Lehrstuhl für Netzwerktheorie und Signalverarbeitung
Technische Universität München

Intercell Interference Robustness

Hans H. Brunner

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Wolfgang Kellerer
Prüfer der Dissertation:
1. Prof. Dr. techn. Dr. h. c. Josef A. Nossek
2. Prof. Giuseppe Caire, Ph.D.

Die Dissertation wurde am 21.04.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 15.09.2016 angenommen.

Berichte aus dem Lehrstuhl für Netzwerktheorie und
Signalverabeitung der Technischen Universität München

Hans H. Brunner

Intercell Interference Robustness

Acknowledgments

First of all, I would like to express my sincere gratitude to Professor Josef A. Nossek for supervising this thesis. What I have learned from him, his advice, trust, and continuous support are of immeasurable value to me. It was a great experience to collaborate with such an excellent mind. I'm also very grateful to the committee members of my examination, Professor Guiseppe Caire and Professor Wolfgang Kellerer.

Before I started to work on this thesis at the Institute for Circuit Theory and Signal Processing (NWS), I received most of my education in signal processing at the Associate Institute for Signal Processing (MSV). I'm very grateful to the extraordinary teachers Professor Wolfgang Utschick, Michael Joham, David A. Schmidt, Pedro Tejera, Raphael Hunger, Christian Guthy, and Frank A. Dietrich, who guided me during the times of my Bachelor's thesis, Hauptseminar, Diploma thesis, and way beyond.

This thesis would not be the same without the input from Andreas Dotzler, Amine Mezghani, Ralf Bendlin, Michel T. Ivrlač, Mario H. Castañeda, Maximilian Riemensberger, Christop Hellings, and all my students who contributed to this thesis. I would like to thank them for all I have learned from them and the fruitful and interesting discussions. It was a pleasure to spend these years in the same office with Israa Slim, Michail Matthaiou, and Haixia Zhang. Many thanks go to Bernhard Lehmeyer, Jawad Munir, Leonardo Baltar, Qing Bai, Rainer Pauli, Felix Antreich, Manuel Stein, Hela Jedda, Kilian Roth, and Fabian Steiner for the successful teamwork and for the memorable time at NWS.

Raphael was a big help to me in the beginning of my time as *Schaltungstechnik 2* assistant. I very much enjoyed the discussions with Michael and Amine on rather theoretical circuits and building real circuits with Bernhard and Christoph. Additionaly, I would like to thank Christoph for our discussions on teaching big audiences, Bernhard for supporting the NKS lab and helping me explore the experimental side of circuit theory, and Mario for handing over *Systeme der Signalverarbeitung* so smoothly.

I want to extend my thanks and acknowledgments to the non-academic staff of NWS: Ulrike Heer, Ali Yilmazcan, Sergey Fedorov, Hartmut Peters, Lidmila Barth, and Elizabeth Söder for their kind assistance and entertaining discussions. Special thanks go to David, Sergey, and Wolfgang Fritz for our tough negotiations and hard work on the IT infrastructure.

My deepest gratitude goes to my family, my wife Susanne, my sons Adam and Emanuel, my parents Helene and Wolf, and my wife's parents Carola and Peter, which supported me all these years and endured me during the stressful times (e.g. while I was writing acknowledgments).

Munich, October 2016

Contents

1. Introduction

In the *downlink* (DL) of a wireless cellular system with full cooperation, the intercell interference at each *mobile device* (MD) can be known and handled by a central processor. However, this requires perfect *channel state information* (CSI) about all channels in the system. Measuring all these channels can occupy all air time in large systems, which leaves no time for data transmission [1]. The assumption of full cooperation is therefore unrealistic. Cooperation is always limited in realistic systems. The measurement of all interference channels and a coordination of all beamformers simply cannot be implemented [2, 3] and some part of the interference has to be regarded as noise. In the main part of this thesis, the base stations do not cooperate or coordinate their beamforming. The interference channels are not measured and all interference has to be regarded as noise. Handling the properties and effects of this interference, which has to be regarded as noise, is the focus of this work.

Even if the not measured interference channels stay constant, the intercell interference variance at a mobile device can change unpredictably whenever a *base station* (BS) in the network changes its transmit processing. A base station does not know the actual intercell interference or the supported rates of the mobile devices it serves. Therefore, base stations use assumed intercell interference variances for the beamforming optimization and link rate adaptation. The mismatch between assumed intercell interference and actual intercell interference leads to the intercell interference blindness problem [4, 5]. When the actual intercell interference is larger than the assumed one, the channel is worse than assumed and the mobile device cannot decode the data. The transmission fails and the achieved rate becomes zero. When the actual intercell interference is smaller than the assumed one, the mobile device can only communicate with the assumed data rate and some resources are wasted.

Intercell interference is the dominant effect limiting the performance of cellular networks. It is assumed, that this limitation can be overcome by letting base stations cooperate (See [6–15] and references therein). The benefits of *cooperative multi-point* (CoMP) techniques sound very promising and a strong research focus is set on how cooperation should be organized. But, very little is known about the drawbacks of cooperation. It is still an open topic, whether cellular systems will benefit from cooperation, if all implementation issues are taken into account. If only the backhaul is limited, cooperation can still show a great gain [16]. But in field trials, the performance of cooperative techniques is even worse compared to traditional uncoordinated methods [17, 18].

The term CoMP covers any transmission technique, which allows more than one link per user. This includes, e.g., soft handover, where a mobile device is connected continuously to multiple base stations and is dynamically served over the strongest link. But, it also includes *network multiple-input multiple-output* (network MIMO, joint transmission CoMP), where

the whole network can be seen as a single broadcast channel with distributed transmit antennas [10, 19]. Usually, cooperation aims at mitigating intercell interference.

The intercell interference problem in cellular MIMO networks is so complex and severe, that a combination of multiple different techniques is necessary for its mitigation [20]. Conventional approaches for handling the intercell interference are frequency reuse, cell splitting, and averaging [21]. With frequency reuse, orthogonal frequency bands are assigned to adjacent base stations to separate the transmissions within these cells. The smaller the fraction of the frequency band at each base station is, the less base stations in the near vicinity use the same frequency band and the distance to the next base station utilizing the same frequency band increases. But, a smaller frequency band per cell dramatically reduces the network capacity. The idea of frequency reuse was generalized to cooperative scheduling in such a way, that all base stations can use the whole frequency band, but, transmissions which could notably disturb each other are assigned different frequency bands or time slots [22, 23]. The use of directive antennas at a base station, which concentrate the transmit power onto the served region, is called sectoring. This not only increases the wanted signal power at the mobile devices in the own cell, it also reduces the not wanted power at mobile devices in other cells. Sectoring is a very effective method, with the drawback of being inflexible with respect to cell load. Distributing the antennas over the cell also concentrates the power in the served region by reducing the distance between the transmit and receive antennas [24]. Interference averaging can be done with spread spectrum techniques or frequency hopping. The idea is to distribute the generated interference as uniformly as possible over the available frequency band. This requires that the load of the network is rather small to limit the total amount of intercell interference.

Intercell interference aware receiver structures can also be used to improve the system performance by means of multiuser detection [25, 26]. This requires in general some channel state information about interfering channels at the receiver. Linear minimum mean square error receivers can suppress some intercell interference depending on the number of receive antennas. The optimal performance of the maximum likelihood multiuser detection can be approached with non-linear receiver structures like a decision feedback equalizer, which successively decodes the detectable data streams to eliminate the intercell interference.

1.1 Contributions and Overview

The thesis is structured as follows.

- **Chapter 2: Upper Bound for Interference Coordination**
  For the downlink of a cellular system, a loose upper bound for the possible network sum utility is introduced and formulated in this chapter. It is assumed that each base station has only channel state information about a restricted subset of mobile devices. The amount of available channel state information is linked to the signaling efficiency in Section 2.2 and the information is prone to an outdating error, which is described in Section 2.3. In Section 2.4, the upper bound is formulated by comparing the investigated system with a system in which the known or measured interference channels do not exist. The remaining intercell interference has to be regarded as noise. With this step, the network sum utility maximization decomposes into individual cell sum utility maximizations, which can be solved with known methods.

The upper bound does not show how much cooperation is optimal, it only states that there is a limit to beneficial cooperation. As one possible approach to interference coordination, the interference temperatures technique is combined with the limited set of measured interference channels, the corresponding efficiency, and outdating error of the channel state information in Section 2.5. A heuristic for the selection of the interference temperatures is motivated, which allows to utilize this technique in a full cellular network. It can be seen that it is advisable to measure only a very limited number of interference channels, if any at all. There will be a vast amount of interference over not measured interference channels. Especially in systems where the complete available bandwidth is used by every base station. This advises to take a closer look at the interference over the not measured interference channels as it is done in the following chapters.

- **Chapter 3: Sum Rate Maximization with Conic Constraints**
  Almost all of the optimizations in this work are based on a weighted sum rate maximization. All interference robustness methods for all different utilized fairness utilities need a weighted sum rate maximization in their core. The problem of the interference temperatures method from Chapter 2 is a weighted sum rate maximization with multiple linear constraints. The covariance shaping method for intercell interference robustness relies on a weighted sum rate maximization with a conic constraint and the loosened covariance shaping idea is the same with an additional linear constraint.
  In Section 3.3, linear and linear conic constraints for the weighted sum rate maximization with respect to the transmit covariance matrices in the downlink are analyzed. The existing minimax uplink-downlink duality in [27] is adapted in Section 3.4 to allow multiple linear conic constraint. A very efficient projected gradient algorithm for solving the uplink saddle-point problem based on the work in [28] is presented and described in detail in Section 3.5 along with the required orthogonal projection of the gradients onto the tangent cone of the constraint set and the projection of the updates onto the constraint set itself. With an analysis of the convergence behavior, it can be shown that the presented algorithm has superior performance not only for the previously untouched problem with conic constraints, but also for multiple linear constraints.

- **Chapter 4: Intercell Interference Robustness Methods**
  The intercell interference blindness problem addressed in [4], which arises in all cellular systems for the part of the intercell interference, which has to be regarded as noise, is discussed in detail in Section 4.1. The characteristics of the associated intercell interference variance are analyzed in Section 4.2 and used to derive a simple single cell system model, where the intercell interference variances are drawn from a gamma distribution, in Section 4.3.
  The target is to make the system robust against random changes of the intercell interference variance. A baseline approach is to accept the intercell interference mismatch and transmit at lower rates with the conservative gambling method described by Ivrlač et al. in [4] and described in Section 4.4. A common back-off factor β is introduced to lower the risk of a failed transmission. The BSs serve the MDs with modest rates. This is extended by Shirani-Mehr et al. in [29] and described in Section 4.5 to conservative gambling with individual backoff factors. The system can be strengthened against unpredictable changes in the intercell interference variance with the expected rate method introduced and described in Section 4.6 [30]. This method takes the actual performance

measure, the expectation of the rate with respect to the instationary intercell interference, as cost function for the optimization of the transmit covariance matrices and transmitted rate.
It is assumed in the genie method that the intercell interference at each mobile device is simply known [4] (Section 4.8). The base stations calculate the transmit covariance matrices and generated intercell interference iteratively until the resulting intercell interference converges. A possibility to make the intercell interference variance available at the base stations is to measure it with additional piloting. If the measurements are limited to a single additional pilot, the method is called the second pilot, which is introduced and described in Section 4.7 [31]. Dotzler et al. proposed the covariance shaping method in [32], where the uncertainty in the intercell interference is eliminated by imposing a shaping constraint on the sum transmit covariance. Then, the intercell interference variances will not change, even, if the other base stations update their beamforming. Although the intercell interference blindness problem is solved, the shaping constraint reduces the region of achievable data rates. The expected rate method and covariance shaping outperform conservative gambling notably and can be combined to the loosened covariance shaping method introduced and described in Section 4.10. Loosened covariance shaping finds the best compromise between stabilizing the intercell interference variance at mobile devices in other cells and optimizing the expected rates of the own mobile devices.
Hybrid Automatic Repeat reQuest (HARQ) can also be used to treat the intercell interference blindness problem [29]. The intercell interference blindness problem is relaxed such, that a transmission will be completed successfully, if the data can be decoded with the combination of several retransmissions. The incorporation of HARQ with soft combining into the optimization of the transmit covariance matrices and transmitted rate with the expected rate method is discussed in Section 4.11 [33]. Simulation results show that the expected rate method can be improved greatly with HARQ.

- **Chapter 5: Intercell Interference Robustness in Fairness Optimizations**
  The implementation of the previously presented intercell interference robustness methods is presented and discussed for different fairness utilities in this chapter. Round robin, maximum sum rate, log fairness, proportional fairness, and max-min fairness are analyzed. In Section 5.2, the fairness utilities are handled by a time slot scheduler combined with a weighted sum rate maximization in each slot. The selection of transmit covariance matrices, which optimize the utility in each single time slot, is discussed in Section 5.3. Detailed algorithms are presented for tackling the considered problems and simulation results show that the performance under fairness utilities greatly increases with the intercell interference robustness methods.

1.2 Notation

Scalars are denoted by lower case italic and capital italic letters, whereas vectors and matrices are denoted by lower case bold italic letters and capital bold italic letters, respectively. Sets are denoted by calligraphic letters. The most frequently used operators, functions, and symbols are listed in Table 1.1. The complete list of used symbols and functions can be found in Appendix A3.

Operator / Symbol	Description
$\mathrm{Re}(\bullet)$, $\mathrm{Im}(\bullet)$	real part, imaginary part
$\mathrm{E}[\bullet]$, $\mathrm{var}(\bullet)$	expectation, variance
$(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$	complex conjugate, transposition, conjugate transposition
x_i	ith entry of vector $\boldsymbol{x}$
x_{ij}	entry on the ith row and jth column of matrix $\boldsymbol{X}$
$x_{1:N}$	set $\{x_1, \ldots, x_N\}$
$\bullet\vert_x$	evaluated at x
$\vert \bullet \vert$	absolute value of a scalar, determinant of a matrix, cardinality of a set
$\Vert \bullet \Vert_2$, $\Vert \bullet \Vert_{\mathrm{F}}$	Euclidean norm of a vector, Frobenius norm of a matrix
$\mathrm{tr}(\bullet)$	trace of a matrix
$\mathrm{eig}(\bullet)$	eigenvalue decomposition
$(\bullet)_+$	element-wise $\max(\bullet, 0)$ for scalars, set all negative eigenvalues to zero for matrices
$(\bullet)_\perp$	orthogonal projection onto constraint set
$(\bullet)_\perp$	projection onto tangent cone of constraint set
$(\bullet)^\perp$	orthogonal subspace
$a \to b$	a converges to b
$a \leftarrow b$	a is set to b
$a \sim b$	a has the same distribution as b
$\boldsymbol{A} \succ \boldsymbol{B}$	$\boldsymbol{A} - \boldsymbol{B}$ is positive-definite
$\boldsymbol{A} \succeq \boldsymbol{B}$	$\boldsymbol{A} - \boldsymbol{B}$ is positive-semidefinite
$P(\bullet)$	probability mass function
$f_X(\bullet)$	probability density function of X
$F_X(\bullet)$	cumulative distribution function of X
$\mathcal{L}(\bullet)$	Lagrangian function
$U(\bullet)$	monotonic utility function
$\mathbf{0}$	all zeros matrix of appropriate size
$\mathbf{I}$	identity matrix of appropriate size
$\boldsymbol{e}_i$	ith column of the identity matrix of appropriate size
$\mathbb{N}$	set of natural numbers including zero
$\mathbb{R}$, $\mathbb{R}_0^+$, $\mathbb{R}^+$	set of all / non-negative / positive rational numbers
$\mathbb{C}$	set of complex numbers

Table 1.1: Most frequently used operators, functions, and symbols

2. Upper Bound for Interference Coordination

In [2], Lozano et al. describe a fundamental limit to any type of cooperation by pointing out that the cluster of base stations, which cooperate, cannot be infinitely large and all cellular systems are interference limited. An upper bound for network MIMO is derived in [34], where the broadcast channel of the network MIMO downlink is compared to interference free, parallel point-to-point channels, which have a common sum power constraint. The high rates achieved with this upper bound advise to take a closer look at cooperation. But, the drawbacks of cooperation have been neglected in this study. For network MIMO, all channels in a cluster of cooperating base stations have to be measured. The complete channel state information has to be communicated to a central processor and the selected beamforming vectors have to be communicated back to the base stations. The data streams for all mobile devices have to be present at all base stations of a coordinated cluster. The upper bound for clustered network MIMO in [1, 35] and the investigations in [36, 37] take the cost of measuring the additional channels for network MIMO into account. In contrast to the optimistic upper bound, the optimal cluster size is identified as a single site even for scenarios with low user mobility and poses the question whether network MIMO should be used at all.

In this thesis, cooperation is restricted to interference coordination, where each mobile device is only served by one base station. Each base station needs to know the channels to the mobile devices it serves and, additionally, to some mobile devices in other cells to perform interference mitigation. Compared to network MIMO, the amount of channel measurements and communicated information in the backhaul is substantially smaller with interference coordination.

In the first part of this chapter, parts of this can also be found in [3,38], no assumptions are made on how the interference coordination is realized. For the upper bound, it is possible, but not necessary, that there are ideal high speed links between the base stations. The methods are applicable in heterogeneous and homogeneous systems with macro, micro, pico and femto cells. The base stations can employ any known cooperative technique, which is covered by interference coordination, like interference alignment [39–41], where the base stations try to spare subspaces of the received signals from interference, or the interference temperatures technique, where each base station limits the interference it causes at mobile devices to a given value [42–44]. Interference pricing is an iterative technique for finding a balance between the signal power at the served mobile devices and the interference caused at other mobile devices [45]. Interference leakage power based optimizations try to maximize the ratio between the received power at the served mobile devices and the caused interference power at other mobile devices [46]. There are many other approaches for interference coordination like

Max-SINR [47, 48] and MMSE algorithms [49]. The possible degrees of freedom of a system with interference coordination are derived in [40] with interference alignment in systems without background interference. But, the gain of interference alignment will disappear or even turn into losses, if it is applied in a large scale system with interference from base stations outside of the cooperation cluster [50–54].

In a network with interference coordination, the base stations face the conflicting goals of serving their associated mobile devices optimally and minimizing the interference they cause. An adaptive, beamforming based interference mitigation can only be performed for known interference channels, whereas the intercell interference over unknown channels has to be regarded as noise. The maximum achievable rates are limited by the signaling efficiency due to acquisition and outdating of channel state information as described in Section 2.2 and 2.3 [3, 55]. Resources have to be spent to measure the channels and the measurements might be perturbed by pilot contamination (See Appendix A2) [56, 57]. On the one hand, an increased number of measured interference channels allows more cooperation, which can lead to higher rates. On the other hand, the signaling efficiency reduces, which deteriorates the rates. An interference channel should only be measured, if the resulting gain through cooperation outperforms the loss through the decreased efficiency. In the presence of intercell interference, the upper bound behaves similarly [58].

The upper bound for interference coordination is formulated by setting the known interference channels simply to zero (See Section 2.4). The upper bound is found in a system, where the measured interference channels do not exist. There is no intercell interference over these channels, although the base stations do not perform any intercell interference mitigation. The base stations can meet both of their goals, because they do not produce any intercell interference over known interference channels and still have all their degrees of freedom to serve the associated mobile devices. The intercell interference decreases and the possible rates increase. Of course, the channels cannot simply disappear, therefore, the upper bound cannot be reached [3]. As the intercell interference over measured channels is set to zero and the intercell interference over not measured channels has to be regarded as noise, the network decomposes into independent broadcast channels and well known techniques for single cell MIMO downlink systems, like the weighted sum rate maximization from Chapter 3, can be used to optimize the transmit covariance matrices.

It will be shown in Section 2.4.1 that even this loose upper bound strongly limits the possible gain of cooperation for scenarios with modest and high user mobility. The number of interference channels, which should be measured to perform interference mitigation is rather small to keep the overhead small and the signaling efficiency high. In other words, there is a rather strict limit to beneficial cooperation. In scenarios with low user mobility, the upper bound leaves quite substantial room for improvement through cooperation. But, the upper bound is not achievable, because the cost of nulling the intercell interference over known interference channels is neglected. To get some insight into the tightness of the upper bound, the framework is applied to interference coordination with interference temperatures in the second part of this chapter (See Section 2.5 [38, 59]). The interference temperature technique is analyzed, where the transmission efficiency is reduced according to the number of measured interference channels.

The performance of the interference temperatures technique is investigated under the maximum sum rate and proportional-fair resource distribution in the scenario with slow moving mobile devices. The limit to beneficial cooperation is similar in both systems, but,

due to different effects. The sum rate is improved because mobile devices close to the site, which are already in a good situation, receive significantly less interference from collocated base stations. In a fair utility, the mobile devices at the cell edge, which suffer the most from intercell interference and, therefore, benefit the most from cooperation, have a stronger influence.

2.1 System Model

The downlink of a cellular network with 19 three faced sites and, therefore, 57 base stations is considered. Each base station serves the mobile devices of the hexagonal shaped cell it covers in the same frequency band (reuse 1). The wrap-around method is used to treat all cells equally. The 57 base stations are copied, including their beamforming, and placed six times around the central cluster. Each mobile device only sees the 57 base stations, which are closest by Euclidean distance. In Figure 2.1, the cellular layout can be seen, where the central cluster is inked slightly darker than the wrap-around clusters. The placement and orientation of the base stations is indicated by small arrows.

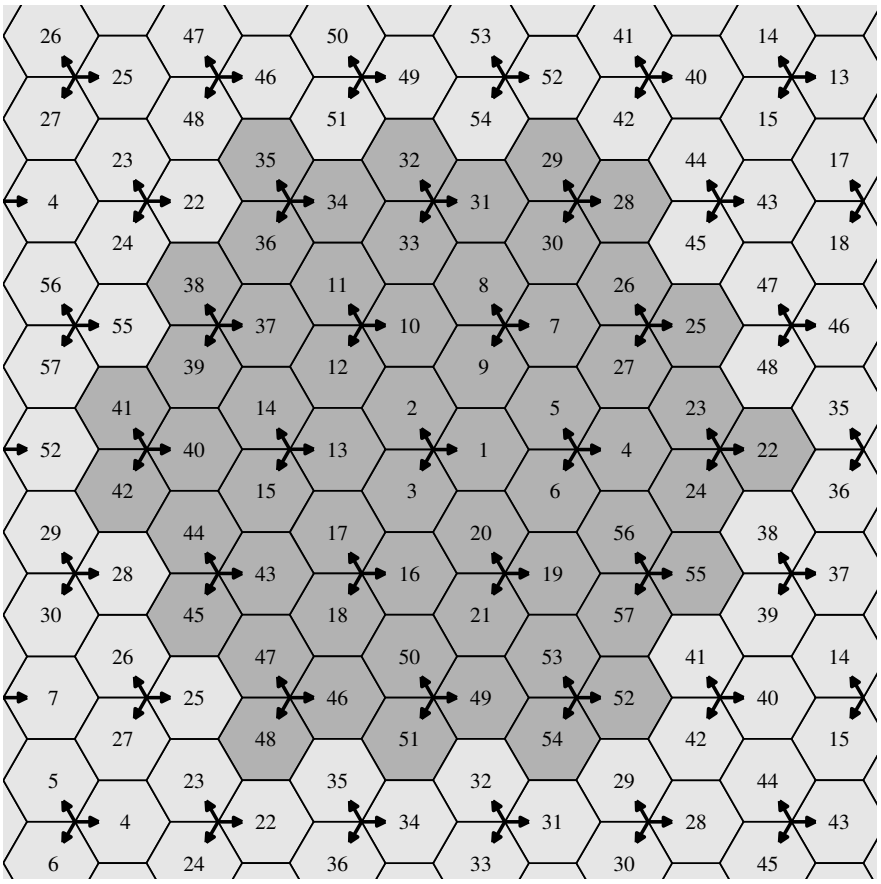


Figure 2.1. Cellular cluster with wrap-around

A mobile device in the set $\mathcal{K}$ of all mobile devices is specified by the tuple $(b, k) \in \mathcal{K}$, where $b \in \mathcal{B}$ identifies the base station in the set $\mathcal{B} = \{1, \ldots, 57\}$ of all base stations and $k \in \mathcal{K}_b$ the mobile device in the set $\mathcal{K}_b = \{1, \ldots, K\}$ of all mobile devices in the cell of base station b. In this chapter, each base station has N antennas and serves $K = |\mathcal{K}_b|$ single

antenna mobile devices, respectively. The vectors $\boldsymbol{h}_{\hat{b},b,k} \in \mathbb{C}^N$ contain the channel coefficients between the antennas of base station $\hat{b}$ and mobile device (b,k).

