive filters follow from an MMSE solution. As shown by simulations in [250], this approach can provide a significant gain over time-sharing of resources between BSs.

To gauge the quality of the rank-one beamforming approximation of the SDR performed in [250], I conduct a worst-case analysis in *[Article 12]*. Moreover, I propose in this article a rank-two multicast precoding approximation based on the Alamouti scheme to reduce the approximation gap; see Section 5.3 for more details.

Multicast Transmit Beamforming for MBMS/MBSFN based V2X in LTE:

We apply a generalized version of the optimization method of *[Article 11]*, which supports joint transmission from multiple BSs within an MBSFN area, for the distribution of ITS related road traffic information in our conference publication [251]. Compared to utilizing the best beamformer from the LTE standard-defined codebook, we achieve a substantial improvement of 3 to 6 dB in the minimum SINR of the multicast UEs in the investigated scenarios. This directly translates to an enhancement in throughput, thereby reducing the transmission latency and the overhead caused by the V2X traffic on the mobile network. Multicast beamforming thus has a significant potential for enhancing the efficiency and dependability of cellular-assisted vehicular communications.

To improve the robustness of multicast beamforming/precoding solutions in high-mobility vehicular scenarios, the angular-domain leakage bounded beamforming/precoding approach of *[Article 7]* can also be adopted for multicast situations. In many vehicular communication scenarios, access points (e.g., BSs) are placed along the roads under LOS to the vehicles; furthermore, vehicles are restricted to move along roads. Both circumstances are favorable for angular-domain based approach, since they limit the angular diversity and allow for angular separability of independent multicast user groups (e.g., vehicles driving in different directions), provided sufficient spatial resolution is achieved by the employed antenna arrays.

5.3. Scientific Contributions and Publications

My research interest in JT multi-user MIMO in DASs with limited feedback was invoked during my research stay with Prof. Robert Heath at the University of Texas at Austin. Thereby, my main focus was on the adaptation of my Grassmannian CSI quantization methods to DASs, which requires to account for the path-loss differences between the signals received from different RRHs of a DAS to achieve an efficient quantization, as I show in [226, 252]. A drawback of classical DASs is their static topology, which prohibits an adaptation to the actual spatio-temporal capacity demands of the network. To circumvent this issue, I recently investigated the potential of dDASs in *[Conf. 3]* to enhance the adaptability of the network, in order to more effectively utilize the available infrastructure.

CoMP techniques came into the focus of my research work through a collaboration with General Motors on cellular-assisted vehicular communications. The core result of this collaboration are the leakage-based coordinated multicasting techniques for the MIMO MCIC, as described in *[Article 11]*, and the application of these methods for the distribution of ITS related road traffic information in our conference paper [251]. My work shows that multicasting is an enabling technology for cellular-assisted vehicular communications,

providing significantly reduced transmission latency while minimizing the load caused by vehicular communications on the mobile network.

5.3.1. Selected Publications

The following publications within the fields of multi-point coordination and joint transmission in unicast and multicast situations contribute to this thesis:

[Article 10] S. Schwarz and M. Rupp, "Exploring coordinated multipoint beamforming strategies for 5G cellular," *IEEE Access*, vol. 2, pp. 930–946, 2014

[Conf. 3] S. Schwarz, "Remote radio head assignment and beamforming in dynamic distributed antenna systems," in *IEEE International Conference on Communications*, Kansas City, Missouri, May 2018, pp. 1–6

[Article 11] S. Schwarz and M. Rupp, "Transmit optimization for the MISO multicast interference channel," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4936 – 4949, 2015

[Article 12] S. Schwarz, "Probabilistic analysis of semidefinite relaxation for leakage-based multicasting," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 742–746, 2016

5.3.2. Summary of Scientific Contributions

The main scientific contributions of these publications are:

1. Conducting a comprehensive performance comparison of coordinated beamforming strategies for interference mitigation in the MIMO UCIC, to provide guidelines for selecting methods depending on the particular transmission situation in *[Article 10]*.
2. Development of a joint RRH assignment and coordination beamforming method for dDASs in *[Conf. 3]*.
3. Derivation of a distributed iterative coordinated multicast precoding method for weighted multicast sum-rate optimization in the MIMO MCIC in *[Article 11]*.
4. Probabilistic analysis of the worst-case approximation ratios achieved by the rank-one and rank-two precoding approximations of the optimal rank-unconstrained transmit covariance matrix of *[Article 11]* in *[Article 12]*.

[Article 10] Exploring coordinated multipoint beamforming strategies for 5G cellular: In this article, I provide an overview and a comprehensive performance comparison of coordinated beamforming strategies for interference mitigation in MIMO broadcast and interference channels. To gain a better understanding and insights into the ergodic behavior of these schemes in terms of the SINR and the achievable transmission rate, I focus on the most simple but nonetheless representative scenario of two transmitters that serve two users. I consider the joint selection of transmit and receive filters, focusing mainly on low-complexity non-iterative schemes. However, I also investigate a relatively simple MSE based iterative scheme. As expected, the best performance is achieved by the iterative scheme; MMSE and SLNR based non-iterative methods, though, do not fall far behind. My investigation provides guidelines for selecting methods depending on the particular transmission situation, in terms of signal and interference strength, the available computational power, as well as, the tolerable signaling overhead.

[Conf. 3] Remote radio head assignment and beamforming in dynamic distributed antenna systems: dDASs are an enhanced form of DASs, where the assignment of RRHs to BSs is not fixed, but can rather be dynamically established on-demand. As noted in Section 5.1.2, finding the optimal assignment of RRHs amongst BSs is a complicated optimization problem, since it is combinatorial in nature. In my recent conference contribution *[Conf. 3]*, I formulate a joint RRH assignment and coordinated beamforming optimization problem with the goal of minimizing the outage probability of the mobile network. I provide a mixed-integer second order cone program (MISOCP) formulation of this optimization problem, which can be approximately solved by applying an integer relaxation. Even though this relaxation facilitates tractability of the problem in terms of computational complexity, it implies a significant performance loss compared to an exhaustive combinatorial search over all possible RRHs assignments. Hence, the topic requires further research work.

[Article 11] Transmit optimization for the MISO multicast interference channel: In this article, I investigate weighted sum-rate maximization for the MISO MCIC. As discussed in Section 5.2.2, this optimization problem is non-convex and I did not succeed in finding a convex reformulation. I therefore resort in *[Article 11]* to an iterative local distributed optimization of the problem, by employing a dual-gradient approach. To this end, I introduce additional interference-leakage parameters that upper-bound the amount of permissible inter-cell interference. This leads to a decoupling of the rate optimization of each BS, subject to interference-leakage constraints. These individual decoupled problems are convex and can therefore be solved efficiently. Yet, the quality of the obtained solution depends on the chosen interference-leakage parameters. To determine a locally optimal solution of the weighted sum-rate optimization problem, I propose an iterative update of the

leakage parameters, by stepping along the gradient of the Lagrangian dual of the weighted sum-rate optimization with respect to the leakage parameters. The corresponding partial derivatives constituting this gradient can be calculated in a distributed manner by the BSs and are exchange by them as coordination information. Depending on the problem size, convergence of this algorithm can be slow. I therefore also propose a heuristic leakage parameter initialization, which provides a performance close to the rate boundary attained by the iterative scheme without requiring any iterations.

[Article 12] Probabilistic analysis of semidefinite relaxation for leakage-based multicasting: Attaining points on the boundary of the rate region of the leakage-based multicasting solution proposed in *[Article 11]* in general requires transmit signals that are distributed according to a multivariate Gaussian distribution with high-rank covariance matrix. In practice, however, simple low-rank transmission schemes, such as, rank-one beamforming in MISO systems, are often preferred, since they can be realized by linear transmit and receive filters. In *[Article 12]*, I investigate the performance of rank-one beamforming and rank-two Alamouti-based space time block coding for leakage-based multicasting. The rank-one and -two transmit processing schemes are obtained from the optimal rank-unconstrained transmit covariance matrix of *[Article 11]* through Gaussian randomization. I derive tight upper bounds on the rank of the optimal transmit covariance matrix of leakage-based multicasting, showing that the rank-one and -two approximations are optimal only for small numbers of multicast users in the system. I moreover analyze the worst-case approximation ratios of these two schemes, exposing that the worst-case approximation ratio of rank-one beamforming scales inversely proportional with the number of multicast users, whereas that of rank-two beamforming scales inversely proportional to the square-root of this number, hence providing a much better approximation performance. The median approximation ratios are much more benign than the worst-cast ratios in both cases, and they improve with a growing number of antennas at the UEs.

IEEE Access
The journal for rapid open access publishing

SPECIAL SECTION ON 5G WIRELESS TECHNOLOGIES: PERSPECTIVES OF THE NEXT GENERATION MOBILE COMMUNICATIONS AND NETWORKING

Received June 13, 2014, accepted August 18, 2014, date of publication August 29, 2014, date of current version September 9, 2014.

Digital Object Identifier 10.1109/ACCESS.2014.2353137

Exploring Coordinated Multipoint Beamforming Strategies for 5G Cellular

STEFAN SCHWARZ, (Member, IEEE), AND MARKUS RUPP, (Senior Member, IEEE)
Institute of Telecommunications, Vienna University of Technology, Vienna 1040, Austria

Corresponding author: S. Schwarz (stefan.schwarz@nt.tuwien.ac.at)

This work was supported in part by the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility, in part by KATHREIN-Werke KG, Rosenheim, Austria, in part by A1 Telekom Austria AG, Vienna, in part by the Federal Ministry of Economy, Family and Youth, and in part by the National Foundation for Research, Technology and Development.

ABSTRACT Cellular networks are a central part of today's communication infrastructure. The global roll-out of 4G long-term evolution is underway, ideally enabling ubiquitous broadband Internet access. Mobile network operators, however, are currently facing an exponentially increasing demand for network capacity, necessitating densification of cellular base stations (keywords: small cells and heterogeneous networks) and causing a strongly deteriorated interference environment. Coordination among transmitters and receivers to mitigate and/or exploit interference is hence seen as a main path toward 5G mobile networks. We provide an overview of existing coordinated beamforming strategies for interference mitigation in broadcast and interference channels. To gain insight into their ergodic behavior in terms of signal to interference and noise ratio as well as achievable transmission rate, we focus on a simplified but representative scenario with two transmitters that serve two users. This analysis provides guidelines for selecting the best performing method depending on the particular transmission situation.

INDEX TERMS Multiple-input multiple-output, beamforming, multiuser gains, coordinated multipoint transmission, ergodic transmission rate, linear transceivers, MMSE receiver.

I. INTRODUCTION AND SYSTEM MODEL

Cellular networks are currently experiencing a tremendous growth of data traffic, due to multimedia and internet applications that are enabled by novel devices such as smart phones and tablet computers [1], [2]. Network densification is considered as one of the most promising means to cope with this increasing demand in the near future [3] and is seen as a main path for advancing 4th generation (4G) cellular networks, i.e., Long Term Evolution (LTE), towards 5G. The resulting heterogeneous networking architectures, however, require coordination of different transmission points to handle the increased inter-cell interference encountered in dense wireless networks [4], [5]. Simple forms of coordination, such as, inter-cell interference coordination (ICIC) (Rel. 8) and its enhanced (Rel. 10) as well as further enhanced (Rel. 11) versions, are already a central part of 4G LTE. The basic idea of ICIC is to keep the inter-cell interference under control by radio resource management (RRM) methods, i.e., restricting the available time-frequency resources on traffic channels to avoid/reduce inter-cell interference at least for cell-edge users. Enhanced ICIC further extends these concepts to control channels and introduces the almost blank subframe (ABS) to mitigate interference in heterogeneous networks. With ABS, however, cell-specific reference symbols still cause significant interference between cells, which is considered in further enhanced ICIC by means of enhanced transceiver processing, such as, interference cancellation receivers; see [6], [7] for more detailed discussions on ICIC.

ICIC based RRM methods have the disadvantage that not all resources can be utilized in all cells at the same time. Spatial coordination schemes, also known as coordinated multi-point (CoMP) transmission, provide a remedy by exploiting the spatial dimension to control the interference between cells. The highest performance can thereby be achieved in combination with multiple transmit antennas at the base stations, employing coordinated beamforming and joint transmission techniques to reduce or even exploit the interference from multiple transmission points, see [8]–[14]. Within the 3rd Generation Partnership Project (3GPP), the first studies on CoMP have been conducted during the preparation of Rel. 10 [15]. The basic enablers

IEEE Access
The journal for rapid open access publishing

for CoMP, such as, extended reference signals and signaling information exchange between base stations, have been standardized with Rel. 11 [16]; however, work is still ongoing to study effects of realistic restrictions, such as, non-ideal backhaul, limited feedback capacity and so on [17]. Hence, although the basics for CoMP are already provided with 4G technology, many practical as well as theoretical questions still have to be answered to unleash the full potential of CoMP as a 5G enabler. Furthermore, with the advent of millimeter wave as a promising 5G technology [18], coordination between multiple transmitters will not only be necessary to avoid interference, but also to improve coverage by exploiting the macro diversity provided by several non-collocated transmitters, in order to mitigate severe shadowing and blockages at such high carrier frequencies.

Since the introduction of multiple antennas at the base stations and the mobile receivers, beamforming and precoding methods are also becoming increasingly popular to enhance the single cell downlink transmission in cellular networks. Especially, multiuser multiple-input multiple-output (MIMO) promises large network capacity gains by spatially multiplexing several users over the same time-frequency resources [19], [20]. Although strictly suboptimal for the MIMO broadcast channel, linear transceiver architectures, see [21], [22], are still considered as practically important for their complexity advantage compared to non-linear techniques, such as vector perturbation [23] and Tomlinson-Harashima precoding based methods [24]. Considering the multiple-input single-output (MISO) interference channel, Jorswieck et.al. investigated the structure of optimal linear beamforming vectors [25]. They showed that in case of two users, any Pareto-optimal beamformer is necessarily a convex combination of the matched filter (MF) vector for its own channel (a.k.a. maximum ratio transmission) and the zero forcing (ZF) vector for the interference channel.[1] However, finding appropriate coefficients for this convex combination, for example such as to maximize the achievable sum rate, is in general non-trivial and implies an exhaustive search over the Pareto-optimal front; only few specific special cases are known in closed-form, e.g., the coefficients that a achieve a Walrasian equilibrium [26]. For more than two users, this necessary condition generalizes such that any Pareto-optimal linear beamformer must be a linear combination of the own channel vector and all interference channel vectors. Similarly, for the MIMO interference channel, the column space of any Pareto-optimal transmit signal covariance matrix is contained in the union of the signal spaces of the channel matrices from the transmitter to all receivers [27].

The aforementioned papers consider only the design of optimal linear beamforming vectors and precoding matrices, without restricting the complexity at the receiver side. In mobile communications, however, the computational capabilities of mobile devices (smart phones, tablets, and more) are typically even more restricted than those of base stations, necessitating linear receive filters to limit complexity. In particular, methods that require large amounts of iterations to achieve an optimum are prohibitive. We hence consider in this paper linear transmission and reception over MIMO channels $\mathbf{H}_{lk} \in \mathbb{C}^{N_R \times N_T}$, from K sources to L users:

$$\mathbf{y}_l = \sum_{k=1}^{K} \mathbf{H}_{lk}\mathbf{f}_k\sqrt{\alpha_k P}s_k + \mathbf{v}_l,$$
$$l \in \{1, 2, ..., L\}, \text{s.t.} \sum_{k=1}^{K} \alpha_k = 1, \quad (1)$$

encompassing both, downlink multiuser MIMO as well as CoMP transmissions from multiple transmitters. To simplify the exposition, we assume that each base station is equipped with N_T transmit antennas and each user has N_R receive antennas. The vector $\mathbf{f}_k \in \mathbb{C}^{N_T \times 1}$ denotes the beamformer of user k. We assume throughout this paper that the beamforming vectors are normalized, i.e., $\|\mathbf{f}_k\|_2 = 1$. Each receiver l observes its intended signal $\mathbf{H}_{ll}\mathbf{f}_l\sqrt{\alpha_l P}s_l$, but also interference caused by the transmission to other users, as well as additive Gaussian noise $\mathbf{v}_l \in \mathcal{N}_{\mathbb{C}}(\mathbf{0}, N_o\,\mathbf{I}_{N_R})$. In this article we confine ourselves to the simplest case of two users ($L = 2$) supported by either a single base station or by two cooperating base stations. Of course this simplified scenario cannot describe all effects in coordinated multipoint transmission/reception; specifically MIMO interference alignment (IA) [28], [29] becomes interesting only with more than two transmitters. However, the most relevant relations to gain an understanding of the basic behavior of the considered transceiver architectures can already be studied in the two user setup. Extensions to more base stations and users are in most cases straightforward. Note also that scenarios with many coordinated coherent transmissions are unfeasible as all channels need to be estimated and exchanged between all base stations within the channel's coherence time. Following the notation of 3GPP LTE, we refer to the base station as evolved Node B (eNodeB) and to the user as user equipment (UE). The power distribution of the sum power P is performed by selecting power fractions α_k. We thus reduce the complex problem by two restrictions:

- BF Beamforming: in multiuser scenarios with a large number of users, limited number of antennas and power limitation will require to use beamforming, e.g., in order to steer the transmit signal of a user along the maximum eigenmode of the channel matrix, performing maximum eigenmode transmission (MET) [12], [30], [31]. It is thus a useful restriction even though in a pure two user setup using only a single stream may not be sufficient and may appear overly restrictive.
- 2U Two users: to study the basic behavior of interference mitigation algorithms, it is sufficient to assume just a second user as the interfering source. This second user may be interpreted as the sole source of interference. Note that in multiuser scenarios with a large number of users, the a-priori knowledge of all interference

[1] Notice that this is only a necessary condition.

channels is not feasible. Hence, in many practical cases of interest a second users will be the single source of major interference.

Here two similar but different scenarios are of interest:

1) Broadcasting (Single eNodeB) scenario:

$$\mathbf{y}_1 = \mathbf{H}_{11}\mathbf{f}_1\sqrt{\alpha P}s_1 + \mathbf{H}_{11}\mathbf{f}_2\sqrt{(1-\alpha)P}s_2 + \mathbf{v}_1, \quad (2)$$

$$\mathbf{y}_2 = \mathbf{H}_{21}\mathbf{f}_1\sqrt{\alpha P}s_1 + \mathbf{H}_{21}\mathbf{f}_2\sqrt{(1-\alpha)P}s_2 + \mathbf{v}_2. \quad (3)$$

2) Interference Channel (Double eNodeB) scenario:

$$\mathbf{y}_1 = \mathbf{H}_{11}\mathbf{f}_1\sqrt{\alpha P}s_1 + \mathbf{H}_{12}\mathbf{f}_2\sqrt{(1-\alpha)P}s_2 + \mathbf{v}_1, \quad (4)$$

$$\mathbf{y}_2 = \mathbf{H}_{21}\mathbf{f}_1\sqrt{\alpha P}s_1 + \mathbf{H}_{22}\mathbf{f}_2\sqrt{(1-\alpha)P}s_2 + \mathbf{v}_2. \quad (5)$$

Although mathematically they both appear very similar as $K = 2$ sources transmit to $L = 2$ users, there is the difference that in the single eNodeB scenario the intended signal as well as the interference are received over the same channel (i.e., $\mathbf{H}_{11} = \mathbf{H}_{12}$ and $\mathbf{H}_{21} = \mathbf{H}_{22}$), whereas different channel matrices are effective in the double eNodeB setup as illustrated in Figure 1. Correspondingly, some of the algorithms we consider in this paper react very differently in both scenarios. Note that the power share α introduces a power control as well as a sum power constraint (1). Such a sum power constraint is natural for the first scenario with a single eNodeB transmitting to multiple receivers. For the double eNodeB scenario, mostly individual power constraints per transmitter are considered in literature. However, we assume that the total transmit power can be shared between the eNodeBs, amounting to just another form of coordination as considered, e.g., in [32] and simplifying the complex problem considerably. We will show in the following that this allows to decouple the problem of finding optimal precoding vectors and optimal power shares.

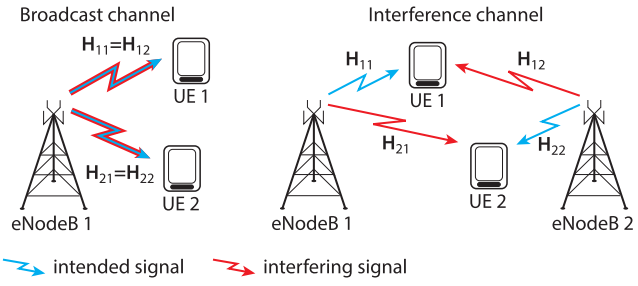


FIGURE 1. Visualization of the considered single and double eNodeB scenarios.

In order to separate the intended signal from the interference and to estimate the transmitted data symbol s_k, user k applies a linear filter $\mathbf{g}_k \in \mathbb{C}^{N_R \times 1}$ to the received signal, e.g.,

$$\hat{s}_1 = \mathbf{g}_1^H\mathbf{H}_{11}\mathbf{f}_1\sqrt{\alpha P}s_1 + \mathbf{g}_1^H\mathbf{H}_{12}\mathbf{f}_2\sqrt{(1-\alpha)P}s_2 + \mathbf{g}_1^H\mathbf{v}_1 \quad (6)$$

Similar expressions are obtained for the single eNodeB scenario and the other user. The corresponding signal to interference and noise ratios (SINRs) of the users are:

$$\mathrm{SINR}_1 = \frac{|\mathbf{g}_1^H\mathbf{H}_{11}\mathbf{f}_1|^2\alpha P}{N_o\|\mathbf{g}_1\|_2^2 + |\mathbf{g}_1^H\mathbf{H}_{12}\mathbf{f}_2|^2(1-\alpha)P} \quad (7)$$

$$= \frac{\alpha P}{N_o\|\mathbf{g}_1\|_2^2}|\mathbf{g}_1^H\mathbf{H}_{11}\mathbf{f}_1|^2 \times \left[1 - \frac{|\mathbf{g}_1^H\mathbf{H}_{12}\mathbf{f}_2|^2(1-\alpha)P}{N_o\|\mathbf{g}_1\|_2^2 + |\mathbf{g}_1^H\mathbf{H}_{12}\mathbf{f}_2|^2(1-\alpha)P}\right], \quad (8)$$

$$\mathrm{SINR}_2 = \frac{|\mathbf{g}_2^H\mathbf{H}_{22}\mathbf{f}_2|^2(1-\alpha)P}{N_o\|\mathbf{g}_2\|_2^2 + |\mathbf{g}_2^H\mathbf{H}_{21}\mathbf{f}_1|^2\alpha P} \quad (9)$$

$$= \frac{(1-\alpha)P}{N_o\|\mathbf{g}_2\|_2^2}|\mathbf{g}_2^H\mathbf{H}_{22}\mathbf{f}_2|^2 \times \left[1 - \frac{|\mathbf{g}_2^H\mathbf{H}_{21}\mathbf{f}_1|^2\alpha P}{N_o\|\mathbf{g}_2\|_2^2 + |\mathbf{g}_2^H\mathbf{H}_{21}\mathbf{f}_1|^2\alpha P}\right], \quad (10)$$

where we assume unit power transmit symbols s_k w.l.o.g. To obtain the SINRs of the single eNodeB scenario we simply set $\mathbf{H}_{12} = \mathbf{H}_{11}$ and $\mathbf{H}_{22} = \mathbf{H}_{21}$, respectively. The alternative forms (8) and (10) will become useful later when we derive general expressions. We already recognize the typical problematic of having a first term that is to maximize (e.g., $|\mathbf{g}_1^H\mathbf{H}_{11}\mathbf{f}_1|^2$) while a second penalty term occurs that needs to be as small as possible, i.e., $|\mathbf{g}_1^H\mathbf{H}_{12}\mathbf{f}_2|^2$. Treating the remaining interference as additional Gaussian noise, the transmission rate achievable with the precoders $\mathbf{f}_1$, $\mathbf{f}_2$ and the receive filters $\mathbf{g}_1$, $\mathbf{g}_2$ is upper bounded by Shannon's famous capacity formula valid for point to point additive white Gaussian noise (AWGN) channels [33]:

$$[R_1, R_2] = \left[\log_2(1+\mathrm{SINR}_1), \log_2(1+\mathrm{SINR}_2)\right]. \quad (11)$$

Notice that this achievable rate region is not equivalent to the capacity region of the broadcast and interference channels. For the broadcast channel, the capacity region is known to be achievable by nonlinear dirty-paper coding based schemes [34]. The capacity region of the interference channel, however, is known only for some special cases, e.g., very strong interference [35]. To optimize the efficiency of the network, the ultimate linear transceiver design goal is to maximize the achievable sum rate of the multiuser transmission. Unfortunately, this problem is hard and we consider instead the following simpler problem:

Two User Problem: Given a power distribution α, find linear receiver filters $\mathbf{g}_1, \mathbf{g}_2$, as well as precoding vectors $\mathbf{f}_1, \mathbf{f}_2$ such that the SINR for each receiver, i.e., (7) resp. (9), is maximized with respect to its own desired signal.

We have a coupled system now for which the maximum of the first user's SINR in general does not show a maximum in the second user's SINR. The solution depends on the level of cooperation. While in a single eNodeB setup, cooperation of the two data streams appears naturally, the double eNodeB setup may require a different level of cooperation. We consider exchanging the power distribution as well as channel state information for cooperation but not the information data, i.e., if two eNodeBs are employed they only support one user each.

Notice that maximizing the SINR is equivalent to minimizing the mean squared error (MSE) of the

estimation [36]–[39]:

$$\mathrm{MSE}_k = \mathrm{E}\left[\left|\hat{s}_k - s_k\right|^2\right]. \tag{12}$$

Hence, minimum MSE (MMSE) transceiver designs play an important role in our investigation.

A. LITERATURE SURVEY AND CONTRIBUTIONS

Due to the difficulty of analyzing the capacity of general interference channels, researchers have resorted to determining the Degrees of Freedom (DoF) of such systems, i.e., the pre-log factor of the sum capacity specifying the slope of the capacity curve at high signal to noise ratio (SNR), and to finding transceivers that achieve these results. Cadambe and Jafar [28] as well as Maddah-Ali et.al. [29] were the first to develop IA techniques for MIMO interference channels, enabling interference-free transmission to a maximum number of users in parallel employing linear precoders and receive filters. Since then, a large body of literature has evolved around IA; see [40] and references therein for an overview.

Similar to IA methods for interference channels, interference cancellation techniques for the MIMO broadcast channel that achieve the maximum DoF, such as zero-forcing beamforming [21] and block-diagonalization precoding [9], are also very popular. Although such schemes can maximize the spatial multiplexing gain at asymptotically high SNR, they suffer a significant SNR gap with respect to optimal performance and compared to other techniques; see [21], [41].

More robust strategies, accounting not only for the multiuser interference but also for the noise, have thus been proposed in literature, which relax the need for perfect interference cancellation/alignment to improve the desired signal power. Regularized zero-forcing beamforming [21] and block-diagonalization precoding [42] improve the low SNR performance by explicitly considering the noise in the precoder design. Also, iterative designs that jointly optimize the precoders and receive filters according to MSE [43], SINR [44] or sum rate [45] criteria exist, gaining in throughput compared to independent transmit/receive filter designs, however, increasing complexity and signaling overhead substantially. Similar approaches have been considered for IA as well, e.g., regularized IA [46] and iterative algorithms, such as the max SINR algorithm [41].

Although we do not consider channel state information acquisition at the receiver and transmitter in this article, it is a central ingredient for the methods and algorithms presented throughout this paper. Specifically, channel estimation at the receiver and channel state information feedback to the transmitter have attracted lots of attention and a vast body of literature exists; see [47]–[51] and references therein.

In this article, we revisit existing linear transceiver design principles and provide a comprehensive performance comparison for scenarios with two transmitters and two receivers. A similar investigation for mostly different beamformers and precoders has been conducted in [52, Ch. 12] for the MIMO broadcast channel. This work is on the one hand more general than ours, because it is not restricted to two users; but on the other hand [52] is also more restrictive than our derivations, because it does not consider the joint selection of transmitter and receive filters, but only evaluates different precoding strategies. We furthermore analyze the approximate ergodic performance of the presented transceiver architectures, providing important insides and guidelines for selecting methods depending on the particular transmission situation and the available computational power as well as the tolerable signaling overhead.

B. NOTATION

We write the singular value decomposition (SVD) of a matrix $\mathbf{H}_{11} \in \mathbb{C}^{N_R \times N_T}$ as $\mathbf{H}_{11} = \mathbf{U}_{11}\Sigma_{11}\mathbf{V}_{11}^H$, with the singular values on the diagonal of $\mathbf{\Sigma}_{11}$ ordered from largest to smallest, that is $\sigma_{11,1} = \sigma_{11,\max}, \sigma_{11,N_*} = \sigma_{11,\min}$. The rank of such matrix is $N_* = \mathrm{rank}(\mathbf{H}_{11})$, which is typically given by $\min\{N_R, N_T\}$. We denote the left pseudo inverse of a matrix by "#": $\mathbf{A}^{\#}\mathbf{A} = \mathbf{I}$. Frequently used variables and parameters are summarized in Table 1.

TABLE 1. Table of frequently used variables and parameters.

notation	support	meaning
$\mathbf{H}_{lk}$	$\mathbb{C}^{N_R \times N_T}$	channel matrix from source k to user l
$\mathbf{f}_l$	$\mathbb{C}^{N_R \times 1}$	precoding vector of user l
$\mathbf{g}_l$	$\mathbb{C}^{N_T \times 1}$	equalizer vector of user l
$\mathbf{v}_l$	$\mathbb{C}^{N_R \times 1}$	additive noise vector of user l
$\mathbf{y}_l$	$\mathbb{C}^{N_R \times 1}$	received vector of user l
SINR_l	$\mathbb{R}_+$	instantaneous SINR of user l
$\overline{\mathrm{SINR}}_l$	$\mathbb{R}_+$	ergodic SINR of user l
PS_l	$\mathbb{R}_+$	desired signal's maximal power of user l
PI_l	$\mathbb{R}_+$	maximal interference power at user l
α	$[0, 1]$	power distribution factor
N_*	$\mathbb{R}_+$	rank of a matrix

C. ORGANIZATION OF THE ARTICLE

In Section II we formulate a set of different algorithms, all supposedly well suited to decrease interference in the considered scenarios. In Section III we derive approximate ergodic performance of these algorithms, that is, their SINR and achievable transmission rate over an ensemble of random channel realization. The predicted values are compared to Monte Carlo (MC) simulations and in most cases we find excellent agreements, explaining the influential terms of the algorithms in question. We derive optimal power distributions that either allow for a fair (= equal) rate or for maximal sum rate. Finally, we conclude our findings in Section IV. Detailed derivations and proofs are provided in the appendix.

II. LINEAR TRANSCEIVER ARCHITECTURES

A. INTERFERENCE IGNORANT MET

Assuming uncoordinated transmission with only local channel state information (CSI) available at the eNodeBs, a reasonable precoding strategy is to select the precoding vector of user l as the right singular vector corresponding to the

largest singular value, i.e., the largest eigenmode of channel matrix $\mathbf{H}_{ll}$. This strategy is known as MET [12], [30], [31] and maximizes the SNR of the received signal, entirely ignoring the amount of interference caused to other users. We apply an SVD to the channel matrices to obtain $\mathbf{H}_{11} = \mathbf{U}_{11}\Sigma_{11}\mathbf{V}_{11}^H$ and $\mathbf{H}_{22} = \mathbf{U}_{22}\Sigma_{22}\mathbf{V}_{22}^H$, with $\mathbf{v}_{11,1}$ and $\mathbf{v}_{22,1}$ denoting the right singular vectors corresponding to the largest singular values $\sigma_{11,1}$ and $\sigma_{22,1}$, respectively. Starting again with Eqs. (7) and (9) and ignoring the interference terms as such, i.e., treating them instead as ordinary white noise, we find a corresponding SNR maximizing equalizer solution in $\mathbf{g}_1, \mathbf{g}_2$ for the given precoding vectors[2]:

$$\mathbf{f}_{1,\text{II}-\text{MET}} = \mathbf{v}_{11,1}, \quad \mathbf{f}_{2,\text{II}-\text{MET}} = \mathbf{v}_{22,1}, \tag{13}$$

$$\mathbf{g}_{1,\text{II}-\text{MET}} = \mathbf{u}_{11,1}, \quad \mathbf{g}_{2,\text{II}-\text{MET}} = \mathbf{u}_{22,1}, \tag{14}$$

which is valid for the double eNodeB scenario; for the single eNodeB we simply set $\mathbf{H}_{22} = \mathbf{H}_{21}$.

Notice that the precoders (13) are MF vectors for the effective channels $\mathbf{g}_{1,\text{II}-\text{MET}}\mathbf{H}_{11}$ and $\mathbf{g}_{2,\text{II}-\text{MET}}\mathbf{H}_{22}$, respectively; hence, they fulfill the necessary condition for Pareto-optimal linear beamformers stated in [25]. Still, these beamformers are known to be non-Pareto-optimal; they only provide an uncoordinated Nash-equilibrium from which a user would not want to deviate, given the receive filters and the precoder(s) of the other user(s) [53].

B. ORTHOGONAL FILTERING

If the users are aware of the interfering precoders, i.e., they can estimate not only the interfering channel but also the precoder applied by the interfering eNodeB, the receive vectors can be selected as orthogonal to the interfering channel to set the interference term equal to zero. Assuming again that the eNodeBs apply MET beamforming, the receive vectors $\mathbf{g}_1, \mathbf{g}_2$ are obtained by enforcing the orthogonality conditions:

$$\mathbf{g}_1^H\mathbf{H}_{12}\mathbf{f}_2 = \mathbf{g}_1^H\mathbf{H}_{12}\mathbf{v}_{22,1} = 0, \tag{15}$$

$$\mathbf{g}_2^H\mathbf{H}_{21}\mathbf{f}_1 = \mathbf{g}_2^H\mathbf{H}_{21}\mathbf{v}_{11,1} = 0. \tag{16}$$

This means that, e.g., $\mathbf{g}_2$ has to lie in the orthogonal complement of $\mathbf{x}_1 = \mathbf{H}_{21}\mathbf{v}_{11,1}$. Hence, employing an orthonormal basis $\mathbf{B}_1$ for the orthogonal complement of $\mathbf{x}_1$, we can write a general solution for the receive filter as

$$\mathbf{g}_2 = \mathbf{B}_1\tilde{\mathbf{g}}_2, \quad \mathbf{B}_1 \in \mathbb{C}^{N_R\times(N_R-1)}, \ \mathbf{B}_1^H\mathbf{B}_1 = \mathbf{I}_{N_R-1}. \tag{17}$$

The vector $\tilde{\mathbf{g}}_2$ is found by additionally maximizing the resulting SNR experienced by user 2:

$$\begin{aligned}\text{SNR}_2 &= \max_{\tilde{\mathbf{g}}_2} \frac{\left\|\mathbf{g}_2^H\mathbf{H}_{22}\mathbf{v}_{22,1}\right\|_2^2}{N_o\left\|\mathbf{g}_2\right\|_2^2} \\ &= \frac{\sigma_{22,1}^2}{N_o}\max_{\tilde{\mathbf{g}}_2}\frac{\tilde{\mathbf{g}}_2^H\mathbf{B}_1^H\mathbf{u}_{22,1}\mathbf{u}_{22,1}^H\mathbf{B}_1\tilde{\mathbf{g}}_2}{\tilde{\mathbf{g}}_2^H\tilde{\mathbf{g}}_2},\end{aligned}$$

$$\Rightarrow \tilde{\mathbf{g}}_{2,\text{OF}} \propto \mathbf{B}_1^H\mathbf{u}_{22,1},$$

$$\Rightarrow \mathbf{g}_{2,\text{OF}} = \mathbf{B}_1\mathbf{B}_1^H\mathbf{u}_{22,1} = \left[\mathbf{I}_{N_R} - \frac{\mathbf{x}_1\mathbf{x}_1^H}{\mathbf{x}_1^H\mathbf{x}_1}\right]\mathbf{u}_{22,1}, \tag{18}$$

i.e., the optimal receive filter vector is obtained by projecting the left singular vector corresponding to largest singular value onto the orthogonal complement of the effective interference channel. Proceeding similarly to determine the filter $\mathbf{g}_1$, we obtain with $\mathbf{x}_2 = \mathbf{H}_{12}\mathbf{v}_{22,1}$:

$$\mathbf{g}_{1,\text{OF}} = \left[\mathbf{I}_{N_R} - \frac{\mathbf{x}_2\mathbf{x}_2^H}{\mathbf{x}_2^H\mathbf{x}_2}\right]\mathbf{u}_{11,1}. \tag{19}$$

C. ORTHOGONAL PRECODING

An analog approach to orthogonal filtering as considered above is to select the receive filters to support maximum gain, $\mathbf{g}_1 = \mathbf{u}_{11,1}, \mathbf{g}_2 = \mathbf{u}_{22,1}$, while the precoding vectors are chosen to complete an orthogonal vector product, resulting in zero interference. Obviously, there is also a multitude of solutions to this problem formulation. The algorithm can be viewed as dual to orthogonal filtering. Surprisingly, it shows a different performance as demonstrated in Section III. With $\mathbf{x}_1 = \mathbf{H}_{12}^H\mathbf{u}_{11,1}$, $\mathbf{x}_2 = \mathbf{H}_{21}^H\mathbf{u}_{22,1}$, the precoders are:

$$\begin{aligned}\mathbf{f}_{1,\text{OP}} &= \left[\mathbf{I}_{N_T} - \frac{\mathbf{x}_2\mathbf{x}_2^H}{\mathbf{x}_2^H\mathbf{x}_2}\right]\mathbf{v}_{11,1}, \\ \mathbf{f}_{2,\text{OP}} &= \left[\mathbf{I}_{N_T} - \frac{\mathbf{x}_1\mathbf{x}_1^H}{\mathbf{x}_1^H\mathbf{x}_1}\right]\mathbf{v}_{22,1}.\end{aligned} \tag{20}$$

In contrast to orthogonal filtering, orthogonal precoding requires coordination of the eNodeBs for CSI exchange and it is hence more difficult to achieve than orthogonal filtering. Still, when the users are equipped with only one antenna $N_R = 1$, orthogonal precoding is the only possibility to eliminate interference. It is also worth mentioning that orthogonal precoding does not include the power distribution αP at all. Hence, it requires less information exchange to facilitate coordination than, for example, the leakage based scheme reviewed further down.

Relating this method to the necessary condition for Pareto-optimal linear beamforming design [25], it corresponds to the extreme case that the specific beamformer is selected as a pure ZF vector; hence, compared to the pure MF strategy of Section II-A, this solution represents the "opposite" convex combination of MF and ZF. However, also this solution is in general (at finite SNR) known to be suboptimal [54].

D. INTERFERENCE AWARE MMSE

Given a fixed set of precoding vectors $\mathbf{f}_1, \mathbf{f}_2$, the interference-aware SINR optimal pair of filter vectors $\mathbf{g}_1, \mathbf{g}_2$ can be computed in closed-form [39], [55], [56].

Theorem 2.1: Given precoding vectors $\mathbf{f}_1, \mathbf{f}_2$ *and the power distribution* α*, the SINR optimal filter vectors* $\mathbf{g}_1, \mathbf{g}_2$

[2]This solution is up to a scaling factor also an MMSE solution under the condition that the interference behaves as white noise, $\mathbf{g}_{1,\text{II}-\text{MMSE}} = \mathbf{H}_{11}\mathbf{f}_1 \propto \mathbf{u}_{11,1}, \mathbf{g}_{2,\text{II}-\text{MMSE}} = \mathbf{H}_{22}\mathbf{f}_2 \propto \mathbf{u}_{22,1}$. The performance is identical in both cases.

are:

$$\mathbf{g}_{1,\mathrm{IA-MMSE}} = [\mathbf{H}_{12}\mathbf{f}_2\mathbf{f}_2^H\mathbf{H}_{12}^H(1-\alpha)P + N_o\mathbf{I}_{N_R}]^{-1}\mathbf{H}_{11}\mathbf{f}_1,$$
$$\mathbf{g}_{2,\mathrm{IA-MMSE}} = [\mathbf{H}_{21}\mathbf{f}_1\mathbf{f}_1^H\mathbf{H}_{21}^H\alpha P + N_o\mathbf{I}_{N_R}]^{-1}\mathbf{H}_{22}\mathbf{f}_2. \quad (21)$$

Their corresponding SINRs are obtained as:

$$\mathrm{SINR}_1 = \mathbf{f}_1^H\mathbf{H}_{11}^H[\mathbf{H}_{12}\mathbf{f}_2\mathbf{f}_2^H\mathbf{H}_{12}^H(1-\alpha)P + N_o\mathbf{I}]^{-1}\mathbf{H}_{11}\mathbf{f}_1\alpha P,$$
$$= \frac{\alpha P}{N_o}\left(\|\mathbf{H}_{11}\mathbf{f}_1\|_2^2 - \frac{(1-\alpha)P\left|\mathbf{f}_1^H\mathbf{H}_{11}^H\mathbf{H}_{12}\mathbf{f}_2\right|^2}{N_o + (1-\alpha)P\|\mathbf{H}_{12}\mathbf{f}_2\|_2^2}\right), \quad (22)$$

$$\mathrm{SINR}_2 = \mathbf{f}_2^H\mathbf{H}_{22}^H[\mathbf{H}_{21}\mathbf{f}_1\mathbf{f}_1^H\mathbf{H}_{21}^H\alpha P + N_o\mathbf{I}]^{-1}\mathbf{H}_{22}\mathbf{f}_2(1-\alpha)P,$$
$$= \frac{(1-\alpha)P}{N_o}\left(\|\mathbf{H}_{22}\mathbf{f}_2\|_2^2 - \frac{\alpha P\left|\mathbf{f}_2^H\mathbf{H}_{22}^H\mathbf{H}_{21}\mathbf{f}_1\right|^2}{N_o + \alpha P\|\mathbf{H}_{21}\mathbf{f}_1\|_2^2}\right). \quad (23)$$

For the single eNodeB setup simply substitute $\mathbf{H}_{12} = \mathbf{H}_{11}$ *and* $\mathbf{H}_{22} = \mathbf{H}_{21}$*, respectively.*

In contrast to the interference-ignorant algorithm above, this method requires perfect knowledge of the interfering channel and precoder at the user, similar to orthogonal filtering. Hence it requires estimation not only of the own channel matrix but also of the interference, thus substantially increasing the receiver complexity. Nonetheless, such interference-aware algorithms are currently in the focus of standardization, owing to their good performance [57].