The achievable, normalized rate of mobile device (b,k), within the capacity region of the MIMO broadcast channel with dirty paper coding, can be expressed as

$$r_{b,k} = \xi \log_2 \left(1 + \frac{\boldsymbol{h}_{b,b,k}^{\mathrm{H}} \boldsymbol{Q}_{b,k} \boldsymbol{h}_{b,b,k}}{\sigma_{b,k}^2 + \theta_{b,k} + \sum_{\hat{k}>k} \boldsymbol{h}_{b,b,k}^{\mathrm{H}} \boldsymbol{Q}_{b,\hat{k}} \boldsymbol{h}_{b,b,k}} \right), \tag{2.1}$$

$$\theta_{b,k} = \sum_{\hat{b} \in \mathcal{B} \setminus b} \boldsymbol{h}_{\hat{b},b,k}^{\mathrm{H}} \boldsymbol{Q}_{\hat{b}} \boldsymbol{h}_{\hat{b},b,k}, \tag{2.2}$$

where $\boldsymbol{Q}_{b,k} \in \mathbb{C}^{N \times N}$ is the transmit covariance matrix for mobile device (b,k) and $\sum_k \boldsymbol{Q}_{b,k} = \boldsymbol{Q}_b \in \mathbb{C}^{N \times N}$ is the sum transmit covariance matrix of base station b. All base stations have to satisfy the transmit power constraint $\mathrm{tr}(\boldsymbol{Q}_b) \leq P$. The signaling overhead reduces the rates through the pre-log efficiency ξ, which is described in Section 2.2.

$\sum_{\hat{k}<k} \boldsymbol{h}_{b,b,k}^{\mathrm{H}} \boldsymbol{Q}_{b,\hat{k}} \boldsymbol{h}_{b,b,k}$ is the variance of the intracell interference with dirty paper coding. Only the signals from base station b to mobile devices with an index $\hat{k} > k$ contribute to the intracell interference at mobile device (b,k). The signals to mobile devices with an index $\hat{k} < k$ can be considered as known interference at the transmitter. Costa showed that interference, which is known to the transmitter, does not reduce the rate of the transmission [60].

$\theta_{b,k}$ is the variance of the received intercell interference, and $\sigma_{b,k}^2 = \sigma_\eta^2 + \sigma_{\mathrm{od},b,k}^2 + \theta_{\mathrm{bg}}$ is the sum of the variance of the thermal noise σ_η^2, the channel state information outdating, and the background interference. The outdating $\sigma_{\mathrm{od},b,k}^2$ depends on the assumed block fading block length T_{block} and the fading scenario and is approximated with Gaussian noise over the serving channel and the measured interference channels as derived in Section 2.3 [3].

Cooperation with a base station far away by distance will surely not improve the performance, as the produced interference rarely harms the rates. The Gaussian background interference θ_{bg} is included in the system, which originates from the base stations further away than the 57 closest base stations. If the wrap-around method is not used and the system is extended to an unlimited number of base stations, this background interference will be the mean sum interference at a mobile device at the central site, from base station 58 to infinity. This background interference variance converges due to the pathloss and a fixed transmit power. With this background interference, the system is again subject to an interference limitation. The system is assumed to be operated in the interference limited region, where an increase in transmit power would not lead to an increase in the sum rate.

With the monotonic increasing utility function $U(r_{b,k})$, the cooperative network sum utility maximization with general interference coordination can be formulated as

$$U_{\mathrm{coop}} = \max_{\boldsymbol{Q}_{b,k}\ \forall (b,k) \in \mathcal{K}} \sum_{(b,k) \in \mathcal{K}} U(r_{b,k}) \qquad \text{s.t. } \mathrm{tr}(\boldsymbol{Q}_b) \leq P\ \forall b \in \mathcal{B}. \tag{2.3}$$

Different utilities $U(\bullet)$, like maximum sum rate, proportional fairness and max-min fairness are discussed and described in Chapter 5. The transmit covariance matrices of all base stations in the central cluster are optimized jointly, but a mobile device is only served by the associated base station.

2.2 Signaling Overhead and Efficiency Upper Bound

To perform adaptive signal processing, instantaneous knowledge of channel state information is required. For a non-cooperative sum rate maximization in each cell, knowledge of the channels from each base station to its own mobile devices is sufficient. If the base stations want to cooperate, additional information about interference channels has to be known.

In [61] it was shown, that there will be always less training symbols in a time division duplex system than in a frequency division duplex system, if further implementation issues are neglected. Therefore, the efficiency of the time division duplex system is higher and such a system will be assumed in the rest of the chapter. In time division duplex systems, the reciprocity of the propagation channels is exploited. In the first step, the channels are measured in the uplink and the gained information is then utilized in the downlink. The mobile devices are split into equally sized subsets $\mathcal{K}_c$. The mobile devices within a subset use pilot sequences, which are orthogonal to each other [58]. But, these pilot sequences are reused in all other subsets. The pilot length has to be at least as large as the number of users in a subset $T_{\text{pilot}} \geq K + L = |\mathcal{K}_c|$. Each base station can measure the channels to its own K mobile devices and L interference channels, additionally. Here, the effect of pilot contamination is neglected, but can be included as described in Appendix A2.

Here, the pilot length T_{pilot} is taken as the dominant overhead, any other overhead is neglected. With the block length T_{block}, $T_{\text{data}} = T_{\text{block}} - T_{\text{pilot}}$ time slots remain for data transmission [55,62]. Data transmission is done during T_{data}. Unlike in [56,57], interference, noise, quantization, and feedback errors during the pilot phase are ignored and the minimal pilot length $T_{\text{pilot}} = K + L$ is optimal. The efficiency of this signaling scheme is

$$\xi = \frac{T_{\text{data}}}{T_{\text{block}}} = \frac{T_{\text{block}} - (K + L)}{T_{\text{block}}}. \tag{2.4}$$

The optimal L is a tradeoff between the reduced efficiency and the possible gain through additional cooperation. This effect is include by the outdating of the channel state information as explained in the next section.

Except for the outdating, all channel measurements are assumed to be perfect. The design of the pilot sequences and the selection of the L interference channels, which should be known to each base station, respectively, are formidable tasks. The conflict arises, that the optimal assignment of the mobile devices to the subsets $\mathcal{K}_c$ can be different from the point of view of each single base station. In this chapter, it is assumed that all these problems are solved optimally. Each base station measures the L interference channels, which have the largest impact on the network sum rate, respectively. Two approaches for distributing the pilot sequences are described in Appendix A2. Because of all these idealistic assumptions, the efficiency has to be regarded as a loose upper bound.

2.3 Outdating of the Channel Measurements

Due to outdating, the measured channel differs from the actual channel at a different time and frequency instance. By modeling the time and frequency variation as a first-order

Markov process [63, Chapter 16], the normalized error variance can be found as

$$\sigma_e^2(t,f) = \frac{\mathrm{E}\left[\|\hat{\boldsymbol{h}}_{\hat{b},b,k} - \boldsymbol{h}_{\hat{b},b,k}(t,f)\|_2^2\right]}{\mathrm{E}\left[\|\hat{\boldsymbol{h}}_{\hat{b},b,k}\|_2^2\right]} = 1 - \rho^2(t,f). \tag{2.5}$$

$\hat{\boldsymbol{h}}_{\hat{b},b,k}$ is the measured channel at a reference time t_0 and frequency f_0, $\boldsymbol{h}_{\hat{b},b,k}(t,f)$ is the outdated channel at a slightly different time $t_0 + t$ and frequency $f_0 + f$, and $\rho(t,f)$ is the correlation coefficient between the two channels.

According to Jakes [64, Chapter 1.5], the correlation can be approximated as

$$\rho(t,f) = \frac{\mathrm{J}_0(2\pi f_D t)}{1 + (2\pi\sigma_{DS} f)^2}, \tag{2.6}$$

where $\mathrm{J}_0(\bullet)$ is the Bessel function of the first kind and zeroth order, f_D is the maximum Doppler frequency, and σ_{DS} is the root mean square delay spread.

The channel state information outdating at mobile device (b,k) is approximated with Gaussian noise over the serving channel and the measured interference channels [3]. With a signal variance per transmit antenna of P/N, the channel state information outdating variance is the sum of the power over the measured channels scaled down with the measurement error:

$$\sigma_{\mathrm{od},b,k}^2 = \bar{\sigma}_e^2 \frac{P}{N} \hat{\boldsymbol{h}}_{b,b,k}^{\mathrm{H}} \hat{\boldsymbol{h}}_{b,b,k} + \bar{\sigma}_e^2 \frac{P}{N} \sum_{\hat{b} \in \mathcal{C}_{b,k} \setminus b} \hat{\boldsymbol{h}}_{\hat{b},b,k}^{\mathrm{H}} \hat{\boldsymbol{h}}_{\hat{b},b,k}, \tag{2.7}$$

where $\mathcal{C}_{b,k}$ is the set containing all base stations, which measured the channel to mobile device (b,k).

As an upper bound approximation, the minimum mean error variance,

$$\bar{\sigma}_e^2 = \frac{E}{T_{\mathrm{block}}}, \tag{2.8}$$

is calculated, which is the minimum sum error E during a block divided by the block length. The block length, the total amount of transmit symbols, is a virtual area in the time-frequency domain and the sum error is the integral over the error in this area. The beamforming is based on only one channel estimation per block. To minimize the sum error, the reference must lie in the center of the block length area. For a given block length this error minimizing area is enclosed by the curve $\hat{\rho}(T_{\mathrm{block}}) = \rho(\hat{t}, \hat{f})$, where $\hat{\rho}(T_{\mathrm{block}})$ is the smallest, occuring correlation for the given block length. Respectively, the block length and minimum sum error can be calculated as

$$T_{\mathrm{block}} = 4 \int_0^{\hat{t}_0} \int_0^{\hat{f}(t)} 1 \,\mathrm{d}f \,\mathrm{d}t, \tag{2.9}$$

$$E = 4 \int_0^{\hat{t}_0} \int_0^{\hat{f}(t)} \sigma_e^2(t,f) \,\mathrm{d}f \,\mathrm{d}t. \tag{2.10}$$

In Figure 2.2 the block length area is shown, where $\hat{t}_0$ and $\hat{f}_0$ denote the maximum time and frequency distance to the reference, respectively.

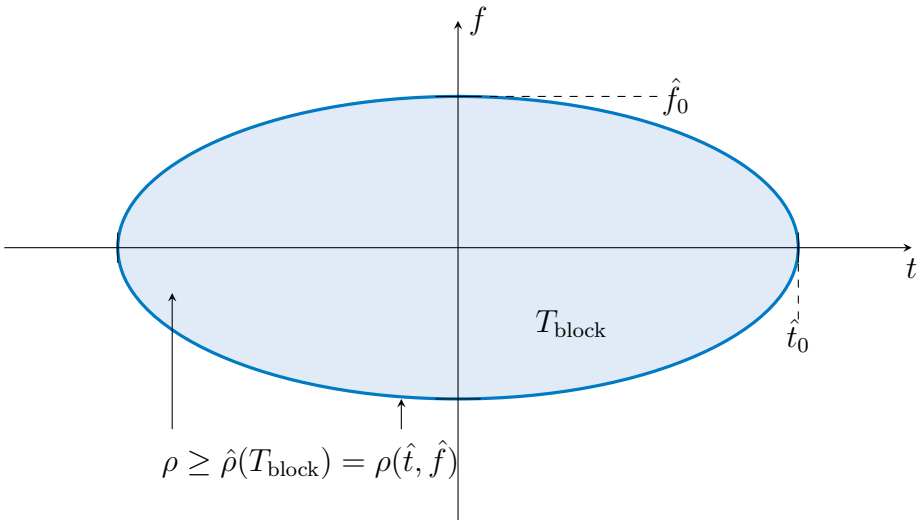


Figure 2.2. Block length in the frequency-time domain

In the simulations, $\hat{\boldsymbol{h}}_{\hat{b},b,k}$ is generated from the channel model. The channel used in the simulations as measured channel is set to

$$\boldsymbol{h}_{\hat{b},b,k} = \sqrt{1-\bar{\sigma}_{\mathrm{e}}^2}\hat{\boldsymbol{h}}_{\hat{b},b,k}. \tag{2.11}$$

This can be seen as the mean channel during the block fading block. With equation (2.11), the norm of the generated channel is preserved in the sum of the average norm of the fading channel and the measurement error variance. If the channel is not assumed to be measured, the generated channel will be used directly in the simulations. Pilot contamination and other errors during the channel measurements are neglected. Except for the outdating, the channel measurements are assumed to be perfect.

2.4 Upper Bound Formulation

As described in Section 2.2, only the L interference channels per base station with the strongest influence on the network sum rate are known. The intercell interference (2.2) can be split into the interference over the measured interference channels and the interference over the unknown interference channels [3]:

$$\theta_{b,k} = \sum_{\hat{b}\in\mathcal{C}_{b,k}\setminus b} \underbrace{\boldsymbol{h}_{\hat{b},b,k}^{\mathrm{H}}}_{\text{known}} \boldsymbol{Q}_{\hat{b}} \boldsymbol{h}_{\hat{b},b,k} + \sum_{\check{b}\in\mathcal{B}\setminus\mathcal{C}_{b,k}} \underbrace{\boldsymbol{h}_{\check{b},b,k}^{\mathrm{H}}}_{\text{unknown}} \boldsymbol{Q}_{\check{b}} \boldsymbol{h}_{\check{b},b,k} \tag{2.12}$$

$$= \sum_{\hat{b}\in\mathcal{C}_{b,k}\setminus b} \underbrace{\boldsymbol{h}_{\hat{b},b,k}^{\mathrm{H}}}_{\text{known}} \boldsymbol{Q}_{\hat{b}} \boldsymbol{h}_{\hat{b},b,k} + \underbrace{\theta_{\text{blind},b,k}}_{\text{known statistics}} . \tag{2.13}$$

The first term in (2.12) is the sum of the interference over all measured and, therefore, known interference channels. The set $\mathcal{C}_{b,k}$ contains all base stations, which know the channel to mobile device (b,k). The second term in (2.12) is the sum interference over the unknown interference channels of the central cluster. This is also depicted in Figure 2.3. BS_b serves $\mathrm{MD}_{b,k}$ while $\mathrm{BS}_{\hat{b}}$ and $\mathrm{BS}_{\check{b}}$ disturb it. $\mathrm{BS}_{\hat{b}}$ has measured the channel to $\mathrm{MD}_{b,k}$ and is therefore in the set $\mathcal{C}_{b,k}$. The channel between $\mathrm{BS}_{\check{b}}$ and $\mathrm{MD}_{b,k}$ remains unknown.

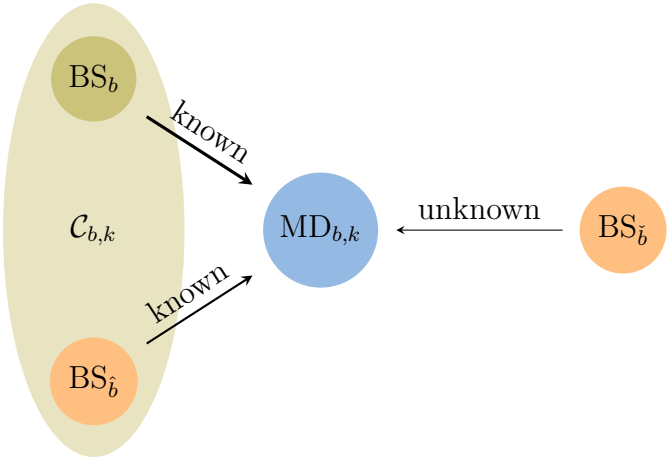


Figure 2.3. Different types of intercell interference

$\theta_{\text{blind},b,k} = \sum_{\hat{b}\in\mathcal{B}\setminus\mathcal{C}_{b,k}} \boldsymbol{h}^{\text{H}}_{\hat{b},b,k}\boldsymbol{Q}_{\hat{b}}\boldsymbol{h}_{\hat{b},b,k}$ is the sum intercell interference over the unknown channels. In this chapter, it is assumed that the variances $\theta_{\text{blind},b,k}$ are known, but the influence of the transmit covariance matrices transmitting over these unknown interference channels cannot be taken into account in the optimization. The upper bound described in this chapter can be formulated for all intercell interference robustness methods from the following chapter, which discuss the problem of predicting $\theta_{\text{blind},b,k}$. As the cooperation is limited to a subset of base stations and the rest of the interference is regarded as noise, the system is interference limited. The power of the interference over the not measured interference channels scales with the transmit power limit [2, 65, 66].

The jointly optimized sum utility U_{coop} is always smaller than the upper bound U_{upper}, which is calculated in a system, where all measured interference channels do not exist (See Figure 2.4). Based on (2.13) and (2.3), the upper bound is formulated with L measured

Figure 2.4. Upper bound system

interference channels per base station as

$$U_{\text{coop}} \leq U_{\text{upper}} = \max_{\boldsymbol{Q}_{b,k}\ \forall(b,k)\in\mathcal{K}} \sum_{(b,k)\in\mathcal{K}} U(r_{\text{upper},b,k}) \qquad \text{s.t. } \text{tr}(\boldsymbol{Q}_b) \leq P\ \forall b \in \mathcal{B}, \tag{2.14}$$

with

$$r_{\text{upper},b,k} = r_{b,k}\big|_{\boldsymbol{h}_{\hat{b},b,k}=\boldsymbol{0}\ \forall\hat{b}\in\mathcal{C}_{b,k}\setminus b} \tag{2.15}$$

$$= \xi \log_2\left(1 + \frac{\boldsymbol{h}^{\text{H}}_{b,b,k}\boldsymbol{Q}_{b,k}\boldsymbol{h}_{b,b,k}}{\sigma^2_{b,k} + \theta_{\text{blind},b,k} + \sum_{\hat{k}>k}\boldsymbol{h}^{\text{H}}_{b,b,k}\boldsymbol{Q}_{b,\hat{k}}\boldsymbol{h}_{b,b,k}}\right). \tag{2.16}$$

This is clearly a non-achievable upper bound as the cost for attenuating the interference is neglected. In contrast to interference coordination techniques like [39–42, 67], all degrees of

freedom are still left at each base station for serving the associated mobile devices in the system, where the measured interference channels do not exist.

As the transmit covariance matrices in $\theta_{\text{blind},b,k}$ cannot be optimized without further channel measurements and are assumed to be known, problem (2.14) breaks down into independent broadcast channels at each base station and can be solved distributively with known techniques, respectively:

$$U_{\text{upper}} = \sum_{b} \left\{ \max_{\boldsymbol{Q}_{b,k}\ \forall (b,k)\in\mathcal{K}_b} \sum_{(b,k)\in\mathcal{K}_b} U(r_{\text{upper},b,k}), \quad \text{s.t. } \operatorname{tr}(\boldsymbol{Q}_b) \leq P \right\}. \tag{2.17}$$

There exist transmit covariance matrices, which would result in a higher sum utility. But, these transmit covariance matrices cannot be found without further measurements. With the given channel state information, U_{upper} is a loose upper bound, which is always maximized with the maximum transmit power.

In (2.13), $\sum_{\hat{b}\in\mathcal{C}_{b,k}\setminus b} \boldsymbol{h}^{\text{H}}_{\hat{b},b,k}\boldsymbol{Q}_{\hat{b}}\boldsymbol{h}_{\hat{b},b,k}$ is the sum of the intercell interference over the measured channels, which is zero in the upper bound system. A residual part due to outdating is taken into account with the afore described $\sigma^2_{\text{od},b,k}$. Therefore, the measured interference channels are not completely set to zero. The outdating error variance is the sum over the scaled down norms of the measured channels as described in Section 2.3. The first term in (2.7) is the error due to the outdating of the serving channel and the second term is the error due to the outdating of the nulled interference channels.

2.4.1 Simulations

The spatial channel model of the 3GPP MIMO urban macro cell with a distance of 500 m between the two closest sites and a center frequency of 2 GHz is utilized [68]. The applied parameters for the simulations can be seen in Table 2.1. The following results are obtained with Monte Carlo simulations. Every base station has $N = 4$ transmit antennas and transmits with $P = 14\,\text{W}$. The antennas are assumed as hypothetical isotropic radiators with a spacing of half a wavelength (0.5λ) between the antennas of the same array. In every cell, $K = 10$ mobile devices are placed uniformly distributed and suffer from a thermal noise variance of $\sigma^2_\eta = 1.4 \cdot 10^{-13}\,\text{W}$ at their receive antenna, respectively.

Extending the base station number to 500, the background interference is computed. At this point, the sum interference from base station 58 to infinity is converged with an acceptable accuracy. All base stations use a scaled identity matrix as transmit covariance for this analysis. The average background interference at the mobile devices in the cells of the central site is found as $\theta_{\text{bg}} = 3.34 \cdot 10^{-12}$ W, which is much higher than the thermal noise. The system is operated on purpose in the interference limited region.

The maximum average cell sum rate is taken as the cost function $\sum_{(b,k)\in\mathcal{K}} U(r_{b,k}) = \sum_{(b,k)\in\mathcal{K}} r_{b,k}$ and is plotted over the block length for different mobile device speeds in Figures 2.5 and 2.6. To make $\theta_{\text{blind},b,k}$ available, the second pilot method from Section 4.7 is used. The upper bound can also be found with any other interference robustness method or utility presented in the Chapters 4 and 5. The optimizations are solved with the methods presented in Chapter 3.

For the simulations, the L interference channels per base station are selected as follows. The initial supported rates of each mobile device are computed with maximum ratio transmission and scaled identity matrices as transmit covariances for the interference producing

scenario	urban macro-cell
center frequency	2 GHz
sectors	$19 \cdot 3 = 57$
inter site distance	500 m
users per sector	$K = \lvert\mathcal{K}_b\rvert$
min distance to site	25 m
antenna configuration	$N \times 1$ MISO
antenna spacing	0.5λ
user speed	slow: 3 km/h fast: 30 km/h
root mean square delay spread	$\sigma_{DS} = 0.5\,\mu$s
transmit power	$P = 14$ W
background interference	$\theta_{\text{bg}} = 3.34 \cdot 10^{-12}$ W
thermal noise	$\sigma_\eta^2 = 1.4 \cdot 10^{-13}$ W

Table 2.1. Simulation parameters

base stations. In parallel, each base station selects the L mobile devices, which will have the largest gain in rate, if the produced interference is nulled. This selection is suboptimal, but, still upper bounding because of the omnisciently given selection.