The alternative forms (22) and (23) provide an equivalent but simplified problem formulation in which the issue of finding the receiver filters $\mathbf{g}_1, \mathbf{g}_2$ has been removed. Here, the complicated problem in terms of the precoding vectors $\mathbf{f}_1, \mathbf{f}_2$ is showing: one can maximize the first terms $\|\mathbf{H}_{11}\mathbf{f}_1\|_2^2$, $\|\mathbf{H}_{22}\mathbf{f}_2\|_2^2$ which relates to the mentioned MET solution just as in Equation (13). However, the so found solutions appear as penalty in the second term. Alternatively, one can try minimizing the second term. A solution of this is to satisfy the orthogonality relations $\mathbf{f}_1^H\mathbf{H}_{11}^H\mathbf{H}_{12}\mathbf{f}_2 = 0$ and $\mathbf{f}_2^H\mathbf{H}_{22}^H\mathbf{H}_{21}\mathbf{f}_1 = 0$, just as in orthogonal precoding. However, this solution may reduce the first term severely. Maximizing (22), (23) thus provides a starting point for many different strategies.

Furthermore, the formulation in (22) and (23) allows to derive upper bounds for all MMSE based algorithms.

Lemma 2.1: The upper performance of the two-user beamforming scenario is obtained if the interference terms are zero and the transmission path obtains highest gain (MET), that is at $\sigma_{11,1}^2$ *and* $\sigma_{22,1}^2$*, respectively. The so obtained achievable rate curve is given by:*

$$\{\overline{R}_1, \overline{R}_2\} = \left\{\log_2\left(1 + \frac{\sigma_{11,1}^2\alpha P}{N_o}\right), \log_2\left(1 + \frac{\sigma_{22,1}^2(1-\alpha)P}{N_o}\right)\right\}, \quad (24)$$

for the double eNodeB scenario; for the single eNodeB we have to replace $\sigma_{22,1}^2$ *by* $\sigma_{21,1}^2$.

Proof: We can maximally select $\mathbf{g}_i = \mathbf{u}_{ii,1}$ and $\mathbf{f}_i = \mathbf{v}_{ii,1}; i = 1, 2$ to select the largest transmission gain on each path. Assuming the interference is zero, we obtain the upper bound. We do not expect to truly achieve such upper bound as selecting $\mathbf{g}_i$ and $\mathbf{f}_i$ in such a way, will result in nonzero interference. □

In contrast to this algorithm, we can randomly choose the precoding vectors and then compute the optimal receive filters. This method will further be referred to as Random-MMSE or RMMSE algorithm and provide lower performance bounds for MMSE based techniques.

E. MAXIMUM INTERFERENCE PRECODING ALGORITHM

In the following algorithm we assume uncoordinated transmission and consider the lack of coordination between the eNodeBs with an appropriate robust receiver design. The receivers are supposed to be aware of the interference by observing the interfering channels, but have no information about the precoding vectors applied at the interfering eNodeBs. Relating this assumption to LTE, it means that a UE is able to estimate the interfering channel matrix from the CSI reference symbols (RS) applied at the interfering base station, but the user is not capable of exploiting the interfering UE-RS to determine the interfering precoder as well.

The goal of this receiver design is to maximize the SINR of a user under a worst case assumption on the interfering precoder, thus the acronym MAximizing interference Precoding algorithm (MAP).[3] Considering, e.g., user 1 in the two user interference channel, the optimization problem for the robust receiver $\mathbf{g}_{1,\mathrm{MAP}}$ is then:

$$\mathbf{g}_{1,\mathrm{MAP}} = \arg\max_{\mathbf{g}_1} \mathrm{SINR}_{1,\mathrm{MAP}} = \arg\max_{\mathbf{g}_1} \min_{\mathbf{f}_2, \|\mathbf{f}_2\|_2=1} \frac{\mathbf{g}_1^H\mathbf{H}_{11}\mathbf{f}_1\mathbf{f}_1^H\mathbf{H}_{11}^H\mathbf{g}_1\alpha P}{\mathbf{g}_1^H[N_o\mathbf{I}_{N_R} + \mathbf{H}_{12}\mathbf{f}_2\mathbf{f}_2^H\mathbf{H}_{12}^H(1-\alpha)P]\mathbf{g}_1}. \quad (25)$$

The worst case is given if the interference channel $\mathbf{H}_{12}$ is excited by the precoding vector that matches its largest singular vector: $\mathbf{f}_2 = \mathbf{v}_{12,1}$. We again assume that the eNodeB serves the intended user along the maximal eigenmode, i.e., it performs MET beamforming according to Equation (13). Based on such worst case interference behavior, the receive filter is selected following the MMSE principle Theorem 2.1. The receive filter of user 2 is obtained similarly. As the interfering eNodeB typically does not select such a worst case precoder, the performance is much better on average.[4]

F. LEAKAGE BASED PRECODING

The question now rises if better precoding performance than with MET is possible, by letting the eNodeBs cooperate in certain ways. One idea in this direction is to maximize the so called signal to leakage and noise ratio (SLNR), that is, the ratio of what is sent to the intended user relative to the

[3]This form of minmax optimization is also referred to as robust filter design.

[4]In fact, the probability of selecting this precoder when the interfering eNodeB applies MET beamforming to its own user is equal to zero.

interference caused to the other user [58]:

$$\mathrm{SLNR}_1 = \max_{\mathbf{f}_1, \|\mathbf{f}_1\|_2=1} \frac{\mathbf{f}_1^H \mathbf{H}_{11}^H \mathbf{H}_{11} \mathbf{f}_1 \alpha P}{\mathbf{f}_1^H \left[\mathbf{H}_{21}^H \mathbf{H}_{21} \alpha P + N_R N_o \mathbf{I}_{N_T}\right] \mathbf{f}_1}, \quad (26)$$

$$\mathrm{SLNR}_2 = \max_{\mathbf{f}_2, \|\mathbf{f}_2\|_2=1} \frac{\mathbf{f}_2^H \mathbf{H}_{22}^H \mathbf{H}_{22} \mathbf{f}_2 (1-\alpha) P}{\mathbf{f}_2^H \left[\mathbf{H}_{12}^H \mathbf{H}_{12} (1-\alpha) P + N_R N_o \mathbf{I}_{N_T}\right] \mathbf{f}_2}. \quad (27)$$

Our formulation including the precoding vectors is indeed equivalent to the original one. Furthermore, we have taken advantage of the fact that precoding vectors are normalized, which allows to embrace also the noise power N_o. Note that we have adapted the original method [58] to account for the power allocation applied by the eNodeBs as power allocation was not excluded in the original formulation. Both terms (26) and (27) are now Rayleigh quotients for the precoding vectors and can thus be solved for[5]:

$$\begin{aligned} \mathbf{f}_{1,\mathrm{L}} &= \max \mathrm{eig} \left[\mathbf{H}_{21}^H \mathbf{H}_{21} \alpha P + N_R N_o \mathbf{I}_{N_T}\right]^{-1} \mathbf{H}_{11}^H \mathbf{H}_{11}, \\ \mathbf{f}_{2,\mathrm{L}} &= \max \mathrm{eig} \left[\mathbf{H}_{12}^H \mathbf{H}_{12} (1-\alpha) P + N_R N_o \mathbf{I}_{N_T}\right]^{-1} \mathbf{H}_{22}^H \mathbf{H}_{22}. \end{aligned} \quad (28)$$

with max eig denoting the eigenvector corresponding to the maximum eigenvalue. Once the precoding vectors are fixed, the receiver filters are obtained via Theorem 2.1. The presented solution is valid for the double eNodeB setup; for the single eNodeB, $\mathbf{H}_{22} = \mathbf{H}_{21}$ and $\mathbf{H}_{12} = \mathbf{H}_{11}$. Note, in order to calculate the SLNR beamformers, the eNodeBs require knowledge of the interfering channels $\mathbf{H}_{21}$ and $\mathbf{H}_{12}$, respectively. In current mobile network standards, such as LTE, UEs can only feedback CSI to their own serving base station. Hence, cooperation in form of CSI information exchange between the eNodeBs is required to enable this kind of precoding. This comment also applies to the remaining transceiver architectures considered below.

We also consider a so called weighted leakage based precoding scheme (WSLNR) in which we reformulate (26)

$$\begin{aligned} &\max_{\mathbf{f}_1} \frac{\mathbf{f}_1^H \mathbf{H}_{11}^H \mathbf{H}_{11} \mathbf{f}_1 \alpha P}{\mathbf{f}_1^H \left[\mathbf{H}_{21}^H \mathbf{H}_{21} \alpha P + N_R N_o \mathbf{I}_{N_T}\right] \mathbf{f}_1} \\ &\quad = \max_{\tilde{\mathbf{f}}_1} \frac{\tilde{\mathbf{f}}_1^H \mathbf{L}^{-\frac{H}{2}} \mathbf{H}_{11}^H \mathbf{H}_{11} \mathbf{L}^{-\frac{1}{2}} \tilde{\mathbf{f}}_1 \alpha P}{\tilde{\mathbf{f}}_1^H \tilde{\mathbf{f}}_1}, \\ &\mathbf{L} = \mathbf{H}_{21}^H \mathbf{H}_{21} \alpha P + N_R N_o \mathbf{I}_{N_T}, \end{aligned} \quad (29)$$

with the relation $\tilde{\mathbf{f}}_1 = \left[\mathbf{H}_{21}^H \mathbf{H}_{21} \alpha P + N_R N_o \mathbf{I}_{N_T}\right]^{\frac{1}{2}} \mathbf{f}_1$. Instead of taking $\mathbf{f}_1$, we take a normalized version of $\tilde{\mathbf{f}}_1$ as precoder. To obtain $\mathbf{f}_2$ we perform a similar weighting. In our simulations in Section III-C (see Figure 3) this WSLNR method exhibits superior behavior in the single eNodeB setup.

Leakage based precoding schemes are very popular in literature for optimizing the performance of interference channels [59]–[61]. Instead of considering the SLNR directly, most proposals optimize the SINR of a user subject to an upper bound on the interference leakage caused to the other users. Selecting appropriate upper bounds, it turns out that this optimization problem results in Pareto-optimal beamforming vectors [60]. However, finding the right upper bounds is another complicated and non-trivial issue, typically requiring large amounts of iterations.

[5]This is equivalent to the solution of [58].

G. MINIMAL INTERFERENCE PRECODING

Another interference-aware precoding method is to select the precoding vectors in such a way that the interference must be sufficiently small but not necessarily zero in order to provide more freedom in selecting optimal filter pairs. A possible strategy for this is achieved by selecting:

$$\mathbf{f}_{1,\mathrm{MIP}} = \mathbf{v}_{21,N_*}, \qquad \mathbf{f}_{2,\mathrm{MIP}} = \mathbf{v}_{12,N_*}, \quad (30)$$

Now the precoding ensures that only a transmission via the smallest singular value causes interference. In case $N_R < N_T$, we can even select a precoding vector in the null space of the matrix and interference is avoided altogether, amounting to ZF beamforming [9]. Otherwise, the remaining interference after applying an optimal receiver vector will be small. In fact, one can view this scheme as the opposite of the MAP scheme (25), assuming the best rather than the worst case interference. The selection of the receiver filters $\mathbf{g}_{1,\mathrm{MIP}}$ and $\mathbf{g}_{2,\mathrm{MIP}}$ simply applies the solutions (21), just for different precoding vectors.

H. ITERATIVE SCHEME

Setting the harsh condition of requiring the interference to completely vanish, as with orthogonal filtering and precoding, may not result in the best performance; this is similar to the performance degradation of a single-user MIMO-ZF equalizer when compared to a MIMO-MMSE equalizer [62]. In order to derive an optimal solution we reformulate our previous approach in terms of MMSE. It is well known that maximizing SINR is equivalent to minimizing the MSE [36]–[39].

$$\begin{aligned} &\min_{\mathbf{f}_1, \mathbf{g}_1} \mathrm{E}\left[\left|\mathbf{g}_1^H \mathbf{y}_1 - s_1\right|^2\right], \quad \text{s.t. } \|\mathbf{f}_1\|_2^2 = 1, \\ &\quad \Rightarrow \min_{\mathbf{f}_1, \mathbf{g}_1} \left|\sqrt{\alpha P} \mathbf{g}_1^H \mathbf{H}_{11} \mathbf{f}_1 - 1\right|^2 + N_o \|\mathbf{g}_1\|_2^2 \\ &\qquad + (1-\alpha) P \left|\mathbf{g}_1^H \mathbf{H}_{12} \mathbf{f}_2\right|^2 + \lambda_1 \left(\|\mathbf{f}_1\|_2^2 - 1\right), \end{aligned} \quad (31)$$

$$\begin{aligned} &\min_{\mathbf{f}_2, \mathbf{g}_2} \mathrm{E}\left[\left|\mathbf{g}_2^H \mathbf{y}_2 - s_2\right|^2\right], \quad \text{s.t. } \|\mathbf{f}_2\|_2^2 = 1, \\ &\quad \Rightarrow \min_{\mathbf{f}_2, \mathbf{g}_2} \left|\sqrt{(1-\alpha) P} \mathbf{g}_2^H \mathbf{H}_{22} \mathbf{f}_2 - 1\right|^2 + N_o \|\mathbf{g}_2\|_2^2 \\ &\qquad + \alpha P \left|\mathbf{g}_2^H \mathbf{H}_{21} \mathbf{f}_1\right|^2 + \lambda_2 \left(\|\mathbf{f}_2\|_2^2 - 1\right). \end{aligned} \quad (32)$$

Here, we follow the approach in [63] and consider the explicit unit power precoder constraint in the Lagrangians (31), (32) using the Lagrange multipliers λ_1, λ_2. An explicit solution for this problem is known only for a single user system [63]

but not for the two-user coupled system under consideration. We can, however, conclude the following result:

Theorem 2.2: The constrained per-user minimization problem (31), (32) results in the following maximization problem containing only the receiver filters:

$$\mathbf{g}_{1,\mathrm{con}} = \max_{\mathbf{g}_1} \frac{\mathbf{g}_1^H \mathbf{H}_{11}\mathbf{H}_{11}^H \mathbf{g}_1}{\mathbf{g}_1^H \left[N_o \mathbf{I} + (1-\alpha) P \mathbf{H}_{12} \frac{\mathbf{H}_{22}^H \mathbf{g}_2 \mathbf{g}_2^H \mathbf{H}_{22}}{\mathbf{g}_2^H \mathbf{H}_{22}\mathbf{H}_{22}^H \mathbf{g}_2} \mathbf{H}_{12}^H \right] \mathbf{g}_1},$$

$$\mathbf{g}_{2,\mathrm{con}} = \max_{\mathbf{g}_2} \frac{\mathbf{g}_2^H \mathbf{H}_{22}\mathbf{H}_{22}^H \mathbf{g}_2}{\mathbf{g}_2^H \left[N_o \mathbf{I} + \alpha P \mathbf{H}_{21} \frac{\mathbf{H}_{11}^H \mathbf{g}_1 \mathbf{g}_1^H \mathbf{H}_{11}}{\mathbf{g}_1^H \mathbf{H}_{11}\mathbf{H}_{11}^H \mathbf{g}_1} \mathbf{H}_{21}^H \right] \mathbf{g}_2}. \quad (33)$$

Once the receive filters are found, the precoding vectors are given by $\mathbf{f}_1 = \mathbf{H}_{11}^H \mathbf{g}_1 / \|\mathbf{H}_{11}^H \mathbf{g}_1\|_2$ *and* $\mathbf{f}_2 = \mathbf{H}_{22}^H \mathbf{g}_2 / \|\mathbf{H}_{22}^H \mathbf{g}_2\|_2$.

Proof: The details including an iterative algorithm are provided in Appendix A.

A closed-form solution to this coupled optimization problem is not available; however, fixing the receive filter of one user, the optimal receiver of the other user can be found in closed-form as it is obtained from a generalized Rayleigh quotient. Hence, we solve the problem by iterating between the two users. Similar iterative approaches have been considered in literature, e.g., in the context of weighted sum-rate maximization for the MIMO broadcast channel [64], [65]. Notice that the precoders correspond to MF vectors; similar to Section II-A, the obtained solution hence only achieves an uncoordinated Nash-equilibrium and in general not a Pareto-optimal beamforming solution. This is because each transmitter obtains the precoder by maximizing only the SINR of its own user, not accounting for the interference caused to the other user(s). Similar to [60], we could consider additional upper bounds on the interference leakage caused to the other user in Equations (31) and (32). However, we do not apply this method here as we observed that it substantially increases the complexity and reduces the convergence speed of the iterative algorithm, which in its original form (45)-(48) converges very rapidly.

III. PREDICTED ERGODIC RATES

As interesting and inspiring the algorithmic designs may be, what counts in the end is their performance. Here, SINRs may be of interest as we intend to maximize them, but even more of importance are achievable transmission rates as they relate to the expected throughput. To compare the various algorithms, we thus compute approximate ergodic means of the SINR and the achievable rate, and set them into relation to their lower and upper bounds. We evaluate their precision on MC runs, and compare the performance of the different algorithms. We furthermore investigate particular solutions of the power distribution α for which we have equal SINR ($\alpha = \alpha_=$), that is, fair transmission, and for which we achieve maximum sum rate ($\alpha = \alpha_{\mathrm{maxrate}}$).

A. PITFALLS

Once nonlinear mappings are included, ergodic mean values can be hard to obtain. For example, calculating the ergodic $\overline{\mathrm{SINR}} = \mathbb{E}\{\mathrm{SINR}\}$ requires to compute $\mathbb{E}\{1/x\}$ terms in which x contains the interference power. As this mapping is convex, replacing the term by the much simpler term $1/\mathbb{E}\{x\}$ results in too large values according to Jensen's inequality. Fortunately, such terms disappear for algorithms that favor zero interference; even more, if the interference power PI_l at user l is below the noise power $PI_l \ll N_o$, as is the case for most of the considered algorithms, accurate predictions are still obtained as shown below. Similarly, the ergodic rate $\overline{R} = \mathbb{E}\{R\}$ maps the SINR by $\log(1+x)$, which is a concave function causing a systematic error in the other direction when approximating $\mathbb{E}\{\log(1+x)\} \approx \log(1+\mathbb{E}\{x\})$. Both operations partially compensate each other but we should not be surprised if for some algorithms SINR prediction works perfectly, but capacity shows a remaining error when such approximations are applied. Given the pdf, we can compute how large our prediction error becomes. Let us consider two extreme cases for signal. In the first case the signal is deterministic and we obtain the classic Shannon formula. If the amplitudes of the signal, however, are random variables, we obtain a different result. Let us consider an other extreme with signal amplitudes described by a random variable x with pdf $f_X(x) = \exp(-x);\, x \geq 0$. We obtain then

$$\mathbb{E}_X \left\{ \log_2 \left(1 + \frac{S}{N_o} x \right) \right\} = \log_2(e) Q\left(\frac{N_o}{S} \right) \exp\left(\frac{N_o}{S} \right)$$

with $Q(x)$ denoting the Gaussian complementary error function. This function lies roughly 2.5dB below $\log_2(1+\frac{S}{N_o})$. In such a case we thus predict too high values. This is the typical source of prediction error in most of the algorithms that successfully suppress the interference. With increasing number of antennas the error becomes smaller and asymptotically vanishes.

A second source of error is the noise in case the algorithm is not suppressing the interference terms. In such interference dominant situations we take the interference power and consider it to be Gaussian, underestimating the rates. Following the same argument as before, we can maximally be 2dB off the true rate. The consequence of the investigation is that our estimations are rather precise in most cases, and in case we over- or underestimate, we can provide simple lower and upper bounds quantifying maximal deviations.

The performance of the considered two user transmission systems can be plotted in parametric form for $0 \leq \alpha \leq 1$ in either SINR or transmission rate $\{R_1(\alpha), R_2(\alpha)\}$ as depicted in Figure 2. The two corner points in Figure 2 correspond to the SINR-pairs $\{\frac{PS_1}{N_o}, 0\}$ and $\{0, \frac{PS_2}{N_o}\}$, or equivalently to the rate-pairs $\{\log_2(1+\frac{PS_1}{N_o}), 0\}$ and $\{0, \log_2(1+\frac{PS_2}{N_o})\}$, with PS_1, PS_2 denoting the maximal signal power of user 1 and 2, respectively. The parametric curve $\{R_1(\alpha), R_2(\alpha)\} = \{\alpha \log_2(1+\frac{PS_1}{N_o}), (1-\alpha)\log_2$

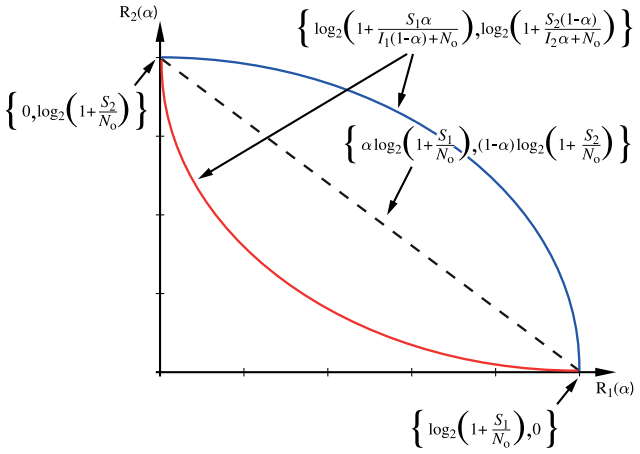


FIGURE 2. **Two user rate region for low interference (continuous blue line) and high interference (continuous red line).**

$\left(1+\frac{PS_2}{N_o}\right)\}$, represented by the dashed line in Figure 2, demonstrates the performance achieved with time-sharing of the channel between the two users. The rates obtained with simultaneous transmission, i.e., including multiuser interference, pass along the lines $\left\{\log_2\left(1+\frac{S_1\alpha P}{I_1(1-\alpha)P+N_o}\right), \log_2\left(1+\frac{S_2(1-\alpha)P}{I_2\alpha P+N_o}\right)\right\}$. For sufficiently small interference power, the curves are concave, (continuous blue line in Figure 2) but for high interference power, they can also be convex (continuous red line in Figure 2). The following lemma provides a simple test which of the two cases is present.

Lemma 3.1: Given a parametric transmission rate description that connects the two corner points $\left\{\log_2\left(1+\frac{PS_1}{N_o}\right), 0\right\}$ *and* $\left\{0, \log_2\left(1+\frac{PS_2}{N_o}\right)\right\}$ *along the curve parametrized by the power distribution factor* α $\left\{\log_2\left(1+\frac{PS_1\alpha}{I_1(1-\alpha)+N_o}\right), \log_2\left(1+\frac{PS_2(1-\alpha)}{I_2\alpha+N_o}\right)\right\}$, *with* PS_1 *and* PS_2 *denoting the maximal desired signal power of the first and second user, respectively, and* PI_1 *and* PI_2 *being the corresponding maximal interference powers, then necessary (but not sufficient) conditions to have a convex curve are:*

$$\gamma_{\mathrm{LB}} < \gamma < \gamma_{\mathrm{UB}}, \quad \gamma = -\frac{\ln(1+PS_2/N_o)}{\ln(1+PS_1/N_o)},$$
$$\gamma_{\mathrm{LB}} = -\frac{S_2}{S_1}\frac{(PI_1+N_o)(PI_2+N_o)}{(PS_2+N_o)},$$
$$\gamma_{\mathrm{UB}} = -\frac{S_2}{S_1}\frac{(PS_1+N_o)N_o}{(PI_1+N_o)(PI_2+N_o)}. \quad (34)$$

The proof of this result is provided in Appendix B. In particular for zero interference schemes (i.e., $I_1 = I_2 = 0$) and/or low receiver noise the conditions are easy to check.

We are particularly interested in fair transmissions ($\alpha = \alpha_=$), that is when both partners obtain the same SINR and in the case where the sum rate is at maximum ($\alpha = \alpha_{\mathrm{maxrate}}$) as we could transmit most bits per resource. Computing such α values relates to either two SINR terms $\{\frac{S_1\alpha P}{I_1(1-\alpha)P+N_o}, \frac{S_2(1-\alpha)P}{I_2\alpha P+N_o}\}$ or their corresponding achievable rates. Finding the desired values of α requires solving a polynomial of second order. However, for most algorithms the quadratic term is either exactly zero (orthogonal schemes) or sufficiently close to zero, so that we only compute the explicit solution of the linear problem in α.

Lemma 3.1: Given a parametric interference model and the corresponding transmission rate description $\left\{\log_2\left(1+\frac{\alpha PS_1}{N_o+(1-\alpha)PI_1}\right), \log_2\left(1+\frac{(1-\alpha)PS_2}{N_o+\alpha PI_2}\right)\right\}$ *the power distribution for equal SINR is given by*

$$\alpha_= \approx \begin{cases} \frac{1/\mathrm{SNR}_1}{1/\mathrm{SNR}_1+1/\mathrm{SNR}_2}, & I_1, I_2 \ll N_o \\ \frac{\sqrt{1/\mathrm{SIR}_1}}{\sqrt{1/\mathrm{SIR}_1}+\sqrt{1/\mathrm{SIR}_2}} & I_1, I_2 \gg N_o, \end{cases}$$
$$\mathrm{SNR}_1 = \frac{PS_1}{N_o}, \ \mathrm{SNR}_2 = \frac{PS_2}{N_o}, \ \mathrm{SIR}_1 = \frac{S_1}{I_1}, \ \mathrm{SIR}_2 = \frac{S_2}{I_2}. \quad (35)$$

In case of a concave behavior, the maximum sum rate solution is found at:

$$\alpha_{\mathrm{maxrate}} = \frac{1}{2} + \frac{N_o}{2P}\frac{S_1 - S_2}{S_1 S_2}, \quad (36)$$

while in a convex curve it turns out to be $\alpha_{\mathrm{maxrate}} = \arg\max_\alpha\{\alpha PS_1, (1-\alpha)PS_2\}$ *which is simply one of the corners.*

Proof: The details are provided in Appendix C.

Note that lemma 3.1 simply offers a harmonic mean solution providing a balancing between the different channel paths, while lemma 3.1 claims that identical power is the best as long as the SNR is sufficiently large. In these latter scenarios, no feedback about the power distribution needs to be provided. Our MC results show very good agreement with such predictions (see Sec.III-C) further ahead.

B. SOME MATHS TERMS

As we will encounter many complex terms in the following, it simplifies the derivations a lot by first investigating typical expressions. We denote n-th order moments of a random variable σ by $\sigma^{(n)}$ and arithmetic means over an ensemble of n random variables σ_i by $\overline{\sigma}$, correspondingly $\overline{\sigma^2}$ for mean squared values. Take, for example two statistically independent and isotropically distributed vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$, e.g., obtained by normalizing statistically independent vectors $\tilde{\mathbf{u}}, \tilde{\mathbf{v}} \in \mathbb{C}^N$ with i.i.d. Gaussian entries, i.e., $\mathbf{u} = \tilde{\mathbf{u}}/\left\|\tilde{\mathbf{u}}\right\|_2$, $\mathbf{v} = \tilde{\mathbf{v}}/\left\|\tilde{\mathbf{v}}\right\|_2$. Then their inner product energy is $\mathbb{E}\{|\mathbf{u}^H\mathbf{v}|^2\} = 1/N$. If a random matrix $\mathbf{H} \in \mathbb{C}^{N\times N}$ for which $\mathbf{v}_i; i = 1, 2, ..., N$ is a right singular vector, then $\mathbb{E}\{|\mathbf{u}^H\mathbf{H}\mathbf{v}_i|^2\} = \sigma_i^{(2)}/N$. Finally, if a random matrix $\mathbf{H} \in \mathbb{C}^{N\times N}$ is not related to the two vectors, we find $\mathbb{E}\{|\mathbf{u}^H\mathbf{H}\mathbf{v}|^2\} = 1/N^2\sum_{i=1}^N \sigma_i^{(2)} = \overline{\sigma^2}/N$. Likewise we obtain $\mathbb{E}\{\mathbf{v}^H\mathbf{H}^H\mathbf{H}\mathbf{v}\} = \overline{\sigma^2}$.

C. COMPARISON OF VARIOUS SETUPS

We will demonstrate the accuracy of our derivations by selected numerical examples in the following. For both cases, the single eNodeB and the double eNodeB setup, we use

$N_R \times N_T = 4 \times 4$ MIMO channels, modeled as uncorrelated i.i.d. complex Gaussian with Frobenius norm four for the first user and one for the second user, thus 6dB difference between both. For the single eNodeB transmissions we select $\mathrm{tr}\left(\mathbb{E}\{\mathbf{H}_{11}^H\mathbf{H}_{11}\}\right) = 4$ and $\mathrm{tr}\left(\mathbb{E}\{\mathbf{H}_{21}^H\mathbf{H}_{21}\}\right) = 1$, while in the double eNodeB setup we additionally employ $\mathrm{tr}\left(\mathbb{E}\{\mathbf{H}_{12}^H\mathbf{H}_{12}\}\right) = 4$ and $\mathrm{tr}\left(\mathbb{E}\{\mathbf{H}_{22}^H\mathbf{H}_{22}\}\right) = 1$. We will focus on results for both scenarios only if they differ. In both scenarios we set the noise power $N_o = 0.01$.[6]

In the following figures we depict mostly ergodic achievable rates of both users with $0 \le \alpha \le 1$ as parameter. The dashed lines depict the results of 4 000 MC runs and are compared to the continuous lines of our analytic predictions. The predicted power distributions are marked by '$\circ$' for fair transmissions $\alpha = \alpha_=$ and '$\times$' for maximal sum rate $\alpha = \alpha_{\text{sumrate}}$.

Below we present the theoretical findings, predicting the algorithmic behavior of all presented algorithms of Section II. As we are interested in the algorithms general behavior we compute ergodic rate results based on Eqs. (8), (10) or in case we have MMSE based algorithms (22), (23). As an example we show the procedure for the Interference Ignorant MET Algorithm (II-MET). In (8) we need to compute $|\mathbf{g}_1^H\mathbf{H}_{11}\mathbf{f}_1|^2 = \sigma_{11,1}^2$ and $|\mathbf{g}_1^H\mathbf{H}_{12}\mathbf{f}_2|^2 = |\mathbf{u}_{11,1}^H\mathbf{H}_{12}\mathbf{v}_{22,1}|^2$. Taking expectations, the first term reads $E[\sigma_{11,1}^2] = \sigma_{11,1}^{(2)}$ while the second becomes $E[|\mathbf{u}_{11,1}^H\mathbf{H}_{12}\mathbf{v}_{22,1}|^2] = \overline{\sigma_{12}^2}/N_R$ resulting in:

$$\overline{\mathrm{SINR}}_{1,\mathrm{II}} = \frac{\sigma_{11,1}^{(2)}\alpha P}{\overline{\sigma_{12}^2}(1-\alpha)P/N_R + N_o},$$
$$\overline{\mathrm{SINR}}_{2,\mathrm{II}} = \frac{\sigma_{22,1}^{(2)}(1-\alpha)P}{\overline{\sigma_{21}^2}\alpha P/N_R + N_o}. \tag{37}$$

for the double eNodeB scenario. For the single eNodeB scenario we cannot simply replace $\overline{\sigma_{12}^2}$ by $\sigma_{11,1}^{(2)}$ and $\overline{\sigma_{21}^2}$ by $\sigma_{21,1}^{(2)}$ but instead we obtain now:

$$\overline{\mathrm{SINR}}_{1,\mathrm{II}} = \frac{\sigma_{11,1}^{(2)}\alpha P}{\sigma_{11,1}^{(2)}(1-\alpha)P/N_R + N_o},$$
$$\overline{\mathrm{SINR}}_{2,\mathrm{II}} = \frac{\sigma_{21,1}^{(2)}(1-\alpha)P}{\sigma_{21,1}^{(2)}\alpha P/N_R + N_o}. \tag{38}$$

The power distribution for fair transmission in the noise dominant case is obtained as:

$$\alpha_= = \frac{1/\sigma_{11,1}^{(2)}}{1/\sigma_{11,1}^{(2)} + 1/\sigma_{22,1}^{(2)}}, \tag{39}$$

[6]Note that we performed a lot more examples to ensure our assumptions are valid in a wide range of operations. The limited size of such an article does not allow to demonstrate more but the reader is invited to download the simulation code under http://www.nt.tuwien.ac.at and evaluate more examples.

while the maximal sum rate is found at:

$$\alpha_{\text{maxrate}} = \frac{1}{2} + \frac{N_o}{2P}\frac{\sigma_{11,1}^{(2)} - \sigma_{22,1}^{(2)}}{\sigma_{11,1}^{(2)}\sigma_{22,1}^{(2)}}. \tag{40}$$

Both, Equations (39) and (40) are provided for the double eNodeB setup; in case of a single eNodeB, $\sigma_{22,1}^{(2)}$ needs to be replaced by $\sigma_{21,1}^{(2)}$. In the following Table 2 we list the ergodic SINR terms of the various algorithms for $N_R = N_T$, out of which we derive the achievable rates. As the single eNodeB case is obtained when replacing $\mathbf{H}_{22} = \mathbf{H}_{12}$ and $\mathbf{H}_{21} = \mathbf{H}_{11}$ in the double eNodeB scenario, in some cases the ensemble averages are obtained by replacing the corresponding moments. If this is the case we did not explicitly provide the result; only in case they differ, we list them in the table. Most expressions are relatively straightforward to obtain as the corresponding vectors are –at least from one side– either eigenvectors to the matrices or simply treated as isotropic random vectors. Only for the leakage algorithm, the corresponding expressions were too difficult to analytically compute and simply obtained by comparing with terms when random vectors for precoding are applied, explaining the factors $\frac{6}{5}$, $\frac{3}{5}$. We used here abbreviations for $\overline{\sigma_{ab}^{(2)}}/N_R = \mathbb{E}\{\mathbf{u}^H\mathbf{H}_{11}^H\mathbf{H}_{12}\mathbf{v}\}$ and $\overline{\sigma_{cd}^{(2)}}/N_R = \mathbb{E}\{\mathbf{u}^H\mathbf{H}_{22}^H\mathbf{H}_{21}\mathbf{v}\}$ with $\mathbf{u}$ and $\mathbf{v}$ being isotropic randomly selected normalized vectors. In the particular case of SLNR for a single eNodeB, the penalty terms are proportional to N_o^2 and thus so small that we simply neglected them.

In the following plots we depict in red lines the ergodic results obtained by MC runs. The blue lines show our predicted curves. In green we depict the upper bounds from Lemma 2.1 to provide a common reference line.

Power distribution: In all but the interference ignorant (II) MET algorithm, the interference becomes sufficiently reduced so that the simple noise dominant scenario is obtained and we find the equal mode of transmission for $\alpha_= = 0.2$. Only in the II-MET algorithm the interference is dominant and here our theory predicts $\alpha_= = 0.5$ which is well achieved in this case. All power distributions for maximum rate are achieved at $\alpha_{\text{maxrate}} = 0.5$ which are in excellent agreement with the theory.

Prediction accuracy: Our predictions are fairly accurate as the following Figures 3 and 4 demonstrate. In most cases we have a noise dominant situation as most, if not all, of the interference has been removed. However the signal amplitudes are not exactly Gaussian distributed and thus our estimates are a bit too high (upper bounds). Our predictions are less accurate only in the II algorithm, which is not surprising as it is an interference dominant scenario. As the interference is dominant and not Gaussian distributed our prediction must be too low in this case. If we lift the curve by 2.5dB as indicated in Section III-A, we recognize that we indeed have obtained now an upper bound. Note also that even for the less accurate prediction of the II algorithm, the particular points for fair transmission mode ('$\circ$') and for maximum sum rate ('$\times$') are always correctly estimated.

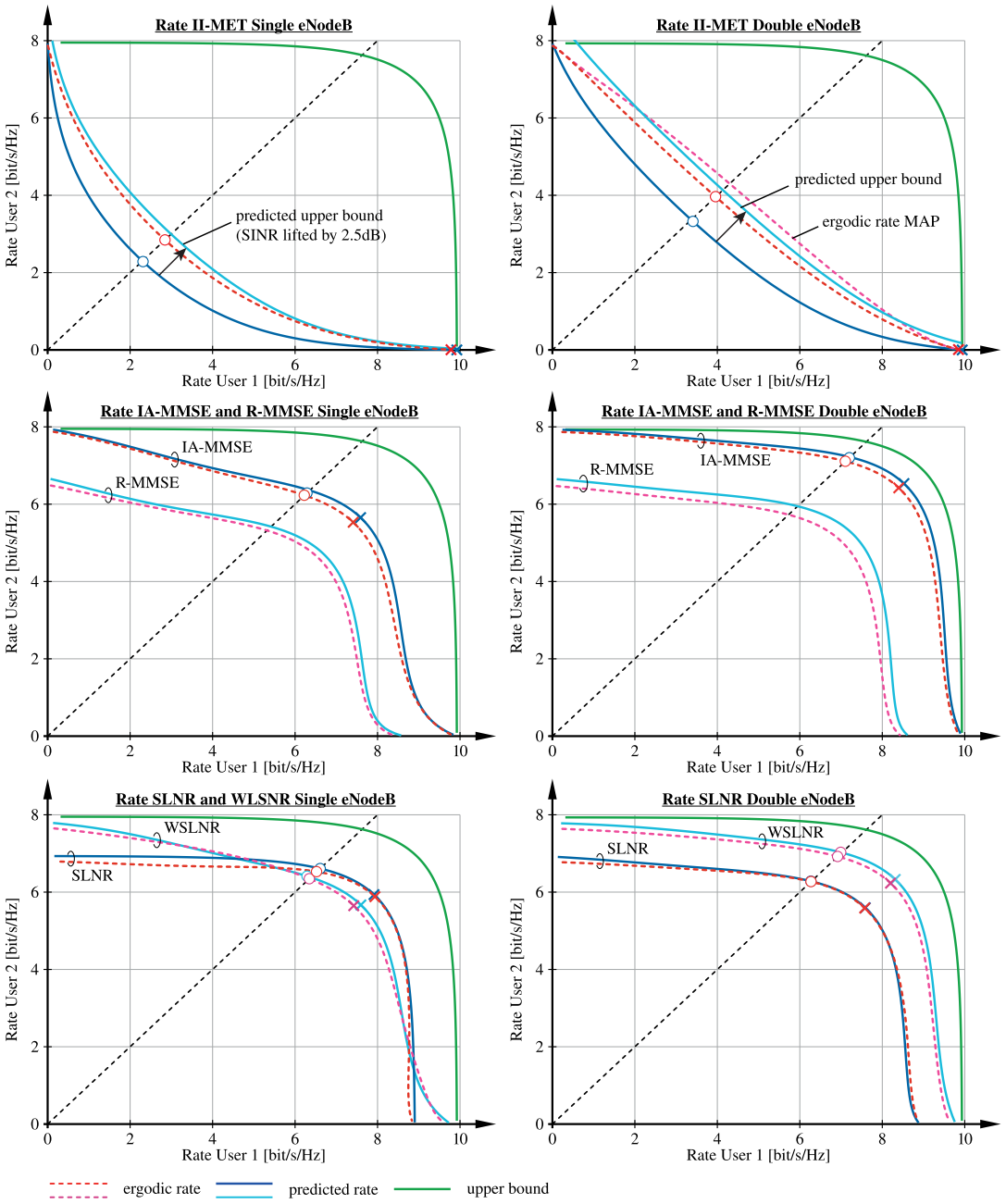


FIGURE 3. **Achievable rates for the single (left) and double (right) eNodeB scenario. top row: II & MAP, middle row: IA-MMSE, bottom row: LSNR-MMSE.**

Asymptotic behavior: All algorithms in Table 2 show that their penalty term goes with $1/N_R$, except for orthogonal precoding where it goes with $1/N_T$. Thus, for all these algorithms the penalty term vanishes with increasing number of receive antennas, respectively transmit antennas for orthogonal precoding. This is particularly interesting for advocating massive MIMO transmission for 5G cellular networks. With such massive antenna numbers even simple low complex algorithms will perform superbly. MC runs with larger values of N_R nicely support this. In fact not only the penalty term becomes smaller and the curves approach the upper bounds but also the accuracy increases as with decreasing interference only Gaussian noise remains. These conclusions, however, hold only true if the number of transmitter-receiver pairs stays constant and does not grow linearly with the number of antennas.

Further observations: In the top-right plot of Fig. 3 we compare the performance of the II-MET algorithm with the results obtained with the MAP algorithm. Remember that II-MET selects the precoder and receive filter to achieve maximal channel gain, without accounting for the interference. MAP, on the other hand, applies MET precoding

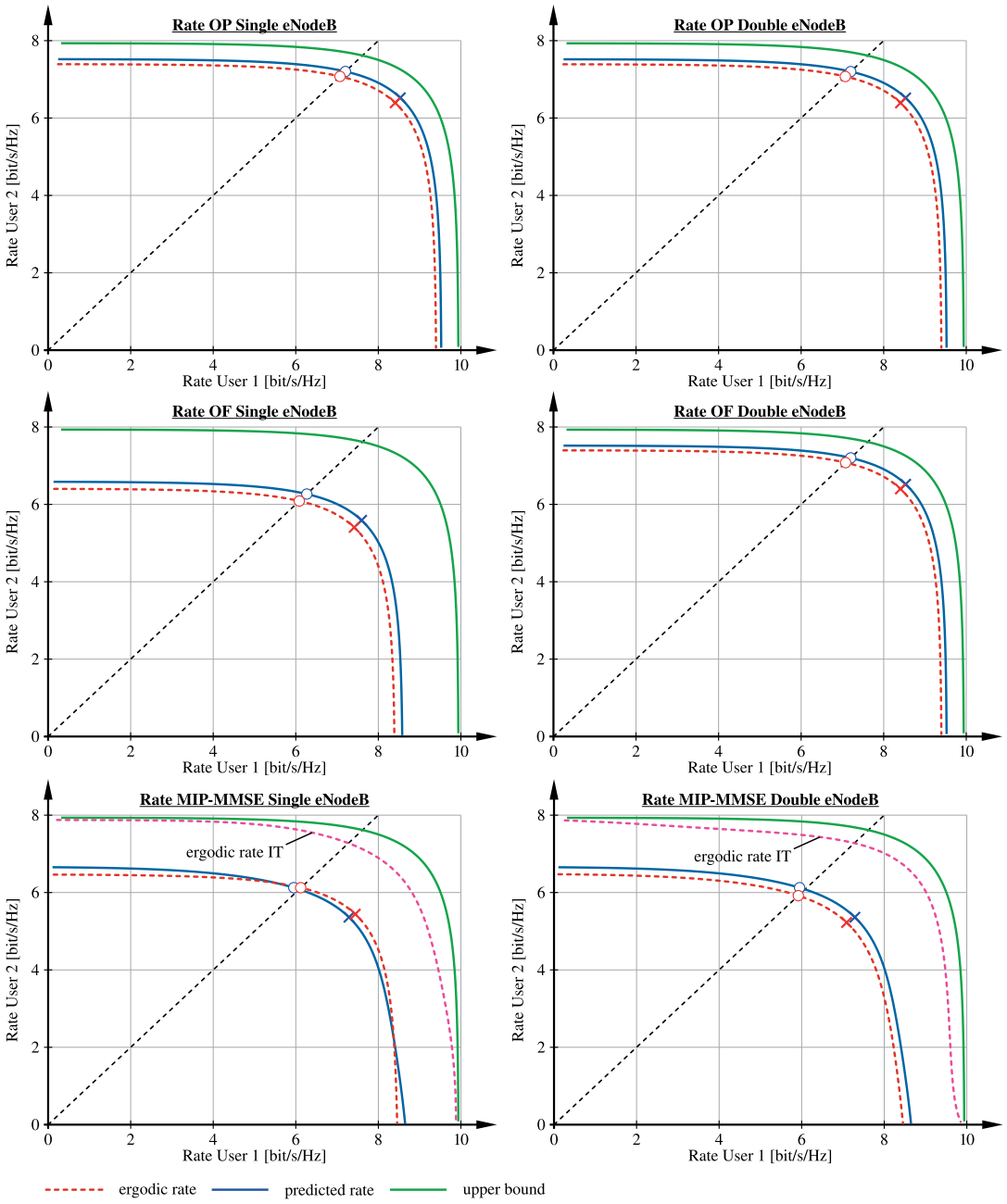


FIGURE 4. Achievable rates for the single (left) and double (right) eNodeB scenario. Top row: OP, middle row: OF, bottom row: MIP-MMSE & IT-MMSE.

at the transmitter, while the receive filters are selected to provide robustness against worst case interference. Although not optimal, the MAP algorithm shows better performance than II-MET. The IA-MMSE algorithm, which combines MET beamforming with interference-aware MMSE filtering, substantially outperforms R-MMSE with randomly selected precoders, due to the improved channel gain provided by MET beamforming. The SLNR algorithm outperforms the simple R-MMSE algorithm only insubstantially, especially in the double eNodeB scenario. This is because the receivers are equipped with enough antennas to be able to cancel the interference; hence, the interference-aware MMSE filter recovers some of the loss caused by the worse random precoder. The WSLNR algorithm shows even better performance then SLNR beamforming in the double eNodeB scenario. We believe that this is because SLNR beamforming puts too much weight on the interference, which is not really necessary as the MMSE receiver can tolerate more interference. If we do not apply the MMSE receiver, but receive over the maximum eigenmode of the channel,

TABLE 2. Overview of the predicted ergodic SINR values for $N_R = N_T$.

Name	$\overline{\text{SINR}_1}$	$\overline{\text{SINR}_2}$
II-MET (1BS)	$\frac{\alpha P\sigma_{11,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\sigma_{11,1}^{(2)}(1-\alpha)P}{N_o+\sigma_{11,1}^{(2)}(1-\alpha)P/N_R}\right]$	$\frac{(1-\alpha)P\sigma_{21,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\sigma_{21,1}^{(2)}\alpha P}{N_o+\sigma_{21,1}^{(2)}\alpha P/N_R}\right]$
II-MET (2BS)	$\frac{\alpha P\sigma_{11,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\overline{\sigma_{12}^2}(1-\alpha)P}{N_o+\overline{\sigma_{12}^2}(1-\alpha)P/N_R}\right]$	$\frac{(1-\alpha)P\sigma_{22,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\overline{\sigma_{21}^2}\alpha P}{N_o+\overline{\sigma_{21}^2}\alpha P/N_R}\right]$
OF (1BS)	$\frac{\alpha P\sigma_{11,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\sigma_{11,1}^{(2)}}{\overline{\sigma_{11}^2}}\right]$	$\frac{(1-\alpha)P\sigma_{21,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\sigma_{21,1}^{(2)}}{\overline{\sigma_{21}^2}}\right]$
OF (2BS)	$\frac{\alpha P\sigma_{11,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\right]$	$\frac{(1-\alpha)P\sigma_{22,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\right]$
OP	$\frac{\alpha P\sigma_{11,1}^{(2)}}{N_o}\left[1-\frac{1}{N_T}\right]$	$\frac{(1-\alpha)P\sigma_{22,1}^{(2)}}{N_o}\left[1-\frac{1}{N_T}\right]$
IA-MMSE (1BS)	$\frac{\alpha P\sigma_{11,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{(1-\alpha)P\sigma_{11,1}^{(2)}}{No+(1-\alpha)P\overline{\sigma_{11}^2}}\right]$	$\frac{(1-\alpha)P\sigma_{21,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\alpha P\sigma_{21,1}^{(2)}}{N_o+\alpha P\overline{\sigma_{21}^2}}\right]$
IA-MMSE (2BS)	$\frac{\alpha P\sigma_{11,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{(1-\alpha)P\overline{\sigma_{12}^2}}{N_o+(1-\alpha)P\overline{\sigma_{12}^2}}\right]$	$\frac{(1-\alpha)P\sigma_{22,1}^{(2)}}{N_o}\left[1-\frac{1}{N_R}\frac{\alpha P\overline{\sigma_{21}^2}}{N_o+\alpha P\overline{\sigma_{21}^2}}\right]$
MIP	$\frac{\alpha P\overline{\sigma_{11}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{(1-\alpha)P\sigma_{12,N_*}^{(2)}}{N_o+(1-\alpha)P\sigma_{12,N_*}^{(2)}}\right]$	$\frac{(1-\alpha)P\overline{\sigma_{22}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\alpha P\sigma_{21,N_*}^{(2)}}{N_o+\alpha P\sigma_{21,N_*}^{(2)}}\right]$
MAP (2BS)	$\frac{\alpha P\overline{\sigma_{11}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{(1-\alpha)P\sigma_{12,1}^{(2)}}{N_o+(1-\alpha)P\sigma_{12,1}^{(2)}}\right]$	$\frac{(1-\alpha)P\overline{\sigma_{22}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\alpha P\sigma_{21,1}^{(2)}}{N_o+\alpha P\sigma_{21,1}^{(2)}}\right]$
SLNR (1BS)	$\frac{\frac{6}{5}\alpha P\overline{\sigma_{11}^2}}{N_o}$	$\frac{\frac{6}{5}(1-\alpha)P\overline{\sigma_{21}^2}}{N_o}$
SLNR (2BS)	$\frac{\frac{6}{5}\alpha P\overline{\sigma_{11}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\frac{5}{6}(1-\alpha)P\overline{\sigma_{ab}^2}/\overline{\sigma_{11}^2}}{N_o+(1-\alpha)P\overline{\sigma_{12}^2}}\right]$	$\frac{\frac{6}{5}(1-\alpha)P\overline{\sigma_{22}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\frac{5}{6}\alpha P\overline{\sigma_{cd}^2}/\overline{\sigma_{22}^2}}{N_o+\alpha P\overline{\sigma_{21}^2}}\right]$
WSLNR (1BS)	$\frac{\frac{13}{5}\alpha P\overline{\sigma_{11}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\frac{15}{26}(1-\alpha)P\sigma_{11}^4/\overline{\sigma_{11}^2}}{N_o+\frac{3}{5}(1-\alpha)P\overline{\sigma_{11}^2}}\right]$	$\frac{\frac{13}{5}(1-\alpha)P\overline{\sigma_{21}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\frac{15}{26}\alpha P\overline{\sigma_{21}^4}/\overline{\sigma_{21}^2}}{N_o+\frac{3}{5}\alpha P\overline{\sigma_{21}^2}}\right]$
WSLNR (2BS)	$\frac{\frac{13}{5}\alpha P\overline{\sigma_{11}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\frac{15}{26}(1-\alpha)P\overline{\sigma_{ab}^2}/\overline{\sigma_{11}^2}}{N_o+\frac{3}{5}(1-\alpha)P\overline{\sigma_{12}^2}}\right]$	$\frac{\frac{13}{5}(1-\alpha)P\overline{\sigma_{22}^2}}{N_o}\left[1-\frac{1}{N_R}\frac{\frac{15}{26}\alpha P\overline{\sigma_{cd}^2}/\overline{\sigma_{22}^2}}{N_o+\frac{3}{5}\alpha P\overline{\sigma_{21}^2}}\right]$

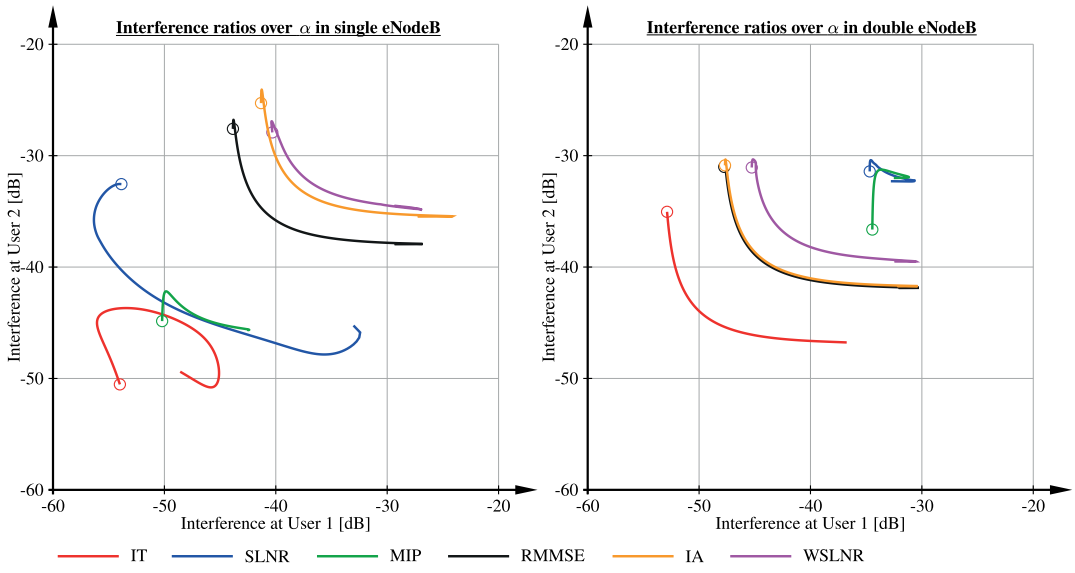


FIGURE 5. Remaining interference power for the single eNodeB (left) and the double eNodeB (right) setup at $N_o = -20$dBW.

we can still obtain very good performance by orthogonal precoding, as shown in the top plots of Fig. 4. Orthogonal precoding keeps the maximum eigenmodes of the users free of interference an thus enables both, high intended signal power and zero interference. Although they both appear very similar, the orthogonal filtering algorithm performs significantly worse than orthogonal precoding in the single eNodeB scenario. This is because in this scenario both, signal and interference, are received over the same channel matrix. Thus, not only the intended signal experiences the maximum eigenmode of the channel matrix, but in general also the interference and thus the receiver has to reject part of the intended signal energy to get rid of the interference. Orthogonal filtering should also be compared to the

IEEE Access
The journal for rapid open access publishing

TABLE 3. Maximum sum rate, left: single eNodeB setup, right: double eNodeB; in the left columns are the experimental results (MC), in the right the predictions (P) for $N_T = N_R = 4$ antenna MIMO.

	Single eNodeB			Double eNodeB		
	$\overline{C}^{(MC)}_{\text{maxrate}}$	$\overline{C}^{(P)}_{\text{maxrate}}$	rel. error[%]	$\overline{C}^{(MC)}_{\text{maxrate}}$	$\overline{C}^{(P)}_{\text{maxrate}}$	rel. error[%]
II-MET	9.80	9.94	-1.4	9.86	9.94	-0.8
OF	12.82	13.18	-2.8	14.80	15.05	-1.7
OP	14.80	15.05	-1.7	14.80	15.05	-1.7
MIP	12.88	12.65	1.8	12.31	12.65	-2.8
IA-MMSE	12.91	13.24	-2.6	14.83	15.06	-1.6
R-MMSE	11.17	11.40	-2.0	12.04	12.51	-3.9
SLNR	13.82	13.85	-0.2	13.16	13.18	-0.1
WSLNR	12.99	12.60	3.0	14.38	14.46	-0.6
IT	14.96	15.88	-6.2	15.19	15.88	-4.6

IA-MMSE approach. Both methods employ the same MET precoders; however, the MMSE receiver clearly outperforms the orthogonal receive filter. The iterative algorithm, shown in the bottom plots of Fig. 4, was run for five iterations, keeping it comparable in complexity to the other algorithms. It did not substantially improve when running longer. Nevertheless it outperforms all other algorithms in all considered scenarios.

An open question is the importance of forcing the interference terms to zero. The OP and OF algorithms have made this a design constraint. But as indicated by the results in [60], this constraint is mostly too harsh and a looser leakage constraint allows for higher achievable rate. In Figure 5 we depict an (I_1, I_2) plot of the remaining interference terms after the receive filter when running the power distribution α from zero to one. The circles denote the $\alpha = 0$ point. The iterative algorithm obviously performs best by reducing the remaining interference terms to a minimum, avoiding however to reduce them to zero. The MIP algorithm behaves excellent if the remaining resultant error is small (single eNodeB) but poor if not (double eNodeB). The weighted leakage method and the IA algorithms show roughly the same behavior.

In Table 3 we list comparisons of predicted values versus observed results for the two setups, facilitating further the interpretation of the presented plots. The relative errors $1 - \overline{C}^{(P)}_{\text{maxrate}}/\overline{C}^{(MC)}_{\text{maxrate}}$ show that our predictions for the maximal achievable sum rate are very accurate. For the iterative algorithm we do not have a predicted value but instead compared to the theoretical upper bound, showing that we are not far off any more.

IV. CONCLUSIONS

We have provided a thorough overview about existing algorithms for transmit coordination and interference cancellation in the context of cellular wireless systems. Such algorithms serve as important enablers for 5G cellular, as they facilitate network densification by avoiding or even exploiting the interference between multiple transmission points. Here we studied broadcasting scenarios, with a single eNodeB transmitting to multiple users, and interference scenarios, where two eNodeBs compete for the same resources. Although most algorithms behave similar in terms of performance, especially with increasing number of antennas, our analysis shows that distinctly different results may be obtained in the broadcasting and the interference scenario. In particular the much favored algorithms based on orthogonal designs (such as interference alignment, zero-forcing beamforming and block-diagonalization precoding), suppressing interference entirely, are not necessarily the best performers. An iterative algorithm that solves an MMSE cost function with side constraints on the precoding vectors turned out to outperform all other methods in every investigated scenario. However, these conclusions depend also on the availability of sufficient receive antennas and interference channel state information at the users. In case user do not have the capability to reduce the impact of interference themselves, sophisticated transmit beamforming becomes more important.

APPENDIX

A. PROOF OF THEOREM 2.2

The proof relates to the derivation of an iterative method as such is known to converge rapidly to the desired solution [66]. Applying this technique, we find the following iterative algorithm for our setup:

$$\mathbf{g}_1 = \left[N_o\mathbf{I}_{N_R} + (1-\alpha)P\mathbf{H}_{12}\mathbf{f}_2\mathbf{f}_2^H\mathbf{H}_{12}^H\right]^{-1}\mathbf{H}_{11}\mathbf{f}_1, \quad (41)$$

$$\mathbf{g}_2 = \left[N_o\mathbf{I}_{N_R} + \alpha P\mathbf{H}_{21}\mathbf{f}_1\mathbf{f}_1^H\mathbf{H}_{21}^H\right]^{-1}\mathbf{H}_{22}\mathbf{f}_2, \quad (42)$$

$$\mathbf{f}_1 = \left[\lambda_1\mathbf{I}_{N_T} + (1-\alpha)P\mathbf{H}_{11}^H\mathbf{g}_1\mathbf{g}_1^H\mathbf{H}_{11}\right]^{-1}\mathbf{H}_{11}^H\mathbf{g}_1, \quad (43)$$

$$\mathbf{f}_2 = \left[\lambda_2\mathbf{I}_{N_T} + \alpha P\mathbf{H}_{22}^H\mathbf{g}_2\mathbf{g}_2^H\mathbf{H}_{22}\right]^{-1}\mathbf{H}_{22}^H\mathbf{g}_2. \quad (44)$$

The algorithms requires the application of the correct Lagrangians λ_1, λ_2 which can be difficult to obtain. Fortunately, Eqs. (43) and (44) simply state that $\mathbf{f}_1 \sim \mathbf{H}_{11}^H\mathbf{g}_1$ and

$\mathbf{f}_2 \sim \mathbf{H}_{22}^H \mathbf{g}_2$.[7] Since both vectors are also unitary, they are now defined and the iterative algorithm simplifies to:

$$\mathbf{g}_1 = \left[N_o \mathbf{I} + (1-\alpha) P \mathbf{H}_{12} \frac{\mathbf{H}_{22}^H \mathbf{g}_2 \mathbf{g}_2^H \mathbf{H}_{22}}{\mathbf{g}_2^H \mathbf{H}_{22} \mathbf{H}_{22}^H \mathbf{g}_2} \mathbf{H}_{12}^H \right]^{-1} \mathbf{H}_{11} \mathbf{f}_1, \quad (45)$$

$$\mathbf{g}_2 = \left[N_o \mathbf{I} + \alpha P \mathbf{H}_{21} \frac{\mathbf{H}_{11}^H \mathbf{g}_1 \mathbf{g}_1^H \mathbf{H}_{11}}{\mathbf{g}_1^H \mathbf{H}_{11} \mathbf{H}_{11}^H \mathbf{g}_1} \mathbf{H}_{21}^H \right]^{-1} \mathbf{H}_{22} \mathbf{f}_2, \quad (46)$$

$$\mathbf{f}_1 = \frac{\mathbf{H}_{11}^H \mathbf{g}_1}{\| \mathbf{H}_{11}^H \mathbf{g}_1 \|_2}, \quad (47)$$

$$\mathbf{f}_2 = \frac{\mathbf{H}_{22}^H \mathbf{g}_2}{\| \mathbf{H}_{22}^H \mathbf{g}_2 \|_2}. \quad (48)$$

Now only (45) and (46) need to be iterated, the last two lines are simply computed once at the end of the iteration. Practically, the algorithm converges in a few (say 10) iterations. □

B. PROOF OF LEMMA 3.1

We connect the two corner points $\left\{ \log_2 \left(1 + \frac{S_1 P}{N_o}\right), 0 \right\}$ and $\left\{ 0, \log_2 \left(1 + \frac{S_2 P}{N_o}\right) \right\}$ via the straight line parametrized by $\beta \in [0, 1]$, that is, $\left\{ \beta \log_2 \left(1 + \frac{S_1 P}{N_o}\right), (1-\beta) \log_2 \left(1 + \frac{S_2 P}{N_o}\right) \right\}$. The derivative along such line is given by the constant $\gamma = -\frac{\ln(1+S_2 P/N_o)}{\ln(1+S_1 P/N_o)}$. This slope is to compare with the derivative along the lines of $\left\{ \log_2 \left(1 + \frac{S_1 \alpha P}{I_1 (1-\alpha) P + N_o}\right), \log_2 \left(1 + \frac{S_2 (1-\alpha) P}{I_2 \alpha P + N_o}\right) \right\}$. The corner point on the x-axis describes the case that $\alpha = 1$; here the derivative needs to be below γ to obtain a convex curve. At the corner point on the y-axis we have $\alpha = 0$; here the derivative needs to be above γ. Evaluating the derivative at these two corner points we find two conditions on γ:

$$\gamma < -\frac{S_2}{S_1} \frac{(S_1 P + N_o) N_o}{(I_1 P + N_o)(I_2 P + N_o)}, \quad (49)$$

$$\gamma > -\frac{S_2}{S_1} \frac{(I_1 P + N_o)(I_2 P + N_o)}{(S_2 P + N_o)}. \quad (50)$$

C. DERIVATION OF POWER DISTRIBUTION FACTORS

Setting $\mathrm{SINR}_1 = \mathrm{SINR}_2$ leads to a quadratic equation in α which does not reveal much of the dependencies due to its many terms. If we on the other hand restrict ourselves to the two most important cases, i.e., low and high SNR or more precisely interference dominance $I \gg N_o$ and noise dominance $N_o \gg I$, then we obtain simple relations.

Noise dominance: We find

$$\frac{\alpha_= P S_1}{N_o} = \frac{(1-\alpha_=) P S_2}{N_o} \quad (51)$$

and thus the harmonic mean of the two SNR values is obtained.

[7] Computing the inverse of $\lambda_1 \mathbf{I}_{N_T} + (1-\alpha) P \mathbf{H}_{11}^H \mathbf{g}_1 \mathbf{g}_1^H \mathbf{H}_{11}$ and multiplying with $\mathbf{H}_{11} \mathbf{f}_1$, we find $\mathbf{f}_1 = \mathbf{H}_{11}^H \mathbf{g}_1 \lambda_1 / [\lambda_1 + \mathbf{g}_1^H \mathbf{H}_{11} \mathbf{H}_{11}^H \mathbf{g}_1] \sim \mathbf{H}_{11}^H \mathbf{g}_1$ and for $\mathbf{f}_2$ correspondingly.

Interference dominance: We find

$$\frac{\alpha_= P S_1}{(1-\alpha_=) P I_1} = \frac{(1-\alpha_=) P S_2}{\alpha_= P I_2} \quad (52)$$

Solving with respect to α and we find the second part of (35).

Finally we are interested in the power distribution for highest data rate. This we find by differentiating

$$\frac{\partial}{\partial \alpha} \ln \left(1 + \frac{\alpha P S_1}{(1-\alpha) P I_1 + N_o}\right) + \ln \left(1 + \frac{(1-\alpha) P S_2}{\alpha P I_2 + N_o}\right) = 0.$$

In general this relates to a relative complex expression ending into a ratio of two polynomials, a polynomial of order two in the numerator and a polynomial of order four in the denominator. In case that interference is small against noise, the quadratic term collapses to the linear term in α with

$$\alpha_{\text{sumrate}} = \frac{1}{2} + \frac{N_o}{2P} \frac{S_1 - S_2}{S_1 S_2}. \quad (53)$$

In case that interference is dominating the noise, we typically have a convex rate plot and we have to pick the user with strongest SINR, thus either $\alpha_{\text{sumrate}} = 0$ or $\alpha_{\text{sumrate}} = 1$.

REFERENCES

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018*, Cisco Systems Inc., San Jose, CA, USA, Feb. 2014.
[2] *Ericsson Mobility Report: On the Pulse of the Networked Society, Interim Update*, Ericsson, Kista, Sweden, Feb. 2014.
[3] N. Bhushan *et al.*, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.
[4] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.
[5] X. Tao, X. Xu, and Q. Cui, "An overview of cooperative communications," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 65–71, Jun. 2012.
[6] *Introducing LTE-Advanced*, Agilent Technologies, Santa Clara, CA, USA, 2011.
[7] *LTE-Advanced (3GPP Rel.11) Technology Introduction*, Rohde & Schwarz, Munich, Germany, 2012.
[8] P. W. Baier, M. Meurer, T. Weber, and H. Troger, "Joint transmission (JT), an alternative rationale for the downlink of time division CDMA using multi-element transmit antennas," in *Proc. IEEE 6th Int. Symp. Spread Spectrum Techn. Appl.*, vol. 1. Sep. 2000, pp. 1–5.
[9] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
[10] C.-B. Chae, I. Hwang, R. W. Heath, Jr., and V. Tarokh, "Interference aware-coordinated beamforming in a multi-cell system," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3692–3703, Oct. 2012.
[11] M. H. A. Khan, M. H. Lee, and M. H. Mustary, "Zero-forcing based multiuser beamforming for coordinated multi-point transmission," in *Proc. Int. Symp. Commun. Inf. Technol.*, Oct. 2012, pp. 589–593.
[12] S. Schwarz, R. W. Heath, Jr., and M. Rupp, "Single-user MIMO versus multi-user MIMO in distributed antenna systems with limited feedback," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 54, 2013.
[13] S. Schwarz, J. C. 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.
[14] S. Schwarz and M. Rupp, "Evaluation of distributed multi-user MIMO-OFDM with limited feedback," *IEEE Trans. Wireless Commun.*, 2014, to be published.
[15] 3GPP TR 36.814. (Mar. 2010). *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36814.htm
[16] 3GPP TR 36.819. (Sep. 2013). *Technical Specification Group Radio Access Network; Coordinated Multi-Point Operation for LTE Physical Layer Aspects*. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36819.htm

[17] 3GPP TR 36.874. (Dec. 2013). *Technical Specification Group Radio Access Network; Coordinated Multi-Point Operation for LTE With Non-Ideal Backhaul (Release 12)*. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36874.htm

[18] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[19] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[20] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C.-B. Chae, and T. Sälzer, "From single user to multiuser communications: Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Oct. 2007.

[21] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[22] S. Schwarz and M. Rupp, "Subspace quantization based combining for limited feedback block-diagonalization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5868–5879, Nov. 2013.

[23] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[24] A. Mezghani, R. Hunger, M. Joham, and W. Utschick, "Iterative THP transceiver optimization for multi-user MIMO systems based on weighted sum-mse minimization," in *Proc. IEEE 7th Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2006, pp. 1–5.

[25] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[26] R. Mochaourab and E. A. Jorswieck, "Exchange economy in two-user multiple-input single-output interference channels," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 151–164, Apr. 2012.

[27] J. Park and Y. Sung, "On the pareto-optimal beam structure and design for multi-user MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5932–5946, Dec. 2013.

[28] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[29] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[30] L. Ping and P. Wang, "Multi-user gain and maximum eigenmode beamforming for MIMO systems with rate constraints," in *Proc. IEEE Inf. Theory Workshop Inf. Theory Wireless Netw.*, Jul. 2007, pp. 1–5.

[31] M. Trivellato, F. Boccardi, and F. Tosate, "User selection schemes for MIMO broadcast channels with limited feedback," in *Proc. 65th IEEE Veh. Technol. Conf.-Spring*, Dublin, Ireland, Apr. 2007, pp. 2089–2093.

[32] N. U. Hassan, C. Yuen, and Z. Zhang, "Optimal power control between two opportunistic cooperative base stations," in *Proc. IEEE 13th Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2012, pp. 194–198.

[33] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[34] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[35] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "Capacity regions and sum-rate capacities of vector Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5030–5044, Oct. 2010.

[36] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. New York, NY, USA: Wiley, 1980.

[37] M. K. Tsatsanis and Z. Xu, "Performance analysis of minimum variance CDMA receivers," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3014–3022, Nov. 1998.

[38] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1968–1983, Sep. 1999.

[39] M. Rupp, "Robust design of adaptive equalizers," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1612–1626, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1109/TSP.2011.2180717

[40] O. El Ayach, S. W. Peters, and R. W. Heath, Jr., "The practical challenges of interference alignment," *IEEE Wireless Commun.*, vol. 20, no. 1, pp. 35–42, Feb. 2013.

[41] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *Proc. IEEE Global Telecommun. Conf.*, Nov./Dec. 2008, pp. 1–6.

[42] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 953–961, Mar. 2008.

[43] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: Duality and sum-MSE minimization," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5436–5446, Nov. 2007.

[44] Q. Liu and C. W. Chen, "Joint transceiver beamforming design and power allocation for multiuser MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 3867–3871.

[45] M. Codreanu, A. Tolli, M. Juntti, and M. Latva-aho, "Joint design of Tx-Rx beamformers in MIMO downlink channel," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4639–4655, Sep. 2007.

[46] S.-H. Park, H. Park, Y.-D. Kim, and I. Lee, "Regularized interference alignment based on weighted sum-MSE criterion for MIMO interference channels," in *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1–5.

[47] M. Simko, C. Mehlführer, M. Wrulich, and M. Rupp, "Doubly dispersive channel estimation with scalable complexity," in *Proc. Int. ITG Workshop Smart Antennas (WSA)*, Feb. 2010, pp. 251–256.

[48] T. Kim, D. J. Love, and B. Clerckx, "MIMO systems with limited rate differential feedback in slowly varying channels," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 1175–1189, Apr. 2011.

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

[50] S. Schwarz, R. W. Heath, Jr., and M. Rupp, "Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems," in *Proc. 38th Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 5021–5025.

[51] S. Schwarz and M. Rupp, "Predictive quantization on the Stiefel manifold," *IEEE Signal Process. Lett.*, 2014, to be published.

[52] B. Clerckx and C. Oestges, *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*, 2nd ed. New York, NY, USA: Academic, 2013.

[53] E. G. Larsson and E. A. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008.

[54] E. G. Larsson, D. Danev, and E. A. Jorswieck, "Asymptotically optimal transmit strategies for the multiple antenna interference channel," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2008, pp. 708–714.

[55] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[56] M. Wrulich, C. Mehlfuhrer, and M. Rupp, "Managing the interference structure of MIMO HSDPA: A multi-user interference aware MMSE receiver with moderate complexity," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1472–1482, Apr. 2010. [Online]. Available: http://dx.doi.org/10.1109/TWC.2010.04.090612

[57] 3GPP, *TSG RAN; Study on Network-Assisted Interference Cancellation and Suppression (NAIC) for LTE Release 12*. [Online]. Available: http://www.3gpp.org/DynaReport/36866.htm, accessed Mar. 2014.

[58] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007. [Online]. Available: http://dx.doi.org/10.1109/TWC.2007.360373

[59] H. Du and P.-J. Chung, "A probabilistic approach for robust leakage-based MU-MIMO downlink beamforming with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1239–1247, Mar. 2012.

[60] J. Park, G. Lee, Y. Sung, and M. Yukawa, "Coordinated beamforming with relaxed zero forcing: The sequential orthogonal projection combining method and rate control," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3100–3112, Jun. 2013.

[61] T. Rüegg, A. U. T. Amah, and A. Wittneben, "On the trade-off between transmit and leakage power for rate optimal MIMO precoding," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2014, pp. 1–5.
[62] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2008–2026, Apr. 2011.
[63] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.
[64] S. S. Christensen, R. Agarwal, E. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.
[65] C. Guthy, W. Utschick, R. Hunger, and M. Joham, "Efficient weighted sum rate maximization with linear precoding," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2284–2297, Apr. 2010.
[66] O. Fresnedo, F. J. Vazquez-Araujo, L. Castedo, and J. Garcia-Frias, "Analog joint source-channel coding for OFDM systems," in *Proc. IEEE 14th Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2013, pp. 704–708.

STEFAN SCHWARZ (M'14) was born in Neunkirchen, Austria, in 1984. He received the B.Sc. degree in electrical engineering and the Dipl.-Ing. (M.Sc. equivalent) (Hons.) degree in telecommunications engineering, and the Dr.techn. (Ph.D. equivalent) (Hons.) degree in telecommunications engineering from the Vienna University of Technology, Vienna, Austria, in 2007, 2009, and 2013, respectively. He received the Honorary Prize of the Austrian Minister of Science and Research for excellent graduates of scientific and artistic universities in 2010. Since 2008, he has been a Project Assistant with the Mobile Communications Group, Institute of Telecommunications, Vienna University of Technology. His research interests are located in the broad fields of wireless communications and signal processing. In his dissertation, he focused on limited feedback single and multiuser MIMO-OFDM communication systems and multiuser scheduling algorithms. He actively serves as a reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

MARKUS RUPP (M'03–SM'06) received the Dipl.-Ing. degree from the University of Saarbrücken, Saarbrücken, Germany, in 1988, and the Dr.Ing. degree from the Technische Universitaet Darmstadt, Darmstadt, Germany, in 1993, where he worked with E. Haensler on designing new algorithms for acoustical and electrical echo compensation. From 1993 to 1995, he held a postdoctoral position with the University of Santa Barbara at California, Santa Barbara, CA, USA, with S. Mitra, where he worked with A. H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From 1995 to 2001, he was a member of the technical staff with the Wireless Technology Research Department, Bell-Labs, Crawford Hill, NJ, USA, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11, and UMTS. Since 2001, he has been a Full Professor of Digital Signal Processing in Mobile Communications with the Vienna University of Technology, Vienna, Austria, where he served as the Dean from 2005 to 2007. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005, is currently an Associate Editor of JASP *EURASIP Journal of Advances in Signal Processing*, JES *EURASIP Journal on Embedded Systems*, and *Research Letters in Signal Processing*, and was elected as an AdCom Member of EURASIP. He authored or co-authored over 250 papers, and holds patents on adaptive filtering, wireless communications, rapid prototyping, and automatic design methods.

Remote Radio Head Assignment and Beamforming in Dynamic Distributed Antenna Systems

Stefan Schwarz
Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion
Institute of Telecommunications, Technische Universität (TU) Wien
Email: sschwarz@nt.tuwien.ac.at

***Abstract*—In this paper, we consider a network architecture that supports dynamic allocation of remote radio heads (RRHs) amongst macro base stations, e.g., over dynamically established wireless fronthaul links or a reconfigurable wired fronthaul network. This architecture is similar to cloud radio access networks, yet with the difference that we rely on existing macro base stations rather than outsourcing them to the cloud. We propose a mixed-integer second order cone program (MISOCP) for RRH assignment amongst base stations and joint coordinated beamforming within the dynamically formed distributed antenna systems (DASs); we denote this network architecture as dynamic distributed antenna system (dDAS). Our optimization focuses on minimizing the number of users that are in outage. We propose several relaxations of our optimization problem to determine suboptimal solutions with much reduced computational complexity. We compare the performance of the dDASs to classical heterogeneous networks that employ coordinated beamforming of independent small cells rather than RRHs, demonstrating substantial improvements in terms of outage probability.**

***Index Terms*—Distributed antenna system, mixed-integer and second-order cone programming, relaxation, cloud RAN**

I. Introduction

Distributed antenna systems (DASs) are wireless communication architectures that utilize multiple spatially (geographically) distributed transmission/reception points to serve users in a wireless cellular network [1]. DASs make use of infrastructure such as remote radio heads (RRHs), with minimal intelligence of their own, to extend the base stations' antenna ports. RRHs are connected to the base station by high-bandwidth low-latency dedicated connections that are non-interfering with the primary radio resource, e.g., dedicated micro-wave links or fiber links using radio-over-fiber technology [2]. Originally, the motivation for DASs was to improve the coverage and reduce the outage probability of the network [3], [4], especially in indoor locations, but they are also effective for enhancing the network capacity and area spectral efficiency (ASE), utilizing coherent single-user and multi-user multiple-input multiple-output (MIMO) transmission [5], [6], [7], [8]. Recently, the application of DASs in cognitive radio has been investigated as a promising technology to further enhance spectrum utilization [9]. Also, in [10] the combination of DAS and simultaneous wireless and power transfer (SWIPT) is considered in a single-cell downlink situation, demonstrating simultaneously large energy and information transfer. In [11], the authors investigate the co-channel interference in DASs with virtual cells, where a few surrounding remote antenna ports of each user form a virtual cell and serve the user.

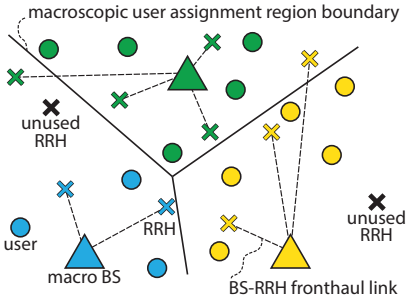


Fig. 1. Illustration of a dDAS; colors and lines illustrate association of users (circles) and RRHs (crosses) amongst BSs (triangles). RRHs can also not be assigned to BSs at all (black crosses), in which case they are deactivated.

In this paper, our focus is on enhancing the coverage and reducing the outage probability of wireless networks, through RRHs that are randomly distributed within the geographic network area. We assume that each RRH can freely be associated to any single one base station (BS) of the network, by dynamically establishing a wireless/wired fronthaul link between them, as illustrated in Fig. 1. This fronthaul link is non-interfering with the primary radio resources. In this way, DASs can be dynamically established within the network; we, hence, denote this architecture as dynamic distributed antenna system (dDAS). This architecture is in principle similar to a cloud radio access network (CRAN) [12], [13], [14]; however, we still rely on conventional macro BSs in the cellular network rather than virtualizing them entirely as base band units on cloud servers. Our architecture can, hence, be seen as an intermediate step between current cellular networks and CRANs, where we add freely assignable RRHs on top of existing macro BSs. It can be realized within the network function virtualization (NFV) framework of fifth generation (5G) mobile networks [15]. We apply a user-plane/control-plane splitting approach in which the signaling information of the control-plane is only radiated from the macro BSs to achieve large-are coverage and efficient support of high mobility, whereas the data transmission over the user-plane is jointly performed from macro BSs and RRHs on the same channel [16], [17]. This implies that the user assignment amongst BSs is determined by the macroscopic signal to interference and noise ratio (SINR) w.r.t. the BSs only, without being affected by the RRHs. Hence, the coverage area of each BS, determining the user association, is independent of the RRH positions and assignment. Only after users have been associated to BSs, we consider the assignment of RRHs

978-1-5386-3180-5/18/$31.00

amongst BSs, with the goal of minimizing the number of users that are in outage. A related architecture has been considered in [18] in the context of SWIPT-DAS; yet, in [18] each user could potentially be served by any RRH, whereas in our case a user can only be served by the RRHs assigned to the user-associated BS, that is, by the dDAS corresponding to the user-associated BS. Also related is the work of Ghauch et al. [19] who consider dynamic formation of antenna domains in CRAN systems; in this work, however, users are a priori associated to RRHs and these RRHs are then assigned to so-called aggregation nodes such that the inter-antenna-domain interference is minimized. When we compare this approach to our work, we can identify our macro BSs with Ghauch's aggregation nodes; in our case, however, users are a priori assigned to macro BSs and RRHs are then assigned amongst BSs to minimize the number of users in outage. We propose a mixed-integer second order cone program (MISOCP) for joint RRH assignment and coordinated beamforming, and provide several relaxations to obtain suboptimal solutions with manageable computational complexity [20]. We compare the outage performance of the proposed dDAS architecture to heterogeneous networks that employ small cells instead of RRHs as independent but coordinated transmission points.

Notation: We denote vectors and matrices by boldface lower- and upper-case letters $\mathbf{x}$ and $\mathbf{X}$. We address the entry in the ith row and jth column of matrix $\mathbf{X}$ by $[\mathbf{X}]_{ij}$. We denote the indicator function as $\delta(c)$, i.e., $\delta(c) = 1$ if the condition c is satisfied and $\delta(c) = 0$ else. The expected value of random variable r is $\mathbb{E}\{r\}$, and the real and imaginary parts of a complex variable z are $\Re(z)$ and $\Im(z)$, respectively.

II. System Model

We consider downlink transmission in a mobile wireless communication network that consists of N_b BSs that serve N_u users. We assume that the BSs broadcast the signaling information of the mobile network and that, based on these signals, each user is attached to the BS that provides the strongest signal; we assume that the association of users amongst BSs is based on the macroscopic signal strength (not accounting for microscopic fading of the channel).

In addition to BSs, we assume that the mobile network also contains N_r RRHs that can be freely associated to any BS, as illustrated in Fig. 1. Each RRH can be assigned to at most one single BS; however, a RRH might also not be assigned to any BS at all, in which case the RRH is switched off. This might be achieved, e.g., by wirelessly connecting RRHs to BSs over millimeter wave (mmWave) fronthaul links, or by providing a wired fronthaul that supports dynamic reconfiguration of the network architecture. Either way, we do not account for the quality (capacity, latency) of the dynamically established fronthaul links between BSs and RRHs in this work, i.e., we assume bandwidth-unlimited zero-latency fronthaul connections that enable coherent beamforming over a BS and its associated RRHs. We denote such a system as a dDAS. Notice, we assume that the signaling information of the mobile network is not transmitted from the RRHs; hence, to point it out again, user association amongst BSs is only based on the macroscopic path loss w.r.t. BSs (though load-balancing could also be applied).