The root mean square delay spread is fixed to $\sigma_{DS} = 0.5\,\mu$s. The maximum Doppler frequency $f_{\text{D}} = f_{\text{C}} v/c$, with the center frequency f_{C} and the speed of light c, directly depends on the mobile device speed v. In Figure 2.5, a low mobility scenario with a common mobile device device speed of $v = 3$ km/h is shown for $K = 10$ mobile devices per cell and $N = 4$ transmit antennas per base station. The different curves display the maximum cell sum rate upper bound $R_{\text{upper}} = U_{\text{upper}}$ per cell over the block length for different choices of L. $L = 0$ stands for no cooperation, an increase in L represents an increase in cooperation. All curves ascend in the beginning for longer block lengths, because the efficiency of the system improves. At some point, each curve starts to descend, because the outdating of the channel degrades the possible rates. For each channel model, we can find a L and a block length which maximizes the sum network throughput. In the low mobility scenario, the optimum lies in the range of $25 \leq L \leq 30$ and $250 \leq T_{\text{block}} \leq 350$. The upper bound with cooperation lies with almost 17 bits per channel usage (bpcu) much higher than the almost 11 bpcu without cooperation.

The upper bound reaches higher values for increasing L in the beginning, because more and more interference is suppressed. Already $L = 5$ with 14 bpcu shows significant improvements compared with no cooperation. But, increasing L above the value of $L = 30$ with the optimal block length reduces the possible rates. The eliminated interference cannot compensate the reduced efficiency due to the increased overhead. In average, 4 mobile devices are served in every cell simultaneously. $L = 24$ could correspond to limiting the cooperation to the mobile devices of the surrounding 6 cells.

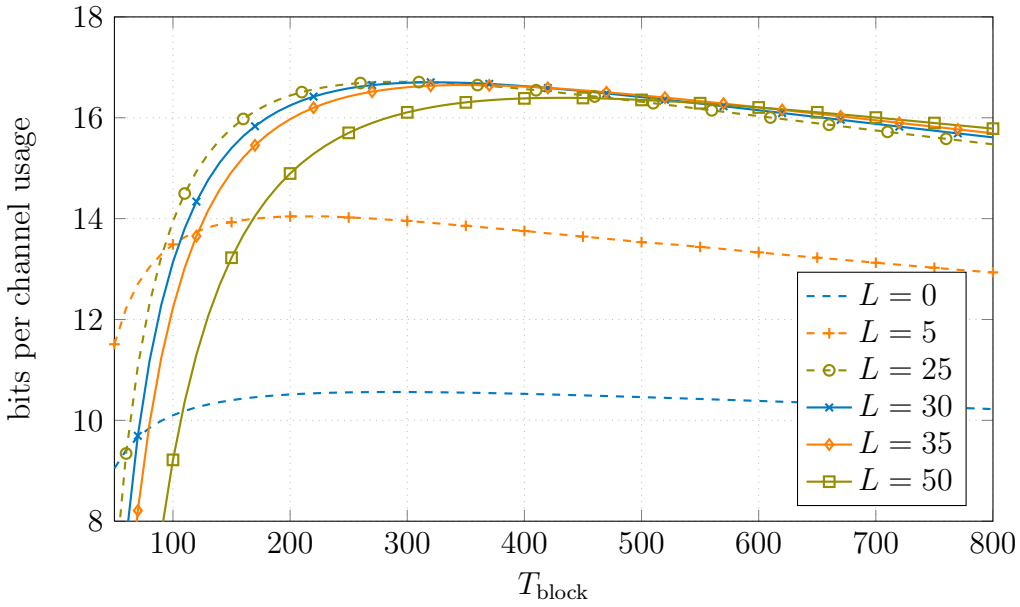


Figure 2.5. Upper bound for the cell sum rate in a low mobility scenario: $v = 3\,\text{km/h}$

The possible improvements through cooperation are substantial in this scenario, and are investigated further in Section 2.5. For this upper bound, no restrictions are made to the backbone network, which connects all base stations. Irrespective of the type of interference coordination and the amount of exchanged information between the base stations, this upper bound will hold, if each mobile device will only be served by its associated base station. The upper bound does not show how much improvement is possible with cooperation. It gives a loose limit to the possible improvement.

A high mobility scenario with a common mobile device device speed of $v = 30\,\text{km/h}$ is shown in Figure 2.6. The optimum rate can be achieved around $L = 15$ and $T_{\text{block}} = 150$. The possible improvements through cooperation are much smaller than in the low mobility scenario. The upper bound with $L = 0$ is close to 9 bpcu and with the best selection of L around 11.5 bpcu. It is rather doubtful that a realistic cooperation can enhance the sum rate achieved in such a scenario.

Only the costs for measuring the channels are taken into account, but not the costs for mitigating the interference over these interference channels. Therefore, this upper bound cannot be reached, especially for a large L. It can be seen, that the possible rates will deteriorate if the number of measured channels grows beyond a limit. To get an idea about a more realistic number of interference channels, which should be measured, the efficiency upper bound and outdating of the measured channels is combined with the interference temperatures technique in the next section.

2.5 Interference Temperatures with CSI Outdating

At the first glance, it seems that interference coordination demands from the base stations to achieve two contradicting goals. On the one hand, the base stations try to serve their own mobile devices with all their degrees of freedom. On the other hand, the base stations try to spend their degrees of freedom to limit the interference they create at mobile devices

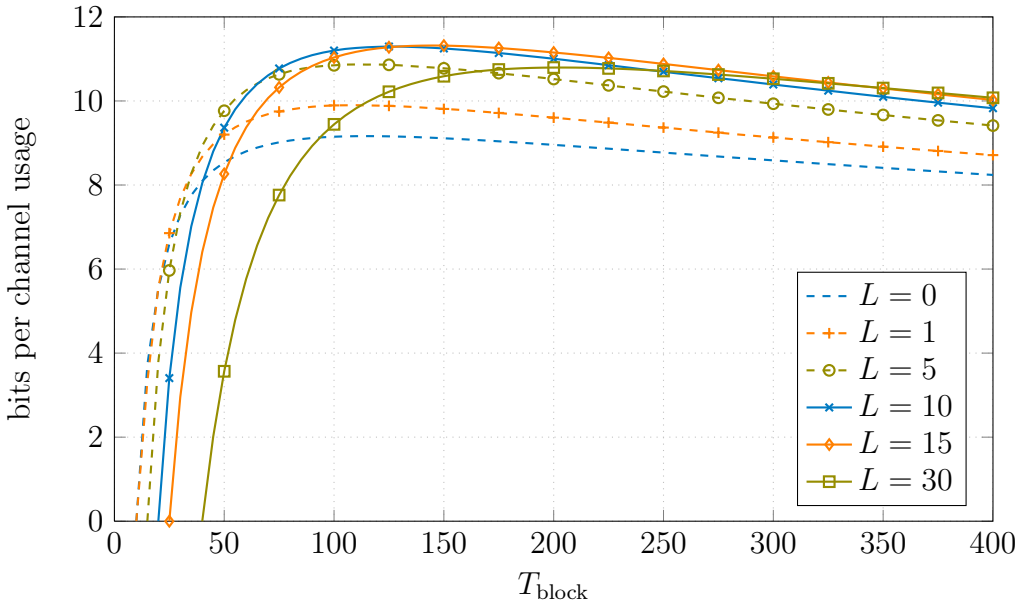


Figure 2.6. Upper bound for the cell sum rate in a high mobility scenario: $v = 30\,\text{km/h}$

in other cells. The previously described upper bound achieves both goals perfectly. The measured interference channels are set to zero. No interference is caused over these channels and not a single degree of freedom has to be spend for mitigating the interference. The base station can still use all their degrees of freedom to serve their own mobile devices.

2.5.1 Lower Bounds for Interference Coordination

With interference coordination, the base stations pursue the higher goal of maximizing the sum rate of the whole network. This will inherently maximize the sum rate of the cell. A compromise between serving the associated mobile devices and limiting the intercell interference at other mobile devices has to be found. The optimal compromise will always be at least as good as if only one of the goals is followed. For a given L, two lower bounds can be defined by looking at the two extremes.

2.5.1.1 No Cooperation

Obviously, no cooperation is always an option. The base stations use all their degrees of freedom to serve their own mobile devices egoistically. The first extreme can be found as follows and is equal to problem (2.17), where no channels are set to zero, i.e., $r_{b,k}$ (2.1) is used instead of $r_{\text{upper},b,k}$ (2.16). $\theta_{\text{blind},b,k}$ consists of the complete interference in the central cluster and $\sigma^2_{\text{od},b,k}$ contains only the serving channels, but L has still an influence on the efficiency ξ:

$$U_{\text{noco}} = \sum_b \left\{ \max_{\boldsymbol{Q}_{b,k}\ \forall(b,k)\in\mathcal{K}_b} \sum_{(b,k)\in\mathcal{K}_b} U(r_{b,k}), \quad \text{s.t. } \operatorname{tr}(\boldsymbol{Q}_b) \leq P \right\}. \tag{2.18}$$

2.5.1.2 Zero Forcing the Interference

The other extreme is to use the degrees of freedom to set the interference caused over measured channels to zero. The number of complex degrees of freedom at a base station are equal to the number of antennas N. If the number of measured interference channels at a base station is smaller than the number of antennas, $N - L$ degrees of freedom will be left for serving the associated mobile devices. This can be implemented by transmitting in the nullspace of the measured interference channels as described in [42]. The transmit covariance matrices are decomposed:

$$\boldsymbol{Q}_{b,k} = \boldsymbol{V}_b \hat{\boldsymbol{Q}}_{b.k} \boldsymbol{V}_b^{\mathrm{H}}, \tag{2.19}$$

where $\boldsymbol{V}_b \in \mathbb{C}^{N \times N-L}$ is the nullspace of the measured interference channels. The lower bound rates can be found by plugging (2.19) into the sum utility optimization (2.3) and optimizing the remaining reduced transmit covariance matrices $\hat{\boldsymbol{Q}}_{b,k} \in \mathbb{C}^{N-L,N-L}$ distributively:

$$U_{\mathrm{zfici}} = \sum_b \left\{ \max_{\hat{\boldsymbol{Q}}_{b,k} \ \forall (b,k) \in \mathcal{K}_b} \sum_{(b,k) \in \mathcal{K}_b} U(r_{\mathrm{zfici},b,k}), \quad \text{s.t. } \mathrm{tr}(\hat{\boldsymbol{Q}}_b) \leq P \right\}, \tag{2.20}$$

$$\begin{aligned} r_{\mathrm{zfici},b,k} &= r_{b,k}|_{\boldsymbol{Q}_{b,k} = \boldsymbol{V}_b \hat{\boldsymbol{Q}}_{b,k} \boldsymbol{V}_b^{\mathrm{H}} \ \forall (b,k) \in \mathcal{K}} \\ &= \xi \log_2 \left(1 + \frac{\boldsymbol{h}_{b,b,k}^{\mathrm{H}} \boldsymbol{V}_b \hat{\boldsymbol{Q}}_{b,k} \boldsymbol{V}_b^{\mathrm{H}} \boldsymbol{h}_{b,b,k}}{\sigma_{b,k}^2 + \theta_{\mathrm{blind},b,k} + \sum_{\hat{k}>k} \boldsymbol{h}_{b,b,k}^{\mathrm{H}} \boldsymbol{V}_b \hat{\boldsymbol{Q}}_{b,\hat{k}} \boldsymbol{V}_b^{\mathrm{H}} \boldsymbol{h}_{b,b,k}} \right). \end{aligned} \tag{2.21}$$

$\theta_{\mathrm{blind},b,k}$ has only of the interference over the not measured channels incorporated and $\sigma_{\mathrm{od},b,k}^2$ reflects the serving and the measured interference channels. If L is equal to or larger than N, zero forcing of the interference will force the base station to shut down.

2.5.2 Interference Temperatures

It is not possible to set the interference channels to zero as described for the upper bound and it is not advisable to ignore or zero force the interference as discussed for the lower bounds. But, the base stations can limit the intercell interference they cause over the L measured interference channels to a certain level, the interference temperatures [44].

Here, the interference temperature technique is combined with the outdating of the channel state information and the signaling efficiency according to the number of measured interference channels as described in the first part of this chapter (See equation (2.1)).

The problem at each base station can be formulated as a sum utility optimization with multiple linear constraints [44]:

$$\begin{aligned} \max_{\boldsymbol{Q}_{b,k} \ \forall (b,k) \in \mathcal{K}_b} & \sum_{(b,k) \in \mathcal{K}_b} U\left(r_{b,k}\right) \\ \text{s.t. } & \mathrm{tr}(\boldsymbol{Q}_b) \leq P \\ & \theta_{b,\hat{b},\hat{k}} \leq \gamma_{b,\hat{b},\hat{k}}, \ \forall (\hat{b},\hat{k}) \in \mathcal{L}_b, \end{aligned} \tag{2.22}$$

where $\mathcal{L}_b$ is the set of mobile devices, which are not associated to base station b, but the channel between these mobile devices and base station b is known. $\theta_{b,\hat{b},\hat{k}} = \boldsymbol{h}_{b,\hat{b},\hat{k}}^{\mathrm{H}} \boldsymbol{Q}_b \boldsymbol{h}_{b,\hat{b},\hat{k}}$ is

the interference base station b causes at mobile device $(\hat{b},\hat{k})$ and $\gamma_{b,\hat{b},\hat{k}}$ is the corresponding interference limit.

For network sum rate optimizations, this results in a weighted sum rate optimization with multiple linear constraints and can be solved with the minimax duality from [69] as shown in [70] (See Chapter 3).

2.5.3 Selection of $\gamma_{b,\hat{b},\hat{k}}$

The selection of the optimal interference constraint limits,

$$\underset{\left\{\boldsymbol{\gamma}_{b,\hat{b},\hat{k}}\ \forall(\hat{b},\hat{k})\in\mathcal{L}_b\ \forall b\in\mathcal{B}\right\}}{\operatorname{argmax}} \sum_b \left\{ \max_{\boldsymbol{Q}_{b,k}\ \forall(b,k)\in\mathcal{K}_b} \sum_{(b,k)\in\mathcal{K}_b} U\left(r_{b,k}\right), \right.$$
$$\text{s.t. } \operatorname{tr}(\boldsymbol{Q}_b)\leq P,$$
$$\left. \theta_{b,\hat{b},\hat{k}}\leq\gamma_{b,\hat{b},\hat{k}},\ \forall(\hat{b},\hat{k})\in\mathcal{L}_b \right\}, \tag{2.23}$$

is a problem with non-polynomial worst-case complexity [43]. The minimal interference constraint limit is zero and above the maximal interference $\theta_{\text{noco},b,\hat{b},\hat{k}}$, the value of the limit makes no difference: $\gamma_{b,\hat{b},\hat{k}}\in\left[0,\theta_{\text{noco},b,\hat{b},\hat{k}}\right]$. The maximal interference is found with the no cooperation lower bound optimization, where the constraint does not exist.

In [44], all constraint limits are set equal to the thermal noise power $\gamma_{b,\hat{b},\hat{k}}=\sigma_\eta^2$, $\forall(\hat{b},\hat{k})\in\mathcal{L}_b$, $\forall b\in\mathcal{B}$. Dotzler et al. proposed to set all constraint limits to the same value $\gamma_{b,\hat{b},\hat{k}}=\gamma\in[0,\theta_{\max}]$, $\forall(\hat{b},\hat{k})\in\mathcal{L}_b$, $\forall b\in\mathcal{B}$ which is found with a line search [43]. $\theta_{\max}=\max_{b\in\mathcal{B},\,(\hat{b},\hat{k})\in\mathcal{L}_b}\ \theta_{\text{noco},b,\hat{b},\hat{k}}$ is the maximum over all maximal interferences in the system. Here, $\gamma_{b,\hat{b},\hat{k}}$ is set to a scaled version of the intercell interference plus noise, which remains at the mobile device $(\hat{b},\hat{k})$ after the intercell interference over the measured channels is subtracted:

$$\gamma_{b,\hat{b},\hat{k}}=\alpha\left(\theta_{\text{blind},\hat{b},\hat{k}}+\sigma_{\hat{b},\hat{k}}^2\right) \tag{2.24}$$

The optimal α is found with a line search.

The described selection of $\gamma_{b,\hat{b},\hat{k}}$ requires that the base stations communicate. The base stations, which measured the interference channel to a mobile device, need to know the intercell interference, which remains at the regarded mobile device after the intercell interference over the measured channels is subtracted. The cost of this communication is neglected in the simulations. For the two extreme values of the scaling $\alpha\geq 0$, the solution of the described algorithm (2.23) converges to the solution with the lower bounds described in Section 2.5.1.1 and 2.5.1.2.

2.5.3.1 $\gamma_{b,\hat{b},\hat{k}}=0,\ \forall(\hat{b},\hat{k})\in\mathcal{L}_b$

If all interference constraint limits are set to zero, the constraints require that the base stations transmit in the nullspace of the interference channels, as they do in the zero forcing lower bound. For $L\geq N$ all rates are zero.

2.5.3.2 $\gamma_{b,\hat{b},\hat{k}} \to \infty, \forall(\hat{b},\hat{k}) \in \mathcal{L}_b$

By setting the interference constraint limits to a very large value, the maximization with multiple linear constraints is equal to the lower bound, where the intercell interference is simply ignored.

2.5.4 Simulations

The same parameters are used as in Section 2.4.1 for the low mobility scenario. But, there are only $K = 4$ mobile devices per cell served with the $N = 4$ transmit antennas per base station. In Figure 2.7 the lower bounds are plotted over the blocklength T_{block} for the sum rate maximization. For $L = 0$, the lower bound with zero forcing of the intercell interference (LB ZF) and the lower bound without cooperation (LB NC) are equal. These curves are also equal to the upper bound with $L = 0$. For increasing L, the LB NC decreases according to the overhead efficiency. The LB ZF increases at first, but degrades at $L = 3$ dramatically and hits zero for $L > 4$.

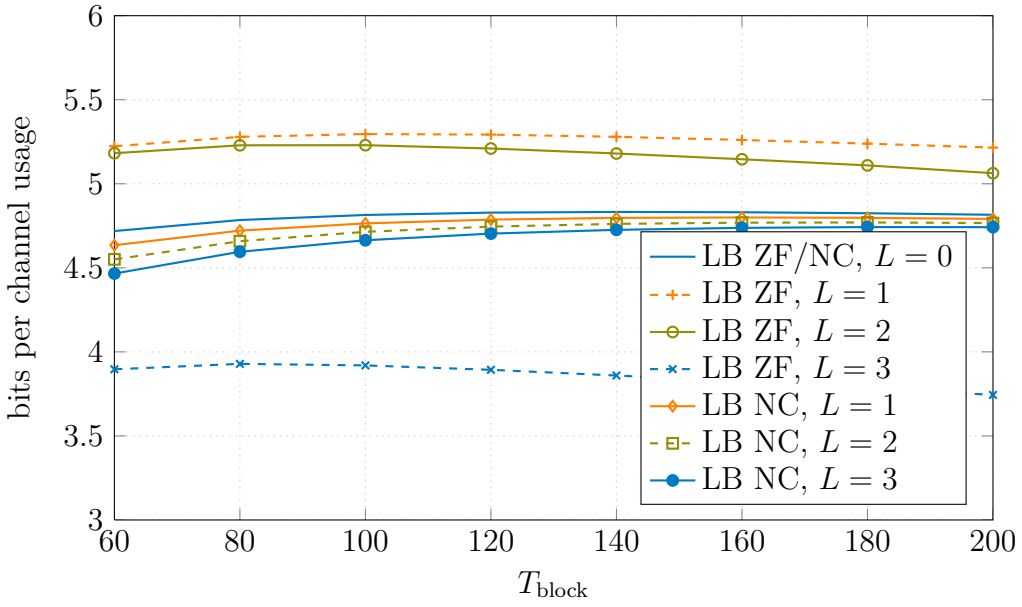


Figure 2.7. Lower bounds for the cell sum rate

The influence of the common scaling α can be seen in Figure 2.8 for $L = \{1, \ldots, 6\}$ under the maximum sum rate utility. The block length is fixed to $T_{\text{block}} = 200$. The lower bound with zero forcing of the interference is marked with an $\times$ at the left border of the plot for the different selections of L and the lower bound which ignores the interference is marked with an $\circ$ at the right border of the plot. For each L, the curve converges to the corresponding solution of the lower bounds for $\alpha \to 0$ and $\alpha \to \infty$, respectively. For any L, the optimal value for α can be found around $\alpha = 1$, where the optimal α is larger for larger L. The influence of choosing α correctly increased for increasing L. For $L = 1$, the gain of the best α compared to the lower bounds is rather small, whereas the correct selection of α for $L = 5$ is crucial. Although the lower bounds decrease for increasing L, the optimal value increases. But, the gain from $L = 3$ to $L = 5$ is very small. A further increase of L will result in worse sum rates at the optimal α.

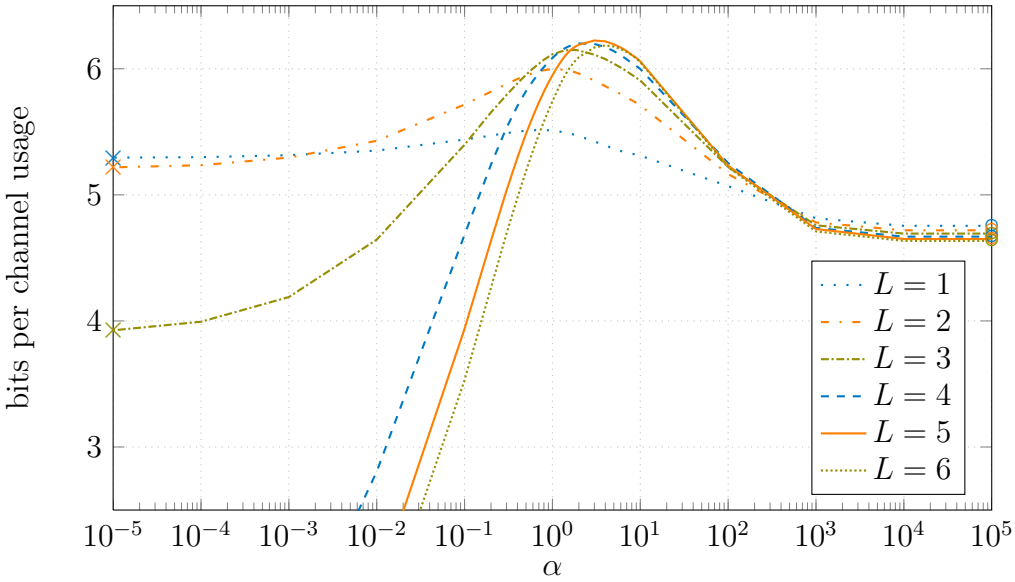


Figure 2.8. Cell sum rate with interference temperatures over the selection of α

The same investigation is done under the proportional-fair utility in Figure 2.9. The idea of this utility is to maximize the sum of the logarithms, here the natural logarithms, of the rates of all mobile devices in a cell (See Chapter 5). To improve the readability of the plot, it does not show the utility itself, but, a simple transformation thereof. The expected value of the logarithms of the mobile device rates is used as exponent for the base of the natural logarithm. It can be seen, that the utility improves in the beginning with increasing cooperation. The optimal L and α are again around fife and one, respectively. The proportional-fair utility finds a compromise between serving all users equally and maximizing the cell sum rate. As this utility is much fairer than the sum rate utility, mobile devices at the cell edge get served more often and have a greater influence on the utility. In contrast, mobile devices with a strong channel are the most important mobile devices for the sum rate utility. It can be seen in Appendix A1 that the situation of both mobile device classes can be improved significantly with cooperation. The sum rate is improved because the collocated base stations limit their interference to the mobile devices close to the site, while in a fairness approach the mobile devices with weak SINR are protected. The results are still optimistic, as many issues of an implementation are neglected. It is most likely that in a realistic system the optimal value of L is smaller and the limit to the amount of beneficial cooperation is stricter.