Considering BS i, we denote the index set of users assigned to BS i as $\mathcal{U}_i \subseteq \{1,\dots,N_u\}$, such that $\mathcal{U}_i \cap \mathcal{U}_j = \emptyset, \forall i \neq j$ and $\bigcup_{i=1}^{N_b} \mathcal{U}_i = \{1,\dots,N_u\}$. To specify the association of RRHs amongst BSs, we use binary indicator variables

$$b_{ki} \in \{0,1\}, \quad 1 \leq k \leq N_r,\ 1 \leq i \leq N_b, \tag{1}$$

such that $b_{ki} = 1$ if RRH k is assigned to BS i and $b_{ki} = 0$ otherwise. Notice, we have $\sum_{i=1}^{N_b} b_{ki} \leq 1$ due to mutually exclusive assignment of RRHs. We assume that each BS is equipped with M_b transmit antennas and has a power constraint of P_b. Similarly, each RRH is equipped with M_r transmit antennas and has a power constraint of P_r. Users are equipped with only a single antenna. If an RRH is activated, i.e., associated to a BS, we assume that it consumes a certain fixed circuit power P_c irrespective of the actual transmit power.

For this system, the input-output relationship of user u is:

$$y_u = \sum_{i=1}^{N_b} \mathbf{h}_{ui}^{\mathsf{H}} \sum_{n\in\mathcal{U}_i} \mathbf{f}_{ni} x_n + \sum_{k=1}^{N_r} \boldsymbol{\eta}_{uk}^{\mathsf{H}} \sum_{i=1}^{N_b} \sum_{n\in\mathcal{U}_i} \boldsymbol{\varphi}_{nk} x_n + n_u, \tag{2}$$

$$\sum_{n\in\mathcal{U}_i} \|\mathbf{f}_{ni}\|^2 \leq P_b, \forall i, \tag{3}$$

$$\sum_{n\in\mathcal{U}_i} \|\boldsymbol{\varphi}_{nk}\|^2 \leq b_{ki}^2 P_r, \forall i,k, \tag{4}$$

where x_n, $\mathbb{E}\{|x_n|^2\} = 1$, denotes the transmit symbol intended for user n, $\mathbf{f}_{ni} \in \mathbb{C}^{M_b\times 1}$ is the beamformer applied at BS i for transmission to user n, and $\mathbf{h}_{ui} \in \mathbb{C}^{M_b\times 1}$ is the channel of user u w.r.t. BS i. Similarly, $\boldsymbol{\varphi}_{nk} \in \mathbb{C}^{M_r\times 1}$ represents the beamformer applied at RRH k for transmission to user n, and $\boldsymbol{\eta}_{uk} \in \mathbb{C}^{M_r\times 1}$ is the channel of user u w.r.t. RRH k. The scalar n_u denotes zero-mean unit-variance complex-Gaussian receiver noise. Notice, we employ some redundant notation in (2), because we define beamformers at each RRH for all users, i.e., also for users that are not served by the corresponding dDAS. We apply this approach to simplify the formulation of the proposed optimization problem below; notice, though, that only the beamformers of the users associated to the dDAS are not equal to zero since the power-constraint (4) involves the binary RRH assignment variables b_{ki}; i.e., beamformers of users not associated to the dDAS are set equal to zero.

The channel vectors $\mathbf{h}_{ui}$ can be decomposed as $\mathbf{h}_{ui} = g_{ui}\tilde{\mathbf{h}}_{ui}$, where $\tilde{\mathbf{h}}_{ui}$ represents the microscopic fading component and satisfies $\mathbb{E}\left\{\left\|\tilde{\mathbf{h}}_{ui}\right\|^2\right\} = M_b$ and g_{ui} is the macroscopic fading; similarly, $\boldsymbol{\eta}_{uk}$ can be decomposed as $\boldsymbol{\eta}_{uk} = \gamma_{uk}\tilde{\boldsymbol{\eta}}_{uk}$. The user association $\mathcal{U}_i, \forall i$ is based purely on g_{ui}. Assuming $u \in \mathcal{U}_i$, the achieved SINR β_u is:

$$\beta_u = \frac{\left|\mathbf{h}_{ui}^{\mathsf{H}}\mathbf{f}_{ui} + \sum_{k=1}^{N_r} \boldsymbol{\eta}_{uk}^{\mathsf{H}}\boldsymbol{\varphi}_{uk}\right|^2}{\sum_{j=1}^{N_b}\sum_{\substack{n\in\mathcal{U}_j\\ n\neq u}} \left|\mathbf{h}_{uj}^{\mathsf{H}}\mathbf{f}_{nj} + \sum_{k=1}^{N_r} \boldsymbol{\eta}_{uk}^{\mathsf{H}}\boldsymbol{\varphi}_{nk}\right|^2 + 1} = \frac{|s_u|^2}{\|\mathbf{i}_u\|^2}, \tag{5}$$

$$s_u = \mathbf{h}_{ui}^{\mathsf{H}} \mathbf{f}_{ui} + \sum_{k=1}^{N_r} \boldsymbol{\eta}_{uk}^{\mathsf{H}} \boldsymbol{\varphi}_{uk},$$

$$\mathbf{i}_u = \left[\left[\mathbf{h}_{uj}^{\mathsf{H}} \mathbf{f}_{nj} + \sum_{k=1}^{N_r} \boldsymbol{\eta}_{uk}^{\mathsf{H}} \boldsymbol{\varphi}_{nk} \right]_{\substack{n \in \mathcal{U}_j, n \neq u \\ j \in \{1,\dots N_b\}}}, 1 \right]^{\mathsf{T}},$$

where, the vector $\mathbf{i}_u \in \mathbb{C}^{N_u \times 1}$ gathers all the sum-terms in the denominator of β_u, i.e., the interference caused by the transmissions to other users, as well as, the noise of unit variance. The scalar s_u denotes the effective channel (including beamforming) of user u. Notice, the signals of a user radiated from all antenna arrays within a dDAS add up coherently. Again, in s_u and $\mathbf{i}_u$ only beamformers $\boldsymbol{\varphi}_{uk}$ of RRHs with $b_{ki} = 1$ are not equal to zero. We assume that user u is in outage if its SINR β_u falls below the user-specific associated threshold value α_u.

III. Joint RRH Assignment and Beamformer Optimization

We formulate in the following an optimization problem for joint RRH assignment and beamforming with the goal of minimizing the number of users in the mobile network that are in outage. That is, our optimization variables are the binary selection variables $b_{ki}, \forall i, k$ and the BS and RRH beamformers $\mathbf{f}_{ui}$ and $\boldsymbol{\varphi}_{uk}$, respectively. If we are able to satisfactorily serve all the users, such that $\beta_u \geq \alpha_u, \forall u$, we additionally attempt to minimize the power consumption of the mobile network, where we account for the transmit power of BSs and RRHs, as well as, the constant circuit power of RRHs that is consumed as soon as an RRH is active; notice, we do not consider the circuit power of BSs since they always need to be active in order to provide the signaling information.

Our proposed optimization problem for joint RRH assignment and beamforming is:

$$\begin{aligned}
&\text{minimize } A_{\text{tot}} + c_{\text{out}} \left\| \mathbf{u}_{\text{out}} \right\|_0 && \text{(6a)}\\
&\quad \text{w.r.t. } \mathbf{u}_{\text{out}} \in \{0,1\}^{N_u},\ b_{ki} \in \{0,1\},\ A_{\text{tot}} \in \mathbb{R}_+,\\
&\qquad \mathbf{f}_{ui} \in \mathbb{C}^{M_b \times 1},\ \boldsymbol{\varphi}_{uk} \in \mathbb{C}^{M_r \times 1},\\
&\forall u \in \{1,\dots,N_u\},\ \forall i \in \{1,\dots,N_b\},\ \forall k \in \{1,\dots,N_r\}\\
&\text{s.t. } [\mathbf{u}_{\text{out}}]_u = \delta(\beta_u < \alpha_u), && \text{(6b)}\\
&\quad \sum_{i=1}^{N_b} \sum_{n \in \mathcal{U}_i} \left(\left\| \mathbf{f}_{ni} \right\|^2 + \sum_{k=1}^{N_r} \left\| \boldsymbol{\varphi}_{nk} \right\|^2 \right) + \dots\\
&\quad \dots + \sum_{i=1}^{N_b} \sum_{k=1}^{N_r} b_{ki}^2 P_c \leq A_{\text{tot}}^2, && \text{(6c)}\\
&\quad \sum_{i=1}^{N_b} b_{ki}^2 \leq 1, && \text{(6d)}\\
&\quad \sum_{n \in \mathcal{U}_i} \left\| \mathbf{f}_{ni} \right\|^2 \leq P_b, && \text{(6e)}\\
&\quad \sum_{n \in \mathcal{U}_i} \left\| \boldsymbol{\varphi}_{nk} \right\|^2 \leq b_{ki}^2 P_r, && \text{(6f)}\\
&\forall u \in \{1,\dots,N_u\},\ \forall i \in \{1,\dots,N_b\},\ \forall k \in \{1,\dots,N_r\}
\end{aligned}$$

with β_u as specified in Eq. (5). Constraint (6b) indicates an outage event of user u. If a user is in outage, it adds a factor c_{out} to the objective function; by setting c_{out} larger than the square-root of the maximal total power consumption, i.e., $c_{\text{out}} \geq \sqrt{(P_c + P_r) N_r + P_b N_b} \geq A_{\text{tot}}$, we ensure that minimizing the number of users that are in outage has priority over minimizing the power consumption $P_{\text{tot}} = A_{\text{tot}}^2$. Constraint (6c) represents the total power consumption of the system; if no user is in outage, we attempt to minimize this total power consumption. Inclusion of the circuit powers in (6c) favors solutions with small number of activated RRHs (depending on the magnitude of P_c). Constraint (6d) ensures that each RRH is assigned to at most one BS and (6e), (6f) represent the individual power constraints of BSs and RRHs.

Problem (6) is non-convex, even if we apply an integer relaxation to the binary variables and an ℓ_1-relaxation to the ℓ_0-norm in the objective function. Yet, the problem can be reformulated to a MISOCP that can be solved by state-of-the-art solvers such as *MOSEK* [21], e.g., with a combination of branch-and-bound and second-order cone programming. To see this, we first of all reformulate constraint (6b). Notice that problem (6) is invariant w.r.t. the absolute phase of the set of beamformers $\{\mathbf{f}_{ui}, \boldsymbol{\varphi}_{uk}\}$ that are applied to serve user u; that is, we can multiply these beamformers of user u with a common phase-factor $e^{j\phi_u}$ without changing the signal or interference powers. This allows us to apply the well-know approach of restricting the effective channel s_u of each user to be purely real-valued [22]. With this, we can bring (6b) into convex form:

$$i_u^{\max} [\mathbf{u}_{\text{out}}]_u \geq \max \left(\left\| \mathbf{i}_u \right\| - \frac{\Re(s_u)}{\sqrt{\alpha_u}}, 0 \right), \quad \Im(s_u) = 0. \quad (7)$$

Here, the constant $i_u^{\max}$ must be chosen to ensure that the left-hand side of (7) is larger than the right-hand side irrespective of the choice of beamformers as soon as $[\mathbf{u}_{\text{out}}]_u$ is set equal to one. This can be achieved by applying a worst-case estimate on $\|\mathbf{i}_u\|$. The following upper bound can be derived by applying the triangle inequality and assuming worst-case interfering beamformers that are aligned with the respective interference channels:

$$i_u^{\max} = \sqrt{\sum_{j=1}^{N_b} \left(\sqrt{P_b} \left\| \mathbf{h}_{uj} \right\| + \sum_{k=1}^{N_r} \sqrt{P_r} \left\| \boldsymbol{\eta}_{uk} \right\| \right)^2 + 1} \geq \left\| \mathbf{i}_u \right\|. \quad (8)$$

The total power constraint (6c) can be written in convex form by stacking all beamforming vectors of BSs and RRHs into one single large beamforming vector $\mathbf{f} \in \mathbb{C}^{(N_u M_b + N_u N_r M_r) \times 1}$, which is further augmented by the circuit powers:

$$\mathbf{f}_{\text{full}} = \left[\mathbf{f}^{\mathsf{T}}, \sqrt{P_c} \sum_{i=1}^{N_b} b_{1i}, \dots, \sqrt{P_c} \sum_{i=1}^{N_b} b_{N_r i} \right]^{\mathsf{T}}. \quad (9)$$

Vector $\mathbf{f}_{\text{full}}$ is of length $N_u M_b + N_u N_r M_r + N_r$; notice, that at most one single term in the sums $\sum_{i=1}^{N_b} b_{ki}$ is not equal to zero, allowing us to pull in (6c) the square out of the summation $\sum_{i=1}^{N_b} b_{ki}^2 = \left(\sum_{i=1}^{N_b} b_{ki}\right)^2$ and, thus, to replace (6c) with:

$$\|\mathbf{f}_{\text{full}}\|^2 \leq A_{\text{tot}}^2. \tag{10}$$

The remaining power constraints (6e), (6f) are anyway convex. The objective function becomes convex by replacing the ℓ_0-norm with the ℓ_1-norm; notice, this does not impact the optimal solution, since the variables $\mathbf{u}_{\text{out}}$ are restricted to be binary. Hence, by replacing (6b) with (7) and (6c) with (10), we obtain an equivalent representation of problem (6) as a MISOCP. We denote this problem as the *full binary* joint RRH assignment and beamforming problem.

IV. Relaxations and Suboptimal Solutions

The full binary optimization problem is computationally demanding, since it requires branch-and-bound optimization over two-sets of binary variables, i.e., the RRH assignment variables b_{ki} and the outage-indicator variables $\mathbf{u}_{\text{out}}$. For problems with a number of RRHs and users that exceeds a few tens, solution of the full binary problem becomes practically infeasible. We, hence, consider below several relaxations of reduced complexity to find a suboptimal solution.

A. L1 Relaxation of Objective Function and Integer Relaxation of Outage Indicators

In problem (6), we can replace the ℓ_0-norm in the objective function with the ℓ_1-norm without changing the optimal solution, since $\mathbf{u}_{\text{out}}$ is a vector of binary random variables. Alternatively, however, we can also apply an integer relaxation to $\mathbf{u}_{\text{out}}$ with the additional vector-valued inequality constraints $0 \leq \mathbf{u}_{\text{out}} \leq 1$ and keep the ℓ_0-norm in the objective function without changing the optimal value of problem (6). Yet, both approaches are computationally demanding.

We, hence, propose to jointly apply both relaxations, the ℓ_1-norm relaxation of the objective function and the binary integer relaxation of the outage indicator variables $\mathbf{u}_{\text{out}}$, even though this will lead to a suboptimal solution. The solution of this relaxed problem is still feasible for the original problem (6); thus, there is no need to post-process the solution. We denote the corresponding optimization problem as *L1 relaxed* joint RRH assignment and beamforming problem.

B. Integer Relaxation of RRH Assignment

Keeping the relaxations from the previous subsection, we additionally consider now an integer relaxation of the RRH assignment variables b_{ki} with the additional constraints $b_{ki} \geq 0$; notice, we do not need the upper-bound $b_{ki} \leq 1$ since (6d) is more strict. With an integer relaxation of b_{ki}, however, $(\sum_{i=1}^{N_b} b_{ki})^2$ is not any longer equivalent to $\sum_{i=1}^{N_b} b_{ki}^2$ and we therefore have to extend (9) and (10) to:

$$\mathbf{f}_{\text{full}}^{(2)} = \Big[\mathbf{f}^{\mathsf{T}}, \sqrt{P_c} b_{11}, \ldots, \sqrt{P_c} b_{1N_b}, \ldots, \sqrt{P_c} b_{N_r 1}, \ldots, \sqrt{P_c} b_{N_r N_b}\Big]^{\mathsf{T}}, \tag{11}$$

$$\left\|\mathbf{f}_{\text{full}}^{(2)}\right\|^2 \leq A_{\text{tot}}^2, \tag{12}$$

with $\mathbf{f}_{\text{full}}^{(2)}$ of length $N_u M_b + N_u N_r M_r + N_r N_b$.

The solution of this problem is in general not feasible for the original problem any longer, because RRHs might be assigned to multiple BSs due to the integer relaxation of the selection variables. One approach to obtain a feasible solution is to apply a simple randomized rounding strategy [23] to the selector variables with a corresponding renormalization of the RRH beamformers. This approach, however, performs poorly because the beamforming solution is not optimized for the RRH allocation after randomized rounding. We therefore propose to utilize only the RRH allocation after randomized rounding (discarding the beamformer solution) and to re-optimize the beamformers with this fixed RRH allocation. This is easily possible by following the optimization approach in Section IV-A while fixing the RRH allocation. We denote the corresponding optimization problem as *integer relaxed* joint RRH assignment and beamforming problem. Since with this approach it is necessary to solve two optimization problems the complexity is increased; yet, both problems are second order cone programs (SOCPs) and thus efficiently solvable.

C. Macro Optimization of RRH Assignment

The joint RRH assignment and beamforming problems described above require perfect knowledge of the channels w.r.t. all transmitters (BSs and RRHs). In practice, this requirement might not be feasible, especially as long as RRHs are not assigned to BSs. We therefore consider next an assignment of RRHs that is based purely on macroscopic channel state information (CSI), i.e., the macroscopic pathloss values g_{ui} and γ_{uk}. This allows for RRH assignment with minimal CSI requirements, whereas for beamforming we still demand full CSI. The corresponding assignment of RRHs is based on the macroscopic SINR:

$$B_u = \frac{g_{ui} p_{ui} + \sum_{k=1}^{N_r} \gamma_{uk} \pi_{uk}}{\sum_{j=1}^{N_b} \sum_{\substack{n \in \mathcal{U}_j \\ n \neq u}} \left(g_{uj} p_{nj} + \sum_{k=1}^{N_r} \gamma_{uk} \pi_{nk}\right) + 1}, \tag{13}$$

where p_{ui} and π_{uk} denote power allocation variables. We can now formulate an optimization problem similar to (6), where we employ the power allocation variables instead of the beamformers as optimization variables. The corresponding power constraints are $\sum_{n \in \mathcal{U}_i} p_{ni} \leq P_b$ and $\sum_{n \in \mathcal{U}_i} \pi_{nk} \leq b_{ki}^2 P_r$. Adopting the approach of Section IV-A we can efficiently determine a suboptimal *macroscopic* RRH assignment. Given this assignment we can then find a beamforming solution, following the same approach as described in Section IV-B.

D. Small Cell Optimization

Rather than employing RRHs, we can place independent small cells at the same positions to serve the users; the total number of small cells hence is $N_s = N_r$. In this case, we associate users to the strongest macro BS or small cell, accounting for the difference in transmit power between the

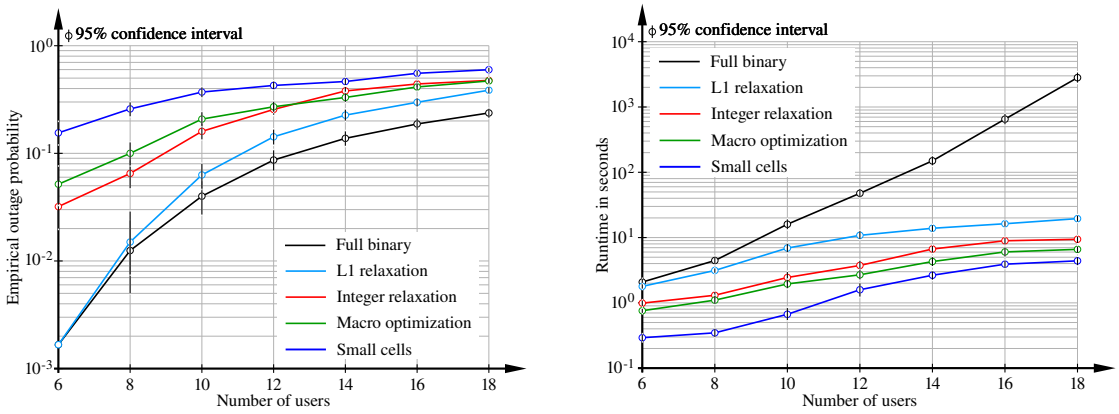


Fig. 2. Outage probability (left) and optimizer runtime (right) with three BSs, four RRHs/small cells and varying number of users.

two device types; we assume that small cells support the same transmit power and number of transmit antennas as RRHs, i.e., $P_s = P_r$ and $M_s = M_r$. The total number of independent BSs thus is $N_b + N_s$ and the corresponding input-output relationship is:

$$y_u = \sum_{i=1}^{N_b+N_s} \mathbf{h}_{ui}^{\mathsf{H}} \sum_{n \in \mathcal{U}_i} \mathbf{f}_{ni} x_n + n_u. \tag{14}$$

Notice, we have the power constraint $\sum_{n \in \mathcal{U}_i} \|\mathbf{f}_{ni}\|^2 \leq P_s, \forall i > N_b$. We can formulate an optimization problem similar to (6) (without RRH assignment variables) to jointly optimize the beamformers of all small cells. An efficient sub-optimal solution, denoted as *small cell* optimization problem, is possible by applying the same strategy as in Section IV-A.

V. Simulations

In this section, we present numerical evaluations of the outage probability achieved by the different RRH assignment and beamformer optimization approaches, as well as, the required runtime of the optimizer (*MOSEK*). We consider a network that consists of three macro BSs that are placed on an isosceles triangle with an edge length of 500 m, as illustrated in Fig. 1. Users and RRHs/small cells are randomly placed within a circle of radius equal to four-thirds times the circumradius of the triangle. We simulate 100 random realizations of user and RRH/small cell positions. To determine the signal to noise ratio (SNR) of a user, we assume a geometric pathloss model with pathloss exponent equal to four and we set the noise variance such that in the center of the triangle a user experiences an average macroscopic SNR of 0 dB (not accounting for beamforming gains).

In our first simulation, we consider a fixed number of $N_r = N_s = 4$ RRHs/small cells and we vary the number of users N_u. We employ two antennas on each antenna array, i.e., $M_b = M_r = M_s = 2$, we assume that the transmit power of BSs is ten times larger than that of RRHs/small cells, that is, $P_b = 10\,P_r = 10\,P_s = 46$ dBm, and we set the circuit power equal to the transmit power $P_c = P_r$. Furthermore, we set the SINR threshold of each user equal to $\alpha_u = 3$ dB, $\forall u$.

In Fig. 2 (left), we show the empirical cumulative distribution function (ecdf) of the outage probability achieved by the different approaches as a function of the number of users N_u served by the network. As expected, the best performance is achieved by the *full binary* joint RRH assignment and beamforming problem. Close to that comes the *L1 relaxed* problem, which applies an integer relaxation to the outage indicators and an ℓ_1-norm relaxation to the objective function. The *integer relaxed* optimization problem, which additionally applies an integer relaxation to the RRH assignment variables, exhibits similar performance to the *macroscopic* SINR based assignment. The worst performance is achieved by replacing RRHs with small cells, even though the beamformers of all small cells and BSs are still jointly optimized.

In terms of average optimizer runtime, as shown in Fig. 2 (right), the behaviour is exactly reversed. The *full binary* optimization is practically only feasible for a small number of users; its runtime grows exponentially with the number of users. The other relaxed problems exhibit much more benign runtime scaling with the number of users. The fastest problem is the small cell optimization, since it does not require any assignment of small cells amongst BSs.

In our second simulation, we consider a fixed number of users and we vary the number of RRHs/small cells; we assume $N_u = 40$ users. The remaining parameters are set as before. In this case, the *full binary* optimization is not feasible any longer in terms of computational complexity; we therefore only investigate the other relaxed approaches. In Fig. 3, we show the corresponding outage performances and optimizer runtimes. We observe that even the *L1 relaxed* problem is only feasible for a small number of RRHs in terms of computational complexity, since it still requires a binary search over the RRH assignment possibilities. The order of the outage and runtime performances of the different optimization problems is the same as in the previous simulation. Notice that adding small cells hardly improves the outage probability, whereas adding RRHs can give a significant gain, especially when applying

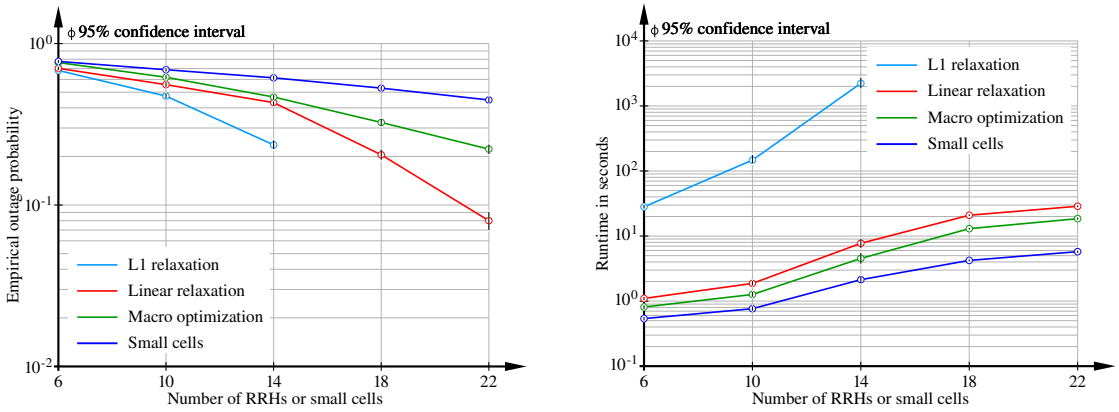


Fig. 3. Outage probability (left) and optimizer runtime (right) with three BSs, forty users and varying number of RRHs/small cells.

only few relaxations/approximations to the *full binary* joint RRH assignment and beamforming problem.

VI. Conclusion

Dynamic association of RRHs amongst BSs, i.e., dynamically forming DASs, has the potential to significantly reduce the outage probability of wireless cellular networks. The corresponding proposed joint RRH assignment and beamforming optimization problem, however, involves binary variables and thus requires complex binary optimization procedures. Such binary optimization is practically feasible only for small number of RRHs and users. We propose several relaxations of reduced complexity; however, these relaxations cause non-negligible performance degradation. Hence, one main future challenge is to find better approximations/relaxations of the full binary optimization problem, which achieve close-to-optimal performance at significantly reduced computational complexity. Furthermore, the impact of the quality of the fronthaul connection between BSs and RRHs must be taken into account to better understand the capabilities of dDASs.

Acknowledgment: The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged.

References

[1] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 161–167, April 2013.

[2] M. Riaz, R. Nielsen, J. Pedersen, N. Prasad, and O. Madsen, "On radio over fiber for heterogeneous wireless networks," in *International Conf. on Wireless and Optical Communications Networks*, April 2009.

[3] K. Kerpez and S. Ariyavisitakul, "A radio access system with distributed antennas," in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 3, Nov. 1994.

[4] 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, Jun 1994.

[5] 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.

[6] R. Heath, Jr., T. Wu, Y. H. Kwon, and A. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, Oct. 2011.

[7] S. Schwarz, R. Heath, Jr., and M. Rupp, "Multiuser MIMO in distributed antenna systems with limited feedback," in *IEEE Globecom Workshops*, Dec 2012, pp. 546–551.

[8] S. Schwarz and M. Rupp, "Evaluation of distributed multi-user MIMO-OFDM with limited feedback," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6081–6094, Aug. 2014.

[9] F. Zhao, S. Ji, and H. Chen, "A spectrum auction algorithm for cognitive distributed antenna systems," *Ad Hoc Networks*, vol. 58, pp. 269 – 277, 2017.

[10] F. Yuan, S. Jin, K. K. Wong, J. Zhao, and H. Zhu, "Wireless information and power transfer design for energy cooperation distributed antenna systems," *IEEE Access*, vol. 5, pp. 8094–8105, 2017.

[11] Z. Liu, Y. Zhao, and H. Wu, "Co-channel interference analysis of large-scale distributed antenna systems with virtual cells," *Electronics Letters*, vol. 53, pp. 968–970, July 2017.

[12] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks: A technology overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, 2015.

[13] Y. Li, T. Jiang, K. Luo, and S. Mao, "Green heterogeneous cloud radio access networks: Potential techniques, performance trade-offs, and challenges," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 33–39, Nov. 2017.

[14] Y. Li, Y. Zhang, K. Luo, T. Jiang, Z. Li, and W. Peng, "Ultra-dense hetnets meet big data: Green frameworks, techniques, and approaches," *CoRR*, vol. abs/1709.08797, 2017, accepted for publication in IEEE Communications Magazine.

[15] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Communications Magazine*, vol. 54, no. 4, pp. 84–91, April 2016.

[16] 4G Americas, "Understanding 3GPP release 12: standards or HSPA+ and LTE enhancements," white paper, Feb. 2015.

[17] S. Schwarz and M. Rupp, "Society in motion: Challenges for LTE and beyond mobile communications," *IEEE Communications Magazine, Feature Topic: LTE Evolution*, vol. 54, no. 5, pp. 76–83, 2016.

[18] Y. Dong, J. Hossain, J. Cheng, and V. C. M. Leung, "Joint RRH selection and beamforming in distributed antenna systems with energy harvesting," in *2017 International Conference on Computing, Networking and Communications*, Jan 2017, pp. 582–586.

[19] H. Ghauch, M. M. U. Rahman, S. Imtiaz, and J. Gross, "Coordination and antenna domain formation in cloud-RAN systems," in *IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–7.

[20] S. Drewes, *Mixed Integer Second Order Cone Programming*. Verlag Dr. Hut, 2009. [Online]. Available: https://books.google.at/books?id=vHP0SAAACAAJ

[21] M. ApS, *MOSEK Optimization Suite - Release 8.0.0.81*, 2017. [Online]. Available: https://mosek.com/resources/doc

[22] R. Zhang and Y. C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 88–102, Feb 2008.

[23] P. Raghavan and C. D. Tompson, "Randomized rounding: A technique for provably good algorithms and algorithmic proofs," *Combinatorica*, vol. 7, no. 4, pp. 365–374, Dec 1987.

Transmit Optimization for the MISO Multicast Interference Channel

Stefan Schwarz, *Member, IEEE*, and Markus Rupp, *Fellow, IEEE*

***Abstract*—In this paper, we consider multiple-input single-output (MISO) physical layer multicasting, where several multiantenna transmitters each simultaneously multicast a common message to a distinct set of single-antenna receivers, causing interference between different multicast messages. To optimize the achievable rate of this MISO multicast interference channel, we propose iterative distributed transmit optimization algorithms that are based on interference leakage control, requiring local channel state information at each transmitter and leakage information exchange among transmitters. We, furthermore, propose extensions of existing coordinated multipoint transmission schemes that have been developed for unicast interference channels, such as signal to leakage and noise ratio beamforming, to the considered multicast system. Such methods are of importance, e.g., for future releases of LTE that will support multiantenna multicasting using MBMS/MBSFN. Numerical simulations confirm the potential of the proposed distributed transmit optimization algorithm.**

***Index Terms*—Physical layer multicasting, leakage based precoding, distributed optimization, semi-definite relaxation, CoMP.**

I. Introduction

THE multiple-input single-output (MISO) multicast interference channel (MC-IC) consists of several multi-antenna transmitters that each simultaneously multicast a common message to several different single-antenna receivers, causing interference among the multicast messages as illustrated in Figure 1. It can be considered as a generalization of the MISO unicast interference channel (UC-IC) in which each transmitter serves only a single user. In wireless communications, the MISO MC-IC has received significant attention in the context of physical layer multicasting. Third generation partnership project (3GPP) universal mobile telecommunications system (UMTS), for instance, supports physical layer multicasting already since 2006 with Release 6 of the standard [1]. This feature, denoted multimedia broadcast multicast service (MBMS), is mainly intended for mobile video broadcasting but, to the best of our knowledge, no real commercial deployments are available to date. Still, MBMS was inherited by the long term evolution (LTE) standard, where it is known as evolved MBMS (eMBMS). Interest in the technology in the context of *venue casting* [2], that is, multicasting/broadcasting at local events, is confirmed by recent technology trials conducted by several operators and equipment manufacturers [3]–[5]. Other application scenarios envisioned include *road safety* information broadcasting [6] and software/firmware/operating-system updates for mobile *Internet of Things* devices [7]. Notice, however, that the considered scenario does not represent multimedia broadcast single frequency network (MBSFN) transmission, because each transmitter multicasts a different message. Yet, the proposed methods can also be applied for joint precoding within an MBSFN area.

In this article, we consider transmit optimization for the MISO MC-IC with perfect local channel state information at the transmitter (CSIT); that is, each transmitter knows perfectly the channel with respect to all receivers in the region of interest. We furthermore consider single-user detection, i.e., each receiver treats interference as additional noise and does not attempt to decode it. This approach is not optimal, especially in case of very strong interference where successive interference cancellation is potentially capable of subtracting interfering signals before decoding the desired signal [8], [9]. Yet, for reasons of complexity, most mobile devices in the foreseeable future will not implement such complex receive processing, thus justifying our approach of treating interference as noise. This situation has been thoroughly investigated in the context of *unicast* transmission in the MISO UC-IC, where it is known that beamforming is the optimal transmit strategy and for which the Pareto-boundary is completely characterized in [10]–[13]. Yet, finding specific points on the Pareto-boundary of the MISO UC-IC, such as, the maximum weighted sum-rate point, is know to be NP-hard [14], which holds true for the more general MISO MC-IC as well. Still, algorithms for local and even global optimization of the weighted sum-rate of the MISO UC-IC are proposed in [15]–[17]. The MISO (and MIMO) UC-IC has also received attention in the context of coordinated multipoint transmission (CoMP) in the standardization of LTE, where interest has mostly focused on low-complexity transmission schemes, such as, signal to leakage and noise ratio (SLNR) beamforming [18] and multi-cell block diagonalization (BD) precoding [19]; see [20] for an overview of various spatial interference-mitigation techniques.

Transmit optimization for *single transmitter* (neglecting interference) multi-antenna physical-layer multicasting is investigated in [21], where beamforming solutions for quality-of-service constrained multicasting as well as max-min signal to interference and noise ratio (SINR) transmission are proposed. It is shown in [21] that these problems are NP-hard

Manuscript received February 17, 2015; revised July 3, 2015 and August 19, 2015; accepted October 22, 2015. Date of publication November 3, 2015; date of current version December 15, 2015. This work was supported by INWITE. The associate editor coordinating the review of this paper and approving it for publication was G. Abreu.

The authors are with the Institute of Telecommunications, Technische Universitaet (TU) Wien, Vienna 1040, Austria (e-mail: sschwarz@nt.tuwien.ac.at; mrupp@nt.tuwien.ac.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2015.2497231

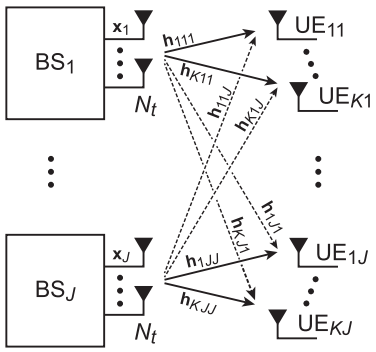


Fig. 1. Illustration of the MISO multicast interference channel with J base stations and K users per base station.

but can be approximately solved employing semi-definite relaxation [22], [23]. From this result it also follows that *beamforming is* in general *not optimal* for MISO multicasting. The authors of [21] also investigated the related problem of maximizing the minimum user throughput, while restricting the available transmit power. Considering multiple co-channel multicast groups, this formulation is extended in [24] to not only guaranteeing a minimum SNR to the served users but also upper-bounding the interference caused to other multicast groups. Furthermore, [25] additionally considered transmit antenna selection to reduce the number of radio frequency chains required at the transmitter. In [26], a stochastic transmit beamforming method is proposed to alleviate the performance loss caused by the rank-one beamforming approximation of semi-definite relaxation and a rank-two approximation based on the Alamouti space-time coding scheme is derived.

In contrast to the UC-IC, the MISO MC-IC has received comparatively little attention in literature. In [27], multi-cell beamforming is proposed based on statistical (long-term) CSIT. The minimum long-term SINR of all users of the network is maximized through joint optimization over all transmitters. This work is extended in [28] by deriving an equivalent quasi-convex form that can be solved efficiently. In contrast, in this paper we consider instantaneous CSIT in combination with distributed transmit optimization. A similar setup as in our article is considered in [29], where the authors propose decentralized algorithms for quality-of-service constrained beamforming as well as global max-min SINR beamforming in the MISO MC-IC. Similar to [11], the authors of [29] apply a sub-gradient projection optimization method to determine optimal solutions of the considered optimization problems. In contrast, in this paper we consider direct rate maximization applying *weighted sum-rate optimization* as well as *Pareto-improvement optimization*, providing the possibility to numerically determine rate regions of the MISO MC-IC and to calculate the input covariances to obtain any rate-tuple on the boundary of this region. In [30] multicast transmit and receive beamforming for the multiple-input multiple-output MC-IC with single-stream transmission is considered. The authors propose heuristic distributed algorithms to iteratively optimize transmit and receive filters. Another important issue related to physical layer multicasting is grouping and scheduling of multicast users. Since in wireless communications different users can experience strongly varying channel conditions, optimized grouping and scheduling of multicast users with similar channel quality can improve transmission efficiency [31], [32]. In this article we do not consider user grouping and scheduling. The proposed method is based on interference leakage control; similar optimization problems as considered here also appear in cognitive radio systems, where the interference leakage from secondary users to primary users has to obey specific leakage masks [33], [34]. In cognitive radio, however, leakage parameters are fixed in contrast to our work.

Contribution: We propose distributed transmit optimization for *weighted sum-rate maximization* as well as *Pareto-improvement optimization* in the MISO MC-IC. Since these problems are non-convex, we consider local optimization employing gradient-ascent algorithms. Our method is inspired by the interference leakage-based distributed optimization proposed in [11] for the MISO UC-IC. We formulate local leakage-based optimization problems for the MISO MC-IC that can be solved by each transmitter individually and hence facilitate distributed optimization; we prove that these problems are equivalent to joint weighted sum-rate maximization, provided the leakage parameters are appropriately chosen. Applying a gradient-ascent algorithm on the dual of the leakage-based optimization problems with respect to the leakage parameters similar to [29], we arrive at a distributed method that converges to a local optimum of weighted sum-rate maximization. We formulate and solve a similar optimization problem for *Pareto-improvement* of the achievable rate-tuple, reducing complexity and improving convergence compared to weighted sum-rate optimization. Furthermore, we enhance two established unicast CoMP transmission schemes to the multicast scenario, that is, multi-cell BD and SLNR beamforming.

Organization: In Section II, we introduce our system model and define the notion of achievable multicast transmission rate. We present problem formulations for Pareto-optimal transmission in the MISO MC-IC in Section III and discuss difficulties associated with finding solutions for these optimization problems; specifically, the presented problems are non-convex and we did not succeed finding equivalent convex forms. This motivates the application of local optimization techniques in Section IV, where we propose iterative distributed transmit optimization. To obtain this algorithm, we first derive the interference leakage-based local optimization problem that must be solved by each transmitter in each iteration of the algorithm with fixed leakage parameters. Then, in Section IV-A we present our dual gradient-based coordinated leakage parameter control to optimize the operating point of the network. Iterating between local optimization and coordinated leakage parameter control ultimately terminates in a locally Pareto-optimal solution as we show in Section IV-B. In Section V, we extend existing CoMP schemes that have been proposed for the UC-IC to the MC-IC. Finally, in Section VI we evaluate the performance of the proposed schemes by means of simulations. We provide concluding remarks in Section VII.

Notation: We denote the set of non-negative and positive reals as $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{+}$, respectively. We use bold lower-case letters $\mathbf{x}$ to represent vectors and bold upper-case letters $\mathbf{X}$ for

matrices. The transpose and conjugate transpose of matrix $\mathbf{X}$ are $\mathbf{X}^{\mathrm{T}}$ and $\mathbf{X}^{\mathrm{H}}$, respectively. With $[\mathbf{X}]_{:,a:b}$ we select columns a to b of matrix $\mathbf{X}$. The notation $\mathbf{X} \succeq 0$ identifies matrix $\mathbf{X}$ as positive semi-definite. We write element-wise vector inequalities as $\mathbf{x} \leq \mathbf{y}$. The operator $[\mathbf{x}]^+$ sets negative elements of $\mathbf{x}$ equal to zero. The expected value of random variable x is $\mathbb{E}(x)$. The circularly-symmetric complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathbf{C})$. The Kronecker-delta is δ_{ab}, that is, $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ otherwise.

II. System Model

We consider J wireless transmitters, e.g., base stations (BSs), that each multicast a common message to K user equipments (UEs) exploiting the broadcast nature of the physical layer of the wireless medium. The BSs are equipped with N_t antennas each (to simplify notation), whereas the UEs have only single antennas. Such a MISO MC-IC is illustrated in Figure 1. We denote user k attached to BS j as UE_{kj}. The input-output relationship of UE_{kj} is

$$y_{kj} = \mathbf{h}_{kjj}^{\mathrm{H}} \mathbf{x}_j + \sum_{\substack{\ell=1 \\ \ell \neq j}}^{J} \mathbf{h}_{kj\ell}^{\mathrm{H}} \mathbf{x}_\ell + n_{kj}. \tag{1}$$

Here, $\mathbf{h}_{kjj} \in \mathbb{C}^{N_t \times 1}$ denotes the MISO channel between UE_{kj} and its serving BS_j, whereas $\mathbf{h}_{kj\ell} \in \mathbb{C}^{N_t \times 1}$ represents the interfering channel with respect to BS_ℓ. We assume the additive noise to be complex Gaussian distributed: $n_{kj} \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma_n^2\right)$. The multicast transmit signal $\mathbf{x}_j \in \mathbb{C}^{N_t \times 1}$ of BS_j is intended for all K users $\mathrm{UE}_{1j}, \ldots, \mathrm{UE}_{Kj}$ attached to base station j.

Assuming that the users apply single-user detection, i.e., interfering signals are treated as additional noise, the achievable rate of UE_{kj} with proper Gaussian signalling is

$$R_{kj} = \log_2 \left(1 + \frac{\mathbf{h}_{kjj}^{\mathrm{H}} \mathbf{C}_j \mathbf{h}_{kjj}}{\sigma_n^2 + \sum_{\substack{\ell=1 \\ \ell \neq j}}^{J} \mathbf{h}_{kj\ell}^{\mathrm{H}} \mathbf{C}_\ell \mathbf{h}_{kj\ell}} \right), \tag{2}$$

with $\mathbf{C}_j = \mathbb{E}\left(\mathbf{x}_j \mathbf{x}_j^{\mathrm{H}}\right) \in \mathbb{C}^{N_t \times N_t}$ representing the covariance matrix of the input signal of BS_j.[1] As we consider multicasting of a common message to K users, the maximum *multicast transmission rate* R_j of BS_j is upper bounded by the minimum achievable rate of the attached users $\mathrm{UE}_{1j}, \ldots, \mathrm{UE}_{Kj}$ to ensure successful reception even by the weakest user

$$R_j = \min_{k \in \{1,\ldots,K\}} R_{kj}. \tag{3}$$

For the special case of a single transmitter $J = 1$ single-user detection is optimal and the capacity of the interference-free multicast channel is obtained as [36]

$$C_j = \max_{\mathbf{C}_j} \min_{k \in \{1,\ldots,K\}} \log_2 \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}_{kjj}^{\mathrm{H}} \mathbf{C}_j \mathbf{h}_{kjj} \right),$$
$$\text{subject to: } \mathbf{C}_j \in \mathbb{C}^{N_t \times N_t},\ \mathbf{C}_j \succeq 0,\ \mathrm{tr}\left(\mathbf{C}_j\right) \leq P_j, \tag{4}$$

[1] Notice that the achievable rate with improper Gaussian signalling reads differently [35]. The extension of the presented results to improper Gaussian input distributions is considered as promising future research direction.

with P_j denoting the transmit power of BS_j. This convex optimization problem can be solved efficiently and hence the capacity-achieving input covariance matrix can be calculated. We denote the corresponding solution as *max-min precoding*. Notice that Equation (4) also represents a semi-definite relaxation of the optimization problem for rate-maximizing multicast transmit beamforming [21]. For the general case $J > 1$ the capacity of the MISO MC-IC is unknown.

III. Pareto-Optimal Problem Formulations

To maximize the efficiency of J parallel multicast transmissions, the J corresponding BSs have to jointly determine optimal input covariance matrices

$$\max_{\{\mathbf{C}_1,\ldots,\mathbf{C}_J\}} [R_1, \ldots, R_J]^{\mathrm{T}}$$
$$= \max_{\{\mathbf{C}_1,\ldots,\mathbf{C}_J\}} \left[\min_{k \in \{1,\ldots,K\}} R_{k1}, \ldots, \min_{k \in \{1,\ldots,K\}} R_{kJ} \right]^{\mathrm{T}},$$
$$\text{subject to: } \mathbf{C}_j \succeq 0,\ \mathrm{tr}\left(\mathbf{C}_j\right) \leq P_j,\ \forall j \in \{1, \ldots, J\}. \tag{5}$$

Since these BSs are mutually interfering, in general there does not exist a single optimum achievable rate tuple $\left[R_1^*, \ldots, R_J^*\right]^{\mathrm{T}} \geq [R_1, \ldots, R_J]^{\mathrm{T}}$ that simultaneously maximizes the achievable multicast rates of all BSs (the inequality $\geq$ is component-wise). We rather have to resort to the concept of *Pareto-optimality* to specify optimal solutions of this multi-objective optimization problem. For the readers convenience we repeat the definition of Pareto-optimality [10]:

Definition III.1 (Pareto-optimal rate-tuples): A rate-tuple $[R_1, \ldots, R_J]^{\mathrm{T}}$ is Pareto-optimal if there exists no other rate-tuple $\left[R_1', \ldots, R_J'\right]^{\mathrm{T}} \geq [R_1, \ldots, R_J]^{\mathrm{T}}$ and $R_i' > R_i$ for at least one i.

If the condition in Definition III.1 holds only in an ϵ-neighbourhood of $[R_1, \ldots, R_J]^{\mathrm{T}}$, then the rate-tuple is called *locally Pareto-optimal*.

Pareto-optimal solutions of multi-objective optimization problems are commonly determined by means of scalarization techniques [37]. One widely-adopted scalarization approach in wireless communications is *weighted sum-rate optimization* [11], [17], [38], [39]

$$\max_{\{\mathbf{C}_1,\ldots,\mathbf{C}_J\}} \sum_{j=1}^{J} w_j R_j = \max_{\{\mathbf{C}_1,\ldots,\mathbf{C}_J\}} \sum_{j=1}^{J} \min_{k \in \{1,\ldots,K\}} w_j R_{kj},$$
$$\text{subject to: } \mathbf{C}_j \succeq 0,\ \mathrm{tr}\left(\mathbf{C}_j\right) \leq P_j,\ \forall j \in \{1, \ldots, J\}. \tag{6}$$

Here, the weights $w_j \geq 0$ determine the point achieved on the Pareto-boundary of the optimization problem. This problem, however, can be shown to be non-convex due to the mutual interference in-between base stations. In [12], weighted sum-rate optimization for the MISO *UC-IC* was successfully reformulated to an equivalent family of convex optimization problems, rendering beamforming as the optimal transmit strategy. Yet, an extension of this approach to the MISO *MC-IC* does not appear feasible, due to the additional minimization over the achievable rates of the multicast users and because beamforming is not optimal in the multicast case [36].

Another approach for determining Pareto-optimal rate-tuples is the *rate profile method* [40]. The formulation of this optimization problem for the MISO MC-IC reads as

$$\max_{\{\mathbf{C}_1,\ldots,\mathbf{C}_J\}} R_{\text{sum}},$$

subject to:

$$\begin{aligned}&\mathbf{C}_j \succeq 0,\ \operatorname{tr}\left(\mathbf{C}_j\right) \leq P_j,\ \ \forall j \in \{1,\ldots,J\},\\ &R_j = \min_{k\in\{1,\ldots,K\}} R_{kj} \geq \alpha_j R_{\text{sum}},\ \ \forall j \in \{1,\ldots,J\},\end{aligned} \quad (7)$$

with $\alpha_j \geq 0,\ \sum_{j=1}^{J} \alpha_j = 1$. Again, this problem is non-convex due to mutual interference between base stations. In case of the MISO UC-IC, a successful transformation of the non-convex rate profile problem to a convex second-order cone program is possible, exploiting the proven fact that beamforming is optimal in the unicast scenario [11], [17]. Unfortunately, this does not hold true for the MISO MC-IC, hindering a similar reformulation in the multicast setup.

Since convex optimization is not directly applicable to problems (6) and (7), determining Pareto-optimal input covariance matrices for the MISO MC-IC calls for a global optimization approach. An efficient framework for global optimization of the weighted sum-rate in the context of the MISO UC-IC is proposed in [17] based on monotonic optimization applying the outer polyblock approximation algorithm [41]. However, not even this framework is applicable to the MISO MC-IC, as it requires the solution of the rate profile (7) in each iteration of the algorithm. Furthermore, the cost function of (6) is not monotonic in the individual user rates R_{kj}; hence, monotonic optimization may not be applicable. Since we did not succeed in convexifying (6) or (7) we thus resort to local solution, which enables efficient distributed implementations as described in the following section.

IV. Distributed Transmit Optimization

In this section, we propose a distributed optimization algorithm to attain locally Pareto-optimal input covariance matrices, requiring minimal communication overhead in-between BSs to achieve coordination of the interfering multicast transmissions. We consider weighted sum-rate maximization below for finding locally Pareto-optimal solutions, although the same approach is also applicable to the rate profile method. The development of the algorithm follows ideas of [11], [29]. We first derive a set of local optimization problems that can be solved by each BS individually, by maximizing the achievable multicast rate of the BS while restricting the interference leakage caused to users of other BSs. We show that this set of problems is equivalent to weighted sum-rate maximization provided interference leakage parameters are appropriately chosen. We then propose coordinated interference leakage control for finding leakage parameters that optimize the performance of the system. Iterating between local optimization and coordinated leakage control ultimately terminates in a locally Pareto-optimal solution.

Due to the mutual interference terms in the denominator of the achievable user rate R_{kj} defined in (2), weighted sum-rate maximization according to (6) requires joint optimization over the set of input covariance matrices $\mathcal{C} = \{\mathbf{C}_1,\ldots,\mathbf{C}_J\}$

$$\mathcal{C}^{(p)} = \arg\max_{\mathcal{C},\mathbf{z}} \sum_{j=1}^{J} z_j,$$

subject to:

$$\begin{aligned}&z_j \leq w_j R_{kj},\ \forall k \in \{1,\ldots,K\},\ \forall j \in \{1,\ldots,J\},\\ &\mathbf{C}_j \succeq 0,\ \operatorname{tr}\left(\mathbf{C}_j\right) \leq P_j,\ \forall j \in \{1,\ldots,J\},\end{aligned} \quad (8)$$

where we introduced slack-variables $\mathbf{z} = [z_1,\ldots,z_J]^{\mathrm{T}}$. We next introduce upper bounds for the interference leakage terms

$$\Gamma_{kj\ell} \geq \operatorname{tr}\left(\mathbf{C}_\ell \mathbf{h}_{kj\ell}\mathbf{h}_{kj\ell}^{\mathrm{H}}\right) \geq 0, \quad (9)$$

$$\begin{aligned}&\forall k \in \{1,\ldots,K\}, \forall j \in \{1,\ldots,J\},\ \forall \ell \in \{1,\ldots,J\}\setminus j,\\ &\boldsymbol{\Gamma} = \big[\Gamma_{112},\ldots,\Gamma_{11J},\Gamma_{211},\ldots,\Gamma_{21J},\ldots,\\ &\qquad \Gamma_{KJ1},\ldots,\Gamma_{KJ(J-1)}\big]^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^{(J-1)KJ},\end{aligned} \quad (10)$$

leading to corresponding achievable rate and SINR lower bounds

$$R_{kj} \geq \bar{R}_{kj} = \log_2\left(1+\overline{\mathrm{SINR}}_{kj}\right), \quad (11)$$

$$\begin{aligned}\mathrm{SINR}_{kj} &= \frac{\operatorname{tr}\left(\mathbf{C}_j \mathbf{h}_{kjj}\mathbf{h}_{kjj}^{\mathrm{H}}\right)}{\sigma_n^2 + \sum_{\substack{\ell=1\\ \ell\neq j}}^{J} \operatorname{tr}\left(\mathbf{C}_\ell \mathbf{h}_{kj\ell}\mathbf{h}_{kj\ell}^{\mathrm{H}}\right)}\\ &\geq \overline{\mathrm{SINR}}_{kj} = \frac{\operatorname{tr}\left(\mathbf{C}_j \mathbf{h}_{kjj}\mathbf{h}_{kjj}^{\mathrm{H}}\right)}{\sigma_n^2 + \sum_{\substack{\ell=1\\ \ell\neq j}}^{J} \Gamma_{kj\ell}}.\end{aligned} \quad (12)$$

Utilizing these bounds, we define the *relaxed weighted sum-rate optimization* problem as

$$\mathcal{C}^{(r)} = \arg\max_{\mathcal{C},\boldsymbol{\Gamma},\mathbf{z}} \sum_{j=1}^{J} z_j,$$

subject to:

$$\begin{aligned}&z_j \leq w_j \bar{R}_{kj},\ \forall k \in \{1,\ldots,K\},\ \forall j \in \{1,\ldots,J\},\\ &\operatorname{tr}\left(\mathbf{C}_\ell \mathbf{h}_{kj\ell}\mathbf{h}_{kj\ell}^{\mathrm{H}}\right) \leq \Gamma_{kj\ell},\ \forall \Gamma_{kj\ell} \in \boldsymbol{\Gamma},\\ &\mathbf{C}_j \succeq 0,\ \operatorname{tr}\left(\mathbf{C}_j\right) \leq P_j,\ \forall j \in \{1,\ldots,J\}.\end{aligned} \quad (13)$$

Notice that problem (13) reduces to problem (8) in case the interference leakage constraints (9) are fulfilled with equality.

Joint optimization of (13) with respect to the input covariance matrices $\mathcal{C}$ and the leakage parameters $\boldsymbol{\Gamma}$ is non-convex. However, for *fixed leakage parameters* $\boldsymbol{\Gamma}$, optimization with respect to $\mathcal{C}$ decomposes into J independent problems that can be solved by each BS j individually

$$\left\{\mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma})\right\} = \arg\max_{\mathbf{C}_j\in\mathbb{C}^{N_t\times N_t}, z_j} z_j,$$

subject to:

$$\begin{aligned}&z_j \leq w_j \bar{R}_{kj},\ \forall k \in \{1,\ldots,K\},\\ &\mathbf{C}_j \succeq 0,\ \operatorname{tr}\left(\mathbf{C}_j\right) \leq P_j,\\ &\operatorname{tr}\left(\mathbf{C}_j \mathbf{h}_{i\ell j}\mathbf{h}_{i\ell j}^{\mathrm{H}}\right) \leq \Gamma_{i\ell j},\\ &\forall i \in \{1,\ldots,K\},\ \forall \ell \in \{1,\ldots,J\}\setminus j.\end{aligned} \quad (14)$$

Notice that BS_j is only impacted by the subset of $2K(J-1)$ leakage parameters $\Gamma_{kj\ell}$, $\forall k \in \{1,\ldots,K\}$, $\forall \ell \in \{1,\ldots,J\} \setminus j$ and $\Gamma_{i\ell j}$, $\forall i \in \{1,\ldots,K\}$, $\forall \ell \in \{1,\ldots,J\} \setminus j$. The set of *leakage-based multicast transmit optimization* problems (14) is convex and thus provides an efficient distributed way to solve problem (13) with fixed leakage parameters. We can establish the following theorem relating the distributed solution of (14) with the optimal point of (8):

Theorem IV.1: Consider the set $\mathcal{C}^{(p)} = \{\mathbf{C}_1^{(p)},\ldots,\mathbf{C}_J^{(p)}\}$ of weighted sum-rate optimal input covariance matrices solving (8) and their corresponding interference leakage powers

$$\Gamma_{i\ell j}^{(p)} = \mathrm{tr}\left(\mathbf{C}_j^{(p)}\mathbf{h}_{i\ell j}\mathbf{h}_{i\ell j}^{\mathrm{H}}\right), \quad (15)$$
$$\forall i \in \{1,\ldots,K\},\ \forall \ell \in \{1,\ldots,J\},\ \forall j \in \{1,\ldots,J\} \setminus \ell.$$

The distributed solution of the J problems (14) with leakage parameters $\Gamma_{i\ell j}$ set equal to the corresponding interference powers $\Gamma_{i\ell j}^{(p)}$ results in the weighted sum-rate optimal solution

$$\mathcal{C}^*\left(\mathbf{\Gamma}^{(p)}\right) = \left\{\mathbf{C}_1^*\left(\mathbf{\Gamma}^{(p)}\right),\ldots,\mathbf{C}_J^*\left(\mathbf{\Gamma}^{(p)}\right)\right\} = \mathcal{C}^{(p)}. \quad (16)$$

We provide the proof of this result in Appendix A.

Given the optimal solution $\mathcal{C}^*(\mathbf{\Gamma})$ of problem set (14) for fixed leakage parameters $\mathbf{\Gamma}$, optimizing (13) with respect to the leakage parameters while the input covariances are fixed to $\mathcal{C}^*(\mathbf{\Gamma})$ is still non-convex. We thus cannot expect to efficiently achieve global optimality of (13). Instead, we propose an iterative gradient-type algorithm in the following section to attain a local optimum of (13) with respect to $\mathbf{\Gamma}$.

A. Gradient-Based Leakage Parameter Control

The J leakage-based multicast transmit optimization problems (14) are convex semi-definite programs. Provided the parameters P_j and $\Gamma_{i\ell j}$, $\forall i \in \{1,\ldots,K\}$, $\forall \ell \in \{1,\ldots,J\} \setminus j$ in (14) are strictly positive, problem (14) is strictly feasible and thus has zero duality gap [37]. In case $\Gamma_{i\ell j} = 0$ for at least N_t leakage parameters, the only feasible solution is $\mathbf{C}_j = \mathbf{0}$, in which case strong duality cannot be guaranteed to hold. This situation, however, is not of interest for us, as it implies that BS_j cannot transmit at all. To improve the efficiency of the multicast network, we follow a similar gradient-based approach as proposed in [11], [29] to vary the leakage parameters smoothly along the gradients of the Lagrangian duals of problem set (14) with respect to $\mathbf{\Gamma}$. We provide details to this approach and derive the gradient of (14) in Appendix B.

a) Weighted sum-rate leakage update: Given the optimal solution $\left\{\mathbf{C}_j^*(\mathbf{\Gamma}), z_j^*(\mathbf{\Gamma})\right\}$ of (14) for fixed leakage parameters $\mathbf{\Gamma}$, we denote the gradient of the Lagrangian of (14) with respect to $\mathbf{\Gamma}$ at this optimal point as $\mathbf{g}_j$

$$\mathbf{g}_j = \nabla_{\mathbf{\Gamma}} L_j(\boldsymbol{\lambda}_j^*(\mathbf{\Gamma}), \mathbf{C}_j^*(\mathbf{\Gamma}), z_j^*(\mathbf{\Gamma}), \mathbf{\Gamma}) \in \mathbb{R}^{(J-1)KJ}. \quad (17)$$

Here, $L_j(\boldsymbol{\lambda}_j^*(\mathbf{\Gamma}), \mathbf{C}_j^*(\mathbf{\Gamma}), z_j^*(\mathbf{\Gamma}), \mathbf{\Gamma})$ represents the Lagrangian of problem (14) as a function of the leakage parameter vector $\mathbf{\Gamma}$, the optimal solutions $\left\{\mathbf{C}_j^*(\mathbf{\Gamma}), z_j^*(\mathbf{\Gamma})\right\}$ as well as the corresponding dual variables $\boldsymbol{\lambda}_j^*(\mathbf{\Gamma})$. The detailed definition of the Lagrangian is provided in Equation (46) of Appendix B and the calculation of $\mathbf{g}_j$ can be found in (53). The local gradient $\mathbf{g}_j$ specifies the direction of largest local improvement of the multicast rate of BS_j; however, a step in this direction will in general decrease the rate of other BSs. Hence, we have to find a global ascent direction that leads to improvement of the weighted sum-rate of all base stations. Since problem (13) with fixed $\mathbf{\Gamma}$ is a linear combination of problem set (14), the gradient of the relaxed weighted sum-rate optimization problem with respect to $\mathbf{\Gamma}$ is

$$\mathbf{g}_{\mathrm{WSR}} = \sum_{j=1}^{J} \mathbf{g}_j. \quad (18)$$

To improve the weighted sum-rate, we propose to update the leakage parameter vector $\mathbf{\Gamma}$ as

$$\tilde{\mathbf{\Gamma}} = \left[\mathbf{\Gamma} + \mu\, \mathbf{g}_{\mathrm{WSR}}\right]^+,\ \mu \geq 0. \quad (19)$$

Algorithm 1. Distributed algorithm for relaxed weighted sum-rate optimization.

Input: Initial leakage parameter vector $\mathbf{\Gamma}[0] \geq 0$; see Section IV-C for suggestions.
Weights $\{w_1,\ldots,w_J\}$ for weighted sum-rate calculation.
Result: Local optimum of the relaxed weighted sum-rate optimization problem (13).
Initialization: Distribute $\mathbf{\Gamma}[0]$ among BSs; notice each BS only requires a subset of $2K(J-1)$ parameters.
Distribute weights $\{w_1,\ldots,w_J\}$ among BSs.
Set the loop counter $n = 0$.
repeat
- Increase the loop counter: $n = n + 1$.
- Solve the leakage-based multicast transmit optimization problem (14) at each BS.
- Calculate the local gradient $\mathbf{g}_j$ according to (53) at each BS.
- Exchange gradient vectors $\mathbf{g}_j$ among base stations; notice each BS only requires a subset of $2K(J-1)$ values impacting the corresponding active leakage parameters of the respective BS.
- **while** *valid step size* $\mu[n]$ *not found* **do**
  - Apply a step size search algorithm (e.g., line-search) among the BSs to determine a valid step size $\mu[n]$ for (19).
  - Notice, each iteration of the search requires local solution of (14) with current step size and exchange of weighted sum-rates to evaluate the stopping criterion.
- **end**
- Update the leakage parameter vector $\mathbf{\Gamma}[n]$ according to (19) applying the valid step size $\mu[n]$

$$\mathbf{\Gamma}[n] = \left[\mathbf{\Gamma}[n-1] + \mu[n]\, \mathbf{g}_{\mathrm{WSR}}\right]^+,$$

- with $\mathbf{g}_{\mathrm{WSR}}$ according to (18).

until *convergence; see Section IV-B*;

Here, step size μ must be selected sufficiently small to ensure that the weighted sum-rate increases, e.g., by applying a line-search method [42], [43]. In (19) $\mathbf{g}_{\mathrm{WSR}}$ can alternatively be replaced with any other ascent direction $\mathbf{g}$ such that $\mathbf{g}_{\mathrm{WSR}}^{\mathrm{T}}\mathbf{g} > 0$; clever selection of $\mathbf{g}$ can improve convergence [42]. Iterating between the solution of problem set (14) and the update of the leakage parameter vector (19) provides a gradient-ascent method that is guaranteed to converge to a local optimum of (13), since the weighted sum-rate increases in each iteration and is ultimately upper bounded by the solution of (8). We present an algorithmic description of the proposed distributed local optimization of the relaxed weighted sum-rate problem (13) in Algorithm 1. The algorithm ultimately terminates in a local optimum of the relaxed weighted sum-rate optimization problem, which is also a local Pareto-optimum. The method requires exchange of the local gradients $\mathbf{g}_j$ at each iteration among BSs. Notice, as for each BS only a subset of leakage parameters is relevant, it is sufficient to exchange only a subset of the entries of $\mathbf{g}_j$; specifically, BS_j and BS_i have to exchange $2K$ values corresponding to leakage parameters $\Gamma_{kj\ell}, \forall k \in \{1, \ldots, K\}$ and $\Gamma_{i\ell j}, \forall i \in \{1, \ldots, K\}$. During the step search the BSs have to exchange their weighted rates to determine if a valid step size has been found.

b) Pareto-improvement leakage update: The leakage parameter update can be further simplified in terms of computational complexity as well as signalling overhead, if we consider finding any *locally Pareto-optimal* rate-tuple instead of the specific weighted sum-rate optimal tuple. To understand this, first of all notice that Pareto-optimality can be determined equivalently from the SINRs instead of the achievable rates, since the two are related monotonically. Hence, we can replace rate optimization as considered in problem (14) with SINR optimization, eliminating the complex *log*-expression from the optimization problem

$$\left\{\mathbf{C}_j^*(\mathbf{\Gamma}), z_j^*(\mathbf{\Gamma})\right\} = \underset{\mathbf{C}_j \in \mathbb{C}^{N_t \times N_t}, z_j}{\arg\max} \; z_j,$$

subject to:

$$\begin{aligned} & z_j \leq \overline{\mathrm{SINR}}_{kj}, \; \forall k \in \{1, \ldots, K\}, \\ & \mathbf{C}_j \succeq 0, \; \mathrm{tr}\left(\mathbf{C}_j\right) \leq P_j, \\ & \mathrm{tr}\left(\mathbf{C}_j \mathbf{h}_{i\ell j} \mathbf{h}_{i\ell j}^{\mathrm{H}}\right) \leq \Gamma_{i\ell j}, \\ & \forall i \in \{1, \ldots, K\}, \; \forall \ell \in \{1, \ldots, J\} \setminus j. \end{aligned} \tag{20}$$

We derive the corresponding local gradient $\boldsymbol{\gamma}_j$ of the Lagrangian dual of (20) in Appendix C. To improve the SINR of each BS, we have to find a valid ascent direction $\boldsymbol{\gamma}_{\mathrm{Par}}$ for the leakage parameters that results in positive inner product with each of the local gradients $\boldsymbol{\gamma}_j$, hence leading to local Pareto-improvement of the multicast rates. In the numerical experiment in Section VI, we pick a valid ascent direction by solving

$$\begin{aligned} & \max_{\boldsymbol{\gamma}_{\mathrm{Par}} \in \mathbb{R}^{(J-1)KJ}} \; \sum_{j=1}^{J} r_j \frac{\boldsymbol{\gamma}_j^{\mathrm{T}} \boldsymbol{\gamma}_{\mathrm{Par}}}{\|\boldsymbol{\gamma}_j\|}, \\ & \text{subject to:} \; \left\|\boldsymbol{\gamma}_{\mathrm{Par}}\right\|^2 \leq 1, \; \boldsymbol{\gamma}_j^{\mathrm{T}} \boldsymbol{\gamma}_{\mathrm{Par}} \geq 0, \; \forall j \end{aligned} \tag{21}$$

with arbitrary weights $0 \leq r_j \leq 1$. These weights can be utilized to impact the outcome of the distributed optimization in favour of specific BSs; i.e., a larger weight r_j for BS_j puts more emphasize on the rate gain achieved by this BS, whereas $r_j = 0$ means that the amount of rate gain achieved by BS_j is irrelevant. Notice, however, that it cannot happen that the rate of a BS decreases due to the constraint $\boldsymbol{\gamma}_j^{\mathrm{T}} \boldsymbol{\gamma}_{\mathrm{Par}} \geq 0$. Given the ascent direction $\boldsymbol{\gamma}_{\mathrm{Par}}$, each BS j then independently applies a step search algorithm to determine its optimal step size $\mu_j[n]$ for the leakage parameter update, that is, the step size that leads to largest improvement of the rate of BS_j. Finally, the BSs exchange their optimal step sizes and employ the minimum step size of all BSs to ensure SINR enhancement for each BS

$$\mathbf{\Gamma}[n] = \left[\mathbf{\Gamma}[n-1] + \min_{j \in \{1, \ldots, J\}} \mu_j[n] \, \boldsymbol{\gamma}_{\mathrm{Par}}\right]^{+}, \tag{22}$$

thus resulting in *Pareto-improvement*. The signalling overhead during step search is reduced compared to Algorithm 1, since information exchange in-between BSs is required only once. Notice that infeasibility of (21) implies local Pareto-optimality of the solution of (20), since then it is not possible to find a valid ascent direction that simultaneously improves the rate of all BSs; hence, if this occurs the iteration can be stopped.

B. Convergence Criteria

Both iterative gradient-based leakage parameter updates described in Section IV-A are guaranteed to convergence to locally Pareto-optimal rate-tuples, since they improve in every step (provided the step size is appropriately chosen) and their optima are upper bounded. To assess convergence of the weighted sum-rate leakage update, the norm of the weighted sum-rate gradient (18) can be examined; at a local optimum of (13) the weighted sum-rate gradient (18) vanishes. Practically, the iteration can be stopped as soon as the norm of (18) falls below a small threshold.

Convergence of the Pareto-improvement leakage update is achieved as soon as no valid ascent direction can be found, i.e., in case it is not possible to improve the rate of all BSs at the same time. Hence, convergence can be determined from the feasibility problem

$$\max_{\boldsymbol{\gamma} \in \mathbb{R}^{(J-1)KJ}} 1, \quad \text{subject to:} \; \boldsymbol{\gamma}_j^{\mathrm{T}} \boldsymbol{\gamma} \geq 0. \tag{23}$$

This criterion, however, is prone to numerical inaccuracies. Consider for example the case with $J = 2$ base stations. Here, local Pareto-optimality implies that the local gradients $\boldsymbol{\gamma}_j, \; j \in \{1, 2\}$ are antipodal; already slightest deviations, e.g., due to rounding errors, result in feasibility of (23). A numerically more robust practical criterion is obtained by solving

$$\max_{\boldsymbol{\gamma} \in \mathbb{R}^{(J-1)KJ}} \sum \boldsymbol{\gamma}_j^{\mathrm{T}} \boldsymbol{\gamma}, \quad \text{subject to:} \; \|\boldsymbol{\gamma}\|^2 \leq 1, \; \boldsymbol{\gamma}_j^{\mathrm{T}} \boldsymbol{\gamma} \geq 0. \tag{24}$$

The iteration can be stopped as soon as the optimal value of (24) falls below a small threshold.

C. Initialization of Leakage Parameters

In the numerical experiment in Section VI, we demonstrate that convergence of the leakage parameter update can be very slow in case the number of users is large. A smart initialization of the leakage parameters that brings the initial solution of problem set (14) close to the border of the rate region achievable with iterative leakage parameter update is thus advantageous. We propose a heuristically optimized leakage parameter initialization below.

Initially, the input covariance matrices of all BSs are unknown. Assuming isotropic input covariances, that is, $\mathbf{C}_j = P_j/N_t\,\mathbf{I}_{N_t}$, the expected SINR of UE_{kj} is

$$\mathbb{E}\left(\mathrm{SINR}_{kj}\right) = \frac{\frac{P_j}{N_t}\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{h}_{kjj}}{\sigma_n^2 + \sum\limits_{\ell\neq j}\frac{P_\ell}{N_t}\mathbf{h}_{kj\ell}^{\mathrm{H}}\mathbf{h}_{kj\ell}} \geq \frac{\frac{P_j}{N_t}\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{h}_{kjj}}{\sigma_n^2 + \sum\limits_{\ell\neq j}\Gamma_{kj\ell}}. \quad (25)$$

Since the performance of the multicast transmission is determined by the weakest of the K UEs attached to BS_j, we propose to initialize the leakage parameters of this UE such that the total interference incurred is proportional to the noise

$$\sum_{\substack{\ell=1\\ \ell\neq j}}^{J}\Gamma_{k^*j\ell} = \alpha_j\,\sigma_n^2, \quad k^* = \underset{k\in\{1,\ldots,K\}}{\arg\min}\ \mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{h}_{kjj}. \quad (26)$$

To distribute the interference among leakage parameters $\Gamma_{k^*j\ell}$, we choose each parameter proportional to the respective channel gain

$$\begin{aligned}&\Gamma_{k^*j\ell} = c\,\mathbf{h}_{k^*j\ell}^{\mathrm{H}}\mathbf{h}_{k^*j\ell}\\ \Rightarrow\ &\Gamma_{k^*j\ell} = \alpha_j\sigma_n^2\frac{\mathbf{h}_{k^*j\ell}^{\mathrm{H}}\mathbf{h}_{k^*j\ell}}{\sum_{n\neq j}\mathbf{h}_{k^*jn}^{\mathrm{H}}\mathbf{h}_{k^*jn}},\ \forall\ell\in\{1,\ldots,J\}\setminus j.\end{aligned} \quad (27)$$

Furthermore, we calculate the leakage parameters of the remaining users such that the same expected SINR lower bound is obtained

$$\begin{aligned}&\frac{\frac{P_j}{N_t}\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{h}_{kjj}}{\sigma_n^2 + \sum\limits_{\ell\neq j}\Gamma_{kj\ell}} \overset{!}{=} \frac{\frac{P_j}{N_t}\mathbf{h}_{k^*jj}^{\mathrm{H}}\mathbf{h}_{k^*jj}}{\sigma_n^2(1+\alpha_j)}\\ &\Rightarrow\ \sum_{\ell\neq j}\Gamma_{kj\ell} = \frac{\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{h}_{kjj}}{\mathbf{h}_{k^*jj}^{\mathrm{H}}\mathbf{h}_{k^*jj}}\sigma_n^2(1+\alpha_j) - \sigma_n^2,\end{aligned} \quad (28)$$

which is non-negative since $\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{h}_{kjj} \geq \mathbf{h}_{k^*jj}^{\mathrm{H}}\mathbf{h}_{k^*jj}$. Setting leakage parameters proportional to channel gains, we obtain

$$\Gamma_{kj\ell} = \frac{\mathbf{h}_{kj\ell}^{\mathrm{H}}\mathbf{h}_{kj\ell}}{\sum_{n\neq j}\mathbf{h}_{kjn}^{\mathrm{H}}\mathbf{h}_{kjn}}\left(\frac{\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{h}_{kjj}}{\mathbf{h}_{k^*jj}^{\mathrm{H}}\mathbf{h}_{k^*jj}}\sigma_n^2(1+\alpha_j) - \sigma_n^2\right), \quad (29)$$

$$\forall k\neq k^*,\ \ \forall\ell\in\{1,\ldots,J\}\setminus j.$$

D. Comments on Channel State Information

The proposed transmission methods require instantaneous CSIT to determine the input covariance matrices at the BSs according to (14), respectively (20). More specifically, each BS_j has to know the local channel vectors $\mathbf{h}_{i\ell j}$, $\forall i\in\{1,\ldots,K\}$, $\forall\ell\in\{1,\ldots,J\}$. In currently prevailing frequency division duplex wireless communications, CSIT is commonly acquired over limited capacity feedback links from the UEs. In this case, efficient CSI quantization is decisive for obtaining accurate CSIT with minimal feedback overhead. For the considered optimization problems a gain-shape vector quantization approach similar to [44] lends itself, independently quantizing magnitude $m_{i\ell j}\in\mathbb{R}_+$ and direction $\mathbf{d}_{i\ell j}$, $\left\|\mathbf{d}_{i\ell j}\right\| = 1$ of each channel vector

$$\mathbf{h}_{i\ell j} = \left\|\mathbf{h}_{i\ell j}\right\|\frac{\mathbf{h}_{i\ell j}}{\left\|\mathbf{h}_{i\ell j}\right\|} = m_{i\ell j}\mathbf{d}_{i\ell j}. \quad (30)$$

Notice, since channel vectors always appear in the form $\mathbf{h}_{i\ell j}\mathbf{h}_{i\ell j}^{\mathrm{H}}$ in (14) and (20), the absolute phase of $\mathbf{d}_{i\ell j}$ is irrelevant and hence Grassmannian quantization is applicable [45], [46].

V. Extension of Existing Transmission Schemes

In this section, we extend two spatial inter-cell interference mitigation and coordination schemes that have been proposed for the UC-IC to the MISO MC-IC. These methods serve as performance benchmarks for comparison with the proposed distributed algorithm in Section VI.

A. Block Diagonal Multicasting

We first propose a simple multicast enhancement of the multi-cell BD technique, which has been introduced for unicast traffic in [19]. Linear BD precoding enables perfect inter-cell interference mitigation [47], provided the feasibility condition $N_t > K(J-1)$ is satisfied. This feasibility condition ensures that each BS can cancel the interference to all UEs not attached to that BS, while leaving at least one degree of freedom for serving the UEs attached to the BS. Specifically, the precoding matrix $\mathbf{F}_j$ employed at BS_j has to satisfy

$$\mathbf{h}_{i\ell j}^{\mathrm{H}}\mathbf{F}_j \overset{!}{=} \mathbf{0},\ \forall i\in\{1,\ldots,K\},\ \forall\ell\in\{1,\ldots,J\}\setminus j. \quad (31)$$

The precoder $\mathbf{F}_j$ is related to the transmit signal $\mathbf{x}_j$ via $\mathbf{x}_j = \mathbf{F}_j\mathbf{s}_j$, with $\mathbf{s}_j$ representing the information symbols; hence, $\mathbf{C}_j = \mathbf{F}_j\mathbf{F}_j^{\mathrm{H}}$. The requirement in Equation (31) can be satisfied by restricting $\mathbf{F}_j$ to the left null-space of the concatenated interference channels

$$\mathbf{F}_j \in \mathrm{null}\left(\bar{\mathbf{H}}_j\right) \Longleftrightarrow \bar{\mathbf{H}}_j^{\mathrm{H}}\mathbf{F}_j = \mathbf{0}, \quad (32)$$

$$\begin{aligned}\bar{\mathbf{H}}_j = \big[&\mathbf{h}_{11j},\ldots,\mathbf{h}_{K1j},\mathbf{h}_{12j},\ldots,\mathbf{h}_{K2j},\ldots,\mathbf{h}_{K(j-1)j},\\ &\mathbf{h}_{1(j+1)j},\ldots,\mathbf{h}_{K(j+1)j},\ldots,\mathbf{h}_{1Jj},\ldots,\mathbf{h}_{KJj}\big]\in\mathbb{C}^{N_t\times K(J-1)}.\end{aligned} \quad (33)$$

By applying a singular value decomposition to matrix $\bar{\mathbf{H}}_j$ we obtain

$$\tilde{\mathbf{F}}_j = [\mathbf{U}]_{:,N_t-K(J-1)+1:N_t}, \quad \bar{\mathbf{H}}_j = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}. \tag{34}$$

with $\tilde{\mathbf{F}}_j$ representing an orthonormal basis for the left null-space of $\bar{\mathbf{H}}_j$. Defining $\mathbf{F}_j = \tilde{\mathbf{F}}_j\mathbf{Q}_j$, $\mathbf{Q}_j \in \mathbb{C}^{N_t-K(J-1)\times N_t-K(J-1)}$ assures that the zero interference requirement is fulfilled. Assuming that all BSs apply such a strategy, interference-free operation of the network is achieved. Matrix $\mathbf{Q}_j$ can then be selected to optimize the interference-free achievable multicast transmission rate

$$\mathbf{Q}_j^* = \arg\max_{\mathbf{Q}_j} \min_{k\in\{1,\ldots,K\}} R_{kj}^{(\mathrm{if})},$$

subject to:

$$\begin{aligned}&\mathrm{tr}\left(\mathbf{Q}_j\mathbf{Q}_j^{\mathrm{H}}\right) \leq P_j,\\ &R_{kj}^{(\mathrm{if})} = \log_2\left(1+\frac{1}{\sigma_n^2}\mathbf{h}_{kjj}^{\mathrm{H}}\tilde{\mathbf{F}}_j\mathbf{Q}_j\mathbf{Q}_j^{\mathrm{H}}\tilde{\mathbf{F}}_j^{\mathrm{H}}\mathbf{h}_{kjj}\right).\end{aligned} \tag{35}$$

This convex optimization problem can be solved efficiently.