2.6 Conclusion

For the downlink of a cellular system with time division duplex, a loose upper bound for the possible network sum utility is formulated in this chapter. The channel state information is restricted to a subset of mobile devices and is prone to an outdating error. By nulling or scaling the known interference variance and treating the unknown interference as noise, the network sum utility maximization is transformed to a known problem, which can be solved. Based on this upper bound on network sum utility, on minimum overhead, on minimum outdating error, and on optimal channel state information measurement and selection, a

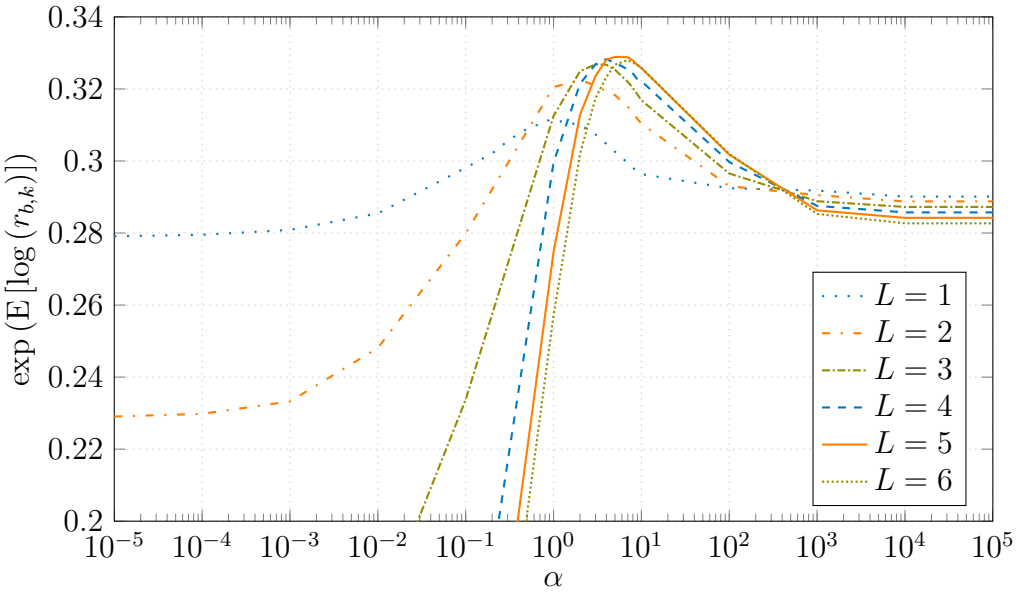


Figure 2.9. Proportional fair utility with interference temperatures over the selection of α

limit to beneficial interference coordination is presented. As the interference coordination is assumed to be optimal for the given channel state information, an increase in cooperation always requires additional measurements of channel state information. The necessary increase in the overhead and the associated lower efficiency of the system absorbs at some point the gain of cooperation. An increase beyond this limit results in lower data rates due to the large overhead.

The upper bound does not show how much cooperation is optimal, it only states that there is a limit to beneficial cooperation. This limit is very small for high mobility scenarios and large for low mobility scenarios, promoting the use of cooperation in the latter case. The interference temperatures technique is combined with the limited set of measured interference channels, the corresponding efficiency and outdating error of the channel state information. This allows the use of the interference temperature methods in a large scale system. The interference temperature algorithm is compared with two lower bounds, namely no cooperation and zero forcing of the interference over the measured interference channels and a heuristic is given for selecting the interference temperatures. It can be seen that the gain of cooperation is rather small for the low mobility scenario with interference temperatures. Therefore, there will be a vast amount of interference over not measured interference channels. Especially in systems where the complete available bandwidth is used by every base station and the cooperation is also limited by other effects besides the optimistic overhead used in this analysis. This advises to take a closer look at the interference over the not measured interference channels as it is done in the following chapters.

3. Sum Rate Maximization with Conic Constraints

Maximizing the weighted sum rate in the downlink of a MU-MIMO system with a single transmit power constraint typically consists of two steps: a transformation of the downlink problem to the dual uplink problem and solving the uplink problem, which can be rewritten as a convex optimization problem. The dual transformation was introduced in [71] and [72] in parallel.

Additionally to generic methods for convex optimizations, there are some methods specifically tailored to the uplink problem. The uplink problem can be solved with iterative waterfilling as described in [73]. A different approach for solving the problem is the steepest ascend algorithm. Hunger et al. showed that an orthogonal projection of the gradient onto the constraint set is required to find the steepest ascend direction [74]. With this projection, the deepest ascend algorithm has a very fast convergence behavior. The polite water-filling from Liu et al. is another possible approach for both, the dual transformation and solving the uplink problem with very fast convergence behavior [75]. The references in [74] and [75] give a broad overview of the existing algorithms for solving the weighted sum rate maximization.

Multiple linear constraints for the weighted sum rate maximization in the uplink were already addressed in [76], where Yu et al. investigated individual transmit power constraints for each user. However, multiple constraints in the downlink could not be handled by the existing dual transformations. In [69], Yu introduced a minimax duality, where the uplink problem is, on the one hand, a maximization of the weighted sum rate with respect to the transmit covariance matrices and, on the other hand, a minimization of the weighted sum rate with respect to the noise covariance matrix. With the minimax duality, constraints on the sum transmit covariance matrix in the downlink are transformed to constraints on the noise covariance matrix in the uplink. Yu et al. used this duality in [77] to tackle the weighted sum rate maximization with per antenna power constraints in the downlink. To solve the resulting minimax problem, Yu et al. updated the transmit and noise covariance matrices simultaneously with an adopted Newton's method. Feasibility of the covariance matrices was assured with the interior-point method.

The weighted sum rate algorithm of Yu et al. was generalized by Huh et al. in [67] for general linear constraints. They alternately solved the maximization of the transmit and noise covariance matrices. The utility was minimized with respect to the noise covariance matrix by a subgradient method, while the maximization with respect to the transmit covariance matrices was done with the adopted Newton's method.

In [70], Zhang et al. relaxed the multiple linear constraints to a single weighted sum constraint. The resulting problem could than be transformed to an uplink problem with the existing duality [71, 72]. Zhang et al. claimed that the optimum could be reached by

alternately maximizing the uplink problem with respect to the transmit covariance matrices and minimizing the downlink problem with respect to the weights. The claim was proven by showing that both, the original and the changed problem, have the same Lagrangian function. The Lagrangian multipliers of the original problem are the weights in the transformed problem.

Designed for more general interference networks, Liu et al. proposed two algorithms based on their polite water-filling to solve the weighted sum rate maximization with multiple linear constraints [78]. The cost function has to be maximized with respect to the transmit covariance matrices and minimized with respect to the Lagrangian multipliers of the constraints. The first algorithm alternates between polite water-filling for the transmit covariance matrices and a subgradient or an heuristic update algorithm for the Lagrangian multipliers. The polite water-filling itself is an alternating algorithm, which rotates between updating the transmit covariances in the uplink and downlink. The second algorithm includes the update of the Lagrangian multipliers into the alternating process of the polite water-filling.

Dotzler et al. introduced in [27] a minimax duality with linear conic constraints, where the uplink noise covariance itself is the Lagrangian multiplier for the constraints on the downlink covariance. In [79], where parts of this chapter can be found, the minimax duality is extended for multiple linear conic constraints. It is shown that the uplink problem can be solved with the rather simple but efficient gradient-projection algorithm. The required projections are discussed and an alternating and a joint update algorithm are proposed. Additionally, the tangent cone projection of the gradients to find the steepest ascend or descend direction are investigated. The proposed algorithms are tailored for conic constraints, which cannot be handled by the previously existing methods. Nevertheless, the proposed algorithms are also very efficient for optimizations with multiple linear constraints.

Next to per antenna power constraints, multiple linear constraints can be used as interference temperatures, which limit the received interference at selected receivers to a given level (See Section 2.5) [59, 77]. In a cellular network with local optimizations of the transmission strategy, a conic constraint, which shapes the sum transmit covariance of a transmitter to a scaled identity matrix, can be useful to remove uncertainty in the interference variance at disturbed users (See Section 4.9) [32]. By combining the shaping constraint with a linear sum power constraint, a controlled uncertainty in the intercell interference variance at the users can be introduced (See Section 4.10) [80]. An example for combining a conic and multiple linear constraints is the combination of the above mentioned shaping and interference temperature constraints.

3.1 System Model

In this chapter, the system model is a single cell MU-MIMO broadcast channel (See Figure 3.1). The base station is equipped with N antennas and serves K mobile devices with M antennas each. The sum transmit signal $\boldsymbol{x} = \sum_k \boldsymbol{x}_k \in \mathbb{C}^N$ contains the transmit signal for all mobile devices. Each of these transmit signals is assumed to be zero-mean, complex Gaussian with covariance matrix $\mathrm{E}\left[\boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{H}}\right] = \boldsymbol{Q}_k$. Transmit signals for different mobile devices are assumed to be independent. The sum transmit covariance matrix is

$$\boldsymbol{Q} = \mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\right] = \sum_k \mathrm{E}\left[\boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{H}}\right] = \sum_k \boldsymbol{Q}_k \in \mathbb{C}^{N \times N}. \tag{3.1}$$

The matrix $\boldsymbol{H}_k^{\mathrm{H}} \in \mathbb{C}^{M \times N}$ contains the channel coefficients between the antennas of the base station and user k. Compared to the single cell with random intercell interference variance from Section 4.3, here the receivers have multiple antennas in general and there is no virtual intercell interference variance. The intercell interference can be seen as a known and fixed part of the noise $\boldsymbol{n}_k \in \mathbb{C}^M$ in this chapter. Without loss of generality, the noise at different receive antennas is assumed to be independent and identically complex Gaussian distributed with zero-mean and normalized variance $\mathrm{E}\left[\boldsymbol{n}_k \boldsymbol{n}_k^{\mathrm{H}}\right] = \mathbf{I}$. Any system with a different noise covariance matrix or individual noise covariance matrices for each mobile device can be transformed to a system with an identity matrix as noise covariance matrix by an according substitution of the channel matrices. Perfect channel state information is assumed.

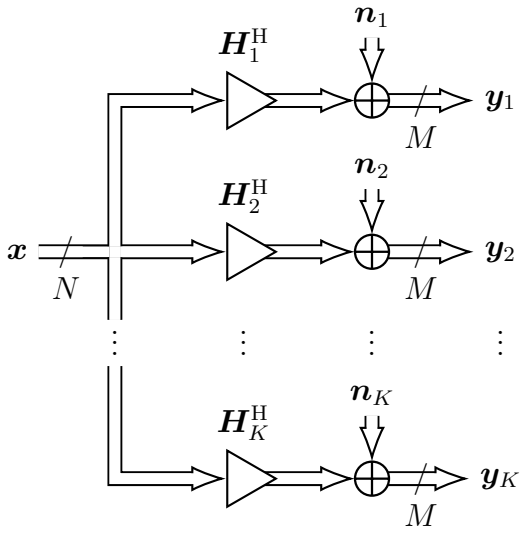


Figure 3.1. Downlink system model

The achievable, normalized rate of user k, within the capacity region of the MU-MIMO downlink with dirty paper coding, can be expressed as [81]

$$r_k^{\text{downlink}} = \log_2 \frac{\left|\mathbf{I} + \boldsymbol{H}_k^{\mathrm{H}} \left(\sum_{\hat{k} \geq k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{H}_k\right|}{\left|\mathbf{I} + \boldsymbol{H}_k^{\mathrm{H}} \left(\sum_{\hat{k} > k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{H}_k\right|}. \tag{3.2}$$

$\boldsymbol{H}_k^{\mathrm{H}} \left(\sum_{\hat{k} > k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{H}_k$ is the variance of the intracell interference with dirty paper coding [60]. In a system with single antenna receivers ($M = 1$), the rate simplifies to

$$r_k^{\text{downlink}} = \log_2 \frac{1 + \boldsymbol{h}_k^{\mathrm{H}} \left(\sum_{\hat{k} \geq k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{h}_k}{1 + \boldsymbol{h}_k^{\mathrm{H}} \left(\sum_{\hat{k} > k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{h}_k}. \tag{3.3}$$

3.2 General Cost Function

The weighted sum rate maximization problem in the downlink reads as

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \quad \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \boldsymbol{Q} \in \mathcal{C}, \tag{3.4}$$

where the sum transmit covariance matrix has to lie in the constraint set $\mathcal{C}$ (See Section 3.3) and all individual transmit covariance matrices have to be positive-semidefinite.

The weighted sum rate,

$$\Psi(\boldsymbol{Q}_{1:K}) = \sum_k w_k r_k^{\text{downlink}} = \sum_k w_k \log_2 \frac{\left|\mathbf{I} + \boldsymbol{H}_k^{\mathrm{H}} \left(\sum_{\hat{k}\geq k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{H}_k\right|}{\left|\mathbf{I} + \boldsymbol{H}_k^{\mathrm{H}} \left(\sum_{\hat{k}>k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{H}_k\right|}, \tag{3.5}$$

consists of the weights w_k and downlink rates r_k^{downlink} of the mobile devices $k \in \mathcal{K}$, where $\mathcal{K}$ is the set of all users. In a system with single antenna receivers ($M = 1$) this simplifies to

$$\Psi(\boldsymbol{Q}_{1:K}) = \sum_k w_k r_k^{\text{downlink}} = \sum_k w_k \log_2 \frac{1 + \boldsymbol{h}_k^{\mathrm{H}} \left(\sum_{\hat{k}\geq k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{h}_k}{1 + \boldsymbol{h}_k^{\mathrm{H}} \left(\sum_{\hat{k}>k} \boldsymbol{Q}_{\hat{k}}\right) \boldsymbol{h}_k}. \tag{3.6}$$

3.3 Constraints on the Transmit Covariance Matrix

The transmit covariance matrices $\boldsymbol{Q}_k \in \mathbb{C}^{N\times N}$ are the optimization variables in the weighted sum rate optimization and have to lie in a convex constraint set. This problem is analyzed with the methods from [82]. Each transmit covariance matrix has to be positive-semidefinite, as this is the nature of covariance matrices. For a bounded optimization, at least the trace of the sum transmit covariance matrix $\boldsymbol{Q} \in \mathbb{C}^{N\times N}$ has to be limited with an additional constraint. The goal of this section is to give an interpretation of linear and conic constraints. At first, the positive-semidefiniteness property of covariance matrices is discussed. Additional linear and conic constraints will be covered in the following.

3.3.1 Positive-Semidefinite

From $\boldsymbol{Q} = \mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\right] = \left(\mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\right]\right)^{\mathrm{H}}$, it can be seen that covariance matrices have to be Hermitian, they are equal to their own conjugate transpose. The lower and upper triangular part of the covariance matrix depend on each other, which reduces the number of independently selectable entries in the matrix. The entry of $\boldsymbol{Q}$ in the ith row and the jth column is $q_{ij} = \mathrm{E}\left[x_i x_j^*\right]$, where $x_i \in \mathbb{C}$ and $x_j \in \mathbb{C}$ are the ith and jth element of $\boldsymbol{x} \in \mathbb{C}^N$, respectively. The entry in the jth row and the ith column is $q_{ji} = \mathrm{E}\left[x_i^* x_j\right]$ and, therefore, the complex conjugate of q_{ij}. The diagonal elements $q_{ii} = \mathrm{E}\left[x_i x_i^*\right] = \mathrm{E}\left[|x_i|^2\right] \geq 0$ are all real valued and non-negative. It is also straight forward to see that $\boldsymbol{Q}$ is not only Hermitian, but also positive- semidefinite[1]:

$$\boldsymbol{h}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{h} = \boldsymbol{h}^{\mathrm{H}} \mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\right] \boldsymbol{h} = \mathrm{E}\left[\left|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{x}\right|^2\right] \geq 0, \quad \forall \boldsymbol{h} \in \mathbb{C}^N \quad \Leftrightarrow \quad \boldsymbol{Q} \succeq \mathbf{0}. \tag{3.7}$$

Regarding $\boldsymbol{h}^{\mathrm{H}}$ as a channel vector between the N transmit antennas and a single receive antenna, $\boldsymbol{h}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{h}$ is the variance of the received signal, which has to be real valued and non-negative for all $\boldsymbol{h} \in \mathbb{C}^N$.

To get a better insight to the properties of covariance matrices and the different constraints in the following, the special case of a real valued two-by-two $N = 2$ covariance

[1] The common definition of positive-semidefiniteness is used, where $\boldsymbol{Q}$ has to be Hermitian to guarantee that $\boldsymbol{h}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{h}, \forall \boldsymbol{h} \in \mathbb{C}^N$ is real valued.

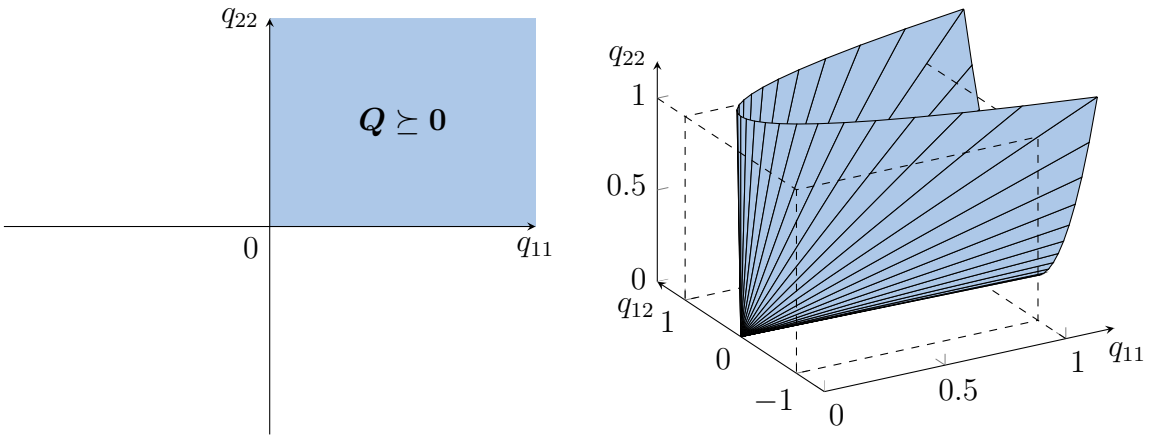
matrix is investigated:

$$\boldsymbol{Q}^{\mathrm{Re}} = \begin{bmatrix} \mathrm{E}\left[x_1^2\right] & \mathrm{E}\left[x_1 x_2\right] \\ \mathrm{E}\left[x_1 x_2\right] & \mathrm{E}\left[x_2^2\right] \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix}. \tag{3.8}$$

Real valued covariance matrices have to be symmetric. The independent entries of all real valued symmetric matrices of size $N \times N$ define a vector space with dimension $N(N+1)/2$. This is the number of entries in the lower or upper triangular part. A two-by-two real valued symmetric matrix has three degrees of freedom. The two diagonal elements (q_{11}, q_{22}) and the off-diagonal element (q_{12}). The corresponding vector space has three dimensions. For real valued two-by-two covariance matrices, the positive-definiteness requirement can be rewritten as

$$\begin{aligned} \boldsymbol{h}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{Re}} \boldsymbol{h} = h_1^2 q_{11} + 2 h_1 h_2 q_{12} + h_2^2 q_{22} \geq 0, \quad \forall h_1, h_2 \in \mathbb{R}, \\ \Leftrightarrow \quad q_{11} \geq 0, \quad q_{22} \geq 0, \quad q_{12}^2 \leq q_{11} q_{22}. \end{aligned} \tag{3.9}$$

This defines the positive-semidefinite cone for real valued two-by-two covariance matrices. See Figure 3.2 (b) for a graphical representation. The diagonal entries have to be positive and the off-diagonal element depends on the selection of the diagonal entries.



(a) Set for diagonal matrices (b) Boundary for real valued matrices

Figure 3.2. Positive-semidefiniteness for two-by-two covariance matrices

A two-dimensional representation can be found by limiting the transmit covariance matrix to a two-by-two diagonal matrix. E.g., for two independently operating transmit antennas, the total transmit covariance matrix is diagonal and has the two real valued diagonal elements as degrees of freedom. The corresponding vector space has two dimensions:

$$\boldsymbol{Q}^{\mathrm{diag}} = \begin{bmatrix} q_{11} & 0 \\ 0 & q_{22} \end{bmatrix}. \tag{3.10}$$

The diagonal elements of a positive-semidefinite covariance matrix have to be non-negative. For the two-by-two diagonal covariance matrix, this is represented by the positive quadrant in the two dimensional vector space (see Figure 3.2 (a)).

3.3.2 Linear Constraints

Without any additional constraints, the entries of the optimal $\boldsymbol{Q}_k$ converge to infinity. Therefore, at least one constraint limiting the trace of the sum transmit covariance matrix is necessary to have a bounded optimization. Treating the variance of a signal as its power, the diagonal entries of the sum transmit covariance matrix stand for the power radiated from each antenna, respectively. The sum of these entries represents the sum power radiated by the transmitter.

A rank one linear constraints limits the transmitted power in the direction $\boldsymbol{a}^*$ to the value a:

$$\mathrm{E}\left[|\boldsymbol{a}^{\mathrm{H}}\boldsymbol{x}|^2\right] = \boldsymbol{a}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{a} = \mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{A}\right) \leq a, \tag{3.11}$$

with $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{H}} \in \mathbb{C}^{N\times N}$ as rank one positive-semidefinite constraint matrix. A linear constraint of higher rank limits the weighted sum of the transmitted power in multiple directions $\boldsymbol{a}_l^*$ to the value a:

$$\sum_l \mathrm{E}\left[|\boldsymbol{a}_l^{\mathrm{H}}\boldsymbol{x}|^2\right] = \sum_l \boldsymbol{a}_l^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{a}_l = \mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{A}\right) \leq a, \tag{3.12}$$

with $\boldsymbol{A} = \sum_l \boldsymbol{a}_l\boldsymbol{a}_l^{\mathrm{H}} \in \mathbb{C}^{N\times N}$ as positive-semidefinite constraint matrix. A linear constraint defines a hyperplane in the vector space of the covariance matrix entries with the corresponding entries of the constraint matrix as normal vector. All points in the vector space of the covariance matrix entries fulfilling the linear constraint with equality lie on this hyperplane. All points in the half-space on one side of the hyperplane are in the constraint set and all points on the other side are not in the constraint set.

The necessary limitation for bounding the optimization can either be achieved by a single linear constraint with a positive-definite constraint matrix. This constraint defines a hyperplane limiting all diagonal entries in the vector space of the constraint set and will always be active in the optimal point. E.g., the commonly utilized sum power constraint is a full rank linear constraint with $\boldsymbol{A} = \mathbf{I}$ and the sum transmit power $a = P$ (See Figure 3.3),

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}} \ \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P. \tag{3.13}$$

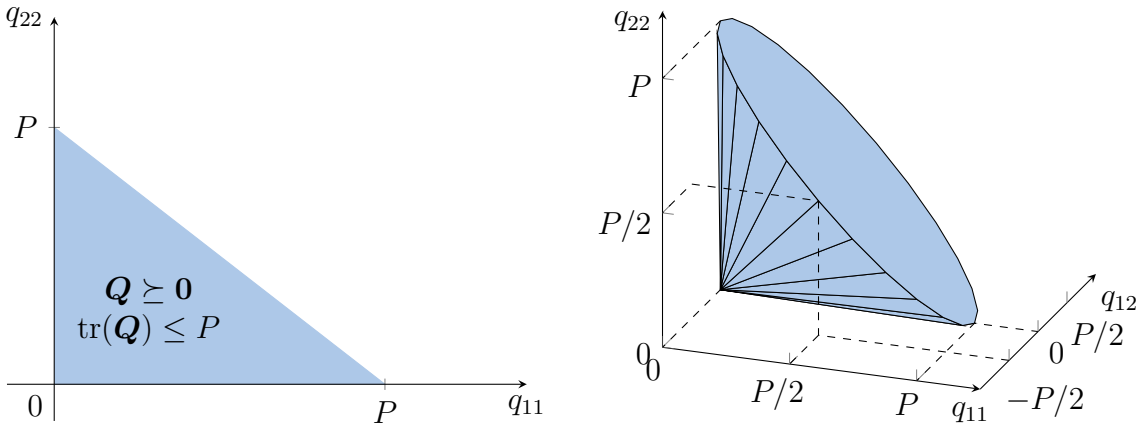
The optimization can also be bounded by combining multiple linear constraints of lower rank,

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \ \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{A}_l\right) \leq a_l, \quad \forall l, \tag{3.14}$$

where the sum of all constraint matrices $\sum_l \boldsymbol{A}_l$ is of full rank. Then, the combination defines multiple hyperplanes, which limit all the diagonal entries in the vector space of the covariance matrix entries.

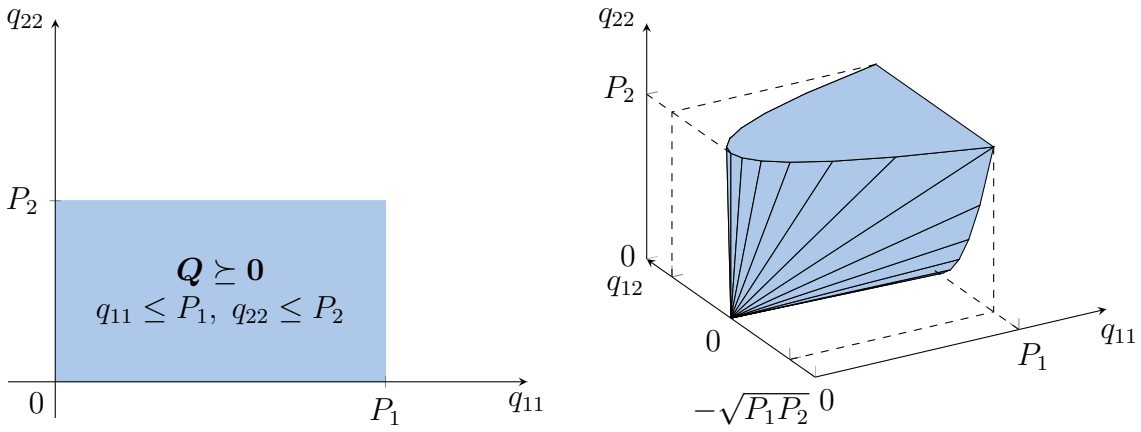
The per antenna power constraints are an example for rank one constraints. Each diagonal entry is limited to a certain power level (See Figure 3.4) [77]:

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \ \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{e}_l\boldsymbol{e}_l^{\mathrm{T}}\right) \leq P_l, \quad \forall l \in \{1,\dots N\}, \tag{3.15}$$



(a) Feasible region for diagonal matrices (b) Boundary for real valued matrices

Figure 3.3. Sum power constraint for two-by-two covariance matrices



(a) Feasible region for diagonal matrices (b) Boundary for real valued matrices

Figure 3.4. Per antenna power constraint for two-by-two covariance matrices

where $\boldsymbol{e}_l$ is the lth unit vector. As the optimal point for a full rank covariance matrix is on the boundary of the set, all per antenna power constraints will be active in this point. If the per antenna power constraints are combined with a looser sum power constraint ($P < \sum_l P_l$) or the covariance matrix is rank deficient, some per antenna constraints will be inactive.

Another example for rank one linear constraints are interference temperatures, where the generated interference power at certain receive antennas are limited [59]. The interference temperatures are typically combined with a general transmit power constraint:

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \begin{aligned} &\mathrm{tr}\left(\boldsymbol{Q}\right) \leq P \\ &\mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{h}_l\boldsymbol{h}_l^{\mathrm{H}}\right) \leq a_l, \quad \forall l. \end{aligned} \tag{3.16}$$

$\boldsymbol{h}_l^{\mathrm{H}}$ is the channel between the base station and the single receive antenna of the disturbed user. If the interference power is limited to $a_l = 0$, the optimization should be projected into the kernel of this or these channels [42]. If the constraints, which demand zero interference, span the whole vector space, the base station has to shut down.

3.3.3 Linear Conic Constraints

A conic constraint describes an infinite cone by its apex and base as feasible region in the vector space of the transmit covariance matrix entries. The positive-semidefinite cone is a ready example for a conic constraint with the apex in the origin and the base opens the cone towards any non-negative diagonal elements (See Figure 3.2). The utilized constraints in this thesis are all based on the positive-semidefinite cone, although, they are referred to as general conic constraints.

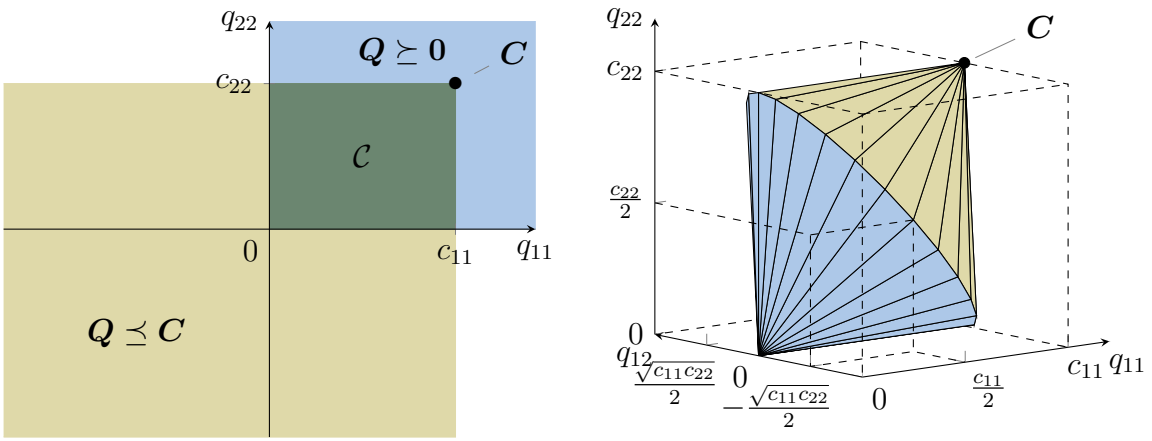
A shaping constraint, a linear conic constraint in the variable $\boldsymbol{Q}$, limits the transmitted power in all directions to the power, which would be transmitted with the shaping matrix $\boldsymbol{C} \in \mathbb{C}^{N\times N}$:

$$\begin{aligned} \mathrm{E}\left[|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{x}|^2\right] = \boldsymbol{h}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{h} &\leq \boldsymbol{h}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathbb{C}^N \\ 0 &\leq \boldsymbol{h}^{\mathrm{H}}\left(\boldsymbol{C}-\boldsymbol{Q}\right)\boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathbb{C}^N \\ \boldsymbol{0} \preceq \boldsymbol{C}-\boldsymbol{Q} \quad &\Leftrightarrow \quad \boldsymbol{Q}-\boldsymbol{C} \preceq \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{Q} \preceq \boldsymbol{C}. \end{aligned} \tag{3.17}$$

Equation (3.17) states that $-\boldsymbol{Q}$ shifted by $\boldsymbol{C}$ has to be positive-semidefinite or $\boldsymbol{Q}$ shifted by $-\boldsymbol{C}$ has to be negative-semidefinite. For the shaping constraint, the positive-semidefinite cone is turned towards negative diagonal entries of the covariance matrix and shifted by the shaping matrix $\boldsymbol{C}$. The shaping constraint needs to limit the transmit power, i.e., the diagonal entries of the covariance matrix. Therefore, the apex has to be a point in the vector space of the covariance matrix entries corresponding to non-negative diagonal entries and the all zeros matrix should be in the feasible set. These conditions are fulfilled with any positive-definite matrix $\boldsymbol{C}$:

$$\max_{\boldsymbol{Q}_k \succeq \boldsymbol{0}\ \forall k} \ \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \boldsymbol{Q} \preceq \boldsymbol{C}. \tag{3.18}$$

In Figure 3.5, the shaping is demonstrated with the two-by-two covariance matrices. c_{11}, c_{22}, and c_{12} are the two diagonal and the off-diagonal entry of a two-by-two shaping matrix, respectively.



(a) Feasible region $\mathcal{C}$ for diagonal matrices (b) Boundary for real valued matrices ($c_{12} = 0$)

Figure 3.5. Shaping constraint for two-by-two covariance matrices

The selection $\boldsymbol{C}^{\text{id}} = c\mathbf{I}$ shapes the transmit covariance to a scaled identity matrix with sum transmit power cN:

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \boldsymbol{Q} \preceq \frac{P}{N}\mathbf{I}. \tag{3.19}$$

If the transmit covariance matrix is of full rank, than the constraint will always be active in the optimal point.

The shaping constraint can be generalized to a linear conic constraint with apex $\boldsymbol{C}$ by adding a subspace $\mathcal{Z}$ to increase the opening of the cone:

$$\exists \boldsymbol{Z} \in \mathcal{Z} : \ \boldsymbol{Q} \preceq \boldsymbol{C} + \boldsymbol{Z}. \tag{3.20}$$

This can be seen as a positive-semidefinite linear conic constraint in the two variables $\boldsymbol{Q}$ and $\boldsymbol{Z} \in \mathcal{Z}$. The optimization is done over the transmit covariance matrices and the matrix $\boldsymbol{Z}$:

$$\max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \boldsymbol{Z} \in \mathcal{Z}}} \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \boldsymbol{Q} \preceq \boldsymbol{C} + \boldsymbol{Z}. \tag{3.21}$$

Any point in the subspace $\mathcal{Z}$ shifted by $\boldsymbol{C}$ can be seen as an apex of a feasible positive-semidefinite cone. The union of all these cones, which is again a cone, is the feasible region of the constraint. The two extreme cases are $\mathcal{Z} = \{\mathbf{0}\}$, which gives the shaping constraint, and $\mathcal{Z} = \mathbb{C}^{N \times N}$, which makes the constraint obsolete. Note, the here reviewed linear conic constraints are always the union of positive-semidefinite cones and, therefore, do not cover all possible conic constraints.

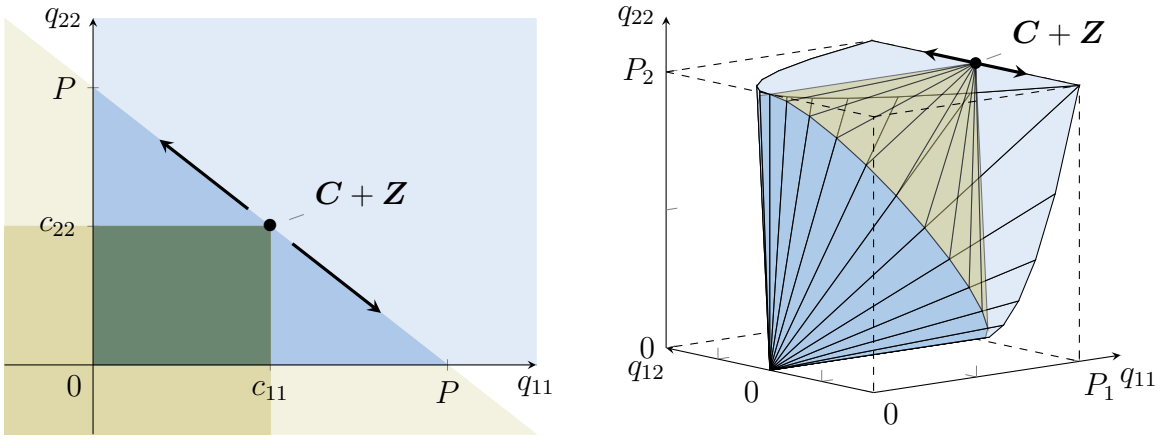
A linear constraint $\operatorname{tr}(\boldsymbol{Q}\boldsymbol{A}) \leq a$ is a conic constraint, where the infinite cone is a half-space. The boundary of the halfspace is a hyperplane with the entries of the constraint matrix $\boldsymbol{A}$ as normal vector. The apex $\boldsymbol{C}$ of the conic constraint can be chosen as any point on the hyperplane of the linear constraint $\operatorname{tr}(\boldsymbol{C}\boldsymbol{A}) = a$. Combined with the subspace $\mathcal{Z} = \{\boldsymbol{Z} : \ \operatorname{tr}(\boldsymbol{Z}\boldsymbol{A}) = 0\}$, all points on the hyperplane of the linear constraint are apexes for negative-semidefinite cones and the union is the feasible half-space of the linear constraint. This can be seen for diagonal two-by-two covariance matrices with the sum power constraint (See equation (3.13)) in Figure 3.6 (a). The sum of the diagonal elements of $\boldsymbol{Z}$ has to be zero and the off-diagonal elements can have any value. The off-diagonal elements can be neglected as the covariance matrix has to be diagonal. Keeping the trace of $\boldsymbol{Z}$ zero, the apex $\boldsymbol{C} + \boldsymbol{Z}$ is moved along $c_{11} + c_{22} = P$, which increases the feasible region $\boldsymbol{C}$ by all points fulfilling $q_{11} + q_{22} \leq P$.

Multiple linear constraints can be combined to one conic constraint, as long as an apex $\boldsymbol{C}$ can be found, which lies on all hyperplanes. The per antenna power constraints (See equation (3.15)) can be combined in such a way with the selection

$$\boldsymbol{C}^{\text{per ant}} = \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_N \end{bmatrix}, \tag{3.22}$$

$$\mathcal{Z}^{\text{per ant}} = \left\{ \boldsymbol{Z}^{\text{per ant}} : \ \operatorname{tr}\left(\boldsymbol{Z}^{\text{per ant}} \boldsymbol{e}_l \boldsymbol{e}_l^{\mathrm{T}}\right) = 0, \quad \forall l \in \{1, \ldots, N\} \right\}. \tag{3.23}$$

The apex is the point $\boldsymbol{C}^{\text{per ant}}$, where all per antenna power constraints are active. The set $\mathcal{Z}^{\text{per ant}}$ restricts all diagonal elements of $\boldsymbol{Z}^{\text{per ant}}$ to be zero, but all off-diagonal elements are unrestricted. In Figure 3.6 (b), this is shown for the case of real valued two-by-two covariance matrices. The apex $\boldsymbol{C} + \boldsymbol{Z}$ can be any point on the line with $q_{11} = c_{11}$ and $q_{22} = c_{22}$, which extends the feasible region to the region already shown in Figure 3.4 (b).



(a) Sum power constraint expressed as a linear conic constraint for diagonal matrices

(b) Boundary for real valued matrices with per antenna power constraints expressed as a linear conic constraint

Figure 3.6. Shaping constraint extended by the subspace $\mathcal{Z}$ for two-by-two covariance matrices

Multiple conic constraints can be formulated as

$$\exists \boldsymbol{Z}_l \in \mathcal{Z}_l : \ \boldsymbol{Q} \preceq \boldsymbol{C}_l + \boldsymbol{Z}_l, \quad \forall l, \tag{3.24}$$

which can be used for example to combine a shaping constraint with a stricter sum power constraint [59] or interference temperature constraints.

3.4 Uplink-Downlink Duality

The weighted sum rate optimization in the downlink (See equation (3.4)) is a non-convex problem and, therefore, hard to solve. But, the downlink problem can be transformed into a dual uplink problem, which is convex and can be solved with standard convex optimization tools. The transmit covariance matrices for the downlink can then be found based on the dual variables. Under a sum power constraint, the capacity region of the broadcast and the multiple access channel are exactly the same [71, 72]. Based on a minimax duality, Yu could show, that the weighted sum rate of the downlink and dual uplink are the same for any linear constraints [69]. This duality was extended to optimizations with linear conic constraints by Dotzler et al. [27]. Wiegart et al. discuss the relevant cases of rank deficient channel and constraint matrices [83], which are excluded in this work.

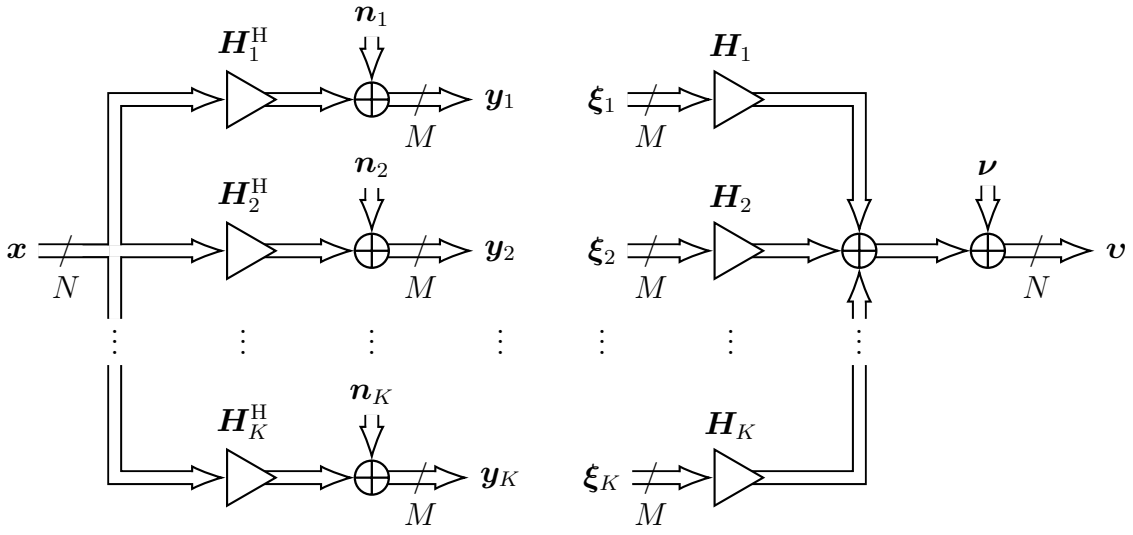
3.4.1 Uplink System Model

The dual uplink channel to the MU-MIMO broadcast channel is a MU-MIMO multiple access channel (See Figure 3.7 (b)). The base station is equipped with N antennas and receives

signals from K mobile devices with M antennas each. The transmit signals are assumed to be zero-mean, complex Gaussian with covariance matrix $\mathrm{E}\left[\boldsymbol{\xi}_k \boldsymbol{\xi}_k^{\mathrm{H}}\right] = \boldsymbol{\Xi}_k$. Transmit signals from different mobile devices are assumed to be independent. The channels in the dual uplink are equal to the flipped channels in the downlink:

$$\boldsymbol{H}_k^{\text{uplink}} = \left(\boldsymbol{H}_k^{\text{downlink}}\right)^{\mathrm{H}} = \boldsymbol{H}_k \in \mathbb{C}^{N \times M}. \tag{3.25}$$

The noise $\boldsymbol{\nu} \in \mathbb{C}^N$ is assumed to be zero-mean, complex Gaussian with covariance matrix $\mathrm{E}\left[\boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{H}}\right] = \boldsymbol{\Omega}$.



(a) Downlink system model (b) Dual uplink system model

Figure 3.7. Duality between broadcast channel and multiple access channel

In the dual multiple access channel, the rate of user k with successive interference cancellation can be found with flipped channels as

$$r_k^{\text{uplink}} = \log_2 \frac{\left|\boldsymbol{\Omega} + \sum_{\hat{k} \leq k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}}\right|}{\left|\boldsymbol{\Omega} + \sum_{\hat{k} < k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}}\right|}. \tag{3.26}$$

$\sum_{\hat{k}<k} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_k$ is the covariance of the intracell interference with successive interference cancellation. In an uplink system with single antenna transmitters ($M = 1$), the rate simplifies to

$$r_k^{\text{uplink}} = \log_2 \frac{\left|\boldsymbol{\Omega} + \sum_{\hat{k} \leq k} q_{\hat{k}} \boldsymbol{h}_{\hat{k}} \boldsymbol{h}_{\hat{k}}^{\mathrm{H}}\right|}{\left|\boldsymbol{\Omega} + \sum_{\hat{k} < k} q_{\hat{k}} \boldsymbol{h}_{\hat{k}} \boldsymbol{h}_{\hat{k}}^{\mathrm{H}}\right|}, \tag{3.27}$$

where q_k are the scalar transmit powers of the K mobile devices, respectively.

3.4.2 Uplink Cost Function

The sum rate with MU-MIMO and successive interference cancellation in the uplink reads as

$$\begin{aligned}\Phi(\boldsymbol{\Xi}_{1:K},\ \boldsymbol{\Omega}) &= \sum_k w_k r_k^{\text{uplink}} \\ &= \sum_k w_k \log_2 \frac{\left|\boldsymbol{\Omega} + \sum_{\hat{k}\le k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}}\right|}{\left|\boldsymbol{\Omega} + \sum_{\hat{k}<k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}}\right|} \\ &= \sum_k \alpha_k \log_2 \left|\boldsymbol{\Omega} + \sum_{\hat{k}\le k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}}\right| - w_1 \log_2 |\boldsymbol{\Omega}|\,, \end{aligned} \tag{3.28}$$

with $\alpha_k = (w_k - w_{k+1})$ for $k = 1 \dots K-1$ and $\alpha_K = w_K$. Without loss of generality, it can be assumed that the mobile devices are handled in the optimal decoding order, i.e., the weights are sorted in non-increasing order $w_k \ge w_{k+1}$ and all α_k are non-negative. The sum rate is a convex function of the transmit covariance matrices $\boldsymbol{\Xi}_k$ and can be optimized with standard convex optimization methods. In an uplink system with single antenna transmitters ($M = 1$), the sum rate simplifies to

$$\Phi(q_{1:K},\ \boldsymbol{\Omega}) = \sum_k w_k r_k^{\text{uplink}} = \sum_k \alpha_k \log_2 \left|\boldsymbol{\Omega} + \sum_{\hat{k}\le k} q_{\hat{k}} \boldsymbol{h}_{\hat{k}} \boldsymbol{h}_{\hat{k}}^{\mathrm{H}}\right| - w_1 \log_2 |\boldsymbol{\Omega}|\,. \tag{3.29}$$

3.4.3 Duality with Linear Constraints

As shown in [69], the dual problem of the weighted sum rate optimization in the downlink with a general linear constraint with positive-definite constraint matrix $\boldsymbol{A} \succ \boldsymbol{0}$,

$$\max_{\boldsymbol{Q}_k \succeq \boldsymbol{0}\ \forall k} \ \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \operatorname{tr}(\boldsymbol{Q}\boldsymbol{A}) \le a, \tag{3.30}$$

is a weighted sum rate optimization in the uplink with a linear constraint on the sum transmit covariance matrix and both achieve the same sum rate. The constraint matrix in the downlink $\boldsymbol{\Omega} = \boldsymbol{A}$ becomes the noise covariance matrix in the uplink (See equation (3.26)):

$$\max_{\boldsymbol{\Xi}_k \succeq 0\ \forall k} \ \Phi(\boldsymbol{\Xi}_{1:K},\ \boldsymbol{A}) \qquad \text{subject to} \quad \operatorname{tr}(\boldsymbol{\Xi}) \le a. \tag{3.31}$$

It can be seen that the constraint matrix has to be of full rank. Otherwise, the downlink constraint would not limit all diagonal entries and parts of the uplink cost function would run towards infinity. The duality with multiple linear constraints is similarly to the duality with a single linear constraint. With multiple linear constraints, only the sum of all constraint matrices has to be of full rank.