B. Max-Min SLNR Precoding

A popular beamforming strategy for the MISO UC-IC, requiring only local CSIT, maximizes the so-called SLNR [18] or equivalently the virtual SINR [48]. These methods have been shown to achieve Pareto-optimality in the case of two unicast links ($J = 2$, $K = 1$) [48]. In [29], SLNR beamforming has been extended to the multicast channel as

$$\begin{aligned}\mathbf{f}_j^* &= \arg\max_{\mathbf{f}_j\in\mathbb{C}^{N_t\times 1},\,\|\mathbf{f}_j\|^2=P_j} \sum_{k=1}^{K}\mathrm{SLNR}_{kj},\\ \mathrm{SLNR}_{kj} &= \frac{\mathbf{f}_j^{\mathrm{H}}\mathbf{h}_{kjj}\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{f}_j}{\sigma_n^2+\sum_{\substack{\ell=1\\\ell\neq j}}^{J}\sum_{i=1}^{K}\mathbf{f}_j^{\mathrm{H}}\mathbf{h}_{i\ell j}\mathbf{h}_{i\ell j}^{\mathrm{H}}\mathbf{f}_j},\end{aligned} \tag{36}$$

that is, maximizing the sum SLNR of all multicast users. We hence denote this method *sum SLNR* beamforming. As demonstrated in the simulations in Section VI, a better but also more computationally complex approach is derived from maximizing the minimum SLNR

$$\mathbf{f}_j^* = \arg\max_{\mathbf{f}_j\in\mathbb{C}^{N_t\times 1},\,\|\mathbf{f}_j\|^2=P_j} \min_{k\in\{1,\ldots,K\}} \mathrm{SLNR}_{kj}. \tag{37}$$

This optimization problem, however, is non-convex. Let us introduce the substitution

$$\tilde{\mathbf{f}}_j = \mathbf{L}_j^{\frac{1}{2}}\mathbf{f}_j, \quad \mathbf{L}_j = \frac{\sigma_n^2}{P_j}+\sum_{\substack{\ell=1\\\ell\neq j}}^{J}\sum_{i=1}^{K}\mathbf{h}_{i\ell j}\mathbf{h}_{i\ell j}^{\mathrm{H}}, \tag{38}$$

leading to the equivalent problem formulation

$$\begin{aligned}\tilde{\mathbf{f}}_j^* = \arg\max_{\tilde{\mathbf{f}}_j\in\mathbb{C}^{N_t\times 1}} \min_{k\in\{1,\ldots,K\}} &\frac{\mathrm{tr}\left(\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j^{\mathrm{H}}\mathbf{L}_j^{-\frac{\mathrm{H}}{2}}\mathbf{h}_{kjj}\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{L}_j^{-\frac{1}{2}}\right)}{\tilde{\mathbf{f}}_j^{\mathrm{H}}\tilde{\mathbf{f}}_j},\\ \text{subject to: } &\mathrm{tr}\left(\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j^{\mathrm{H}}\mathbf{L}_j^{-1}\right) = P_j,\end{aligned} \tag{39}$$

which is still non-convex due to the transmit power constraint and the restriction to beamforming. To obtain a convex optimization problem, we thus allow for precoding, i.e., we apply the semi-definite relaxation $\tilde{\mathbf{C}}_j = \tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j^{\mathrm{H}}$ without imposing a rank constraint on $\tilde{\mathbf{C}}_j$

$$\begin{aligned}\tilde{\mathbf{C}}_j^* = \arg\max_{\tilde{\mathbf{C}}_j\in\mathbb{C}^{N_t\times N_t}} \min_{k\in\{1,\ldots,K\}} &\frac{\mathrm{tr}\left(\tilde{\mathbf{C}}_j\,\mathbf{L}_j^{-\frac{\mathrm{H}}{2}}\mathbf{h}_{kjj}\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{L}_j^{-\frac{1}{2}}\right)}{\mathrm{tr}\left(\tilde{\mathbf{C}}_j\right)},\\ \text{subject to: } &\mathrm{tr}\left(\tilde{\mathbf{C}}_j\mathbf{L}_j^{-1}\right) = P_j.\end{aligned} \tag{40}$$

Relaxing the non-convex power constraint leads to the convex optimization problem

$$\bar{\mathbf{C}}_j^* = \arg\max_{\bar{\mathbf{C}}_j\in\mathbb{C}^{N_t\times N_t}} \min_{k\in\{1,\ldots,K\}} \frac{\mathrm{tr}\left(\bar{\mathbf{C}}_j\,\mathbf{L}_j^{-\frac{\mathrm{H}}{2}}\mathbf{h}_{kjj}\mathbf{h}_{kjj}^{\mathrm{H}}\mathbf{L}_j^{-\frac{1}{2}}\right)}{\mathrm{tr}\left(\bar{\mathbf{C}}_j\right)}, \tag{41}$$

which can be solved efficiently. As shown in Appendix D, with the post-normalization

$$\tilde{\mathbf{C}}_j^* = P_j\frac{\bar{\mathbf{C}}_j^*}{\mathrm{tr}\left(\bar{\mathbf{C}}_j^*\mathbf{L}_j^{-1}\right)}, \tag{42}$$

$\tilde{\mathbf{C}}_j^*$ solves (40). Back-substituting according to (38), we obtain the *max-min SLNR* solution

$$\mathbf{C}_j = \mathbf{L}_j^{-\frac{\mathrm{H}}{2}}\tilde{\mathbf{C}}_j^*\mathbf{L}_j^{-\frac{1}{2}}. \tag{43}$$

VI. Numerical Examples

In this section, we conduct Monte-Carlo simulations to evaluate the performance of the physical layer multicast transmission schemes proposed in Sections IV and V. We consider $J = 2$ BSs, each equipped with $N_t = 8$ transmit antennas and multicasting a common message to $K \in \{4, 6, 7\}$ UEs. Channels are i.i.d. Rayleigh fading $\mathbf{h}_{kj\ell} \in \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \mathbf{I}_{N_t}\right)$, $\forall k, j, \ell$

$$\begin{aligned}&\mathbb{E}\left(\mathbf{h}_{kj\ell}\mathbf{h}_{inm}^{\mathrm{H}}\right) = \delta_{ki}\delta_{jn}\delta_{\ell m}\mathbf{I}_{N_t},\\ &\forall k, i \in \{1,\ldots,K\}, \forall j, \ell, n, m \in \{1,\ldots,J\},\end{aligned}$$

with δ_{ab} denoting the Kronecker-delta. Hence, without precoding each UE receives the signals from all BSs equally strong on average. The average SNR is $P_j/\sigma_n^2 = 10\,\mathrm{dB}$. For reasons of complexity, we simulate only a single randomly drawn set of channel realizations.[2]

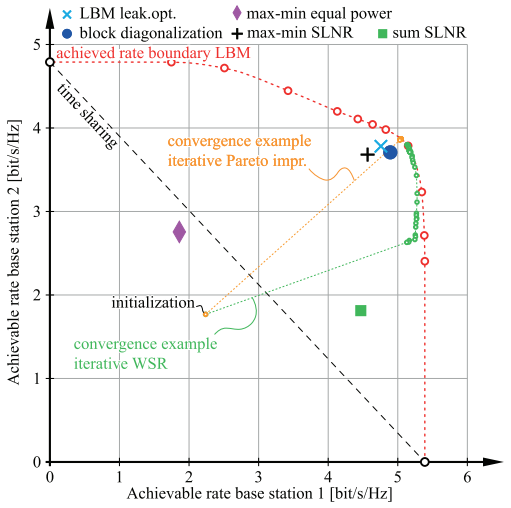
We present the rate region achieved (without time-sharing) with the proposed *leakage-based multicast transmit optimization* with *iterative gradient-ascent leakage parameter control* according to Section IV-A. The achieved rate region is obtained by iterating until convergence between (20) and the *Pareto-improvement* leakage update (21); different points on the boundary of the achieved rate region are attained by varying

[2]We have simulated several independent channel realizations during our investigations showing very similar trends as presented here. To simplify reproduction of the results, we provide our *MATLAB* code online at [49].

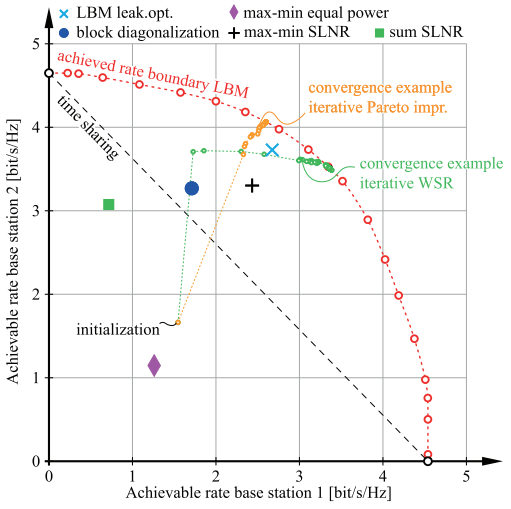
the weights r_j in (21). We furthermore determine the sum-rate optimal transmission point of the achieved rate region by applying the *weighted sum-rate* leakage update described in Algorithm 1 with equal weights $w_j = 1$. The optimal step size is obtained with a simple backtracking line-search applying the Armijo termination criterion [50]. We compare the performance of these iterative algorithms to the methods presented in Section V, to interference unaware *max-min precoding* according to (4) as well as to *leakage-based multicasting* (14) with fixed leakage parameters calculated as proposed in Section IV-C setting $\alpha_j = 1$; notice that each of these methods realizes one specific point in the rate region, i.e., they do not enable varying the operating point of the network.

Figure 2 shows the results when $K = 4$, $K = 6$ and $K = 7$ UEs are served per BS. Notice that different channel realizations are effective for different number of UEs; hence, the results should only be cross-compared qualitatively and not quantitatively among Figures 2a–2c. In the figure, markers represent actual simulation outcomes, whereas dashed lines are linear interpolations. We summarize our observations:

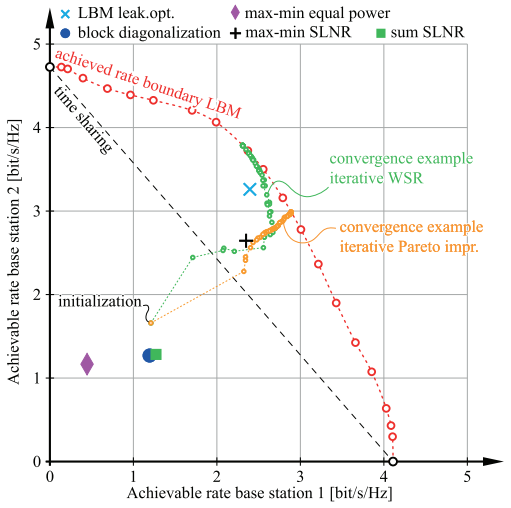
- If only one BS transmits at a time, *max-min precoding* according to (4) is optimal; this situation corresponds to the achievable rates obtained on the axes of Figs. 2a, 2b, 2c. However, as soon as both BSs transmit at the same time with equal power, *max-min precoding* (magenta-diamonds) performs much worse than time-sharing (black-dashed).
- *Leakage-based multicasting* (LBM) with iterative leakage update (red-dashed) achieves a rate region that is substantially larger than time-sharing and that coincides with max-min precoding if only a single BS transmits. With growing number of multicast users the achieved rate region shrinks, since more and more leakage constraints must be satisfied, leaving less degrees of freedom to maximize the signal power of intended users.
- *Sum SLNR* beamforming (36) (green-squares) as proposed in [29] provides a gain over time-sharing only with a small number of multicast users. As soon as the total number of users $K \cdot J$ is larger than the number of antennas per BS N_t, the method is outperformed by time-sharing.
- *Block diagonal multicasting* (blue-circles) as presented in Section V-A outperforms *sum SLNR* beamforming and operates close to the achieved rate region of our iterative schemes when K is small. Yet, for large number of multicast users it performs worse than time-sharing. Notice in all considered cases the BSs have enough transmit antennas to eliminate the interference caused to all other UEs, that is, block diagonalization is feasible. However, with growing number of UEs the power of the intended signal diminishes.
- *Max-min SLNR* precoding (black-pluses) as proposed in Section V-B provides a gain over *block diagonal multicasting* for $K \in \{6, 7\}$; yet with $K = 4$, block diagonalization is slightly better. Since the method is less restrictive in terms of interference than block diagonalization, it can



(a) $K = 4$ users per base station.



(b) $K = 6$ users per base station.



(c) $K = 7$ users per base station.

Fig. 2. Achievable multicast rates of $J = 2$ base stations each equipped with $N_t = 8$ antennas serving $K \in \{4, 6, 7\}$ users.

obtain a favourable trade-off between minimization of interference leakage and maximization of intended signal power. Also, *max-min SLNR* precoding can be utilized for efficient initialization of iterative LBM, enabling fast convergence of the algorithm.

- Even better performance than with *max-min SLNR* precoding is achieved by *leakage based multicasting* with fixed leakage parameters that are set as proposed in Section IV-C (blue-crosses). This method operates not too far from the rate boundary achieved with the iterative scheme even when $K = 7$, while requiring leakage parameter exchange among BSs only once after initialization.
- To demonstrate the convergence behaviour of the iterative schemes, we plot example rate traces of the *weighted sum-rate* (green-dashed) and the *Pareto-improvement* (orange-dashed) leakage updates. Notice that the algorithms have been initialized with all leakage parameters set equal to one; the rate achieved with this initialization is denoted by *initialization*. It falls below time-sharing in all cases, since interference is very high. The *Pareto-improvement* update rule converges uniformly in the sum-rate and the rate of each BS. It converges very quickly when K is small in comparison to N_t (with $K = 4$ virtually within 1 step); however, with growing number of users it becomes more and more difficult to find an ascent direction that significantly improves the rate of all UEs at the same time and thus convergence gets very slow. The *weighted sum-rate* method exhibits worse convergence behaviour when K is small, since it has to find the specific weighted sum-rate optimal point on the boundary of the rate region and not just any point as *Pareto-improvement*.

Since we did not succeed in reformulating problem (13) to a convex optimization problem, we cannot ascertain that the achieved rate regions shown in Figure 2 coincide with the global Pareto-optimal front of the multi-objective optimization (5); we can only claim local Pareto-optimality. Still, multiple iterative solutions with different random initial leakage parameters always terminated in the same region; also, both leakage-update rules proposed in Section IV-A terminated on the same boundary, thus indicating global Pareto-optimality.

As soon as the number of users is increased to $K \geq N_t = 8$, it is not possible to keep interference leakage low while at the same time providing good signal quality to the intended users. More generally, when all BSs are received equally strong, performance of the presented multicasting schemes falls below time-sharing when the number of leakage constraints per BS $K(J-1)$ is larger than the number of transmit antennas N_t. This is because the number of nulls that can be formed with a given transmit antenna array is limited by the number of antenna elements N_t; hence, if there are more than N_t users at other BSs we cannot minimize the interference to all of them at the same time. The condition $N_t > K(J-1)$ is necessary for *block diagonalization* precoding to exist. For all other methods, we can still obtain precoders even if this condition is violated; however, performance degrades quickly with growing number of interference leakage constraints. In practice, the intended BS will in most situations be received stronger than the interfering BSs due to pathloss differences; then it does not matter so much if a UE is hit by the transmit beam of an interfering BS, since the power is anyway low. We consider such situations in the following simulation below. We, furthermore, plan system level simulations as future work to investigate the performance of the proposed algorithms in realistic cellular networks. Also, in case UEs are equipped with multiple receive antennas, it is possible to further increase the number of users while still outperforming time-sharing, since UEs can eliminate interference partly. We have investigated this situation in [30], though with fixed leakage parameters. In this case, additional iterations between transmitter and receiver filters are required to achieve alignment of the interference from multiple interfering BSs.

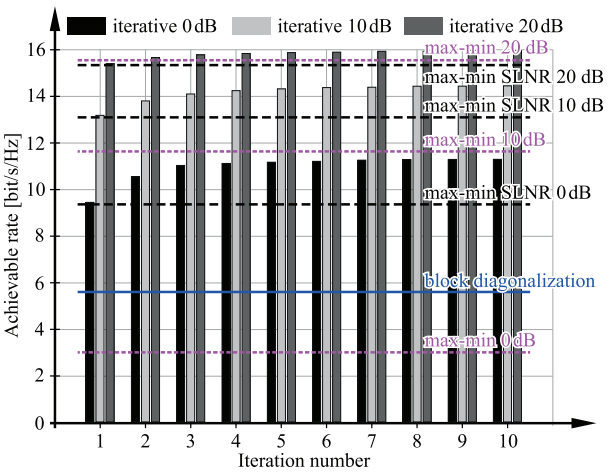


Fig. 3. Convergence behaviour of iterative LBM with $J = 3$ BSs each serving $K = 11$ UEs over $N_t = 24$ transmit antennas.

We next investigate the convergence behaviour of *leakage-based multicasting* with *Pareto-improvement* leakage update with $J = 3$ BSs each serving $K = 11$ UEs over $N_t = 24$ transmit antennas. Thus, the total number of interference leakage constraints per BS is $K(J-1) = 22$, leaving two degrees of freedom to optimize the intended signal power. We assume a signal to interference ratio (SIR) without precoding of $\{0, 10, 20\}$ dB; that is, the intended BS is received on average with unit power $\mathbb{E}\left(\mathbf{h}_{kjj}\mathbf{h}_{kjj}^{\mathrm{H}}\right) = \mathbf{I}_{N_t}$, $\forall k, j$, whereas the interfering BSs are received correspondingly weaker $\mathbb{E}\left(\mathbf{h}_{kj\ell}\mathbf{h}_{kj\ell}^{\mathrm{H}}\right) = \frac{1}{J-1}10^{-\mathrm{SIR}}\,\mathbf{I}_{N_t}, \forall k, j, \ell \neq j$. For reasons of complexity, we cannot simulate the full three-dimensional rate region for this case. We average results over 10 statistically independent Monte Carlo runs.

In Figure 3, we demonstrate the convergence behaviour of the *Pareto-improvement* leakage update for this situation and compare with *block diagonal multicasting*, *max-min* precoding, as well as *max-min SLNR* precoding. We initialize the leakage parameters of *iterative LBM* according to the interference leakage obtained with *max-min SLNR* precoding, which hence determines the performance after the first iteration. We observe that *iterative LBM* practically converges within five iterations in the considered scenario and that it is able to outperform the other methods irrespective of the SIR; the gain compared to the other methods, however, is largest at low SIR.

The rate of *block diagonalization* is independent of the SIR, since the applied precoder perfectly eliminates the interference in all cases. As expected, *max-min* precoding, which does not account for interference, performs well only if the SIR is high; more specifically, with SIR $\rightarrow \infty$, *max-min* precoding becomes optimal. This also explains why *iterative LBM* cannot improve much at SIR $= 20$ dB, because *max-min* precoding is already close to optimal. Hence, in such circumstances coordination of the BSs does not pay off.

VII. Conclusion

In this paper, we have proposed two gradient-based distributed transmit optimization algorithms for the multiple-input single-output multicast interference channel. Both algorithms can achieve locally Pareto-optimal transmission rate-tuples, yet global optimality cannot be guaranteed. We have provided heuristic initialization methods for the proposed iterative algorithms, enabling fast convergence or, alternatively, close-to-optimal operation without any iterations. We have investigated rate regions that can be achieved by the proposed methods in numerical experiments and demonstrated the superiority of the proposed methods compared to multicast enhancements of existing coordinated multipoint transmission schemes. Future work includes the extension of the proposed methods to multiple-input multiple-output transmission, the investigation of the robustness with respect to channel state information imperfections at the transmitters as well as the evaluation under more realistic conditions by means of large-scale system level simulations. The proposed physical layer multicasting techniques will become of interest for future releases of LTE, as soon as MBMS is extended to support multiple antennas. In this context, another topic for future work is joint precoding over multiple base stations to enhance multi-antenna multicasting in LTE MBSFN.

Acknowledgments

The authors appreciate the support of the LTE research group of the Institute of Telecommunications, Technische Universität Wien. The authors are thankful for valuable comments of Dr. Tal Philosof, Wireless Enablers Lab, General Motors R&D ATCI.

Appendix A
Proof of Theorem IV.1

It can be shown that the distributed solution of the set of J optimization problems (14) is optimal for the weighted sum-rate problem (8), provided the leakage parameters are set correctly, and hence the distributed solution results in a Pareto-optimal solution of problem (5).

For the proof, we consider the optimal rate-tuple $[R_1^{(p)}, \ldots, R_J^{(p)}]^{\mathrm{T}}$ of the weighted sum-rate optimization problem (8) with corresponding input covariance matrices $\mathcal{C}^{(p)} = \{\mathbf{C}_1^{(p)}, \ldots, \mathbf{C}_J^{(p)}\}$ and interference leakage powers

$$\Gamma_{i\ell j}^{(p)} = \operatorname{tr}\left(\mathbf{C}_j^{(p)} \mathbf{h}_{i\ell j} \mathbf{h}_{i\ell j}^{\mathrm{H}}\right),$$
$$\forall i \in \{1, \ldots, K\},\ \forall \ell \in \{1, \ldots, J\},\ \forall j \in \{1, \ldots, J\} \setminus \ell.$$

Provided the leakage parameters in (14) satisfy $\Gamma_{i\ell j} \geq \Gamma_{i\ell j}^{(p)}$, the optimal solution $\mathbf{C}_j^{(p)}$ is *feasible* for (14). Next, assume that all leakage parameters $\Gamma_{i\ell j}$ are set equal to the corresponding optimal interference leakage powers $\Gamma_{i\ell j}^{(p)}$ and suppose, however, that the multicast rate R_j^* obtained with the solution $\mathbf{C}_j^*$ from (14) is larger than $R_j^{(p)}$: $R_j^* > R_j^{(p)}$. The achievable rate of any other BS_ℓ when all BSs apply weighted sum-rate optimal precoders is

$$R_\ell^{(p)} = \min_{i \in \{1,\ldots,K\}} \log_2\left(1 + \mathrm{SINR}_{i\ell}^{(p)}\right), \tag{44}$$

$$\mathrm{SINR}_{i\ell}^{(p)} = \frac{\operatorname{tr}\left(\mathbf{C}_\ell^{(p)} \mathbf{h}_{i\ell\ell} \mathbf{h}_{i\ell\ell}^{\mathrm{H}}\right)}{\sigma_n^2 + \sum\limits_{n \neq \{\ell, j\}} \operatorname{tr}\left(\mathbf{C}_n^{(p)} \mathbf{h}_{i\ell n} \mathbf{h}_{i\ell n}^{\mathrm{H}}\right) + \underbrace{\operatorname{tr}\left(\mathbf{C}_j^{(p)} \mathbf{h}_{i\ell j} \mathbf{h}_{i\ell j}^{\mathrm{H}}\right)}_{\Gamma_{i\ell j}^{(p)}}}.$$

Since we have $\operatorname{tr}\left(\mathbf{C}_j^* \mathbf{h}_{i\ell j} \mathbf{h}_{i\ell j}^{\mathrm{H}}\right) \leq \Gamma_{i\ell j}^{(p)}$, due to the interference leakage constraint imposed in (14), the achievable rate of BS_ℓ when BS_j employs $\mathbf{C}_j^*$ is at least as good as $R_\ell^{(p)}$

$$R_\ell^* = \min_{i \in \{1,\ldots,K\}} \log_2\left(1 + \mathrm{SINR}_{i\ell}^*\right) \geq R_\ell^{(p)}, \tag{45}$$

$$\mathrm{SINR}_{i\ell}^* = \frac{\operatorname{tr}\left(\mathbf{C}_\ell^{(p)} \mathbf{h}_{i\ell\ell} \mathbf{h}_{i\ell\ell}^{\mathrm{H}}\right)}{\sigma_n^2 + \sum\limits_{n \neq \{\ell, j\}} \operatorname{tr}\left(\mathbf{C}_n^{(p)} \mathbf{h}_{i\ell n} \mathbf{h}_{i\ell n}^{\mathrm{H}}\right) + \underbrace{\operatorname{tr}\left(\mathbf{C}_j^* \mathbf{h}_{i\ell j} \mathbf{h}_{i\ell j}^{\mathrm{H}}\right)}_{\leq \Gamma_{i\ell j}^{(p)}}}.$$

As this relation holds true for all $\ell \neq j$, we have a *contradiction* to the assumption that $[R_1^{(p)}, \ldots, R_J^{(p)}]^{\mathrm{T}}$ solves (8) and is Pareto-optimal. Consequently, the solution of (14) with interference leakage parameters set according to the optimal solution of (8) cannot result in a larger rate than the weighted sum-rate optimal value, even though the leakage constraints have been relaxed to inequality constraints in (9). Conversely, setting the leakage parameters in problem (13) equal to $\Gamma_{i\ell j}^{(p)}$, the optimization with respect to $\mathcal{C}$ and $\mathbf{z}$ results in a solution that is at least as good as the weighted sum-rate optimal rate-tuple, since the leakage equality constraints have been relaxed to inequality constraints. Furthermore, as problem set (14) is equivalent to problem (13) when the leakage parameters are fixed, it follows that the solution of (14) with leakage parameters set according to the optimal solution of (8) is also at least as good as the weighted sum-rate optimal rate-tuple. Combining both arguments establishes the proof.

Appendix B
Gradient for Weighted Sum-Rate Optimization

We formulate the Lagrangian function of problem (14)

$$L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma}) = z_j + \sum_{k=1}^{K} \lambda_{kj}\left(w_j \bar{R}_{kj} - z_j\right)$$
$$+ \sum_{\substack{\ell=1 \\ \ell \neq j}}^{J} \sum_{i=1}^{K} \lambda_{i\ell j}\left(\Gamma_{i\ell j} - \operatorname{tr}\left(\mathbf{C}_j \mathbf{h}_{i\ell j} \mathbf{h}_{i\ell j}^{\mathrm{H}}\right)\right) + \lambda_j\left(P_j - \operatorname{tr}\left(\mathbf{C}_j\right)\right)$$

$$= \sum_{k=1}^{K} \lambda_{kj} w_j \bar{R}_{kj} + \sum_{\substack{\ell=1 \\ \ell \neq j}}^{J} \sum_{i=1}^{K} \lambda_{i\ell j} \Gamma_{i\ell j} + f_j(\boldsymbol{\lambda}_j, z_j, \mathbf{C}_j), \quad (46)$$

$$\boldsymbol{\lambda}_j = \big[\lambda_j, \lambda_{1j}, \ldots, \lambda_{Kj}, \lambda_{11j}, \ldots, \lambda_{K1j}, \lambda_{12j}, \ldots, \lambda_{K(j-1)j}, \lambda_{1(j+1)j}, \ldots, \lambda_{KJj}\big]^{\mathrm{T}} \in \mathbb{R}^{KJ+1}. \quad (47)$$

with $\bar{R}_{kj}$ as defined in (11), $\boldsymbol{\Gamma}$ as specified in (10) and $f_j(\boldsymbol{\lambda}_j, z_j, \mathbf{C}_j)$ containing those parts of the Lagrangian that are independent of the leakage parameters. The Lagrangian dual function is

$$d_j(\boldsymbol{\lambda}_j, \boldsymbol{\Gamma}) = \sup_{\mathbf{C}_j, z_j} L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma}). \quad (48)$$

Since strong duality holds under the assumptions mentioned in Section IV-A, we have

$$z_j^*(\boldsymbol{\Gamma}) = \min_{\boldsymbol{\lambda}_j \geq 0} d_j(\boldsymbol{\lambda}_j, \boldsymbol{\Gamma}) = \min_{\boldsymbol{\lambda}_j \geq 0} \sup_{\mathbf{C}_j, z_j} L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma}), \quad (49)$$

with $z_j^*(\boldsymbol{\Gamma})$ denoting the optimal value of (14) for given leakage parameters $\boldsymbol{\Gamma}$. We denote the corresponding optimal dual variables as $\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma})$. We next apply a first order Taylor approximation of the Lagrangian around $\boldsymbol{\Gamma}$

$$\begin{aligned} L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \tilde{\boldsymbol{\Gamma}}) &\approx \hat{L}_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \tilde{\boldsymbol{\Gamma}}) \\ &= L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma}) + \nabla_{\boldsymbol{\Gamma}} L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma}) \left(\tilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right), \end{aligned} \quad (50)$$

with $\nabla_{\boldsymbol{\Gamma}} L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma})$ denoting the gradient of the Lagrangian with respect to the leakage parameters. In the regime around $L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma})$ where this first order Taylor approximation holds or underestimates the actual behaviour, i.e., $L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \tilde{\boldsymbol{\Gamma}}) \geq \hat{L}_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \tilde{\boldsymbol{\Gamma}})$, selecting

$$\tilde{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma} + \mu \nabla_{\boldsymbol{\Gamma}} L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma}), \ \mu > 0 \quad (51)$$

strictly improves the Lagrangian. Provided the step size μ is selected sufficiently small to validate the first order approximation, it thus follows

$$\min_{\boldsymbol{\lambda}_j \geq 0} \sup_{\mathbf{C}_j, z_j} \hat{L}_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \tilde{\boldsymbol{\Gamma}}) \geq \min_{\boldsymbol{\lambda}_j \geq 0} \sup_{\mathbf{C}_j, z_j} L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma}). \quad (52)$$

A valid ascent direction of problem (14) at $\{\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma}), \mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma}), \boldsymbol{\Gamma}\}$ with respect to $\boldsymbol{\Gamma}$ is

$$\mathbf{g}_j = \nabla_{\boldsymbol{\Gamma}} L_j(\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma}), \mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma}), \boldsymbol{\Gamma}), \quad (53)$$

$$\frac{\partial}{\partial \Gamma_{kj\ell}} L_j(\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma}), \mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma}), \boldsymbol{\Gamma}) = -\lambda_{kj}^* w_j \frac{\mathrm{tr}\left(\mathbf{C}_j^*(\boldsymbol{\Gamma}) \mathbf{h}_{kjj} \mathbf{h}_{kjj}^{\mathrm{H}}\right) / \left(\sigma_n^2 + \sum_{\substack{\ell=1 \\ \ell \neq j}}^{J} \Gamma_{kj\ell}\right)}{\sigma_n^2 + \sum_{\substack{\ell=1 \\ \ell \neq j}}^{J} \Gamma_{kj\ell} + \mathrm{tr}\left(\mathbf{C}_j^*(\boldsymbol{\Gamma}) \mathbf{h}_{kjj} \mathbf{h}_{kjj}^{\mathrm{H}}\right)}, \quad \forall k \in \{1, \ldots, K\}, \ \forall \ell \in \{1, \ldots, J\} \setminus j, \quad (54)$$

$$\frac{\partial}{\partial \Gamma_{i\ell j}} L_j(\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma}), \mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma}), \boldsymbol{\Gamma}) = \lambda_{i\ell j}^*, \quad \forall i \in \{1, \ldots, K\}, \ \forall \ell \in \{1, \ldots, J\} \setminus j. \quad (55)$$

The partial derivatives with respect to the remaining leakage parameters are zero. Alternatively, as considered in [29], one can employ any sub-gradient of the Lagrangian instead of the gradient. Due to the strong non-linearity of $L_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \tilde{\boldsymbol{\Gamma}})$ a sub-gradient, however, is difficult to find.

Appendix C
Gradient for Pareto-Improvement

The local gradient of problem (20) is obtained analogously to the derivation in Appendix B. Denoting the Lagrangian of (20) as $\Lambda_j(\boldsymbol{\lambda}_j, \mathbf{C}_j, z_j, \boldsymbol{\Gamma})$, the result is

$$\boldsymbol{\gamma}_j = \nabla_{\boldsymbol{\Gamma}} \Lambda_j(\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma}), \mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma}), \boldsymbol{\Gamma}), \quad (56)$$

$$\frac{\partial}{\partial \Gamma_{kj\ell}} \Lambda_j(\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma}), \mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma}), \boldsymbol{\Gamma}) = -\lambda_{kj}^* \frac{\mathrm{tr}\left(\mathbf{C}_j^*(\boldsymbol{\Gamma}) \mathbf{h}_{kjj} \mathbf{h}_{kjj}^{\mathrm{H}}\right)}{\left(\sigma_n^2 + \sum_{\substack{\ell=1 \\ \ell \neq j}}^{J} \Gamma_{kj\ell}\right)^2}, \quad \forall k \in \{1, \ldots, K\}, \ \forall \ell \in \{1, \ldots, J\} \setminus j, \quad (57)$$

$$\frac{\partial}{\partial \Gamma_{i\ell j}} \Lambda_j(\boldsymbol{\lambda}_j^*(\boldsymbol{\Gamma}), \mathbf{C}_j^*(\boldsymbol{\Gamma}), z_j^*(\boldsymbol{\Gamma}), \boldsymbol{\Gamma}) = \lambda_{i\ell j}^*, \quad \forall i \in \{1, \ldots, K\}, \ \forall \ell \in \{1, \ldots, J\} \setminus j. \quad (58)$$

Appendix D
Max-Min SLNR Precoding

The input covariance matrix $\bar{\mathbf{C}}_j^*$ solves the convex optimization problem

$$\bar{\mathbf{C}}_j^* = \underset{\bar{\mathbf{C}}_j \in \mathbb{C}^{N_t \times N_t}}{\arg\max} \min_{k \in \{1, \ldots, K\}} \mathrm{tr}\left(\bar{\mathbf{C}}_j \mathbf{L}_j^{-\frac{\mathrm{H}}{2}} \mathbf{h}_{kjj} \mathbf{h}_{kjj}^{\mathrm{H}} \mathbf{L}_j^{-\frac{1}{2}}\right), \quad \text{subject to: } \mathrm{tr}\left(\bar{\mathbf{C}}_j\right) = 1. \quad (59)$$

Plugging the post-normalized solution $\tilde{\mathbf{C}}_j^*$ as defined in Equation (42) into one of the terms of problem (40), we obtain

$$\frac{\frac{P_j}{\mathrm{tr}\left(\bar{\mathbf{C}}_j^* \mathbf{L}_j^{-1}\right)} \mathrm{tr}\left(\bar{\mathbf{C}}_j^* \mathbf{L}_j^{-\frac{\mathrm{H}}{2}} \mathbf{h}_{kjj} \mathbf{h}_{kjj}^{\mathrm{H}} \mathbf{L}_j^{-\frac{1}{2}}\right)}{\frac{P_j}{\mathrm{tr}\left(\bar{\mathbf{C}}_j^* \mathbf{L}_j^{-1}\right)} \mathrm{tr}\left(\bar{\mathbf{C}}_j^*\right)}, \quad \text{subject to: } \frac{P_j}{\mathrm{tr}\left(\bar{\mathbf{C}}_j^* \mathbf{L}_j^{-1}\right)} \mathrm{tr}\left(\bar{\mathbf{C}}_j^* \mathbf{L}_j^{-1}\right) = P_j. \quad (60)$$

Thus, $\tilde{\mathbf{C}}_j^*$ satisfies the power constraint and, since $\bar{\mathbf{C}}_j^*$ maximizes the minimum over $k \in \{1, \ldots, K\}$ of (60), $\tilde{\mathbf{C}}_j^*$ solves problem (40).

References

[1] 3GPP. (2006, Sep.). *TSG RAN; Introduction of the Multimedia Broadcast Multicast Service (MBMS) in the Radio Access Network; Stage 2 (Release 6)* [Online]. Available: http://www.3gpp.org/DynaReport/25346.htm

[2] Ericsson, "LTE broadcast: A revenue enabler in the mobile media era," White Paper, Feb. 2013.

[3] D. Lewis (Verizon), "Verizon delivers LTE multicast over commercial 4G LTE network in Indy," Press Release, May 21, 2014.

[4] V. D'Avino (Ericsson), "Ericsson and Polkomtel test LTE broadcast in Poland," Press Release, Sep. 1, 2014.

[5] B. Smith (AT&T), "AT&T LTE broadcast demo at college football national championship showcases exciting network future," Press Release, Jan. 16, 2015.

[6] D. Valerio, F. Ricciato, P. Belanovic, and T. Zemen, "UMTS on the road: Broadcasting intelligent road safety information via MBMS," in *Proc. 67th IEEE Veh. Technol. Conf. (VTC'08-Spring)*, Singapore, May 2008, pp. 3026–3030.

[7] T. Lohmar, M. Slssingar, S. Puustinen, and V. Kenehan, "Delivering content with LTE broadcast," *Ericsson Rev.: Commun. Technol. J. 1924*, vol. 2013, no. 1, pp. 1–8, Feb. 2013.

[8] A. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, Jan. 1978.

[9] J. Lindblom, E. Karipidis, and E. Larsson, "Efficient computation of Pareto optimal beamforming vectors for the MISO interference channel with successive interference cancellation," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4782–4795, Oct. 2013.

[10] E. Jorswieck, E. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[11] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[12] X. Shang, B. Chen, and H. Poor, "Multiuser MISO interference channels with single-user detection: Optimality of beamforming and the achievable rate region," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4255–4273, Jul. 2011.

[13] R. Mochaourab and E. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1128–1141, Mar. 2011.

[14] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "On the complexity of optimal coordinated downlink beamforming," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar. 2010, pp. 3274–3277.

[15] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Distributed resource allocation schemes," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 53–63, Sep. 2009.

[16] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2508–2523, May 2012.

[17] L. Liu, R. Zhang, and K.-C. Chua, "Achieving global optimality for weighted sum-rate maximization in the K-user Gaussian interference channel with multiple antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1933–1945, May 2012.

[18] M. Sadek, A. Tarighat, and A. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[19] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.

[20] S. Schwarz and M. Rupp, "Exploring coordinated multipoint beamforming strategies for 5G cellular," *IEEE Access*, vol. 2, pp. 930–946, 2014.

[21] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[22] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[23] A. Gershman, N. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.

[24] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Mar. 2008.

[25] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, May 2013.

[26] S. Wu, W.-K. Ma, and A.-C. So, "Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time coding," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4230–4245, Sep. 2013.

[27] G. Dartmann, X. Gong, and G. Ascheid, "Low complexity cooperative multicast beamforming in multiuser multicell downlink networks," in *Proc. 6th Int. ICST Conf. Cognit. Radio Oriented Wireless Netw. Commun.*, Jun. 2011, pp. 370–374.

[28] G. Dartmann and G. Ascheid, "Equivalent quasi-convex form of the multicast max-min beamforming problem," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4643–4648, Nov. 2013.

[29] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, Jan. 2013.

[30] S. Schwarz, T. Philosof, and M. Rupp, "Leakage-based multicast transmit beamforming," in *Proc. Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 2405–2410.

[31] H. Won *et al.*, "Multicast scheduling in cellular data networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4540–4549, Sep. 2009.

[32] J. Chen, M. Chiang, J. Erman, G. Li, K. Ramakrishnan, and R. Sinha, "Fair and optimal resource allocation for lte multicast (eMBMS): Group partitioning and dynamics," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2015, pp. 1266–1274.

[33] K. Phan, S. Vorobyov, N. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via QoS-aware secondary multicast beamforming," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2323–2335, Jun. 2009.

[34] M. Beko, S. Tomic, R. Dinis, and V. Lipovac, "Convex optimization-based beamforming in cognitive radio multicast transmission," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2012, pp. 1–5.

[35] Y. Zeng, C. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2899–2913, Jun. 2013.

[36] N. Jindal and Z.-Q. Luo, "Capacity limits of multiple antenna multicast," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 1841–1845.

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[38] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[39] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[40] M. Mohseni, R. Zhang, and J. Cioffi, "Optimized transmission for fading multiple-access and broadcast channels with multiple antennas," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1627–1639, Aug. 2006.

[41] A. Rubinov, H. Tuy, and H. Mays, "An algorithm for monotonic global optimization problems," *Optim.: J. Math. Prog. Oper. Res.*, vol. 49, no. 3, pp. 205–221, 2001.

[42] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[43] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 95–110.

[44] S. Schwarz and M. Rupp, "Predictive quantization on the Stiefel manifold," *IEEE Signal Proc. Lett.*, vol. 22, no. 2, pp. 234–238, 2015.

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

[46] S. Schwarz, R. Heath Jr., and M. Rupp, "Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems," in *Proc. 38th Int. Conf. Acoust. Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 5021–5025.

[47] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[48] R. Zakhour and D. Gesbert, "Distributed multicell-MISO precoding using the layered virtual SINR framework," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2444–2448, Aug. 2010.

[49] S. Schwarz and M. Rupp, "Transmit optimization for the MISO multicast interference channel with single-user detection" *IEEE Trans. Commun.*, Institute of Telecommunications, TU Wien, 2015 [Online]. Available: http://www.nt.tuwien.ac.at/about-us/staff/stefan-schwarz/reproducible-research/, MATLAB code.

[50] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives." *Pac. J. Math.*, vol. 16, no. 1, pp. 1–3, 1966.

Stefan Schwarz (M'14) was born in Neunkirchen, Austria, in 1984. He received the B.Sc. degree in electrical engineering and the Dipl.-Ing. degree (M.Sc. equivalent) (with highest distinctions) in telecommunications engineering, from TU Wien, Wien, Austria, the Dr. Techn. degree (Ph.D. equivalent) (with highest distinctions) in telecommunications engineering at TU Wien, in 2007, 2009, 2013, respectively. From 2008 to 2014, he worked as a Project Assistant with the Mobile Communications Group of Prof. Markus Rupp, Institute of Telecommunications, TU Wien, focusing on link and system level simulation of LTE/LTE-A networks and acting as one of the lead developers of the Vienna LTE Simulators. In 2012, he visited the University of Texas at Austin, Austin, TX, USA, as Research Scholar on limited feedback algorithms for distributed antenna systems. In his dissertation, he focused on limited feedback single- and multiuser MIMO–OFDM communication systems and multiuser scheduling algorithms. Since 2015, he has been employed as a Postdoctoral Researcher (University Assistant) with the Institute of Telecommunications. His research interests include wireless communications and signal processing. He serves as an Associate Editor of *EURASIP Journal on Wireless Communications and Networking*. In 2010, he was the recipient of the Honorary Price of the Austrian Minister of Science and Research, for excellent graduates of scientific and artistic universities and in 2014, the INiTS Award in the category Information and Communication Technologies for innovative scientific works with prospect for economic applicability.

Markus Rupp (F'15) was born in 1963 in Völklingen, Germany. He received the Dipl.-Ing. degree from the University of Saarbrücken, Saarbrücken, Germany, and the Dr.-Ing. degree from the Technische Universität Darmstadt, Darmstadt, Germany, in 1988 and 1993, respectively. He worked with Eberhardt Hänsler on designing new algorithms for acoustical and electrical echo compensation at the Technische Universität Darmstadt. From November 1993 to July 1995, he had a postdoctoral position with the University of California, Santa Barbara, Santa Barbara, CA, USA, where he worked on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 to August 2001, he was a member of Technical Staff with the Wireless Technology Research Department, Bell-Labs, Crawford Hill, NJ, USA, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11, and UMTS. Since October 2001, he has been a Full Professor for digital signal processing in mobile communications with the Technical University of Vienna, Vienna, Austria, where he served as the Dean from 2005 to 2007 and from 2016 to 2017. He has authored and coauthored more than 500 scientific papers, including 15 patents on adaptive filtering and wireless communications. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005, is currently an Associate Editor of *EURASIP Journal of Advances in Signal Processing* and *EURASIP Journal on Embedded Systems*. He was Elected AdCom Member of EURASIP from 2004 to 2012, serving as the President of EURASIP from 2009 to 2010.

Probabilistic Analysis of Semidefinite Relaxation for Leakage-Based Multicasting

Stefan Schwarz, *Member, IEEE*

***Abstract*—In this letter, we derive worst-case approximation results for rank one and rank two leakage-based multicasting (LBM) as recently proposed by Schwarz and Rupp ["Transmit optimization for the MISO multicast interference channel," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4936–4949, Dec. 2015]. Specifically, we provide worst-case lower bounds on the approximation ratios achieved with rank one/two Gaussian randomization of the optimal solution as obtained from a semidefinite relaxation (SDR). We demonstrate the validity of the derived bounds through Monte Carlo simulations and we show that good approximation ratios are achieved even for very large number of multicast users and leakage constraints.**

***Index Terms*—Alamouti coding, approximation bound, beamforming (BF), multicast, semidefinite relaxation (SDR).**

I. Introduction

PHYSICAL-LAYER multicasting in wireless networks has received significant attention over the past decade, e.g., [2]–[4], since it enables efficient transmission of common information to many users in parallel. The 3GPP introduced physical layer multicasting with Release 6 of UMTS [5] and extended the technology within LTE to enhanced multimedia broadcast/multicast service (eMBMS) [6]. Envisioned use cases include amongst others *venue casting* at local events [7], support of *road safety* information broadcasting [8], [9], and software updates for mobile *Internet of Things* devices [10].