3.4.4 Duality with Linear Conic Constraints

The minimax duality with a conic constraint from [27] shows that the sum rate of

$$\max_{\substack{\boldsymbol{Q}_k \succeq \boldsymbol{0}\ \forall k \\ \boldsymbol{Z} \in \mathcal{Z}}} \ \Psi(\boldsymbol{Q}_{1:K}) \qquad \text{subject to} \quad \boldsymbol{Q} \preceq \boldsymbol{C} + \boldsymbol{Z}, \tag{3.32}$$

with the positive-defininte covariance matrix $\boldsymbol{C} \succ \boldsymbol{0}$, is the same as the sum rate of the dual saddle point problem,

$$\min_{\substack{\boldsymbol{\Omega}\succeq\boldsymbol{0},\, \boldsymbol{\Omega}\in\mathcal{Z}^{\perp} \\ \mathrm{tr}(\boldsymbol{C}\boldsymbol{\Omega})=P}} \quad \max_{\substack{\boldsymbol{\Xi}_k\succeq\boldsymbol{0}\ \forall k \\ \mathrm{tr}(\boldsymbol{\Xi})=P}} \Phi(\boldsymbol{\Xi}_{1:K},\ \boldsymbol{\Omega}), \tag{3.33}$$

where $P \in \mathbb{R}_+$ is any positive constant and $\mathcal{Z}^{\perp}$ is the orthogonal subspace to $\mathcal{Z}$:

$$\mathcal{Z}^{\perp} = \{\boldsymbol{\Omega} : \mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{Z}) = 0, \forall \boldsymbol{Z} \in \mathcal{Z}\}\,. \tag{3.34}$$

The inner maximization is concave with respect to the transmit covariance matrices and the outer minimization is convex with respect to the noise covariance matrix.

In the following, this duality is is extended for multiple (L) conic constraints. Because of strong duality[1] and the variables being selected from closed sets, the downlink problem,

$$\max_{\substack{\boldsymbol{Q}_k\succeq\boldsymbol{0}\ \forall k \\ \boldsymbol{Z}_l\in\mathcal{Z}_l\ \forall l}} \Psi(\boldsymbol{Q}_{1:K}) \quad \text{subject to} \quad \boldsymbol{Q} \preceq \boldsymbol{C}_l + \boldsymbol{Z}_l,\ \forall l, \tag{3.35}$$

can be expressed by its Lagrangian dual,

$$\min_{\boldsymbol{\Omega}_l\succeq\boldsymbol{0}\ \forall l} \quad \max_{\substack{\boldsymbol{Q}_k\succeq\boldsymbol{0}\ \forall k \\ \boldsymbol{Z}_l\in\mathcal{Z}_l\ \forall l}} \Psi(\boldsymbol{Q}_{1:K}) - \sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l(\boldsymbol{Q} - \boldsymbol{C}_l - \boldsymbol{Z}_l)\right), \tag{3.36}$$

where $\boldsymbol{\Omega}_l \in \mathbb{C}^{N\times N}$ is the Lagrangian multiplier of constraint l. The inner problem of (3.36) is unbounded, unless $\boldsymbol{\Omega}_l \in \mathcal{Z}_l^{\perp},\ \forall l$, where

$$\mathcal{Z}_l^{\perp} = \{\boldsymbol{\Omega}_l : \mathrm{tr}(\boldsymbol{\Omega}_l\boldsymbol{Z}_l) = 0, \forall \boldsymbol{Z}_l \in \mathcal{Z}_l\}\,, \tag{3.37}$$

and can be rearranged as

$$\min_{\boldsymbol{0}\preceq\boldsymbol{\Omega}_l\in\mathcal{Z}_l^{\perp}\ \forall l} \quad \max_{\boldsymbol{Q}_k\succeq\boldsymbol{0}\ \forall k} \Psi(\boldsymbol{Q}_{1:K}) - \sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l\boldsymbol{Q}\right) + \sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l\boldsymbol{C}_l\right). \tag{3.38}$$

Due to the complementary slackness condition of all constraints,

$$\sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l\boldsymbol{Q}\right) = \sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l\boldsymbol{C}_l\right) \tag{3.39}$$

has to be fulfilled. A joint scaling of all dual variables does not change the solution. Without loss of generality, both sides of equation (3.39) can be fixed to some value P. This relaxation is equal to fulfilling all individual complementary slackness conditions in the optimal point. Having equation (3.39) as constraint to the weighted sum rate maximization has the same Lagrangian function with some arbitrary joint scaling. With this relaxation, the saddle point problem reads as

$$\min_{\substack{\boldsymbol{0}\preceq\boldsymbol{\Omega}_l\in\mathcal{Z}_l^{\perp}\ \forall l \\ \sum_l \mathrm{tr}(\boldsymbol{\Omega}_l\boldsymbol{C}_l)=P}} \quad \max_{\substack{\boldsymbol{Q}_k\succeq\boldsymbol{0}\ \forall k \\ \mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{Q})=P}} \Psi(\boldsymbol{Q}_{1:K}), \tag{3.40}$$

[1] The strong duality for multiple constraints can be shown with the same methods presented in [27] for a single constraint

where $\boldsymbol{\Omega} = \sum_l \boldsymbol{\Omega}_l$ is the sum of all Lagrangian multipliers. The inner maximization is a weighted sum rate maximization with a single linear constraint and, therefore, can be replaced by its dual maximization in the uplink:

$$\min_{\substack{\mathbf{0} \preceq \boldsymbol{\Omega}_l \in \mathcal{Z}_l^{\perp} \ \forall l \\ \sum_l \mathrm{tr}(\boldsymbol{\Omega}_l \boldsymbol{C}_l) = P}} \ \max_{\substack{\boldsymbol{\Xi}_k \succeq \mathbf{0} \ \forall k \\ \mathrm{tr}(\boldsymbol{\Xi}) = P}} \ \Phi\left(\boldsymbol{\Xi}_{1:K}, \sum_l \boldsymbol{\Omega}_l\right). \tag{3.41}$$

The dual uplink noise covariance matrix $\boldsymbol{\Omega}$ is equal to the sum of the Lagrangian multiplier matrices of the downlink problem.

In this thesis, it is assumed that the constraint matrices $\boldsymbol{C}_l$ and the joint channel matrix $[\boldsymbol{H}_1, \ \ldots \ , \boldsymbol{H}_K]$ are of full rank. The problem and treatment of rank deficient matrices is described and handled in [27, 83]. Expressing the linear interference temperatures as conic constraints with the requirement of zero interference in a direction leads to a rank deficient $\boldsymbol{C}_l$. For this case, convergence of the optimization cannot be guaranteed with this duality. The problem has to be transformed into the kernel of the forbidden directions before the duality is applied [42].

3.4.5 Recovery of the Downlink Transmit Covariance Matrices

The individual transmit covariance matrices in the downlink can be computed based on the optimal solution to the dual problem:

$$\boldsymbol{Q}_k = \frac{1}{\mu \ln(2)} \left(\left(\boldsymbol{\Omega} + \sum_{\hat{k}<k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}} \right)^{-1} - \left(\boldsymbol{\Omega} + \sum_{\hat{k} \leq k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}} \right)^{-1} \right), \tag{3.42}$$

where μ is the Lagrangian multiplier of the power constraint in the uplink and the uplink noise covariance matrix $\boldsymbol{\Omega} = \sum_l \boldsymbol{\Omega}_l$ is the sum of all Lagrangian multipliers of the downlink constraints. If μ is unknown, but the power constraint should be active, the correct scaling can always be found such, that the sum of the traces fulfills the power constraint. In the algorithms, this calculation is called MAC2BC($\bullet$), while the conversion from the downlink covariance matrices to the uplink covariance matrices is called BC2MAC($\bullet$).

3.5 Alternating Projected Gradient Algorithm

Recall the optimization problem (3.41),

$$\min_{\substack{\mathbf{0} \preceq \boldsymbol{\Omega}_l \in \mathcal{Z}_l^{\perp} \ \forall l \\ \sum_l \mathrm{tr}(\boldsymbol{\Omega}_l \boldsymbol{C}_l) = P}} \ \max_{\substack{\boldsymbol{\Xi}_k \succeq \mathbf{0} \ \forall k \\ \mathrm{tr}(\boldsymbol{\Xi}) = P}} \ \sum_k \alpha_k \log_2 \left| \boldsymbol{\Omega} + \sum_{\hat{k} \leq k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}} \right| - w_1 \log_2 |\boldsymbol{\Omega}|, \tag{3.43}$$

with L linear conic constraints. As shown in [74], the inner maximization of problem (3.43) can be solved efficiently with the iterative scaled gradient algorithm with an orthogonal projection onto the constraint set.

Here, it is proposed to solve the complete minimax problem with a joint or alternating projected gradient algorithm [79, 84]. The algorithm alternates between finding the optimal

transmit covariances for a fixed noise covariance and vice versa as depicted in Algorithm 3.1. In lines 6 to 14, the transmit covariance matrices are updated with steepest ascend steps until convergence and, in lines 15 to 23, the noise covariance matrices are updated with steepest descend steps until convergence is reached. The steepest ascend and descend steps for the transmit and noise covariance matrices are discussed in the following sections for different types of constraints. The algorithm switches between the two inner loops until a convergence of the outer loop is reached. This kind of alternating structure, with different realizations of the inner loops, corresponds to existing algorithms in the literature for optimizations with multiple linear constraints [67, 70].

Algorithm 3.1 Alternating gradient-projection algorithm—switch at convergence

Require: Accuracy ε, constraints $(C_l,\ \mathcal{Z}_l)$, $\forall l$
1: $\boldsymbol{\Xi}_k \leftarrow \frac{1}{MK}\mathbf{I}$, $\forall k$ ▷ initialize transmit covariances
2: $\boldsymbol{\Omega}_{1:L} \leftarrow \left(\frac{P}{L\,\mathrm{tr}(\boldsymbol{C}_{1:L})}\mathbf{I}\right)_{\perp}$ ▷ initial noise covariances (3.51)
3: $d_\mathrm{t} \leftarrow 1$, $d_\mathrm{n} \leftarrow 1$, $i \leftarrow 0$ ▷ initialize step-size
4: $\Phi_\mathrm{n}^{(0)} \leftarrow \Phi(\boldsymbol{\Xi}_{1:K},\ \sum_l \boldsymbol{\Omega}_l)$ ▷ initial objective (3.28)
5: **repeat**
6: $\quad\Phi_\mathrm{t}^{(0)} \leftarrow \Phi_\mathrm{n}^{(i)}$, $\boldsymbol{\Xi}_k^{(0)} \leftarrow \boldsymbol{\Xi}_k$, $\forall k$, $i \leftarrow 0$
7: $\quad$**repeat**
8: $\quad\quad\boldsymbol{\Xi}_{1:K}^{(i+1)} \leftarrow \mathrm{TransCovStep}\left(\boldsymbol{\Xi}_{1:K}^{(i)}\right)$ ▷ Algorithm 3.3
9: $\quad\quad\Phi_\mathrm{t}^{(i+1)} \leftarrow \Phi(\boldsymbol{\Xi}_{1:K}^{(i+1)},\ \sum_l \boldsymbol{\Omega}_l)$ ▷ evaluate objective (3.28)
10: $\quad\quad\mathrm{cost\_increase} \leftarrow \Phi_\mathrm{t}^{(i+1)} - \Phi_\mathrm{t}^{(i)}$
11: $\quad\quad i \leftarrow i + 1$ ▷ inner iteration counter
12: $\quad$**until** cost_increase $\leq \varepsilon$ ▷ break if convergence is reached
13: $\quad\boldsymbol{\Xi}_k \leftarrow \boldsymbol{\Xi}_k^{(i)}$, $\forall k$ ▷ new transmit covariance matrices
14: $\quad\mathrm{total\_cost\_increase} \leftarrow \Phi_\mathrm{t}^{(i)} - \Phi_\mathrm{t}^{(0)}$
15: $\quad\Phi_\mathrm{n}^{(0)} \leftarrow \Phi_\mathrm{t}^{(i)}$, $\boldsymbol{\Omega}_l^{(0)} \leftarrow \boldsymbol{\Omega}_l$, $\forall l$, $i \leftarrow 0$
16: $\quad$**repeat**
17: $\quad\quad\boldsymbol{\Omega}_{1:L}^{(i+1)} \leftarrow \mathrm{NoiseCovStep}\left(\boldsymbol{\Omega}_{1:L}^{(i)}\right)$ ▷ Algorithm 3.4
18: $\quad\quad\Phi_\mathrm{n}^{(i+1)} \leftarrow \Phi(\boldsymbol{\Xi}_{1:K},\ \sum_l \boldsymbol{\Omega}_l^{(i+1)})$ ▷ evaluate objective (3.28)
19: $\quad\quad\mathrm{cost\_decrease} \leftarrow \Phi_\mathrm{n}^{(i+1)} - \Phi_\mathrm{n}^{(i)}$
20: $\quad\quad i \leftarrow i + 1$ ▷ inner iteration counter
21: $\quad$**until** cost_decrease $\geq -\varepsilon$ ▷ break if convergence is reached
22: $\quad\boldsymbol{\Omega}_l \leftarrow \boldsymbol{\Omega}_l^{(i)}$, $\forall l$ ▷ new noise covariance matrices
23: $\quad\mathrm{total\_cost\_decrease} \leftarrow \Phi_\mathrm{n}^{(i)} - \Phi_\mathrm{n}^{(0)}$
24: **until** total_cost_increase $\leq \varepsilon$ **and** total_cost_decrease $\geq -\varepsilon$
25: $\boldsymbol{Q}_{1:K} \leftarrow \mathrm{MAC2BC}(\boldsymbol{\Xi}_{1:K},\ \mu,\ \sum_l \boldsymbol{\Omega}_l)$ ▷ downlink recovery (3.42)

It is also possible and much more efficient to alternate after each projected gradient step between the transmit and noise covariance matrix updates as it is shown in Algorithm 3.2. In each iteration of the algorithm, the transmit and noise covariance matrices are updated with a steepest ascend/descend step. Line 6 describe the update of the transmit covariance matrices according to Section 3.5.1. The description of the noise covariance matrix update

in line 9 can be found in Section 3.5.2. Any other fixed number of steps into one direction or the other before switching between the two tasks can also be implemented.

Algorithm 3.2 Joint gradient-projection algorithm—switch in every iteration

Require: Accuracy ε, constraints $(C_l, \mathcal{Z}_l)$, $\forall l$

1: $\boldsymbol{\Xi}_k^{(0)} \leftarrow \frac{P}{MK}\mathbf{I}$, $\forall k$ ▷ initialize transmit covariances

2: $\boldsymbol{\Omega}_{1:L}^{(0)} \leftarrow \left(\frac{P}{L\,\mathrm{tr}(\boldsymbol{C}_{1:L})}\mathbf{I}\right)_{\perp}$ ▷ initial noise covariances (3.51)

3: $d_\mathrm{t} \leftarrow 1$, $d_\mathrm{n} \leftarrow 1$, $i \leftarrow 0$ ▷ initialize step-size

4: $\Phi_\mathrm{n}^{(0)} \leftarrow \Phi(\boldsymbol{\Xi}_{1:K}^{(0)},\ \sum_l \boldsymbol{\Omega}_l^{(0)})$ ▷ initial objective (3.28)

5: **repeat**

6: $\quad\boldsymbol{\Xi}_{1:K}^{(i+1)} \leftarrow \mathrm{TransCovStep}\left(\boldsymbol{\Xi}_{1:K}^{(i)}\right)$ ▷ Algorithm 3.3

7: $\quad\Phi_\mathrm{t}^{(i+1)} \leftarrow \Phi(\boldsymbol{\Xi}_{1:K}^{(i+1)},\ \sum_l \boldsymbol{\Omega}_l^{(i)})$ ▷ evaluate objective (3.28)

8: $\quad\mathrm{cost\_increase} \leftarrow \Phi_\mathrm{t}^{(i+1)} - \Phi_\mathrm{n}^{(i)}$

9: $\quad\boldsymbol{\Omega}_{1:L}^{(i+1)} \leftarrow \mathrm{NoiseCovStep}\left(\boldsymbol{\Omega}_{1:L}^{(i)}\right)$ ▷ Algorithm 3.4

10: $\quad\Phi_\mathrm{n}^{(i+1)} \leftarrow \Phi(\boldsymbol{\Xi}_{1:K}^{(i+1)},\ \sum_l \boldsymbol{\Omega}_l^{(i+1)})$ ▷ evaluate objective (3.28)

11: $\quad\mathrm{cost\_decrease} \leftarrow \Phi_\mathrm{n}^{(i+1)} - \Phi_\mathrm{t}^{(i+1)}$

12: $\quad i \leftarrow i + 1$ ▷ iteration counter

13: **until** cost_increase $\leq \varepsilon$ **and** cost_decrease $\geq -\varepsilon$

14: $\boldsymbol{Q}_{1:K} \leftarrow \mathrm{MAC2BC}(\boldsymbol{\Xi}_{1:K},\ \mu,\ \sum_l \boldsymbol{\Omega}_l)$ ▷ downlink recovery (3.42)

In line 1 of Algorithm 3.1 and 3.2, the transmit covariance matrices are initialized simply and feasible with scaled identity matrices, which fulfill the joint trace limit. In line 2, the noise covariance matrices are initialized. To find a feasible initialization, all $\boldsymbol{\Omega}_l$ have to be selected from the set $\mathcal{Z}_l^\perp$, respectively. Jointly, they have to fulfill the trace constraint. Such a selection is sketched out with an orthogonal projection $(\bullet)_\perp$ of scaled identity matrices. The steps in (3.51) can be followed in general, but have to be adapted for the constraints at hand. In the last line of the two algorithms, respectively, the primal optimization variables, the downlink transmit covariances, are calculated based on (3.42).

3.5.1 Projected Gradient Update of the Transmit Covariance Matrices

Following [28, 74], the gradient with respect to the individual transmit covariance matrices is

$$\begin{aligned}\hat{\boldsymbol{G}}_{\mathrm{t},k} &= \frac{\partial \Phi\left(\boldsymbol{\Xi}_{1:K},\ \sum_l \boldsymbol{\Omega}_l\right)}{\partial \boldsymbol{\Xi}_k^{\mathrm{T}}} \\ &= \frac{1}{\ln(2)} \sum_{\hat{k}\geq k} \alpha_{\hat{k}} \boldsymbol{H}_k^{\mathrm{H}} \left(\boldsymbol{\Omega} + \sum_{\check{k}\leq\hat{k}} \boldsymbol{H}_{\check{k}} \boldsymbol{\Xi}_{\check{k}} \boldsymbol{H}_{\check{k}}^{\mathrm{H}}\right)^{-1} \boldsymbol{H}_k, \end{aligned} \tag{3.44}$$

which is always positive-semidefinite. In each iteration (i) of the gradient update algorithm, the gradient $\hat{\boldsymbol{G}}_{\mathrm{t},k}^{(i)}$, $\forall k$, is evaluated with the transmit covariance matrices $\boldsymbol{\Xi}_k^{(i)}$, $\forall k$, and noise covariance matrices $\boldsymbol{\Omega}_l^{(i)}$, $\forall l$, of this step. In general, the steepest ascend needs a projection

of the gradient onto the tangent cone $(\bullet)_{\perp\!\!\!\perp}$:

$$\boldsymbol{G}_{\mathrm{t},1:K}^{(i)} = \left(\hat{\boldsymbol{G}}_{\mathrm{t},1:K}^{(i)}\right)_{\perp\!\!\!\perp}. \tag{3.45}$$

A description of the orthogonal projection of the gradient onto the tangent cone of the constraint set $(\bullet)_{\perp\!\!\!\perp}$ can be found in [28, Section 5.3.2.2]. Hunger shows that this projection indeed yields the steepest ascent, but has a negligible effect on the convergence speed for the optimization of the covariances. It can simply be replaced with the identity mapping. With the update matrices $\boldsymbol{G}_{\mathrm{t},k}^{(i)}$, $\forall k$, the unconstrained updates of the covariance matrices $\hat{\boldsymbol{\Xi}}_k^{(i+1)}$, $\forall k$, for the next iteration $(i+1)$ are calculated as

$$\hat{\boldsymbol{\Xi}}_k^{(i+1)} = \boldsymbol{\Xi}_k^{(i)} + s_{\mathrm{t}}^{(i)} \boldsymbol{G}_{\mathrm{t},k}^{(i)}, \tag{3.46}$$

which will be positive-semidefinite, if the tangent cone projection is replaced with the identity mapping or a simple scaling. The joint step-size $s_{\mathrm{t}}^{(i)}$ for all transmit covariance matrix updates is discussed in more detail in Section 3.5.3. The transmit covariance matrices $\boldsymbol{\Xi}_k^{(i)}$, $\forall k$, in step i are in the constraint set. But, the unconstrained updates of the transmit covariance matrices $\hat{\boldsymbol{\Xi}}_k^{(i+1)}$, $\forall k$ might not be in the constraint set. To find the steepest ascend update, which gives the transmit covariance matrices in the next iteration $\boldsymbol{\Xi}_k^{(i+1)}$, $\forall k$, the joint projection of all unconstrained updates onto the constraints set has to be done:

$$\boldsymbol{\Xi}_{1:K}^{(i+1)} = \left(\hat{\boldsymbol{\Xi}}_{1:K}^{(i+1)}\right)_{\perp}, \tag{3.47}$$

which is discussed in more detail in Section 3.5.4. The algorithm for a single gradient update step of the transmit covariance matrices can be seen for the on demand step-size in Algorithm 3.3.

Algorithm 3.3 TransCovStep

Projected gradient step of the transmit covariance matrices with diminishing and on demand step-size reduction

Require: previous or initial $\boldsymbol{\Xi}_k$, $\forall k$, $\boldsymbol{\Omega}_l$, $\forall l$, cost_old, inverse step-size d_{t}, accuracy ε

$\boldsymbol{G}_{\mathrm{t},k} \leftarrow \left(\partial\Phi\left(\boldsymbol{\Xi}_{1:K}, \sum_l \boldsymbol{\Omega}_l\right)\right)\left(\partial\boldsymbol{\Xi}_k\right)^{-1}$, $\forall k$ ▷ gradient computation (3.44)

$p_{\mathrm{t}} \leftarrow P\left(\sum_k \mathrm{tr}(\boldsymbol{G}_{\mathrm{t},k})\right)^{-1}$ ▷ preconditioning (3.54)

repeat

  $s_{\mathrm{t}} \leftarrow p_{\mathrm{t}} d_{\mathrm{t}}^{-1} i^{-1/2}$ ▷ set step-size (3.53)

  $\hat{\boldsymbol{\Xi}}_k \leftarrow \boldsymbol{\Xi}_k + s_{\mathrm{t}} \boldsymbol{G}_{\mathrm{t},k}$, $\forall k$ ▷ unconstr. update (3.46)

  $\tilde{\boldsymbol{\Xi}}_{1:K} \leftarrow \left(\hat{\boldsymbol{\Xi}}_{1:K}\right)_{\perp}$ ▷ joint projection (3.47)

  cost_new $\leftarrow \Phi(\tilde{\boldsymbol{\Xi}}_{1:K}, \sum_l \boldsymbol{\Omega}_l)$ ▷ evaluate objective (3.28)

  cost_increase $\leftarrow$ cost_new $-$ cost_old

  if cost_increase $< -\epsilon$ **then**

    $d_{\mathrm{t}} \leftarrow d_{\mathrm{t}} + 1$ ▷ decrease step-size

  end if

until cost_increase $\geq -\varepsilon$

$\boldsymbol{\Xi}_k \leftarrow \tilde{\boldsymbol{\Xi}}_k$, $\forall k$ ▷ new covariances

3.5.2 Projected Gradient Update of the Uplink Noise Covariance Matrices

The gradient update of the noise covariance matrix follows the same steps as the update of the transmit covariance matrices except that it is a minimization instead of a maximization. The gradient $\hat{\boldsymbol{G}}_{\mathrm{n}}$ of $\Phi\left(\boldsymbol{\Xi}_{1:K},\ \sum_l \boldsymbol{\Omega}_l\right)$ with respect to the noise covariance matrix, or each of its summands, is

$$\begin{aligned}\hat{\boldsymbol{G}}_{\mathrm{n},l} &= \frac{\partial \Phi\left(\boldsymbol{\Xi}_{1:K},\ \sum_l \boldsymbol{\Omega}_l\right)}{\partial \boldsymbol{\Omega}_l^{\mathrm{T}}} \\ &= \sum_k \frac{\alpha_k}{\ln(2)} \left(\boldsymbol{\Omega} + \sum_{\hat{k} \leq k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}} \right)^{-1} - \frac{w_1}{\ln(2)} \boldsymbol{\Omega}^{-1} = \hat{\boldsymbol{G}}_{\mathrm{n}},\end{aligned} \tag{3.48}$$

which is always negative-semidefinite. The gradient is the same for all summands of the noise covariance matrix, which results in a slow convergence behavior for multiple constraints. The gradient $\hat{\boldsymbol{G}}_{\mathrm{n}}^{(i)}$ in iteration (i) is evaluated with the transmit covariance matrices $\boldsymbol{\Xi}_k^{(i)},\ \forall k$, and noise covariance matrices $\boldsymbol{\Omega}_l^{(i)},\ \forall l$, of this step. The projection of the gradient onto the tangent cone $(\bullet)_{\perp}$ reads as

$$\boldsymbol{G}_{\mathrm{n},1:L}^{(i)} = \left(\hat{\boldsymbol{G}}_{\mathrm{n}}\right)_{\perp} \tag{3.49}$$

and is discussed for different types of constraints in Section 3.5.5. The projection onto the tangent cone is optional, but can drastically increase the convergence speed. The unconstrained updates of the noise covariance matrix summands are

$$\hat{\boldsymbol{\Omega}}_l^{(i+1)} = \boldsymbol{\Omega}_l^{(i)} - s_{\mathrm{n}}^{(i)} \boldsymbol{G}_{\mathrm{n},l}^{(i)}. \tag{3.50}$$

The joint step-size for the noise covariance matrix updates $s_{\mathrm{n}}^{(i)}$ follows the same rules as the step-size for the transmit covariance updates as described in Section 3.5.3. The gradient is subtracted from the previous value, because it is a minimization instead of the maximization in equation (3.46). The steepest descend update of the noise covariance matrix summands,

$$\boldsymbol{\Omega}_{1:L}^{(i+1)} = \left(\hat{\boldsymbol{\Omega}}_{1:L}^{(i+1)}\right)_{\perp}, \tag{3.51}$$

is found with a joint orthogonal projection step, which is done with the generalized water-spilling presented in Section 3.5.6 for different types of constraints. The algorithm for a single gradient update step of the noise covariance matrices can be seen for the on demand step-size in Algorithm 3.4.

3.5.3 Step-size

A detailed discussion of different step-size strategies for the gradient-projection algorithm can be found in [28, Section 5.1.3]. Goldstein and Levitin showed, that the projected-gradient algorithm converges with a constant step-size to a stationary point under certain conditions [85, 86]. These conditions are fulfilled for the optimization of the transmit covariance matrices and, due to the similar structure, also for the optimization of the noise covariance matrices [28, Section 5.1.3.1]. The joint update of all transmit covariance matrices is a concave problem for fixed noise covariance matrices and the joint update of all noise covariance

Algorithm 3.4 NoiseCovStep
Projected gradient step of the noise covariance matrices with diminishing and on demand step-size reduction

Require: previous or initial $\boldsymbol{\Xi}_k$, $\forall k$, $\boldsymbol{\Omega}_l$, $\forall l$, cost_old, inverse step-size d_n, accuracy ϵ
$\hat{\boldsymbol{G}}_\mathrm{n} \leftarrow \left(\partial\Phi\left(\boldsymbol{\Xi}_{1:K},\ \sum_l \boldsymbol{\Omega}_l\right)\right)\left(\partial\boldsymbol{\Omega}\right)^{-1}$ ▷ gradient computation (3.48)
$p_\mathrm{n} \leftarrow P\Big(\sum_l \mathrm{tr}(\boldsymbol{C}_l\hat{\boldsymbol{G}}_\mathrm{n})\Big)^{-1}$ ▷ preconditioning (3.54)
$\boldsymbol{G}_{\mathrm{n},1:L} \leftarrow \left(\hat{\boldsymbol{G}}_\mathrm{n}\right)_\perp$ ▷ tangent cone projection (3.49)
repeat
  $s_\mathrm{n} \leftarrow p_\mathrm{n} d_\mathrm{n}^{-1} i^{-1/2}$ ▷ set step-size (3.53)
  $\hat{\boldsymbol{\Omega}}_l \leftarrow \boldsymbol{\Omega}_l - s_\mathrm{n}\boldsymbol{G}_{\mathrm{n},l},\ \forall l$ ▷ unconstr. update (3.50)
  $\tilde{\boldsymbol{\Omega}}_{1:L} \leftarrow \left(\hat{\boldsymbol{\Omega}}_{1:L}\right)_\perp$ ▷ joint projection (3.51)
  cost_new $\leftarrow \Phi(\boldsymbol{\Xi}_{1:K},\ \sum_l \tilde{\boldsymbol{\Omega}}_l)$ ▷ evaluate objective (3.28)
  cost_decrease $\leftarrow$ cost_new $-$ cost_old
  if cost_decrease $> \epsilon$ **then**
    $d_\mathrm{n} \leftarrow d_\mathrm{n} + 1$ ▷ decrease step-size
  end if
until cost_decrease $\leq \epsilon$
$\boldsymbol{\Omega}_l \leftarrow \tilde{\boldsymbol{\Omega}}_l,\ \forall l$ ▷ new covariances

matrices is a convex problem with fixed transmit covariance matrices. For given gradient matrices, the step-size, which gives the optimal result in the iteration at hand, can be found with a line search at high computational cost [87]. The line search can be approximated with the Armijo rule, which sets the normalized step-size into relation with the increase or decrease of the utility [88]. A simple but slow approach is the diminishing step-size [89]:

$$\lim_{i\to\infty} s^{(i)} = 0, \quad \text{and} \quad \sum_{i=0}^{\infty} s^{(i)} = \infty. \tag{3.52}$$

The convergence of the projected-gradient algorithm for saddle point problems was investigated for constant step-sizes in [84]. A more efficient implementation with diminishing step-sizes and on-demand step-size reduction was evaluated in [90] and is used in this work.

The utilized step-size consists of three factors for the transmit and noise covariance matrices, respectively:

$$s_\mathrm{t}^{(i)} = p_\mathrm{t}^{(i)} \cdot \frac{1}{d_\mathrm{t}} \cdot \frac{1}{\sqrt{i}}, \qquad s_\mathrm{n}^{(i)} = p_\mathrm{n}^{(i)} \cdot \frac{1}{d_\mathrm{n}} \cdot \frac{1}{\sqrt{i}}. \tag{3.53}$$

The first factor is the preconditioning scalar,

$$p_\mathrm{t}^{(i)} = \frac{P}{\sum_k \mathrm{tr}\left(\hat{\boldsymbol{G}}_{\mathrm{t},k}^{(i)}\right)}, \qquad p_\mathrm{n}^{(i)} = \frac{P}{\sum_l \mathrm{tr}\left(\boldsymbol{C}_l\hat{\boldsymbol{G}}_\mathrm{n}^{(i)}\right)}, \tag{3.54}$$

which normalizes the sum of the gradient traces to the chosen P and makes the gradient almost independent of the selection of P.

The second factor assures, that the utility is never changed for the worse. d_t and d_n are initialized with one. The newly calculated transmit and noise covariance matrices with the current step size are saved into temporary variables $\tilde{\boldsymbol{\Xi}}_k^{(i+1)}$, $\forall k$ and $\tilde{\boldsymbol{\Omega}}_l^{(i+1)}$, $\forall l$, respectively. These will only be used as the covariance matrices for the next step, if the utility after the update,

$$\mathrm{cost\_new}_\mathrm{t}^{(i+1)} = \Phi\left(\tilde{\boldsymbol{\Xi}}_{1:K}^{(i+1)},\ \sum_l \boldsymbol{\Omega}_l^{(i)}\right), \tag{3.55}$$

$$\mathrm{cost\_new}_\mathrm{n}^{(i+1)} = \Phi\left(\boldsymbol{\Xi}_{1:K}^{(i+1)},\ \sum_l \tilde{\boldsymbol{\Omega}}_l^{(i+1)}\right), \tag{3.56}$$

is better than the utility before the update,

$$\mathrm{cost\_old}_\mathrm{t}^{(i+1)} = \mathrm{cost\_new}_\mathrm{n}^{(i)} = \Phi\left(\boldsymbol{\Xi}_{1:K}^{(i)},\ \sum_l \boldsymbol{\Omega}_l^{(i)}\right), \tag{3.57}$$

$$\mathrm{cost\_old}_\mathrm{n}^{(i+1)} = \mathrm{cost\_new}_\mathrm{t}^{(i+1)} = \Phi\left(\boldsymbol{\Xi}_{1:K}^{(i+1)},\ \sum_l \boldsymbol{\Omega}_l^{(i)}\right), \tag{3.58}$$

respectively. The utility has to increase for the transmit covariance matrices and decrease for the noise covariance matrices,

$$\mathrm{cost\_increase}_\mathrm{t}^{(i+1)} = \mathrm{cost\_new}_\mathrm{t}^{(i+1)} - \mathrm{cost\_old}_\mathrm{t}^{(i+1)} > 0, \tag{3.59}$$

$$\mathrm{cost\_decrease}_\mathrm{n}^{(i+1)} = \mathrm{cost\_new}_\mathrm{n}^{(i+1)} - \mathrm{cost\_old}_\mathrm{n}^{(i+1)} < 0. \tag{3.60}$$

If the utility with the temporary covariance matrices is worse than with the covariance matrices from the previous step, d_t or d_n will be increased by one, to decrease d_t^{-1} or d_n^{-1} and the temporary transmit or noise covariance matrices will be calculated again. Based on the newly calculated temporary covariance matrices, equation (3.59) or (3.60) is checked again. This is repeated until the utility improves. This on-demand step-size reduction has a very fast convergence behavior, but convergence cannot be assured in general. It can happen, that the algorithm oscillates in the vicinity of the optimum. To guarantee convergence and prevent the algorithm from oscillating, the third factor of the step-size is the diminishing step-size $i^{-1/2}$. The step-size selection can be seen in Algorithm 3.3 for the transmit covariance matrices and in Algorithm 3.4 for the noise covariance matrices.

The original algorithm from Arrow et al. makes a joint step with a joint step-size for the transmit and noise covariance matrices based on the covariance matrices from the previous iteration (i) [84]. Here, the algorithm alternates between an update of the transmit covariance matrices and an update of the noise covariance matrices utilizing the updated covariance matrices from the previous step with independent step-sizes, respectively. More precisely, the old cost for the noise covariance matrix update in equation (3.58) and the update of the noise covariance matrices are calculated with the already updated transmit covariance matrices. In the simulations, this has shown a much faster convergence behavior.

3.5.4 Orthogonal Projection of the Transmit Covariance Matrices

The orthogonal projection $(\bullet)_\perp$ onto the constraint set is done with the water-spilling algorithm from [74]. A summary of the steps is given in this section to show the link to

the orthogonal projection of the noise covariance matrices in Section 3.5.6. The goal of the orthogonal projection onto the constraint set is to minimize the Euclidean distance between all the unconstrained updates of the transmit covariance matrices and the constraint set simultaneously. The corresponding optimization is

$$\boldsymbol{\Xi}_{1:K} = \underset{\substack{\boldsymbol{\Xi}_k \succeq \mathbf{0}\ \forall k \\ \sum_k \mathrm{tr}(\boldsymbol{\Xi}_k)=P}}{\operatorname{argmin}} \quad \sum_k \left\| \boldsymbol{\Xi}_k - \hat{\boldsymbol{\Xi}}_k \right\|_{\mathrm{F}}^2, \tag{3.61}$$

where $\|\bullet\|_{\mathrm{F}}$ is the Frobenius norm. The iteration index is dropped to increase the readability.

The Lagrangian function can be constructed with the dual variables $\boldsymbol{S}_k \in \mathbb{C}^{M\times M}$, $\forall k$, for the positive semidefiniteness constraints and μ for the joint trace constraint:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Xi}_{1:K},\ \boldsymbol{S}_{1:K}, \mu) = &\sum_k \mathrm{tr}\left(\left(\boldsymbol{\Xi}_k - \hat{\boldsymbol{\Xi}}_k\right)\left(\boldsymbol{\Xi}_k - \hat{\boldsymbol{\Xi}}_k\right)^{\mathrm{H}}\right) \\ &- \sum_k \mathrm{tr}\left(\boldsymbol{\Xi}_k \boldsymbol{S}_k\right) + \mu\left(\sum_k \mathrm{tr}\left(\boldsymbol{\Xi}_k\right) - P\right).\end{aligned} \tag{3.62}$$

The dual variables $\boldsymbol{S}_k$ have to be positive-semidefinite and with an orthogonal eigenbasis to $\boldsymbol{\Xi}_k$, $\forall k$, respectively. The eigenvalues of $\boldsymbol{S}_k$ are the Lagrangian multipliers of the non-negative constraints for the eigenvalues of $\boldsymbol{\Xi}_k$. Setting the derivative of the Lagrangian function with respect to $\boldsymbol{\Xi}_k^{\mathrm{T}}$ to zero yields

$$\boldsymbol{\Xi}_k = \hat{\boldsymbol{\Xi}}_k + \boldsymbol{S}_k - \mu\mathbf{I},\ \forall k, \tag{3.63}$$

or equivalently

$$\boldsymbol{\Xi}_k = \left(\hat{\boldsymbol{\Xi}} - \mu\mathbf{I}\right)_+,\ \forall k, \tag{3.64}$$

where the influence of $\boldsymbol{S}_k$ is replaced by the operation $(\bullet)_+$, which sets all negative eigenvalues to zero. The Lagrangian multiplier μ has the same value for all k. The left and the right side of equation (3.64) have to have the same eigenbasis. From the eigen-decomposition $\hat{\boldsymbol{\Xi}}_k = \hat{\boldsymbol{U}}_k \hat{\boldsymbol{\Lambda}}_k \hat{\boldsymbol{U}}_k^{\mathrm{H}}$ with eigenbasis $\hat{\boldsymbol{U}}_k$ and diagonal eigenvalue matrix $\hat{\boldsymbol{\Lambda}}_k$ follows $\boldsymbol{\Xi}_k = \hat{\boldsymbol{U}}_k \boldsymbol{\Lambda}_k \hat{\boldsymbol{U}}_k^{\mathrm{H}}$ with diagonal eigenvalue matrix $\boldsymbol{\Lambda}_k$. Therefore, equation (3.64) can be diagonalized as

$$\boldsymbol{\Lambda}_k = \left(\hat{\boldsymbol{\Lambda}}_k - \mu\mathbf{I}\right)_+, \tag{3.65}$$

$$\lambda_{k,m} = \left(\hat{\lambda}_{k,m} - \mu\right)_+. \tag{3.66}$$

Here, $\lambda_{k,m}$, $\forall m$, and $\hat{\lambda}_{k,m}$, $\forall m$, are the eigenvalues of $\boldsymbol{\Xi}_k$ and $\hat{\boldsymbol{\Xi}}_k$, respectively. The water level μ is found by plugging equation (3.66) into the trace constraint:

$$\sum_k \mathrm{tr}\left(\boldsymbol{\Xi}_k\right) = \sum_k \mathrm{tr}\left(\boldsymbol{\Lambda}_k\right) = \sum_k \sum_m \left(\hat{\lambda}_{k,m} - \mu\right)_+ = P. \tag{3.67}$$

The joint set $\hat{\mathcal{M}}$ of the indices of the positive eigenvalues of all $\boldsymbol{\Xi}_k$ has to be identified to determine μ. It can be found in an iterative search. The set is initializing with the set

of indices of all eigenvalues and, in each iteration, the index corresponding to the smallest eigenvalue $\hat{\lambda}_{k,m}$ is dropped, one at a time, until the termination criterion,

$$\min_{(k,m)\in\hat{\mathcal{M}}} \hat{\lambda}_{k,m} - \mu > 0, \tag{3.68}$$

holds. With the cardinality $\hat{M}$ of $\hat{\mathcal{M}}$, the water level can be computed as

$$\mu = \frac{\sum_{(k,m)\in\hat{\mathcal{M}}} \hat{\lambda}_{k,m} - P}{\hat{M}}. \tag{3.69}$$

Finally, the orthogonal projected transmit covariance matrices after the gradient update read as

$$\boldsymbol{\Xi}_k = \hat{\boldsymbol{U}}_k \Big(\hat{\boldsymbol{\Lambda}}_k - \mu \mathbf{I} \Big)_+ \hat{\boldsymbol{U}}_k^{\mathrm{H}}. \tag{3.70}$$

As discussed before for the unconstrained update in equation (3.46), $\hat{\boldsymbol{\Xi}}_k$, $\forall k$, will be the sum of the transmit covariance matrix from the step before plus a scaled version of the positive-semidefinite gradient, if the tangent cone projection is neglected. The transmit covariances from the step before have to fulfill the joint trace constraint. Adding positive-semidefinite matrices can only increase the sum trace and therefore $\sum_k \mathrm{tr}(\hat{\boldsymbol{\Xi}}_k) \geq P$ will hold. From this follows, that the water spilling level μ will be non-negative. With the projection onto the tangent cone, the gradient matrices can be any Hermitian matrix. Eigenvalues of $\hat{\boldsymbol{\Xi}}_k$ might turn negative and $\sum_k \mathrm{tr}(\hat{\boldsymbol{\Xi}}_k) \geq P$ cannot be assured anymore. This could lead to a negative water level μ, which can still be found with the described algorithm. In contrast to [28, Section 5.3.2.1], here the trace constraint has to be fulfilled with equality.

3.5.5 Tangent Cone Projection of the Noise Covariance Matrix Gradient

The steps in [28, Section 5.3.2.2] are followed to find the tangent cone projection. The gradient is the same for all noise covariance matrices, although the constraints can be very different. The tangent cone projection of the gradient can increase the convergence speed dramatically. With the tangent cone projection, the gradient $\hat{\boldsymbol{G}}_{\mathrm{n}}$ is projected for each noise covariance matrix onto a feasible update direction $\boldsymbol{G}_{\mathrm{n},l}$:

$$\exists \bar{\epsilon} > 0 : \quad \forall \epsilon \in [0, \bar{\epsilon}] : \quad \left\{ \boldsymbol{\Omega}_1^{(i)} - \epsilon \boldsymbol{G}_{n,1},\ \ldots\ , \boldsymbol{\Omega}_L^{(i)} - \epsilon \boldsymbol{G}_{n,L} \right\} \in \mathcal{C}_{\mathrm{n}}, \tag{3.71}$$

where

$$\mathcal{C}_{\mathrm{n}} = \left\{ \boldsymbol{\Omega}_{1:L} :\ \mathbf{0} \preceq \boldsymbol{\Omega}_l \in \mathcal{Z}_l^{\perp},\ \forall l,\ \sum_l \mathrm{tr}(\boldsymbol{Q}_l \boldsymbol{C}_l) = P \right\} \tag{3.72}$$

is the constraint set for the noise covariance matrices.

For inequality constraints, the tangent cone projection will be only different from the identity mapping, if the constraint is active. This leads to a distinction of cases. The sum trace constraint is an equality constraint and always active. The positive-semidefiniteness constraints are inequality constraints and a distinction has to be made for each eigenvalue between the case with a positive and a zero eigenvalue. For the subset constraints $\boldsymbol{\Omega}_l \in \mathcal{Z}_l^{\perp}$,

$\forall l$, it depends on the subset. In [28, Section 5.3.2.3] it is shown, that the tangent cone projection makes no difference for the transmit covariance matrices. The projected gradient step of the noise covariance matrix with a scaled identity shaping constraint is an equivalent problem and, therefore, the tangent cone projection makes no difference as well for this problem. But, for linear constraints it makes a crucial difference.

3.5.5.1 Tangent Cone Projection with Multiple Linear Constraints

For linear constraints,

$$\operatorname{tr}(\boldsymbol{Q}\boldsymbol{A}_l) \leq a_l, \tag{3.73}$$

the subset $\mathcal{Z}_l^{\text{lin}}$ and corresponding orthogonal subset $\mathcal{Z}_l^{\perp,\text{lin}}$ can be found as (See Section 3.3.3)

$$\mathcal{Z}_l^{\text{lin}} = \left\{ \boldsymbol{Z}^{\text{lin}} : \ \operatorname{tr}(\boldsymbol{Z}_l^{\text{lin}}\boldsymbol{A}_l) = 0 \right\}, \tag{3.74}$$

$$\mathcal{Z}_l^{\perp,\text{lin}} = \left\{ \boldsymbol{\Omega}_l^{\text{lin}} : \boldsymbol{\Omega}_l^{\text{lin}} = \omega_l \boldsymbol{A}_l, \ \forall \omega_l \in \mathbb{R}_0^+ \right\}. \tag{3.75}$$

$\boldsymbol{\Omega}_l^{\text{lin}} = \omega_l \boldsymbol{A}_l$ has to be a scaled version of the positive-semidefinite constraint matrix $\boldsymbol{A}_l$ with the scaling variable $\omega_l \in \mathbb{R}_0^+$. From

$$\boldsymbol{\Omega}_l^{\text{lin},(i)} = \omega_l^{(i)} \boldsymbol{A}_l \quad \text{and} \quad \boldsymbol{\Omega}_l^{\text{lin},(i)} - \epsilon \boldsymbol{G}_{\text{n},l} = \omega_l \boldsymbol{A}_l \succeq \boldsymbol{0} \tag{3.76}$$

follows

$$\boldsymbol{G}_{\text{n},l} = g_l \boldsymbol{A}_l \quad \text{and} \quad \omega_l = \omega_l^{(i)} - \epsilon g_l \geq 0. \tag{3.77}$$

To lie on the tangent cone, the update direction has to be a scaled versions of the constraint matrix $\boldsymbol{A}_l$. The update scalar g_l can be positive, if $\omega_l^{(i)}$ is positive. But, it has to be non-positive, if $\omega_l^{(i)}$ is zero. All updates have to fulfill jointly the sum trace constraint,

$$\sum_l \operatorname{tr}\left(\boldsymbol{C}_l^{\text{lin}} (\boldsymbol{\Omega}_l^{\text{lin},(i)} - \epsilon \boldsymbol{G}_{\text{n},l}) \right) = \sum_l \operatorname{tr}\left(\boldsymbol{C}_l^{\text{lin}} \boldsymbol{\Omega}_l^{\text{lin},(i)} \right) - \epsilon \sum_l g_l \operatorname{tr}\left(\boldsymbol{C}_l^{\text{lin}} \boldsymbol{A}_l \right) = P. \tag{3.78}$$

The previous noise covariance matrices $\boldsymbol{\Omega}_l^{\text{lin},(i)}$ have to fulfill the sum trace constraint and, as described in Section 3.3.3, $\boldsymbol{C}_l^{\text{lin}}$ is any matrix that fulfills $a_l = \operatorname{tr}(\boldsymbol{A}_l \boldsymbol{C}_l^{\text{lin}})$. Therefore, (3.78) becomes

$$\sum_l g_l a_l = 0. \tag{3.79}$$

The projection of the gradient onto the tangent cone is done by minimizing the distance between the gradient for all summands $\boldsymbol{G}_{\text{n}}$ and the update for each summand $g_l \boldsymbol{A}_l$:

$$g_{1:L} = \underset{\substack{\sum_l g_l a_l = 0 \\ g_l \geq 0, \, \forall l \in \tilde{\mathcal{L}}}}{\operatorname{argmin}} \ \sum_l \left\| g_l \boldsymbol{A}_l - \boldsymbol{G}_{\text{n}} \right\|_{\text{F}}^2, \tag{3.80}$$

where $\tilde{\mathcal{L}} = \left\{ l \in \{1, \ \ldots \ , \ L\} : \ \omega_l^{(i)} = 0 \right\}$ is the set of indices belonging to the $\omega_l^{(i)}$, which are zero. Solving (3.80) yields

$$g_l = \left(\operatorname{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_{\text{n}} \right) - \nu a_l \right)_+, \ \forall l \in \tilde{\mathcal{L}} \quad \text{and} \quad \operatorname{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_{\text{n}} \right) - \nu a_l, \ \forall l \notin \tilde{\mathcal{L}}. \tag{3.81}$$