In [1] and [11], we proposed leakage-based multicasting (LBM) as an efficient interference-aware transmission scheme for the multicast interference channel (MC-IC). LBM aims at maximizing the achievable multicast transmission rate of a base station, while restricting the interference leakage caused to users of other base stations. Combined with coordinated leakage parameter control among base stations, the method significantly extends the achievable rate region of the network. Attaining points on the boundary of the LBM rate region in general requires transmit signals that are distributed according to a multivariate Gaussian distribution with high-rank covariance matrix. In practice, however, simple transmit beamforming (BF) schemes are often preferred over high-rank transmission; then, LBM can be interpreted as the semidefinite relaxation (SDR) of rank-constrained transmit optimization and randomization can provide suboptimal low-rank BF solutions [12], [13]. Notice that multicasting in cognitive radio transmission has a very similar system model to LBM, with the difference that leakage constraints cannot be optimized in cognitive radio but are prescribed by the primary system [14]–[17]. SDR has been successfully adopted in many diverse applications, e.g., multiple-input multiple-output (MIMO) detection [18] and localization in sensor networks [19].

In this letter, we analyze the approximation performance of rank one/two BF obtained through randomization of the LBM solution, demonstrating good performance even for large numbers of multicast users. Specifically, we derive bounds on the approximation ratio, extending existing results [17], [20]–[22] to MIMO transmission.

Notation: The trace of matrix $\mathbf{A}$ is $\mathrm{tr}\left(\mathbf{A}\right)$, the conjugate-transpose is $\mathbf{A}^{\mathrm{H}}$, the Frobenius norm is $\left\|\mathbf{A}\right\|_F$, and the vector obtained by stacking the columns is $\mathrm{vec}\left(\mathbf{A}\right)$. The expected value of random variable x is $\mathbb{E}\left\{x\right\}$. The vector-valued complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$ is $\mathcal{N}_{\mathbb{C}}\left(\boldsymbol{\mu},\mathbf{C}\right)$. The probability of random event $\mathcal{A}$ is $\mathbb{P}\left\{\mathcal{A}\right\}$.

II. System Model

We consider a transmitter (e.g., base station), equipped with N_t transmit antennas, that multicasts a common message to K users, each having N_r receive antennas. The input–output relationship of user k over r consecutive time instants is

$$\mathbf{Y}_k = \mathbf{H}_k\mathbf{X} + \mathbf{Z}_k \in \mathbb{C}^{N_r\times r}, \quad k\in\{1,\ldots,K\} \tag{1}$$

with $\mathbf{H}_k \in \mathbb{C}^{N_r\times N_t}$ denoting the channel matrix of user k and $\mathbf{Z}_k$ representing noise and interference. We assume that the channel stays constant during the transmission of one space-time/frequency code block $\mathbf{X}\in\mathbb{C}^{N_t\times r}$ of length r. We consider low-complexity receivers that are not able to cancel interference, i.e., we assume that users treat interference as additional noise. We assume that the effective noise $\mathbf{Z}_k$ is independent and identically distributed (i.i.d) of power σ_k^2.

We consider two transmit strategies; that is, rank one BF and rank two transmit beamformed Alamouti space-time/frequency coding (ABF) [17], [22], [23]. In case of BF, the transmit signal is $\mathbf{x}=\mathbf{f}_1 s_1 \in \mathbb{C}^{N_t\times 1}$ with $\mathbf{f}_1$ denoting the applied beamformer. For ABF, the transmit signal is

$$\mathbf{X} = \left[\mathbf{f}_2,\mathbf{g}_2\right]\begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^*\end{bmatrix} \in \mathbb{C}^{N_t\times 2} \tag{2}$$

with $s_i\in\mathbb{C}$, $\mathbb{E}\left\{\left|s_i\right|^2\right\}=1$ denoting the information symbols. Assuming that user k applies in both cases a (space-time) matched filter, the obtained signal-to-noise ratios (SNRs) can be calculated from a simple extension of [24, Sec.III.B]

$$\begin{aligned}\mathrm{SNR}_k\left(\mathbf{f}_1\right) &= \frac{1}{\sigma_k^2}\left\|\mathbf{H}_k\mathbf{f}_1\right\|_F^2, \\ \mathrm{SNR}_k\left(\mathbf{f}_2,\mathbf{g}_2\right) &= \frac{1}{\sigma_k^2}\left\|\mathbf{H}_k\left[\mathbf{f}_2,\mathbf{g}_2\right]\right\|_F^2.\end{aligned} \tag{3}$$

Manuscript received December 26, 2015; revised March 22, 2016; accepted April 02, 2016. Date of publication April 05, 2016; date of current version April 20, 2016. The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technolgy and Development is gratefully acknowledged. This work has been co-funded by A1 Telekom Austria AG, Kathrein Werke KG and Nokia Solutions and Networks. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jiaheng Wang.

The author is with the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion, Institute of Telecommunications, TU Wien, 1040 Vienna, Austria (e-mail: sschwarz@nt.tuwien.ac.at).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2016.2550618

III. Leakage-Based Multicasting

To optimize the achievable rate of the multicast transmission, while restricting the amount of interference caused to N unintended users of other base stations, we consider the following LBM optimization problem [1], which maximizes the worst-user SNR under interference-leakage constraints

$$\begin{aligned} v^*_{r_{\max}} = \max_{\mathbf{C}\succeq 0,\, \mathbf{C}\in\mathbb{C}^{N_t\times N_t}} \min_{k\in\{1,\ldots,K\}} & \operatorname{tr}(\mathbf{C}\,\mathbf{R}_k) \\ \text{s.t.} \quad & \operatorname{tr}(\mathbf{C}\,\mathbf{S}_n) \le 1 \quad \forall n \in \{1,\ldots,N+1\} \\ & \operatorname{rank}(\mathbf{C}) \le r_{\max}. \end{aligned} \tag{4}$$

Here, the rank-constrained optimization variable $\mathbf{C}$ represents the transmit covariance matrix and, thus, $\operatorname{tr}(\mathbf{C}\,\mathbf{R}_k)$ is the SNR of user k. The optimal value $v^*_{r_{\max}}$ denotes the achieved worst-user SNR under rank constraint $r_{\max}$; e.g., v^*_1 is the optimal value with $r_{\max}=1$. In (4), we employ the abbreviations

$$\mathbf{R}_k = \frac{1}{\sigma_k^2}\mathbf{H}_k{}^{\mathrm{H}}\mathbf{H}_k \in \mathbb{C}^{N_t\times N_t} \tag{5}$$

$$\mathbf{S}_n = \frac{1}{N_r\Gamma_n}\mathbf{G}_n{}^{\mathrm{H}}\mathbf{G}_n \in \mathbb{C}^{N_t\times N_t} \quad \forall n\le N, \quad \mathbf{S}_{N+1} = \frac{1}{P}\mathbf{I}_{N_t} \tag{6}$$

with $\mathbf{G}_n \in \mathbb{C}^{N_r\times N_t}$ being the channel matrix with respect to unintended user n and Γ_n denoting its interference leakage constraint per receive antenna, so that $\operatorname{tr}\left(\mathbf{C}\,\mathbf{G}_n{}^{\mathrm{H}}\mathbf{G}_n\right) \le N_r\Gamma_n$. Since the unintended user n does not apply the matched filter corresponding to the interference channel $\mathbf{G}_n$, but rather some optimal receive filter with respect to its own transmitter, the total interference collected grows linearly with the number of receive antennas. This is reflected by the factor N_r in the denominator of $\mathbf{S}_n$. The matrix $\mathbf{S}_{N+1}$ accounts for the power constraint P of the transmitter, such that $\operatorname{tr}(\mathbf{C}) \le P$. To obtain the optimal beamformers for BF or ABF, we set $r_{\max}$ equal to one or two, respectively, and decompose the optimal $\mathbf{C}^*_{r_{\max}}$ as

$$\mathbf{C}^*_1 = \mathbf{f}_1\mathbf{f}_1{}^{\mathrm{H}}, \quad \mathbf{C}^*_2 = [\mathbf{f}_2,\mathbf{g}_2]\,[\mathbf{f}_2,\mathbf{g}_2]^{\mathrm{H}}. \tag{7}$$

The LBM solution of [1] corresponds to the SDR $r_{\max}=N_t$. Notice that problem (4) is nonconvex whenever $r_{\max}<N_t$. Furthermore, employing an extra slack variable in (4), we can reformulate the minimization of the SNR over k as K additional linear inequality constraints on $\operatorname{tr}(\mathbf{C}\,\mathbf{R}_k)$.

Following [25] and [26], it can easily be verified that the rank of the optimal solution $\mathbf{C}^*_{N_t}$ of the SDR is upper bounded as

$$r^* = \operatorname{rank}\left(\mathbf{C}^*_{N_t}\right) \le \sqrt{K+N+1} = r^{(\mathrm{ub})}. \tag{8}$$

Hence, as long as $K+N\le 2$ or $K+N\le 7$, the optimal $\mathbf{C}^*_{N_t}$ is equal to $\mathbf{C}^*_1$ or $\mathbf{C}^*_2$, respectively. Furthermore, whenever $r^*>\{1,2\}$, feasible beamformers for $r_{\max}=1$ and $r_{\max}=2$ can be obtained from the two Gaussian randomizations

$$\hat{\mathbf{f}}_1^{(\ell)} = \frac{\boldsymbol{\varphi}}{\sqrt{\max_n \operatorname{tr}(\boldsymbol{\varphi}\boldsymbol{\varphi}^{\mathrm{H}}\mathbf{S}_n)}}, \tag{9}$$

$$\begin{aligned} \left[\hat{\mathbf{f}}_2^{(\ell)},\hat{\mathbf{g}}_2^{(\ell)}\right] &= \frac{\boldsymbol{\Theta}}{\sqrt{\max_n \operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{H}}\mathbf{S}_n)}}, \\ \hat{v}_1^{(\ell)} = \mathrm{SNR}_k\left(\hat{\mathbf{f}}_1^{(\ell)}\right), \hat{v}_2^{(\ell)} &= \mathrm{SNR}_k\left(\hat{\mathbf{f}}_2^{(\ell)},\hat{\mathbf{g}}_2^{(\ell)}\right) \\ \boldsymbol{\Theta} = \frac{1}{\sqrt{2}}[\boldsymbol{\varphi},\boldsymbol{\gamma}], \quad \boldsymbol{\varphi},\boldsymbol{\gamma} &\sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0},\mathbf{C}^*_{N_t}\right), \ell\in\mathcal{L}=\{1,\ldots,L\} \end{aligned} \tag{10}$$

where the normalization in the denominator assures that the tightest leakage or power constraint is fulfilled with equality and ℓ denotes the randomization index. To obtain good performance, we select the applied beamformers $\hat{\mathbf{f}}_1$ resp. $[\hat{\mathbf{f}}_2,\hat{\mathbf{g}}_2]$ as the best solutions from L independent randomizations; we denote the corresponding worst-user SNRs as $\hat{v}_1^{(\mathcal{L})}$ and $\hat{v}_2^{(\mathcal{L})}$

$$\hat{\mathbf{f}}_1 = \max_{\ell\in\{1\ldots L\}} \min_k \mathrm{SNR}_k\left(\hat{\mathbf{f}}_1^{(\ell)}\right), \tag{11}$$

$$\begin{aligned} [\hat{\mathbf{f}}_2,\hat{\mathbf{g}}_2] &= \max_{\ell\in\{1\ldots L\}} \min_k \mathrm{SNR}_k\left(\hat{\mathbf{f}}_2^{(\ell)},\hat{\mathbf{g}}_2^{(\ell)}\right), \\ \hat{v}_1^{(\mathcal{L})} = \mathrm{SNR}_k\left(\hat{\mathbf{f}}_1\right), \hat{v}_2^{(\mathcal{L})} &= \mathrm{SNR}_k\left(\hat{\mathbf{f}}_2,\hat{\mathbf{g}}_2\right). \end{aligned} \tag{12}$$

Alternatively, feasible solutions can be obtained from the $r_{\max}$ eigenvectors of $\mathbf{C}^*_{N_t}$ corresponding to the largest eigenvalues. Yet, this method in general performs worse than Gaussian randomization, provided L is sufficiently large.

IV. Probabilistic Analysis

In this section, we provide worst-case approximation bounds for the randomizations (9), (10) with respect to the performance achieved with $\mathbf{C}^*_{N_t}$. Specifically, the goal is to derive worst case lower bounds on the approximation ratios

$$\rho_1 = \frac{\hat{v}_1^{(\mathcal{L})}}{v^*_{N_t}}, \quad \rho_2 = \frac{\hat{v}_2^{(\mathcal{L})}}{v^*_{N_t}} \tag{13}$$

in the sense that a single randomization achieves a value larger than the lower bound with positive probability. Hence, with L independent randomizations, the probability of achieving a value above the lower bound tends to one exponentially in L. Similar to [17], [20]–[22], such bounds are in general not tight, since they are valid for arbitrary positive semidefinite Gramian matrices $\mathbf{R}_k, \mathbf{S}_n$. The important information conveyed by such worst-case bounds, however, is the scaling of the approximation ratio with the number of served users K and interfering users N, representing the performance degradation of randomization with growing problem dimensions. Below, we provide derivations for ABF and only state final results for BF, since the calculation procedure is very similar for both.

We start our derivation with the following valid inequalities:

$$v^*_{N_t} \ge v^*_2 \ge \hat{v}_2^{(\mathcal{L})} \ge \hat{v}_2^{(\ell)} = \frac{\min_k \operatorname{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{H}}\mathbf{R}_k\right)}{\max_n \operatorname{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{H}}\mathbf{S}_n\right)} = \frac{\min_k P_k}{\max_n L_n} \tag{14}$$

where the SDR $v^*_{N_t}$ is better than the rank constrained solution v^*_2; further, v^*_2 outperforms the randomized solution $\hat{v}_2^{(\mathcal{L})}$, which in turn is better than the solution of the ℓth randomization $\hat{v}_2^{(\ell)}$. We target finding constants $\mu_2,\gamma_2\in\mathbb{R}_+$ satisfying

$$\hat{v}_2^{(\ell)} \ge \frac{\mu_2}{\gamma_2} v^*_{N_t} \Rightarrow \rho_2 \ge \frac{\hat{v}_2^{(\ell)}}{v^*_{N_t}} \ge \frac{\mu_2}{\gamma_2} \tag{15}$$

with positive probability. We thus require

$$\mathcal{P}_2 = \mathbb{P}\left\{\min_k P_k \ge \mu_2 v^*_{N_t} \cap \max_n L_n \le \gamma_2\right\} > 0. \tag{16}$$

Consider the following chain of inequalities:

$$\begin{aligned} \mathcal{P}_2 &= 1 - \mathbb{P}\left\{P_k < \mu_2 v^*_{N_t} \text{ for some } k \cup L_n > \gamma_2 \text{ for some } n\right\} \\ &\ge 1 - \sum_k \mathbb{P}\left\{P_k < \mu_2 v^*_{N_t}\right\} - \sum_n \mathbb{P}\left\{L_n > \gamma_2\right\} \\ &\ge 1 - \sum_k \mathbb{P}\left\{\operatorname{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{H}}\mathbf{R}_k\right) < \mu_2 \operatorname{tr}\left(\mathbf{C}^*_{N_t}\mathbf{R}_k\right)\right\} \\ &\quad - \sum_n \mathbb{P}\left\{\operatorname{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{H}}\mathbf{S}_n\right) > \gamma_2 \operatorname{tr}\left(\mathbf{C}^*_{N_t}\mathbf{S}_n\right)\right\} \end{aligned} \tag{17}$$

where the first two steps follow from basic probability calculus and the last step is due to $v^*_{N_t} \le \operatorname{tr}\left(\mathbf{C}^*_{N_t}\mathbf{R}_k\right)$ and $\operatorname{tr}\left(\mathbf{C}^*_{N_t}\mathbf{S}_n\right) \le 1$. We thus have to find bounds on the terms within the summations. Notice, for BF, such bounds are derived in [20, Lemmas 3 and 5].

We first consider $\mathbb{P}\left\{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{H}}\mathbf{S}_n\right) > \gamma_2\,\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{S}_n\right)\right\}$ and apply a square-root decomposition to symmetric positive semidefinite matrix $\mathbf{C}_{N_t}^{*} = \mathbf{S}\mathbf{S}$, with $\mathbf{S}$ denoting the symmetric positive semidefinite square root. We next apply an eigen-decomposition $\mathbf{S}\mathbf{S}_n\mathbf{S} = \mathbf{Q}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{Q}$, with unitary $\mathbf{Q} \in \mathbb{C}^{N_t \times N_t}$, $\tilde{r}_n = \mathrm{rank}\,(\mathbf{\Lambda}) = \min\left(r^*, \mathrm{rank}\,(\mathbf{S}_n)\right)$, and obtain

$$\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{S}_n\right) = \mathrm{tr}\left(\mathbf{Q}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{Q}\right) = \sum_{i=1}^{\tilde{r}_n}\lambda_i. \tag{18}$$

Here, λ_i denotes the ith strictly positive eigenvalue of $\mathbf{S}\mathbf{S}_n\mathbf{S}$ and we assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\tilde{r}_n}$. Next, we construct vectors $\boldsymbol{\varphi}, \boldsymbol{\gamma} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \mathbf{C}_{N_t}^{*}\right)$, constituting matrix $\mathbf{\Theta}$, as

$$[\boldsymbol{\varphi}, \boldsymbol{\gamma}] = \mathbf{S}\mathbf{Q}^{\mathrm{H}}[\mathbf{x}, \mathbf{y}] \tag{19}$$

with $\mathbf{x}, \mathbf{y} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \mathbf{I}_{N_t}\right)$, where the multiplication with $\mathbf{Q}$ leaves the distribution of $[\mathbf{x}, \mathbf{y}]$ unchanged and $\mathbf{S}$ causes the intended correlation. With this, we obtain the probability as

$$\begin{aligned}
&\mathbb{P}\left\{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{H}}\mathbf{S}_n\right) > \gamma_2\,\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{S}_n\right)\right\} \\
&\quad= \mathbb{P}\left\{\mathrm{tr}\left(\frac{1}{2}\begin{bmatrix}\mathbf{x}^{\mathrm{H}}\\ \mathbf{y}^{\mathrm{H}}\end{bmatrix}\mathbf{\Lambda}[\mathbf{x}, \mathbf{y}]\right) > \gamma_2\sum_{i=1}^{\tilde{r}_n}\lambda_i\right\} \\
&\quad= \mathbb{P}\left\{\sum_{i=1}^{\tilde{r}_n}\frac{\bar{\lambda}_i}{2}(|x_i|^2 + |y_i|^2) > \gamma_2\right\} \\
&\bar{\lambda}_i = \frac{\lambda_i}{\sum_{j=1}^{\tilde{r}_n}\lambda_j}, \quad \sum_{i=1}^{\tilde{r}_n}\bar{\lambda}_i = 1
\end{aligned} \tag{20}$$

with x_i, y_i denoting the ith element of $\mathbf{x}$, $\mathbf{y}$, respectively. With these decompositions, we find the following upper bound:

$$\begin{aligned}
&\mathbb{P}\left\{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{H}}\mathbf{S}_n\right) > \gamma_2\,\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{S}_n\right)\right\} \\
&\quad\leq \mathbb{P}\left\{|x_1|^2 + |y_1|^2 > 2\gamma_2 \cup \ldots \cup |x_{\tilde{r}_n}|^2 + |y_{\tilde{r}_n}|^2 > 2\gamma_2\right\} \\
&\quad\leq \sum_{i=1}^{\tilde{r}_n}\mathbb{P}\left\{|x_i|^2 + |y_i|^2 > 2\gamma_2\right\} = \tilde{r}_n(2\gamma_2 + 1)e^{-2\gamma_2} \\
&\quad\leq \sum_{i=1}^{\tilde{r}_n}\mathbb{P}\left\{|x_i|^2 > \gamma_2\right\} + \mathbb{P}\left\{|y_i|^2 > \gamma_2\right\} = 2\tilde{r}_n e^{-\gamma_2}
\end{aligned} \tag{21}$$

where the first inequality follows since at least one term $|x_i|^2 + |y_i|^2$ must be larger than $2\gamma_2$ for the convex combination $\sum_{i=1}^{\tilde{r}_n}\bar{\lambda}_i(|x_i|^2 + |y_i|^2)$ to be larger than $2\gamma_2$; similar argumentation applies to the last inequality. The equalities are valid since $2|x_i|^2$, $2|y_i|^2$ are both chi-squared random variables with two degrees of freedom; see [20, eq. (14)].

We next consider $\mathbb{P}\left\{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{H}}\mathbf{R}_k\right) < \mu_2\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{R}_k\right)\right\}$ and apply the same strategy as above, i.e., eigen-decomposition of $\mathbf{S}\mathbf{R}_k\mathbf{S} = \mathbf{Q}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{Q}$, with unitary $\mathbf{Q} \in \mathbb{C}^{N_t \times N_t}$, $\bar{r}_k = \mathrm{rank}\,(\mathbf{\Lambda}) = \min\left(r^*, \mathrm{rank}\,(\mathbf{R}_k)\right)$ and square root $\mathbf{S}$ of $\mathbf{C}_{N_t}^{*}$, to obtain

$$\begin{aligned}
&\mathbb{P}\left\{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{H}}\mathbf{R}_k\right) < \mu_2\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{R}_k\right)\right\} \\
&\quad= \mathbb{P}\left\{\sum_{i=1}^{\bar{r}_k}\frac{\bar{\lambda}_i}{2}(|x_i|^2 + |y_i|^2) < \mu_2\right\}
\end{aligned} \tag{22}$$

with $\bar{\lambda}_i$ now representing the strictly positive normalized eigenvalues of $\mathbf{S}\mathbf{R}_k\mathbf{S}$. Consider an arbitrary α in the open interval $(0, 1)$ and assume $\bar{\lambda}_1 \geq \alpha$; this is definitely the case if $\bar{r}_k = 1$ since then $\bar{\lambda}_1 = 1$. For such $\bar{\lambda}_1 \geq \alpha$, we obtain

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{i=1}^{\bar{r}_k}\bar{\lambda}_i(|x_i|^2 + |y_i|^2) < 2\mu_2\right\} \\
&\quad\leq \mathbb{P}\left\{\bar{\lambda}_1(|x_1|^2 + |y_1|^2) < 2\mu_2\right\} \\
&\quad\leq \mathbb{P}\left\{2(|x_1|^2 + |y_1|^2) < \frac{4\mu_2}{\alpha}\right\} = 1 - \left(\frac{2\mu_2}{\alpha} + 1\right)e^{-\frac{2\mu_2}{\alpha}}
\end{aligned}$$

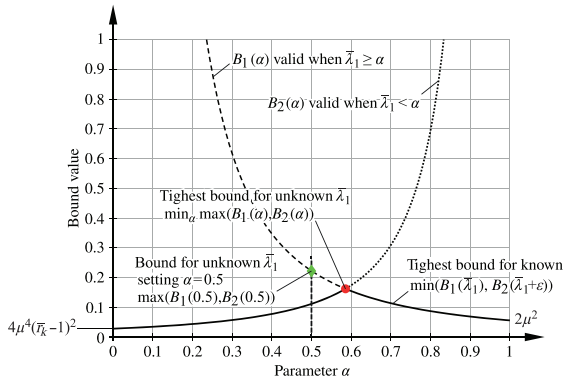


Fig. 1. Illustration of the two upper bounds $B_1(\alpha)$, $B_2(\alpha)$.

$$\leq 2\left(\frac{\mu_2}{\alpha}\right)^2 = B_1(\alpha), \bar{\lambda}_1 \geq \alpha \tag{23}$$

where the last two steps follow because $2(|x_1|^2 + |y_1|^2)$ is chi-squared distributed with four degrees-of-freedom. If, on the other hand, $\bar{\lambda}_1 < \alpha$ we have

$$\begin{aligned}
(\bar{r}_k - 1)\bar{\lambda}_2 \geq \sum_{i=2}^{\bar{r}_k}\bar{\lambda}_i = 1 - \bar{\lambda}_1 > 1 - \alpha \\
\Rightarrow \quad \bar{\lambda}_1 \geq \bar{\lambda}_2 > \frac{1-\alpha}{\bar{r}_k - 1}.
\end{aligned} \tag{24}$$

This delivers the upper bound

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{i=1}^{\bar{r}_k}\bar{\lambda}_i(|x_i|^2 + |y_i|^2) < 2\mu_2\right\} \\
&\leq \mathbb{P}\left\{\bar{\lambda}_1(|x_1|^2 + |y_1|^2) < 2\mu_2 \cap \bar{\lambda}_2(|x_2|^2 + |y_2|^2) < 2\mu_2\right\} \\
&= \mathbb{P}\left\{\bar{\lambda}_1(|x_1|^2 + |y_1|^2) < 2\mu_2\right\}\mathbb{P}\left\{\bar{\lambda}_2(|x_2|^2 + |y_2|^2) < 2\mu_2\right\} \\
&\leq 4\left(\frac{\mu_2}{\bar{\lambda}_1}\right)^2\left(\frac{\mu_2}{\bar{\lambda}_2}\right)^2 \leq 4\mu_2^4\left(\frac{\bar{r}_k - 1}{1-\alpha}\right)^2 = B_2(\alpha), \quad \bar{\lambda}_1 < \alpha.
\end{aligned} \tag{25}$$

The two bounds $B_1(\alpha), B_2(\alpha)$ are illustrated in Fig. 1; notice that $B_1(\alpha)$ is strictly monotonically decreasing in α and $B_2(\alpha)$ is strictly monotonically increasing. If $\bar{\lambda}_1$ is known, we find the tightest bound as $\min\left(B_1(\bar{\lambda}_1), B_2(\bar{\lambda}_1 + \epsilon)\right)$, with vanishingly small ϵ. Since we are, however, interested in a general upper bound that is independent of the actual value of $\bar{\lambda}_1$, we have to take the maximum of (23) and (25); the tightest upper bound valid $\forall\bar{\lambda}_1$ is then obtained as

$$\mathbb{P}\left\{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{H}}\mathbf{R}_k\right) < \mu_2\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{R}_k\right)\right\} \leq B(\alpha) \quad \forall\bar{\lambda}_1 \tag{26}$$

$$B(\alpha) = \min_{\alpha\in(0,1)}\max\left(B_1(\alpha), B_2(\alpha)\right). \tag{27}$$

Notice that the minimum with respect to α depends on $\bar{r}_k$ and, thus, implicitly on N_t, N_r, N, and K. To obtain a closed-form solution, we can fix any $\alpha = \bar{\alpha}$ and take the maximum of $B_1(\bar{\alpha}), B_2(\bar{\alpha})$ as less tight upper bound, yet still conserving the scaling laws of interest; e.g., with $\bar{\alpha} = 1/2$ we get

$$\begin{aligned}
&\mathbb{P}\left\{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{H}}\mathbf{R}_k\right) < \mu_2\mathrm{tr}\left(\mathbf{C}_{N_t}^{*}\mathbf{R}_k\right)\right\} \\
&\quad\leq \max\left(8\mu_2^2, 16\left(\bar{r}_k - 1\right)^2\mu_2^4\right).
\end{aligned} \tag{28}$$

With these intermediate results, we now turn our attention back to $\mathcal{P}_2$ of (16). Substituting (21) and (26) into (16) and assuming for simplicity of notation $\mathrm{rank}\,(\mathbf{R}_k) = \mathrm{rank}\,(\mathbf{S}_n)\ \forall k\ \forall n \leq N$,[1] implying $\bar{r}_k = \tilde{r}_n = \bar{r}\ \forall k\ \forall n \leq N$, and $\bar{r}_{N+1} = \min\left(r^*, N_t\right)$, corresponding to $\mathbf{S}_{N+1}$, we get

[1]This is fulfilled when channels are not rank deficient.

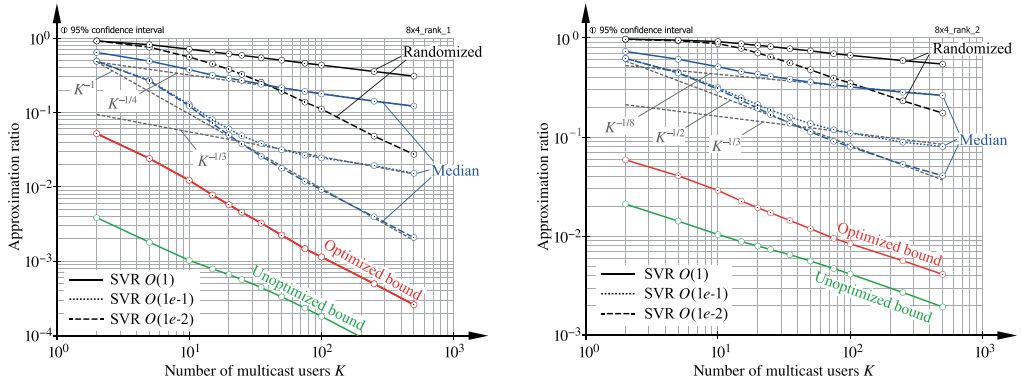


Fig. 2. Approximation ratio of rank one (left) and rank two (right) BF versus number of multicast users K shown for three different SVRs.

$$\mathcal{P}_2 \geq 1 - KB(\alpha) - 2\left(N\bar{r} + \bar{r}_{N+1}\right) e^{-\gamma_2} = U(\mu_2, \gamma_2, \alpha). \quad (29)$$

The largest lower bound on ρ_2 achievable with positive probability can be found by maximizing the ratio μ_2/γ_2, while constraining $1 \geq U(\mu_2, \gamma_2, \alpha) > 0$. This, however, leads to a nonconvex optimization problem, for which we were not able to find an efficient solution. Yet, in our simulations, we applied an exhaustive search to determine this *optimized lower bound*.

We can still obtain a less tight *unoptimized* closed-form *lower bound* by utilizing (28)

$$\mathcal{P}_2 \geq 1 - K \max\left(8\mu_2^2, 16(\bar{r}-1)^2\mu_2^4\right) - 2(N\bar{r} + \bar{r}_{N+1})e^{-\gamma_2} = \mathcal{P}_2^{(\mathrm{lb})}. \quad (30)$$

Consider first $K \max\left(8\mu_2^2, 16(\bar{r}-1)^2\mu_2^4\right)$ and assume $\bar{r}$ as fixed. If the first factor in the maximization is active, μ_2 must scale with $K^{-1/2}$ to obtain a bound that is independent of K, whereas scaling with $K^{-1/4}$ is sufficient if the second factor is active. However, it can easily be seen that with growing K for both choices of μ_2 the first factor in the maximization is active, since the second factor decreases faster with K. With the following choice of μ_2, the factor $8\mu_2^2$ is active for all K

$$\mu_2 = \frac{1}{3\sqrt{2}(\bar{r}-1)\sqrt{K}}, \quad \bar{r} > 1. \quad (31)$$

Moreover, $K \max\left(8\mu_2^2, 16(\bar{r}-1)^2\mu_2^4\right) \leq \frac{4}{9}$. If N_t, N_r scale with the number of users K, N (as in massive MIMO) such that $\bar{r} = r^* < \min(N_t, N_r)$, $\bar{r}$ can at worst grow proportionally to the square-root of the problem dimensions according to (8); then (31) would actually scale inversely with K. Yet in this case, we can insert $r^{(\mathrm{ub})}$ into (30) to obtain an even looser lower bound on $\mathcal{P}_2$; then the choice

$$\mu_2 = \frac{1}{3\sqrt{2}\sqrt{K+N+1}} \quad (32)$$

activates the first factor $8\mu_2^2$ for all K and N; hence, scaling inversely proportional to $\sqrt{K}$ is still retained. Both factors (31) and (32) deliver positive probability $\mathcal{P}_2$, which one is favourable (delivers the larger μ_2) depends on $K, N, \bar{r}$.

Consider next the term $2(N\bar{r} + \bar{r}_{N+1})e^{-\gamma_2}$; with

$$\gamma_2 = \log\left(2(N\bar{r} + \bar{r}_{N+1})c_\gamma\right) \quad (33)$$

the term gets independent of $N, \bar{r}, \bar{r}_{N+1}$. Furthermore, with the choice $c_\gamma = 9/4$ we have

$$\mathcal{P}_2 \geq \mathcal{P}_2^{(\mathrm{lb})} \geq \frac{1}{9} \quad (34)$$

and hence achieve positive probability irrespective of $N, K, \bar{r}, \bar{r}_{N+1}$. Thus, using (31), (32), and (33), the approximation ratio μ_2/γ_2 is achieved with probability at least $1 - (8/9)^L$ when L independent Gaussian randomizations are evaluated.

In case of BF, we can follow a very similar derivation to arrive at the lower bound

$$\mathcal{P}_1 \geq 1 - K \max\left(\frac{4}{3}\mu_1, 16(\bar{r}-1)^2\mu_1^2\right) - (N\bar{r} + \bar{r}_{N+1})e^{-\gamma_1} = \mathcal{P}_1^{(\mathrm{lb})}. \quad (35)$$

To achieve positive probability $\forall K$, μ_1 must scale inversely proportional to K and γ_1 must scale logarithmically with N, similar to (33). Thus, we expect that performance degradation with K is more severe in case of BF than with ABF.

V. Numerical Experiments and Conclusion

In this section, we evaluate the proposed worst-case bounds through Monte Carlo simulations. We consider a transmitter with $N_t = 8$ transmit antennas and users with $N_r = 4$ receive antennas. We evaluate the approximation ratio with varying number of multicast users K, fixing the number of other users to $N = 10$, their leakage constraints to $\Gamma_n = 10^{-2}\ \forall n$, and the noise power to $\sigma_k^2 = 10^{-2}\ \forall k$.

As channel model, we consider Kronecker correlated Gaussian channels [27] with

$$\mathbb{E}\left\{\mathrm{vec}\left(\mathbf{H}_k\right)\mathrm{vec}\left(\mathbf{H}_k\right)^{\mathrm{H}}\right\} = \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{G}_n\right)\mathrm{vec}\left(\mathbf{G}_n\right)^{\mathrm{H}}\right\} = \mathbf{K}_{\mathrm{r}} \otimes \mathbf{K}_{\mathrm{t}}, \quad [\mathbf{K}_{\mathrm{r}}]_{i,j} = \beta^{\frac{|i-j|}{N_r-1}}, \quad [\mathbf{K}_{\mathrm{t}}]_{i,j} = \beta^{\frac{|i-j|}{N_t-1}}. \quad (36)$$

We select the parameter β such that we achieve on average a specific ratio of largest to smallest singular value of the channel matrices, denoted singular value ratio (SVR).

In Fig. 2, we show the achieved approximation ratios with $L = 1000$ randomizations and the corresponding worst-case bounds. We evaluate the approximation ratio for BF and ABF only when $r^* > \{1, 2\}$, respectively, as otherwise $\mathbf{C}_{N_t}^* = \{\mathbf{C}_1^*, \mathbf{C}_2^*\}$ rendering randomization unnecessary. We show the performance attained with the best of L randomizations as well as their median value. Our proposed worst-case bounds scale with K^{-1} and $K^{-1/2}$ for BF and ABF. The actual achieved approximation ratio strongly depends on the SVR. For small SVR, scaling follows the predictions of our bounds. Yet, if the SVR lies in the order of one much better scaling is obtained; we conjecture from further ample simulations that the scaling with SVR $\sim \mathcal{O}(1)$ is proportional to K^{-1/N_r} and $K^{-1/(2N_r)}$ for BF and ABF, respectively. Our worst-case approximation bounds, however, are valid for any positive semidefinite channel matrices; thus, they have to cover the case SVR $\to 0$ as well.

REFERENCES

[1] S. Schwarz and M. Rupp, "Transmit optimization for the MISO multicast interference channel," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4936–4949, Dec. 2015.

[2] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[3] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Mar. 2008.

[4] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, May 2013.

[5] 3GPP, "TSG RAN; Introduction of the multimedia broadcast multicast service (MBMS) in the radio access network (RAN); stage 2 (release 6)," [Online]. Available: http://www.3gpp.org/DynaReport/25346.htm, Sep. 2006

[6] 3GPP, "Techn. Spec. Group Radio Access Network; (E-UTRA) and E-UTRAN); Overall description; Stage 2 (release 10)," Dec. 2010 [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36300.htm.

[7] Ericsson, "LTE broadcast: A revenue enabler in the mobile media era," white paper, Feb. 2013

[8] D. Valerio, F. Ricciato, P. Belanovic, and T. Zemen, "UMTS on the road: Broadcasting intelligent road safety information via MBMS," in *Proc. 67th IEEE Veh. Technol. Conf. (VTC'08-Spring)*, Singapore, May 2008, pp. 3026–3030.

[9] I. Safiulin, S. Schwarz, and M. Rupp, "System level simulation of LTE MBSFN networks with high mobility users," in *Proc. 22nd Int. Conf. Syst. Signals Image Process.*, London, U.K., 2015, pp. 21–24.

[10] T. Lohmar, M. Slssingar, S. Puustinen, and V. Kenehan, "Delivering content with LTE broadcast," *Ericsson Rev. Commun. Technol. J.*, vol. 2013, no. 1, pp. 1–8, Feb. 2013.

[11] S. Schwarz, T. Philosof, and M. Rupp, "Leakage-based multicast transmit beamforming," in *Proc. Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 2405–2410.

[12] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[13] Z.-Q. Luo and T.-H. Chang, "SDP relaxation of homogeneous quadratic optimization: Approximation and applications," in *Convex Optimization in Signal Processing and Communications*, D. Palomar and Y. Eldar, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2010

[14] K. Phan, S. Vorobyov, N. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via QoS-aware secondary multicast beamforming," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2323–2335, Jun. 2009.

[15] Y. Huang, Q. Li, W.-K. Ma, and S. Zhang, "Robust multicast beamforming for spectrum sharing-based cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 527–533, Jan. 2012.

[16] A. H. Phan, H. D. Tuan, H. H. Kha, and D. T. Ngo, "Nonsmooth optimization for efficient beamforming in cognitive radio multicast transmission," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2941–2951, Jun. 2012.

[17] S. Ji, S. Wu, A.-C. So, and W.-K. Ma, "Multi-group multicast beamforming in cognitive radio networks via rank-two transmit beamformed Alamouti space-time coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2013, pp. 4409–4413.

[18] S. Yang, T. Lv, and L. Hanzo, "Semidefinite programming relaxation based virtually antipodal detection for MIMO systems using gray-coded high-order QAM," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1667–1677, May 2013.

[19] G. Wang, "A semidefinite relaxation method for energy-based source localization in sensor networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2293–2301, Jun. 2011.

[20] Z. Luo, N. D. Sidiropoulos, P. Tseng, and S. Zhang, "Approximation bounds for quadratic optimization with homogeneous quadratic constraints," *SIAM J. Optim.*, vol. 18, no. 1, pp. 1–28, 2007.

[21] T.-H. Chang, Z.-Q. Luo, and C.-Y. Chi, "Approximation bounds for semidefinite relaxation of max-min-fair multicast transmit beamforming problem," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3932–3943, Aug. 2008.

[22] S. Wu, A.-C. So, and W.-K. Ma, "Rank-two transmit beamformed Alamouti space-time coding for physical-layer multicasting," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar. 2012, pp. 2793–2796.

[23] S. Wu, W.-K. Ma, and A.-C. So, "Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time coding," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4230–4245, Sep. 2013.

[24] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[25] A. M.-C. So, Y. Ye, and J. Zhang, "A unified theorem on SDP rank reduction," *Math. Oper. Res.*, vol. 33, no. 4, pp. 1–11, 2008.

[26] Y. Huang and D. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 664–678, Feb. 2010.

[27] C. Oestges, "Validity of the Kronecker model for MIMO correlated channels," in *Proc. IEEE 63rd Veh. Technol. Conf. (VTC'06-Spring)*, May 2006, vol. 6, pp. 2818–2822.

6. Conclusions

The *Society in Motion* poses challenging requirements on mobile wireless communication systems: It demands large network capacity and dependable wireless connectivity not only for quasi-static situations, but even when large numbers of users (humans and machines) are moving through the network coverage area at potentially very high velocity. In this thesis, I addressed these demands by developing accurate and rate-distortion efficient CSIT acquisition techniques and by utilizing this CSIT for spectrally efficient and reliable multi-user, multi-antenna and multi-point transceiver optimization.

Within the context of CSIT acquisition, I concentrated mainly on CSI quantization for limited feedback systems. I succeeded in devising quantization methods that enable efficient CSIT acquisition by relying on the following three fundamental principles:

- Dimensionality reduction by projection onto a topological manifold that captures the essential algebraic structure underlying the considered transmission strategy.
- Predictive quantization to exploit the temporal channel correlation.
- Parametric channel decomposition to enable an accurate sparse approximation with a small number of expansion terms.

Building on top of these CSI feedback algorithms, I developed robust multi-antenna transceivers that are able to cope with the unavoidable CSIT imperfections that are caused, e.g., by estimation/quantization errors and channel aging. I obtained the following main insights concerning robust transceiver designs:

- It is essential to tailor the channel quality feedback to the applied precoding strategy and the achieved CSIT quality, to support reliable transmission rate adaptation and effective multi-user scheduling.
- Interference leakage-bounded angular-domain precoding is a potent solution for spectrally efficient and robust multi-user MIMO operation even at very high mobility. The method, however, requires relatively large antenna arrays with sufficient angular resolution to effectively separate different users, as well as, operation in a limited scattering propagation environment.

- In non-static scenarios in particular at higher carrier frequencies (e.g., in the mmWave band), it is reasonable to conservatively assume that the relative phases of the scattering components of the channel vary randomly; cf. TWDP fading for the special case of two scattering components. For the general case, my results showed that it is possible to reduce the outage probability by suitable beamforming, even with unknown relative phases of the scattering components; however, determining a globally outage-optimal beamforming solution is in general computationally hard.

In multi-cell systems, controlling the inter-cell-interference is vital to support large network capacities and dependable operation, especially at the cell edge. Convex optimization and relaxation tools are indispensable for multi-point transmission coordination and cooperation. I successfully applied such techniques to weighted sum-rate optimization for the MISO MCIC and to joint RRH assignment and coordinated beamforming in dDASs. Both problem are non-convex initially, but can be approximately solved by applying suitable relaxations:

- The weighted sum-rate optimization for the MISO MCIC can be locally solved by employing a dual gradient descent based on a Lagrangian relaxation.
- The joint RRH assignment and coordinated beamforming MISOCP can be approximately solved by applying an integer relaxation combined with randomized rounding.

The multicasting technology has applications in cellular-assisted vehicular communications, where it can provide a spectrally efficient solution for ITS. My investigations showed that, compared to unicasting, substantial transmission latency reductions can be achieved, which is a fundamental requirement when supporting road safety-relevant information exchange.

The combination of multi-point transmissions with dynamic network reconfiguration, as supported by dDASs and CRANs, exhibits great potential for improving the scalability of mobile networks and exploiting macroscopic diversity. My analysis of dDASs demonstrated that a dynamic RRH assignment can significantly reduce outage probabilities as compared to static DASs, thus supporting enhanced dependability.

In conclusion, the scientific results achieved within this thesis pave the way towards mobile communication systems that are capable of providing high performance dependable wireless connectivity for the envisioned *Society in Motion*. To gauge the practical value of the developed techniques, I benchmarked them against standardized techniques within link and system level simulation environments, substantiating their potential. Nevertheless, some open research questions remain, in particular regarding the practical feasibility of the more computationally demanding methods. In this respect, further work is required to devise (sub-optimal) low complexity methods that draw on the insights obtained in this thesis to achieve a sensible performance-complexity trade-off.

A. List of Abbreviations

3GPP	Third Generation Partnership Project
2D	two dimensional
3D	three dimensional
4G	fourth generation
5G	fifth generation
ACK	acknowledgment
ADC	analog to digital converter
AMC	adaptive modulation and coding
ARQ	automatic repeat request
ASE	area spectral efficiency
AWGN	additive white Gaussian noise
BBU	base band unit
BC	broadcast channel
BD	block diagonalization
BEP	bit error probability
BLER	block error ratio
BICM	bit-interleaved coded modulation
BS	base station
CoMP	coordinated multi-point transmission
CP	cyclic prefix
CRAN	cloud radio access network
CQI	channel quality indicator
CRC	cyclic redundancy check
CS	compressed sensing
CSI	channel state information
CSIT	CSI at the transmitter
DAC	digital to analog converter
DAS	distributed antenna system
dDAS	dynamic distributed antenna system
DCT	discrete cosine transform
DFT	discrete Fourier transform
DoF	degrees of freedom

DPS	dynamic point selection
ESM	effective SINR mapping
FBMC	filter bank multi-carrier
FD-MIMO	full-dimension MIMO
FDD	frequency division duplex
FEC	forward error correction
FFR	fractional frequency reuse
FFT	fast Fourier transform
HARQ	hybrid automatic repeat request
IA	interference alignment
IC	interference channel
ICI	inter-carrier interference
ICIC	inter-cell interference coordination
eICIC	inter-cell interference coordination
IEEE	Institute of Electrical and Electronic Engineers
iid	independent and identically distributed
ISI	inter-symbol interference
ITS	intelligent transportation systems
JT	joint transmission
KLT	Karhunen Loeve transform
LOS	line of sight
LP	linear program
LTE	long term evolution
LTE-A	LTE-Advanced
MBMS	multimedia broadcast multicast service
MBSFN	multimedia broadcast single frequency network
MCIC	multicast interference channel
MCS	modulation and coding scheme
MEC	maximum eigen-mode combining
MIMO	multiple-input multiple-output
MIESM	mutual information ESM
MISO	multiple-input single-output
MISOCP	mixed-integer second order cone program
MERC	maximum expected achievable rate combining
MSE	mean squared error
MMSE	minimum mean squared error
mmWave	millimeter wave
NLOS	non line of sight
NACK	not ACK
NOMA	non-orthogonal multiple access

NP	non-deterministic polynomial-time
NR	new radio
OFDM	orthogonal frequency division multiplexing
CP-OFDM	cyclic prefix orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OOB	out-of-band
OMP	orthogonal matching pursuit
PAPR	peak to average power ratio
PDCCH	physical downlink control channel
PDP	power delay profile
PF	proportional fair
PMI	precoding matrix indicator
PSN	phase shifting network
PHY	physical layer
PWS	public warning system
QAM	quadrature amplitude modulation
QBC	quantization based combining
QCI	QoS class identifier
QCQP	quadratically constrained quadratic program
QoS	quality of service
RAB	robust adaptive beamforming
RBD	regularized block diagonalization
RE	resource element
RF	radio frequency
RI	rank indicator
RRH	remote radio head
RRU	remote radio unit
SCFDMA	single carrier frequency division multiple access
SCM	spatial channel model
SDMA	space division multiple access
SDP	semi-definite programming
SDR	semi-definite relaxation
SIC	successive interference cancellation
SINR	signal to interference and noise ratio
SISO	single-input single-output
SLNR	signal to leakage and noise ratio
SNR	signal to noise ratio
SQBC	subspace quantization-based combining
SVD	singular value decomposition
SWIPT	simultaneous wireless information and power transfer

TDD	time division duplex
TTI	transmission time interval
TWDP	two wave with diffuse power
UCIC	unicast interference channel
UE	user equipment
UFMC	universal filtered multi-carrier
V2V	vehicle to vehicle
V2X	vehicle to everything
VCCS	Vienna Cellular Communications Simulators
WINNER	Wireless World Initiative New Radio
ZF	zero forcing

B. Notation

The following notation is employed throughout this thesis:

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
$\|\mathbf{a}\|$	l_2 norm of a vector
$\|\mathbf{A}\|$	Frobenius norm of a matrix
$\|\mathcal{A}\|$	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
$\mathbb{E}\left(a\right)$	expected value of a
$\mathrm{tr}\left(\mathbf{A}\right)$	trace of a matrix
$\det\left(\mathbf{A}\right)$	determinant of a matrix
$\log_{\mathrm{b}}\left(a\right)$	base b logarithm of a
$\mathrm{span}\left(\mathbf{A}\right)$	space spanned by the columns of $\mathbf{A}$
$[\mathbf{A}]_{i:j}$	selects the i-th to j-th column of $\mathbf{A}$

Bibliography

[1] Ericsson, "Ericsson mobility report," white paper, June 2018.

[2] United Nations, "World urbanization prospects: The 2018 revision [key facts]," 2018, [Online]. https://esa.un.org/unpd/wup/Publications/Files/WUP2018-KeyFacts.pdf.

[3] Credo Business Consulting LLP and Siemens AG, "The mobility opportunity: Improving public transport to drive economic growth," 2014, [Online]. www.siemens.com/press/mobility-opportunity.

[4] G. Araniti, C. Campolo, M. Condoluci, A. Iera, and A. Molinaro, "LTE for vehicular networking: a survey," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 148–157, May 2013.

[5] C. Mecklenbräuker, A. Molisch, J. Karedal, F. Tufvesson, A. Paier, L. Bernado, T. Zemen, O. Klemp, and N. Czink, "Vehicular channel characterization and its implications for wireless system design and performance," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1189–1212, July 2011.

[6] S. Schwarz and M. Rupp, "Society in motion: Challenges for LTE and beyond mobile communications," *IEEE Communications Magazine, Feature Topic: LTE Evolution*, vol. 54, no. 5, pp. 76–83, 2016.

[7] S. Schwarz, T. Philosof, and M. Rupp, "Signal processing challenges in cellular assisted vehicular communications," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 47–59, March 2017.

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

[9] S. Caban, J. A. García Naya, and M. Rupp, "Measuring the physical layer performance of wireless communication systems," *IEEE Instrumentation & Measurement Magazine*, vol. 14, no. 5, pp. 8–17, 2011.

[10] M. Rupp, S. Schwarz, and M. Taranetz, *The Vienna LTE-Advanced Simulators: Up and Downlink, Link and System Level Simulation*, 1st ed. Springer, 2016.

[11] 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 (JASP) special issue on Reproducible Research*, vol. 2011, no. 1, pp. 1–14, 2011.

[12] 3GPP, "TSG RAN; study on 3D channel model for LTE (release 12)," http://www.3gpp.org/DynaReport/36873.htm, Mar. 2015.

[13] J. Baillieul, G. Grenier, and G. Setti, "Reflections on the future of research curation and research reproducibility," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 779–783, May 2018.

[14] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 15)," June 2018, [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36211.htm.

[15] 3GPP, "Technical Specification Group Radio Access Network; NR; Physical Channels and Modulation (Release 15)," June 2018, [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/38211.htm.

[16] IEEE, "Std 802.11-2016 – ieee standard for information technology," Dec 2016, [Online]. Available: https://ieeexplore.ieee.org/servlet/opac?punumber=7786993.

[17] E. Zöchmann, S. Schwarz, S. Pratschner, L. Nagel, M. Lerch, and M. Rupp, "Exploring the physical layer frontiers of cellular uplink – The Vienna LTE-A uplink simulator," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 118, 2016.

[18] L. Marijanovic, S. Schwarz, and M. Rupp, "Inter-carrier interference of multiple access UFMC with flexible subcarrier spacings," in *25th European Signal Processing Conference*, Kos, Greece, Aug. 2017, pp. 1–5.

[19] R. Nissel, S. Schwarz, and M. Rupp, "Filter bank multicarrier modulation schemes for future mobile communications," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1768–1782, Aug 2017.

[20] V. D. Nguyen and H.-P. Kuchenbecker, "Intercarrier and intersymbol interference analysis of OFDM systems on time-invariant channels," in *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, Sept 2002, pp. 1482–1487.

[21] M. Russell and G. L. Stuber, "Interchannel interference analysis of OFDM in a mobile environment," in *IEEE 45th Vehicular Technology Conference*, vol. 2, Jul 1995, pp. 820–824.

[22] Y.-S. Choi, P. Voltz, and F. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 49, no. 8, pp. 1375–1387, 2001.

[23] T. Wang, J. Proakis, E. Masry, and J. Zeidler, "Performance degradation of OFDM systems due to Doppler spreading," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1422–1432, 2006.

[24] A. F. Molisch, *Wireless Communications*. Wiley-IEEE Press, Dec. 2010.

[25] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, Oct 2017.

[26] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transaction on Information Theory*, vol. 44, No. 3, May 1998.

[27] J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," *RLE Technical Reports*, vol. 216, 1952.

[28] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *16th IEEE International Conference on Electronics, Circuits and Systems*, Dec 2009, pp. 683–686.

[29] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8131–8146, Dec 2017.

[30] L. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE Journal on Selected Topics in Signal Processing*, vol. pp, pp. 1–15, 2018.

[31] M. R. Khanzadi, R. Krishnan, and T. Eriksson, "Estimation of phase noise in oscillators with colored noise sources," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2160–2163, November 2013.

[32] C.-P. Liang, J.-H. Jong, W. E. Stark, and J. R. East, "Nonlinear amplifier effects in communications systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1461–1466, Aug 1999.

[33] C. Rapp, "Effects of HPA-nonlinearity on a 4-DPSK/OFDM-signal for a digital sound broadcasting signal," in *ESA Special Publication*, P. S. Weltevreden, Ed., vol. 332, Oct. 1991.

[34] S. Buzzi, C. D'Andrea, T. Foggi, A. Ugolini, and G. Colavolpe, "Single-carrier modulation versus OFDM for millimeter-wave wireless MIMO," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1335–1348, March 2018.

[35] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, January 2016.

[36] E. Zöchmann, S. Schwarz, and M. Rupp, "Comparing antenna selection and hybrid precoding for millimeter wave wireless communications," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janerio, Brazil, July 2016, pp. 1–5.

[37] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog-digital transceivers for single- and multi-carrier large antenna array systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1980–1995, Sept 2017.

[38] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[39] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in *Proc. 71st Vehicular Technology Conference VTC2010-Spring*, 2010.

[40] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: An open-source framework," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, pp. 498–513, Feb 2011.

[41] A. Virdis, G. Stea, and G. Nardini, "Simulating LTE/LTE-advanced networks with SimuLTE," in *Simulation and Modeling Methodologies, Technologies and Applications*, M. S. Obaidat, T. Ören, J. Kacprzyk, and J. Filipe, Eds. Cham: Springer International Publishing, 2015, pp. 83–105.

[42] 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.

[43] L. Wan, S. Tsai, and M. Almgren, "A Fading-Insensitve Performance Metric for a Unified Link Quality Model," in *Proc. IEEE Wireless Communications & Networking Conference WCNC*, vol. 4, 2006, pp. 2110–2114.

[44] 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.

[45] 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.

[46] 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.

[47] S. Schwarz, "Limited feedback transceiver design for downlink MIMO OFDM cellular networks," Ph.D. dissertation, Technische Universität Wien, 2013.

[48] F. Ademaj, M. Taranetz, and M. Rupp, "3GPP 3D MIMO channel model: a holistic implementation guideline for open source simulation tools," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–14, 2016.

[49] W. I. WP1, "Winner ii channel models," http://www.3gpp.org/DynaReport/36996.htm, Sept. 2007.

[50] 3GPP, "TSG RAN; spatial channel model for multiple input multiple output (MIMO) simulations," http://www.3gpp.org/DynaReport/36996.htm, Sept. 2003.

[51] 3GPP, "TSG RAN; study on channel model for frequency spectrum above 6GHz," http://www.3gpp.org/DynaReport/38900.htm, Sept. 2016.

[52] H. Claussen, "Efficient modelling of channel maps with correlated shadow fading in mobile radio systems," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1, Sept 2005, pp. 512–516.

[53] W. Ni, W. Zou, and H. Wang, "Modeling of spatially cross-correlated shadow fading in distributed radio access networks," in *IEEE International Conference on Communications*, May 2008, pp. 4472–4476.

[54] S. Schwarz, I. Safiulin, T. Philosof, and M. Rupp, "Gaussian modeling of spatially correlated LOS/NLOS maps for mobile communications," in *IEEE 84th Vehicular Technology Conference*, Montreal, Canada, Sep. 2016, pp. 1–5.

[55] F. Ademaj, M. K. Müller, S. Schwarz, and M.Rupp, "Modeling of spatially correlated geometry-based stochastic channels," in *IEEE 86th Vehicular Technology Conference*, Toronto, Canada, Sept. 2017, pp. 1–6.

[56] F. Ademaj, S. Schwarz, K. Guan, and M. Rupp, "Ray-tracing based validation of spatial consistency for geometry-based stochastic channels," in *IEEE 88th Vehicular Technology Conference*, Chicago, USA, Aug. 2018, pp. 1–5.

[57] S. Pratschner, B. Tahir, L. Marijanovic, M. Mussbah, K. Kirev, R. Nissel, S. Schwarz, and M. Rupp, "Versatile mobile communications simulation: The Vienna 5G link level simulator," *EURASIP Journal on Wireless Communications and Networking*, pp. 1–18, 2018, accepted for publication.

[58] M. Müller, F. Ademaj, T. Dittrich, A. Fastenbauer, B. Ramos Elbal, A. Nabavi, L. Nagel, S. Schwarz, and M. Rupp, "Flexible multi-node simulation of cellular mobile communications: The Vienna 5G system level simulator," *EURASIP Journal on Wireless Communications and Networking*, pp. 1–17, 2018, accepted for publication.

[59] 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.

[60] M. Taranetz, T. Blazek, T. Kropfreiter, M. Müller, S. Schwarz, and M. Rupp, "Runtime precoding: Enabling multipoint transmission in LTE-advanced system-level simulations," *IEEE Access*, vol. 3, pp. 725–736, 2015.

[61] M. Simko, P. S. R. Diniz, Q. Wang, and M. Rupp, "Adaptive pilot-symbol patterns for MIMO OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4705–4715, September 2013.

[62] L. Marijanovic, S. Schwarz, and M. Rupp, "Optimal numerology in OFDM systems based on imperfect channel knowledge," in *IEEE 87th Vehicular Technology Conference*, Porto, Portugal, Jun. 2018, pp. 1–5.

[63] M. Simko, S. Pendl, S. Schwarz, Q. Wang, J. Ikuno, and M. Rupp, "Optimal pilot symbol power allocation in LTE," in *IEEE Vehicular Technology Conference (VTC Fall)*, Sept. 2011, pp. 1–5.

[64] P. W. C. Chan, E. S. Lo, R. R. Wang, E. K. S. Au, V. K. N. Lau, R. S. Cheng, W. H. Mow, R. D. Murch, and K. B. Letaief, "The evolution path of 4G networks: FDD or TDD?" *IEEE Communications Magazine*, vol. 44, no. 12, pp. 42–50, Dec 2006.

[65] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[66] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[67] M. Petermann, M. Stefer, F. Ludwig, D. Wubben, M. Schneider, S. Paul, and K. Kammeyer, "Multi-user pre-processing in multi-antenna OFDM TDD systems with non-reciprocal transceivers," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3781–3793, September 2013.

[68] D. Liu, W. Ma, S. Shao, Y. Shen, and Y. Tang, "Performance analysis of TDD reciprocity calibration for massive MU-MIMO systems with ZF beamforming," *IEEE Communications Letters*, vol. 20, no. 1, pp. 113–116, Jan 2016.

[69] W. Ma, D. Liu, Y. Liu, W. Pan, and H. Zhao, "On the capacity of ZF beamforming in massive MIMO systems with imperfect reciprocity calibration," in *IEEE Globecom Workshops*, Dec 2017, pp. 1–5.

[70] K. Truong and R. Heath, "Effects of channel aging in massive MIMO systems," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, Aug 2013.

[71] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1428–1442, Feb 2017.

[72] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, August 2011.

[73] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO – 5G system," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 905–923, 2016.

[74] H. Ahmadi, A. Farhang, N. Marchetti, and A. MacKenzie, "A game theoretic approach for pilot contamination avoidance in massive MIMO," *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 12–15, Feb 2016.

[75] V. Saxena, G. Fodor, and E. Karipidis, "Mitigating pilot contamination by pilot reuse and power control schemes for massive MIMO systems," in *IEEE 81st Vehicular Technology Conference*, May 2015, pp. 1–6.

[76] D. Samardzija and N. Mandayam, "Unquantized and uncoded channel state information feedback in multiple-antenna multiuser systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1335–1345, July 2006.

[77] O. E. Ayach and R. W. Heath, "Interference alignment with analog CSI feedback," in *Military communications conference (MILCOM)*, Oct 2010, pp. 1644–1648.

[78] S. Schwarz, "Limited feedback for 4G and beyond," in *Advances in Mobile Computing and Communications: 4G and Beyond*, M. Bala Krishna and J. Lloret Mauri, Eds. CRC Press Taylor & Francis Group, 2016.

[79] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 461 – 471, Feb. 2004.

[80] E. Zöchmann, S. Caban, C. F. Mecklenbräuker, S. Pratschner, M. Lerch, S. Schwarz, and M. Rupp, "Better than Rician: Modelling millimetre wave channels as Two-Wave with Diffuse Power," *ArXiv e-prints*, Apr. 2018, under review at EURASIP JWCN.

[81] M. Trivellato, F. Boccardi, and F. Tosate, "User selection schemes for MIMO broadcast channels with limited feedback," in *65th IEEE Vehicular Technology Conference, Spring 2007*, Dublin, April 2007.

[82] A. Durán, M. Toril, F. Ruiz, and A. Mendo, "Self-optimization algorithm for outer loop link adaptation in LTE," *IEEE Communications Letters*, vol. 19, no. 11, pp. 2005–2008, Nov 2015.

[83] B. Clerckx and C. Oestges, *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*, ser. Academic Press. Elsevier, 2013.

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

[85] G. Caire, G. Taricco, and E. Biglieri, "Capacity of bit-interleaved channels," *Electron. Lett.*, vol. 32, issue 12, pp. 1060–1061, June 1996.

[86] D. Love, R. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, Oct. 2003.

[87] 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.

[88] A. D. Dabbagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2190–2202, May 2006.

[89] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.

[90] 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.

[91] V. Raghavan, R. Heath, and A. Sayeed M., "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1298–1310, Sept. 2007.

[92] IEEE, "IEEE Standard 802.16-2004," October 2004.

[93] C. Sun, J. Cheng, and T. Ohira, *Handbook on Advancements in Smart Antenna Technologies for Wireless Networks*. Hershey, PA: Information Science Reference - Imprint of: IGI Publishing, 2008.

[94] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, Nov. 1999.

[95] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE Journal Sel. Areas in Communications*, 2007.

[96] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1473 –1482, Oct. 2008.

[97] M. Rezaeekheirabadi and M. Guillaud, "Limited feedback for interference alignment in the K-user MIMO interference channel," in *Proc. Information Theory Workshop*, Lausanne, Schweiz, September 2012, pp. 1–5.

[98] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, p. 5, Nov. 2006.

[99] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, Oct. 2008.

[100] S. Schwarz and M. Rupp, "Subspace quantization based combining for limited feedback block-diagonalization," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5868–5879, 2013.

[101] S. Schwarz and M. Rupp, "Maximum expected achievable rate combining for limited feedback block-diagonalization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Brisbane, Australia, April 2015, pp. 1–5.

[102] S. Schwarz and M. Rupp, "Subspace versus eigenmode quantization for limited feedback block-diagonalization," in *6th International Symposium on Communications, Control and Signal Processing*, Athens, Greece, May 2014, pp. 1–4.

[103] 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.

[104] K. Gomadam, V. Cadambe, and S. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-dec. 4 2008, pp. 1 –6.

[105] M. Sadek, A. Tarighat, and A. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[106] C.-B. Chae, D. Mazzarese, N. Jindal, and R. Heath Jr., "Coordinated beamforming with limited feedback in the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1505–1515, October 2008.

[107] D. Sacristan-Murga and A. Pascual-Iserte, "Differential feedback of MIMO channel Gram matrices based on geodesic curves," *IEEE Trans. on Wireless Communications*, vol. 9, no. 12, pp. 3714–3727, 2010.

[108] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," *IEEE Transactions on Signal Processing*, vol. 61, no. 15, pp. 3908–3917, Aug 2013.

[109] 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.

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

[111] 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, pp. 5021 – 5025.

[112] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov 1993, pp. 40–44.

[113] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge Univ. Press, 2012.

[114] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: A compressive sensing approach," *IEEE Signal Processing Magazine*, vol. 35, no. 3, pp. 40–58, May 2018.

[115] P. Kuo, H. T. Kung, and P. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *IEEE Wireless Communications and Networking Conference*, April 2012, pp. 492–497.

[116] J. Joung and S. Sun, "SCF: sparse channel-state-information feedback using Karhunen-Loeve transform," in *IEEE Globecom Workshops*, Dec 2014, pp. 314–319.

[117] M. S. Sim, J. Park, C. Chae, and R. W. Heath, "Compressed channel feedback for correlated massive MIMO systems," *Journal of Communications and Networks*, vol. 18, no. 1, pp. 95–104, Feb 2016.

[118] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, A. Molisch, C. Oestges, and H. Özcelik, "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, p. 19, 2007.

[119] H. Özcelik, "Indoor MIMO channel models," 2004, Wien, Technische Universität, Dissertation, 2005.

[120] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[121] S. Schwarz and M. Rupp, "Limited feedback based double-sided full-dimension MIMO for mobile backhauling," in *50th Asilomar Conference on Signals Systems and Computers*, Asilomar, CA, 2016.

[122] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "Quadriga: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, June 2014.

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

[124] S. Schwarz and M. Rupp, "Predictive quantization on the Stiefel manifold," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 234–238, 2015.

[125] S. Schwarz, "Robust full-dimension MIMO transmission based on limited feedback angular-domain CSIT," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 1–20, Mar 2018.

[126] E. Zöchmann, S. Pratschner, S. Schwarz, and M. Rupp, "Limited feedback in OFDM systems for combating ISI/ICI caused by insufficient cyclic prefix length," in *48th Asilomar Conference on Signals, Systems and Computers*, Asilomar, CA, November 2014, pp. 1–5.

[127] S. Homayouni, S. Schwarz, M. Müller, and M. Rupp, "CQI mapping optimization in spatial wireless channel prediction," in *IEEE 87th Vehicular Technology Conference*, Porto, Portugal, Jun. 2018, pp. 1–5.

[128] S. Homayouni, S. Schwarz, and M. Rupp, "On CQI estimation for mobility and correlation properties of Gaussian process regression," in *IEEE 88th Vehicular Technology Conference*, Chicago, USA, Aug. 2018, pp. 1–5.

[129] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 11–15, Jan 2007.

[130] R. Kwan, C. Leung, and J. Zhang, "Multiuser Scheduling on the Downlink of an LTE Cellular System," *Research Letters in Communications*, 2008, Hindawi Publishing Corporation.

[131] L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless networks," in *Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, June 2002, pp. 763–772.

[132] 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, September 1984.

[133] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, 1997.

[134] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *INFOCOM, 2010 Proceedings IEEE*, San Diego, CA, March 2010.

[135] S. Schwarz, C. Mehlführer, and M. Rupp, "Throughput maximizing multiuser scheduling with adjustable fairness," in *International Conference on Communications ICC 2011*, Kyoto, Japan, June 2011, pp. 1–5.

[136] E. B. Rodrigues and F. Casadevall, "Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation," *IEEE Communications Magazine*, vol. 49, no. 9, pp. 90–98, September 2011.

[137] A. Asadi and V. Mancuso, "A survey on opportunistic scheduling in wireless communications," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1671–1688, Fourth 2013.

[138] Technical Specification Group Services and System Aspects, "Policy and charging control architecture," 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 23.203 Version 15.3.0, Jun. 2018.

[139] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, Feb 2001.

[140] R. Basukala, H. A. M. Ramli, and K. Sandrasegaran, "Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system," in *First Asian Himalayas International Conference on Internet*, Nov 2009, pp. 1–5.

[141] G. Piro, L. A. Grieco, G. Boggia, R. Fortuna, and P. Camarda, "Two-level downlink scheduling for real-time multimedia services in LTE networks," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 1052–1065, Oct 2011.

[142] M. Müller, S. Schwarz, and M. Rupp, "QoS investigation of proportional fair scheduling in LTE networks," in *IFIP Wireless Days (WD)*, Nov 2013, pp. 1–4.

[143] V. S. Annapureddy, D. V. Marathe, T. R. Ramya, and S. Bhashyam, "Outage probability of multiple-input single-output (MISO) systems with delayed feedback," *IEEE Transactions on Communications*, vol. 57, no. 2, pp. 319–326, February 2009.

[144] C. Guthy, J. Brehmer, and W. Utschick, "Physical layer characterization of a multi-user MISO system by efficient outage probabilities," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, Sept 2005, pp. 2171–2175.

[145] E. Stathakis, J. Jalden, L. K. Rasmussen, and M. Skoglund, "Outage region characterization for beamforming in MISO interference networks with imperfect CSI," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2378–2382, Dec 2015.

[146] S. Schwarz, "Outage-based admission control for multi-user MISO transmission with imperfect CSIT," in *IEEE 86th Vehicular Technology Conference*, Toronto, Canada, Sept. 2017, pp. 1–5.

[147] G. L. Turin, "The characteristic function of Hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, no. 1/2, pp. 199–201, 1960.

[148] P. Bello, "Binary error probabilities over selectively fading channels containing specular components," *IEEE Transactions on Communication Technology*, vol. 14, no. 4, pp. 400–406, August 1966.

[149] J. Proakis, "On the probability of error for multichannel reception of binary signals," *IEEE Transactions on Communication Technology*, vol. 16, no. 1, pp. 68–71, February 1968.

[150] D. Raphaeli, "Distribution of noncentral indefinite quadratic forms in complex normal variables," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 1002–1007, May 1996.

[151] T. T. Tjhung, C. C. Chai, and X. Dong, "Outage probability for lognormal-shadowed Rician channels," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 2, pp. 400–407, May 1997.

[152] H. Ji, Y. Kim, J. Lee, E. Onggosanusi, Y. Nam, J. Zhang, B. Lee, and B. Shim, "Overview of full-dimension MIMO in LTE-advanced pro," *IEEE Communications Mag.*, vol. PP, no. 99, pp. 2–11, October 2016.

[153] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4445–4459, July 2017.

[154] O. E. Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 626–636, February 2012.

[155] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[156] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4847–4861, Nov 2016.

[157] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4418–4431, July 2012.

[158] P. de Kerret, D. Gesbert, J. Zhang, and P. Elia, "Optimally bridging the gap from delayed to perfect CSIT in the K-user MISO BC," in *2016 IEEE Information Theory Workshop*, Sept 2016, pp. 300–304.

[159] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2228–2240, August 2011.

[160] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, Feb 2003.

[161] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1931–1946, July 2004.

[162] A. Pascual-Iserte, D. P. Palomar, A. I. Perez-Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 346–360, Jan 2006.

[163] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3086–3100, Aug 2009.

[164] S. Loyka and C. D. Charalambous, "On the compound capacity of a class of MIMO channels subject to normed uncertainty," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2048–2063, April 2012.

[165] J. Wang, M. Bengtsson, B. Ottersten, and D. P. Palomar, "Robust MIMO precoding for several classes of channel uncertainty," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3056–3070, June 2013.

[166] N. Vucic and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 714–725, Feb 2009.

[167] M. B. Shenouda and T. N. Davidson, "Convex conic formulations of robust downlink precoder designs with quality of service constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 714–724, Dec 2007.

[168] M. Payaro, A. Pascual-Iserte, and M. A. Lagunas, "Robust power allocation designs for multiuser and multiantenna downlink communication systems through convex optimization," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1390–1401, September 2007.

[169] X. He and Y. Wu, "Probabilistic QoS constrained robust downlink multiuser MIMO transceiver design with arbitrarily distributed channel uncertainty," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6292–6302, December 2013.

[170] E. Chiu, V. K. N. Lau, H. Huang, T. Wu, and S. Liu, "Robust transceiver design for K-pairs quasi-static MIMO interference channels via semi-definite relaxation," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3762–3769, December 2010.

[171] Y. Xie, C. N. Georghiades, and A. Arapostathis, "Minimum outage probability transmission with imperfect feedback for MISO fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1084–1091, May 2005.

[172] K. Wang, A. M. So, T. Chang, W. Ma, and C. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5690–5705, Nov 2014.

[173] A. Ben-Tal and A. Nemirovski, "On safe tractable approximations of chance-constrained linear matrix inequalities," *Mathematics of Operations Research*, vol. 34, no. 1, pp. 1–25, 2009.

[174] J. Park, Y. Sung, D. Kim, and H. V. Poor, "Outage probability and outage-based robust beamforming for MIMO interference channels with imperfect channel state information," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3561–3573, October 2012.

[175] S. Schwarz, "Signal outage optimized beamforming for MISO TWDP fading channels," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Bologna, Italy, Sep. 2018, pp. 1–6.

[176] G. D. Durgin, T. S. Rappaport, and D. A. de Wolf, "New analytical models and probability density functions for fading in wireless communications," *IEEE Trans. on Comm.*, vol. 50, no. 6, pp. 1005–1015, Jun 2002.

[177] S. A. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints," *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5719–5724, Nov 2008.

[178] 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.

[179] S. Schwarz and M. Rupp, "Evaluation of distributed multi-user MIMO-OFDM with limited feedback," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6081–6094, Aug. 2014.

[180] S. Schwarz, "Outage investigation of beamforming over random-phase finite-scatterer MISO channels," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1029–1033, July 2017.

[181] W. Webb, *Wireless communications: The future*. Wiley, 2007.

[182] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, September 2008.

[183] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 20–27, June 2013.

[184] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Comm. Mag.*, vol. 52, no. 2, pp. 82–89, February 2014.

[185] S. F. Yunas, M. Valkama, and J. Niemelae, "Spectral and energy efficiency of ultra-dense networks under different deployment strategies," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 90–100, January 2015.

[186] X. Zhang and J. G. Andrews, "Downlink cellular network analysis with multi-slope path loss models," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1881–1894, May 2015.

[187] M. Ding, P. Wang, D. Lopez-Perez, G. Mao, and Z. Lin, "Performance impact of LoS and NLoS transmissions in dense cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2365–2380, March 2016.

[188] M. Taranetz, R. W. Heath, and M. Rupp, "Analysis of urban two-tier heterogeneous mobile networks with small cell partitioning," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 7044–7057, Oct 2016.

[189] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): a primer," *IEEE Network*, vol. 29, no. 1, pp. 35–41, Jan 2015.

[190] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[191] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, March 2017.

[192] G. N. Kamga, M. Xia, and S. Aïssa, "Spectral-efficiency analysis of massive MIMO systems in centralized and distributed schemes," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1930–1941, May 2016.

[193] M. M. U. Rahman, H. Ghauch, S. Imtiaz, and J. Gross, "RRH clustering and transmit precoding for interference-limited 5G CRAN downlink," in *2015 IEEE Globecom Workshops*, Dec 2015, pp. 1–7.

[194] H. Ghauch, M. M. U. Rahman, S. Imtiaz, and J. Gross, "Coordination and antenna domain formation in cloud-RAN systems," in *IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–7.

[195] O. Chabbouh, S. B. Rejeb, Z. Choukair, and N. Agoulmine, "A two-stage RRH clustering mechanism in 5G heterogeneous C-RAN," in *5th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2017)*, Evry, France, Jan. 2017.

[196] G. R. MacCartney, T. S. Rappaport, and A. Ghosh, "Base station diversity propagation measurements at 73 GHz millimeter-wave for 5G coordinated multipoint (CoMP) analysis," in *2017 IEEE Globecom Workshops*, Dec 2017, pp. 1–7.

[197] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37 328–37 348, 2018.

[198] T. D. Novlan, R. K. Ganti, A. Ghosh, and J. G. Andrews, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4294–4305, December 2011.

[199] Z. Xiong, M. Zhang, M. Baker, and H. Sun, "Dynamic ICIC in LTE-advanced networks for inter-cell interference mitigation," in *IEEE 82nd Vehicular Technology Conference*, Sept 2015, pp. 1–5.

[200] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 148–155, February 2012.

[201] R. Agrawal, A. Bedekar, R. Gupta, S. Kalyanasundaram, H. Kroener, and B. Natarajan, "Dynamic point selection for LTE-advanced: Algorithms and performance," in *IEEE Wireless Communications and Networking Conference*, April 2014, pp. 1392–1397.

[202] A. Zakrzewska, D. Lopez-Perez, S. Kucera, and H. Claussen, "Dual connectivity in LTE HetNets with split control- and user-plane," in *IEEE Globecom Workshops*, Dec 2013, pp. 391–396.

[203] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, Dec 2008.

[204] E. Jorswieck, E. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5292–5296, Oct 2008.

[205] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5450–5458, Oct 2010.

[206] X. Shang, B. Chen, and H. Poor, "Multiuser MISO interference channels with single-user detection: Optimality of beamforming and the achievable rate region," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4255–4273, July 2011.

[207] R. Mochaourab and E. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1128–1141, March 2011.

[208] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "On the complexity of optimal coordinated downlink beamforming," in *IEEE International Conference on Acoustics Speech and Signal Processing*, March 2010, pp. 3274–3277.

[209] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2508–2523, May 2012.

[210] L. Liu, R. Zhang, and K.-C. Chua, "Achieving global optimality for weighted sum-rate maximization in the K-user Gaussian interference channel with multiple antennas," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1933–1945, May 2012.

[211] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1435–1445, December 2010.

[212] J. Xu and R. He, "Leakage-based probabilistic robust beamforming over MISO interference channel," in *IEEE/CIC International Conference on Communications in China*, Oct 2017, pp. 1–5.

[213] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425 –3441, Aug. 2008.

[214] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3457 –3470, Aug. 2008.

[215] O. Gonzalez, C. Beltran, and I. Santamaria, "A feasibility test for linear interference alignment in MIMO channels with constant coefficients," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1840–1856, March 2014.

[216] S. W. Peters and R. W. Heath, "Interference alignment via alternating minimization," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 2445–2448.

[217] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, Nov 2009, pp. 1106–1110.

[218] S. W. Peters and R. W. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 206–218, Jan 2011.

[219] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 161–167, April 2013.

[220] M. Riaz, R. Nielsen, J. Pedersen, N. Prasad, and O. Madsen, "On radio over fiber for heterogeneous wireless networks," in *International Conf. on Wireless and Optical Communications Networks*, April 2009.

[221] K. Kerpez and S. Ariyavisitakul, "A radio access system with distributed antennas," in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 3, Nov. 1994.

[222] P. Chow, A. Karim, V. Fung, and C. Dietrich, "Performance advantages of distributed antennas in indoor wireless communication systems," in *IEEE 44th Vehicular Techn. Conf.*, vol. 3, Jun 1994.

[223] W. Choi and J. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. on Wireless Communications*, vol. 6, no. 1, Jan. 2007.

[224] R. Heath, Jr., T. Wu, Y. H. Kwon, and A. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Trans. on Signal Proc.*, vol. 59, no. 10, Oct. 2011.

[225] F. Yuan, S. Jin, K. K. Wong, J. Zhao, and H. Zhu, "Wireless information and power transfer design for energy cooperation distributed antenna systems," *IEEE Access*, vol. 5, pp. 8094–8105, 2017.

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

[227] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, Aug 2006.

[228] X. Hong, Y. Jie, C. Wang, J. Shi, and X. Ge, "Energy-spectral efficiency trade-off in virtual MIMO cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2128–2140, October 2013.

[229] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2112–2127, October 2013.

[230] Ericsson, "LTE broadcast: a revenue enabler in the mobile media era," Feb. 2013, white paper.

[231] Deby Lewis (Verizon), "Verizon delivers LTE multicast over commercial 4G LTE network in Indy," May 21. 2014, press release.

[232] Valter D'Avino (Ericsson), "Ericsson and Polkomtel test LTE broadcast in Poland," Sept. 1. 2014, press release.

[233] Bill Smith (AT&T), "AT&T LTE broadcast demo at college football national championship showcases exciting network future," Jan. 16. 2015, press release.

[234] T. Lohmar, M. Slssingar, S. Puustinen, and V. Kenehan, "Delivering content with LTE broadcast," *Ericsson review: the communications technology journal since 1924*, vol. 2013, no. 1, Feb. 2013.

[235] D. Valerio, F. Ricciato, P. Belanovic, and T. Zemen, "UMTS on the Road: Broadcasting Intelligent Road Safety Information via MBMS," in *67th IEEE Vehicular Technology Conference (VTC2008-Spring)*, Singapore, May 2008, pp. 3026–3030.

[236] I. Safiulin, S. Schwarz, T. Philosof, and M. Rupp, "Latency and resource utilization analysis for V2X communication over LTE MBSFN transmission," in *Proceedings of the ITG Worshop on Smart Antennas*, Munich, Germany, March 2016, pp. 1–6.

[237] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, June 2006.

[238] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[239] A. Gershman, N. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, May 2010.

[240] T.-H. Chang, Z.-Q. Luo, and C.-Y. Chi, "Approximation bounds for semidefinite relaxation of max-min-fair multicast transmit beamforming problem," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3932–3943, Aug 2008.

[241] S. Wu, W.-K. Ma, and A.-C. So, "Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time coding," *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4230–4245, Sept 2013.

[242] Z. Xiang, M. Tao, and X. Wang, "Massive MIMO multicasting in noncooperative multicell networks," in *IEEE International Conference on Communications*, June 2014, pp. 4777–4782.

[243] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1268–1279, March 2008.

[244] B. Du, Y. Jiang, X. Xu, and X. Dai, "Transmit beamforming for MIMO multicast channels," in *IEEE International Conference on Communications*, June 2012, pp. 3800–3805.

[245] B. Du, X. Xu, Z. Zhang, and X. Dai, "Precoder design for dual-stream MIMO multicasting," in *International Conference on Wireless Communications and Signal Processing*, Oct 2013, pp. 1–6.

[246] Z. Zhang, X. Xu, B. Du, and Z. Chen, "A novel precoder design for MIMO multicasting with MMSE-DFE receivers," in *80th Vehicular Technology Conference*, Sept 2014, pp. 1–5.

[247] G. Dartmann, X. Gong, and G. Ascheid, "Low complexity cooperative multicast beamforming in multiuser multicell downlink networks," in *Sixth International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications*, June 2011, pp. 370–374.

[248] G. Dartmann and G. Ascheid, "Equivalent quasi-convex form of the multicast max-min beamforming problem," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4643–4648, Nov 2013.

[249] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 12–21, January 2013.

[250] S. Schwarz, T. Philosof, and M. Rupp, "Leakage-based multicast transmit beamforming," in *International Conference on Communications*, London, UK, June 2015, pp. 1–6.

[251] I. Safiulin, S. Schwarz, and M. Rupp, "Multicast beamforming capabilities of LTE MBSFN for V2X communications," in *IEEE 86th Vehicular Technology Conference*, Toronto, Canada, Sept. 2017, pp. 1–5.

[252] S. Schwarz, R. Heath, Jr., and M. Rupp, "Multiuser MIMO in distributed antenna systems with limited feedback," in *IEEE Globecom Workshops*, Dec 2012, pp. 546–551.

[253] S. Schwarz and M. Rupp, "Exploring coordinated multipoint beamforming strategies for 5G cellular," *IEEE Access*, vol. 2, pp. 930–946, 2014.

[254] S. Schwarz, "Remote radio head assignment and beamforming in dynamic distributed antenna systems," in *IEEE International Conference on Communications*, Kansas City, Missouri, May 2018, pp. 1–6.

[255] S. Schwarz and M. Rupp, "Transmit optimization for the MISO multicast interference channel," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4936 – 4949, 2015.

[256] S. Schwarz, "Probabilistic analysis of semidefinite relaxation for leakage-based multicasting," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 742–746, 2016.