Note, setting $\boldsymbol{\Omega}_l = \omega_l \boldsymbol{A}_l$ in the objective (3.43) and taking the derivation directly with respect to ω_l also leads to

$$\begin{aligned}\hat{g}_l &= \frac{\partial \Phi\left(\boldsymbol{\Xi}_{1:K},\ \sum_l \boldsymbol{\Omega}_l\right)}{\partial \omega_l} \\ &= \operatorname{tr}\left(\boldsymbol{A}_l \sum_k \frac{\alpha_k}{\ln(2)} \left(\boldsymbol{\Omega} + \sum_{\hat{k}\leq k} \boldsymbol{H}_{\hat{k}} \boldsymbol{\Xi}_{\hat{k}} \boldsymbol{H}_{\hat{k}}^{\mathrm{H}}\right)^{-1} - \boldsymbol{A}_l \frac{w_1}{\ln(2)} \boldsymbol{\Omega}^{-1}\right) \\ &= \operatorname{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_{\mathrm{n}}\right). \end{aligned} \tag{3.82}$$

$\hat{g}_l$ is always non-positive. Therefore, ν is also always non-positive. Plugging (3.81) into the sum constraint $\sum_l g_l a_l = 0$ gives

$$\sum_{l\in\tilde{\mathcal{L}}} \left(a_l \operatorname{tr}(\boldsymbol{A}_l \boldsymbol{G}_{\mathrm{n}}) - \nu a_l^2\right)_+ + \sum_{l\notin\tilde{\mathcal{L}}} \left(a_l \operatorname{tr}(\boldsymbol{A}_l \boldsymbol{G}_{\mathrm{n}}) - \nu a_l^2\right) = 0 \tag{3.83}$$

$$\sum_{l\in\hat{\mathcal{L}}} \left(a_l \operatorname{tr}(\boldsymbol{A}_l \boldsymbol{G}_{\mathrm{n}}) - \nu a_l^2\right) = 0, \tag{3.84}$$

where $\hat{\mathcal{L}}$ is the set of all indices belonging to non-zero g_l. With this, the water level ν can be found as

$$\nu = \frac{\sum_{l\in\hat{\mathcal{L}}} a_l \operatorname{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_{\mathrm{n}}\right)}{\sum_{l\in\hat{\mathcal{L}}} a_l^2}. \tag{3.85}$$

The set $\hat{\mathcal{L}}$ can be found by initializing it with the set of all indices. One after another, the index belonging to the smallest $\hat{g}_l a_l^{-1}$ with $l \in \tilde{\mathcal{L}}$ is dropped, until either

$$\min_{l\in\tilde{\mathcal{L}}\cap\hat{\mathcal{L}}} \hat{g}_l - \nu a_l > 0 \quad \text{or} \quad \hat{\mathcal{L}} \cap \tilde{\mathcal{L}} = \emptyset. \tag{3.86}$$

Now, the gradient matrices projected onto the tangent cone can be calculated by plugging (3.85) into (3.81) and (3.77):

$$\boldsymbol{G}_{\mathrm{n},l} = \left(\operatorname{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_{\mathrm{n}}\right) - \nu a_l\right)_+ \boldsymbol{A}_l,\ \forall l \in \tilde{\mathcal{L}} \quad \text{and} \quad \boldsymbol{G}_{\mathrm{n},l} = \left(\operatorname{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_{\mathrm{n}}\right) - \nu a_l\right) \boldsymbol{A}_l,\ \forall l \notin \tilde{\mathcal{L}}. \tag{3.87}$$

See Algorithm 3.5 for a pseudo code of the described steps.

3.5.6 Orthogonal Projection of the Uplink Noise Covariance Matrices

As already described in Section 3.5.4 for the transmit covariance matrices, the orthogonal projection onto the constraint set has to minimize the Euclidean distance between all the unconstrained update steps and the constraint set simultaneously:

$$\min_{\substack{\boldsymbol{\Omega}_l \succeq \boldsymbol{0},\ \boldsymbol{\Omega}_l \in \mathcal{Z}_l^{\perp}\ \forall l \\ \sum_l \operatorname{tr}(\boldsymbol{\Omega}_l \boldsymbol{C}_l) = P}} \sum_l \left\| \boldsymbol{\Omega}_l - \hat{\boldsymbol{\Omega}}_l \right\|_{\mathrm{F}}^2, \tag{3.88}$$

but with different constraints compared to the optimization in equation (3.61). With the positive-semidefinite matrices $\boldsymbol{S}_l \in \mathbb{C}^{N\times N}$, $\forall l$, for the positive semidefiniteness constraints,

Algorithm 3.5 Tangent cone projection with multiple linear constraints

Require: gradients $\hat{g}_l = \mathrm{tr}(\boldsymbol{A}_l \boldsymbol{G}_\mathrm{n})$, $\forall l$, constraint values a_l, $\forall l$, scaling variables $\omega_l^{(i)}$, $\forall l$

$\tilde{\mathcal{L}} = \left\{ l \in \{1, \ldots, L\} : \omega_l^{(i)} = 0 \right\}$

$\hat{\mathcal{L}} = \{1, \ldots, L\}$ ▷ initialize index set

$\nu = \left(\sum_{l \in \hat{\mathcal{L}}} a_l \hat{g}_l\right) \left(\sum_{l \in \hat{\mathcal{L}}} a_l^2\right)^{-1}$ ▷ find initial water level (3.85)

while $\min_{l \in \tilde{\mathcal{L}} \cap \hat{\mathcal{L}}} \hat{g}_l - \nu a_l \leq 0 \quad$ or $\quad \hat{\mathcal{L}} \cap \tilde{\mathcal{L}} \neq \emptyset$ **do** ▷ check for (3.86)

  $\hat{\mathcal{L}} \leftarrow \hat{\mathcal{L}} \setminus \mathrm{argmin}_{l \in \tilde{\mathcal{L}} \cap \hat{\mathcal{L}}} \hat{g}_l a_l^{-1}$ ▷ drop smallest $\hat{g}_l a_l^{-1}$

  $\nu = \left(\sum_{l \in \hat{\mathcal{L}}} a_l \hat{g}_l\right) \left(\sum_{l \in \hat{\mathcal{L}}} a_l^2\right)^{-1}$ ▷ find water level (3.85)

end while

$\boldsymbol{G}_{\mathrm{n},l} = \left(\mathrm{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_\mathrm{n}\right) - \nu a_l\right)_+ \boldsymbol{A}_l$, $\forall l \in \tilde{\mathcal{L}}$

$\boldsymbol{G}_{\mathrm{n},l} = \left(\mathrm{tr}\left(\boldsymbol{A}_l \boldsymbol{G}_\mathrm{n}\right) - \nu a_l\right) \boldsymbol{A}_l$, $\forall l \notin \tilde{\mathcal{L}}$ ▷ projected gradient matrices (3.87)

$\boldsymbol{T}_l \in \mathbb{C}^{N \times N}$, $\forall l$, for the subspace constraints, and μ for the joint trace constraint, the Lagrangian function is

$$\begin{aligned} \mathcal{L} = &\sum_l \mathrm{tr}\left(\left(\boldsymbol{\Omega}_l - \hat{\boldsymbol{\Omega}}_l\right)\left(\boldsymbol{\Omega}_l - \hat{\boldsymbol{\Omega}}_l\right)^\mathrm{H}\right) \\ &- \sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l \boldsymbol{S}_l\right) - \sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l \boldsymbol{T}_l\right) + \mu \left(\sum_l \mathrm{tr}\left(\boldsymbol{\Omega}_l \boldsymbol{C}_l\right) - P\right). \end{aligned} \tag{3.89}$$

The derivative of the Lagrangian function with respect to $\boldsymbol{\Omega}_l^\mathrm{T}$ is

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\Omega}_l^\mathrm{T}} = \boldsymbol{\Omega}_l - \hat{\boldsymbol{\Omega}}_l - \boldsymbol{S}_l - \boldsymbol{T}_l + \mu \boldsymbol{C}_l. \tag{3.90}$$

Setting the derivative to zero yields

$$\boldsymbol{\Omega}_l = \hat{\boldsymbol{\Omega}}_l + \boldsymbol{S}_l + \boldsymbol{T}_l - \mu \boldsymbol{C}_l. \tag{3.91}$$

The complementary slackness conditions for the subspace constraints $\mathrm{tr}(\boldsymbol{\Omega}_l \boldsymbol{T}_l) = 0$, $\forall l$, have to hold. As $\boldsymbol{\Omega}_l$ has to be element of the set $\mathcal{Z}_l^\perp$, $\boldsymbol{T}_l$ has to be from the orthogonal subspace $\mathcal{Z}_l$, which is the subspace defined for $\boldsymbol{Z}_l$ in the downlink constraint. The projection depends on the type of constraint. In the following, the orthogonal projection for a single shaping constraint with scaled identity matrix and arbitrary positive-definite matrix, multiple linear constraints, and the combination of linear and shaping constraints is discussed.

The noise covariance matrices from the previous iteration fulfill the trace constraint. Without the tangent cone projection, the unconstrained updates $\hat{\boldsymbol{\Omega}}_l$, $\forall l$, are calculated by subtracting the negative-semidefinite gradient matrices. Therefore, the sum trace can only increase leading to a positive μ. If the tangent cone projection is used, the water level μ might turn negative.

3.5.6.1 Single Scaled Identity Shaping Constraint

For a single shaping constraint, $\mathcal{Z}^\mathrm{shape}$ contains only the all zero matrix. Therefore, $\boldsymbol{T}^\mathrm{shape}$ is zero and equation (3.91) becomes

$$\boldsymbol{\Omega}^\mathrm{shape} = \left(\hat{\boldsymbol{\Omega}}^\mathrm{shape} - \mu \boldsymbol{C}^\mathrm{shape}\right)_+, \tag{3.92}$$

where the influence of $\boldsymbol{S}^{\text{shape}}$ is replaced by the operation $(\bullet)_+$, which sets all negative eigenvalues to zero. In equation (3.92), the water level μ and the eigenvectors of $\boldsymbol{\Omega}^{\text{shape}}$ depend on each other and can only be found iteratively for arbitrary $\boldsymbol{C}^{\text{shape}}$.

If $\boldsymbol{C}^{\text{shape}} = \boldsymbol{C}^{\text{id}} = c\mathbf{I}$ is a scaled identity matrix,

$$\boldsymbol{\Omega}^{\text{id}} = \left(\hat{\boldsymbol{\Omega}}^{\text{id}} - \mu c\mathbf{I}\right)_+, \tag{3.93}$$

$\boldsymbol{\Omega}^{\text{id}}$ and $\hat{\boldsymbol{\Omega}}^{\text{id}}$ have to have the same eigenbasis $\hat{\boldsymbol{U}}^{\text{id}}$. Therefore, equation (3.93) can be diagonalized:

$$\boldsymbol{\Lambda}^{\text{id}} = \left(\hat{\boldsymbol{\Lambda}}^{\text{id}} - \mu c\mathbf{I}\right)_+, \tag{3.94}$$

where $\boldsymbol{\Lambda}^{\text{id}}$ and $\hat{\boldsymbol{\Lambda}}^{\text{id}}$ are diagonal matrices with the eigenvalues λ_n^{id}, $\forall n$, and $\hat{\lambda}_n^{\text{id}}$, $\forall n$ of $\boldsymbol{\Omega}^{\text{id}}$ and $\hat{\boldsymbol{\Omega}}^{\text{id}}$ as their diagonal elements, respectively. The water level μ is found by plugging equation (3.94) into the trace constraint:

$$c\,\mathrm{tr}\left(\boldsymbol{\Omega}^{\text{id}}\right) = c\sum_n \left(\hat{\lambda}_n^{\text{id}} - \mu c\right)_+ = P. \tag{3.95}$$

The water level is

$$\mu = \frac{\sum_{n\in\hat{\mathcal{N}}} \hat{\lambda}_n^{\text{id}} c - P}{\hat{N}c^2}, \tag{3.96}$$

with the set $\hat{\mathcal{N}}$ of the indices of the positive eigenvalues of $\boldsymbol{\Omega}^{\text{id}}$ and its cardinality $\hat{N}$. This set can be found by initializing it with the set of the indices of all eigenvalues. One after another the eigenvalue indices are removed, beginning with the index corresponding to the smallest $\hat{\lambda}_n^{\text{id}}$ and in non-decreasing order, until the termination criterion,

$$\min_{n\in\hat{\mathcal{N}}} \hat{\lambda}_n^{\text{id}} - \mu c > 0, \tag{3.97}$$

holds. The orthogonal projected noise covariance matrix is

$$\boldsymbol{\Omega}^{\text{id}} = \hat{\boldsymbol{U}}^{\text{id}}\left(\hat{\boldsymbol{\Lambda}}^{\text{id}} - \mu c\mathbf{I}\right)_+ \hat{\boldsymbol{U}}^{\text{id,H}}. \tag{3.98}$$

The required water-spilling (See Algorithm 3.6) is exactly the same as the water-spilling presented for the orthogonal projection of the transmit covariance matrices onto the constraint set with a sum trace constraint as explained in Section 3.5.4 [74].

3.5.6.2 Positive-Definite Shaping Constraint

As discussed in the last section, the orthogonal projection with an arbitrary, positive-definite shaping matrix requires an iterative algorithm to solve equation (3.92). But, an optimization with arbitrary, positive-definite shaping matrix $\boldsymbol{C}^{\text{shape}} = \boldsymbol{C}^{1/2}\boldsymbol{C}^{1/2,\text{H}}$ can always be transformed to an optimization with a scaled identity as shaping matrix [83]. With the

Algorithm 3.6 Orthogonal projection with a single scaled identity constraint

Require: unconstrained update $\hat{\boldsymbol{\Omega}}^{\mathrm{id}}$, constraint scaling c, joint limit P

1: $\hat{\boldsymbol{U}}^{\mathrm{id}}\hat{\boldsymbol{\Lambda}}^{\mathrm{id}}\hat{\boldsymbol{U}}^{\mathrm{id,H}} \leftarrow \mathrm{eig}(\hat{\boldsymbol{\Omega}}^{\mathrm{id}})$ ▷ eigenvalue decomposition

2: $\hat{\mathcal{N}} \leftarrow \left\{n \in \{1, \ldots, N\} : \hat{\lambda}_n^{\mathrm{id}} > 0\right\}$, $\hat{N} \leftarrow |\hat{\mathcal{N}}|$ ▷ initialize index set

3: $\mu \leftarrow \left(\sum_{n\in\hat{\mathcal{N}}} \hat{\lambda}_n^{\mathrm{id}} c - P\right)\left(\hat{N}c^2\right)^{-1}$ ▷ find initial water level (3.96)

4: **while** $\min_{n\in\hat{\mathcal{N}}} \hat{\lambda}_n^{\mathrm{id}} - \mu c \leq 0$ **do** ▷ check for (3.97)

5: $\quad \hat{\mathcal{N}} \leftarrow \hat{\mathcal{N}} \setminus \mathrm{argmin}_{n\in\hat{\mathcal{N}}} \hat{\lambda}_n^{\mathrm{id}}$, $\hat{N} \leftarrow \hat{N} - 1$ ▷ drop smallest eigenvalue

6: $\quad \mu \leftarrow \left(\sum_{n\in\hat{\mathcal{N}}} \hat{\lambda}_n^{\mathrm{id}} c - P\right)\left(\hat{N}c^2\right)^{-1}$ ▷ find water level (3.96)

7: **end while**

8: $\boldsymbol{\Omega}^{\mathrm{id}} \leftarrow \hat{\boldsymbol{U}}^{\mathrm{id}}\left(\hat{\boldsymbol{\Lambda}}^{\mathrm{id}} - \mu c\mathbf{I}\right)_+ \hat{\boldsymbol{U}}^{\mathrm{id,H}}$ ▷ new worst case noise (3.98)

substitutions $\tilde{\boldsymbol{Q}} = \boldsymbol{C}^{-1/2}\boldsymbol{Q}\boldsymbol{C}^{-1/2,\mathrm{H}}$, $\tilde{\boldsymbol{Q}}_k = \boldsymbol{C}^{-1/2}\boldsymbol{Q}_k\boldsymbol{C}^{-1/2,\mathrm{H}}$, $\forall k$, and $\tilde{\boldsymbol{H}}_k = \boldsymbol{C}^{1/2}\boldsymbol{H}_k$, $\forall k$, the downlink optimization with a single shaping constraint,

$$\max_{\boldsymbol{Q}_k\succeq\mathbf{0}\ \forall k} \sum_k w_k \log_2 \frac{\left|\mathbf{I} + \boldsymbol{H}_k^{\mathrm{H}}\left(\sum_{\hat{k}\geq k}\boldsymbol{Q}_{\hat{k}}\right)\boldsymbol{H}_k\right|}{\left|\mathbf{I} + \boldsymbol{H}_k^{\mathrm{H}}\left(\sum_{\hat{k}>k}\boldsymbol{Q}_{\hat{k}}\right)\boldsymbol{H}_k\right|} \quad \text{subject to} \quad \boldsymbol{Q} \preceq \boldsymbol{C}^{\mathrm{shape}}, \tag{3.99}$$

can be written with the substituted values as

$$\max_{\tilde{\boldsymbol{Q}}_k\succeq\mathbf{0}\ \forall k} \sum_k w_k \log_2 \frac{\left|\mathbf{I} + \tilde{\boldsymbol{H}}_k^{\mathrm{H}}\left(\sum_{\hat{k}\geq k}\tilde{\boldsymbol{Q}}_{\hat{k}}\right)\tilde{\boldsymbol{H}}_k\right|}{\left|\mathbf{I} + \tilde{\boldsymbol{H}}_k^{\mathrm{H}}\left(\sum_{\hat{k}>k}\tilde{\boldsymbol{Q}}_{\hat{k}}\right)\tilde{\boldsymbol{H}}_k\right|} \quad \text{subject to} \quad \tilde{\boldsymbol{Q}} \preceq \mathbf{I}. \tag{3.100}$$

The optimization in the uplink can now be done with a scaled identity as constraint matrix, but with the substituted channels. The orthogonal projection of the uplink noise covariance matrix can be done as described in the previous section. After the recovery of the transmit covariance matrices $\tilde{\boldsymbol{Q}}_k$, $\forall k$, in the downlink, the actual transmit covariance matrices $\boldsymbol{Q}_k$, $\forall k$, can be found by undoing the substitution.

3.5.6.3 Multiple Linear Constraints

As discussed in Section 3.5.5, $\boldsymbol{\Omega}_l^{\mathrm{lin}} = \omega_l \boldsymbol{A}_l$ has to be a scaled version of the constraint matrix $\boldsymbol{A}_l$. It follows, that $\boldsymbol{T}_l^{\mathrm{lin}} \in \mathcal{Z}_l^{\mathrm{lin}}$ has to fulfill $\mathrm{tr}(\boldsymbol{A}_l\boldsymbol{T}_l^{\mathrm{lin}}) = 0$. Equation (3.91) can be multiplied from the left with $\boldsymbol{A}_l$:

$$\begin{aligned} \boldsymbol{A}_l\boldsymbol{\Omega}_l^{\mathrm{lin}} &= \boldsymbol{A}_l\hat{\boldsymbol{\Omega}}_l^{\mathrm{lin}} + \boldsymbol{A}_l\boldsymbol{S}_l^{\mathrm{lin}} + \boldsymbol{A}_l\boldsymbol{T}_l^{\mathrm{lin}} - \mu\boldsymbol{A}_l\boldsymbol{C}_l^{\mathrm{lin}}, \\ \omega_l\boldsymbol{A}_l^2 &= \hat{\omega}_l\boldsymbol{A}_l^2 + \boldsymbol{A}_l\boldsymbol{S}_l^{\mathrm{lin}} + \boldsymbol{A}_l\boldsymbol{T}_l^{\mathrm{lin}} - \mu\boldsymbol{A}_l\boldsymbol{C}_l^{\mathrm{lin}}. \end{aligned} \tag{3.101}$$

Without loss of generality, it can be assumed that $\mathrm{tr}(\boldsymbol{A}_l^2) = 1$ and $\mathrm{tr}(\boldsymbol{A}_l\boldsymbol{C}_l^{\mathrm{lin}}) = a_l$, $\forall l$. Both sides of equation (3.73) can be scaled arbitrarily without changing the constraint. Taking the trace of equation (3.101) gives

$$\omega_l = \hat{\omega}_l + \mathrm{tr}(\boldsymbol{A}_l\boldsymbol{S}_l^{\mathrm{lin}}) - \mu a_l. \tag{3.102}$$

As $\boldsymbol{\Omega}_l^{\text{lin}} = \omega_l \boldsymbol{A}_l$ holds and $\boldsymbol{A}_l$ has to be a positive semidefinite matrix, the dual variable $\boldsymbol{S}_l^{\text{lin}}$ has only to make sure that ω_l is positive:

$$\omega_l = (\hat{\omega}_l - \mu a_l)_+. \tag{3.103}$$

This can also be found by using $\boldsymbol{\Omega}_l = \omega_l \boldsymbol{A}_l$, $\forall l$, in the cost function directly. Equation (3.103) can be transformed, by multiplying with $\boldsymbol{A}_l$ again, to

$$\boldsymbol{\Omega}_l^{\text{lin}} = \omega_l \boldsymbol{A}_l = (\hat{\omega}_l - \mu a_l)_+ \boldsymbol{A}_l. \tag{3.104}$$

If all constraints are linear constraints, the water level μ is found by plugging all $\boldsymbol{\Omega}_l^{\text{lin}}$ into the sum trace constraint:

$$\mu = \frac{\sum_{l \in \hat{\mathcal{L}}} \hat{\omega}_l a_l - P}{\sum_{l \in \hat{\mathcal{L}}} a_l^2}, \tag{3.105}$$

where the set $\hat{\mathcal{L}}$ of cardinality $\hat{L}$ contains the indices of all positive ω_l. To find $\hat{\mathcal{L}}$, it is initialized with the indices of all ω_l and the index corresponding to the smallest $\hat{\omega}_l a_l^{-1}$ is dropped, one at a time, until

$$\min_{l \in \hat{\mathcal{L}}} \frac{\hat{\omega}_l}{a_l} - \mu > 0 \tag{3.106}$$

holds. The updated noise covariance matrices are

$$\boldsymbol{\Omega}_l^{\text{lin}} = \omega_l \boldsymbol{A}_l = (\hat{\omega}_l - \mu a_l)_+ \boldsymbol{A}_l, \ \forall l. \tag{3.107}$$

See Algorithm 3.7 for a pseudo code of the described steps.

Algorithm 3.7 Orthogonal projection with multiple linear constraints

Require: unconstrained updates $\hat{\omega}_l$, $\forall l$, constraint values a_l, $\forall l$, joint limit P

1: $\hat{\mathcal{L}} \leftarrow \{l \in \{1, \ldots, L\} : \hat{\omega}_l > 0\}$ ▷ initialize index set
2: $\mu \leftarrow \left(\sum_{l \in \hat{\mathcal{L}}} \hat{\omega}_l a_l - P\right) \left(\sum_{l \in \hat{\mathcal{L}}} a_l^2\right)^{-1}$ ▷ find initial water level (3.105)
3: **while** $\min_{l \in \hat{\mathcal{L}}} \hat{\omega}_l - \mu a_l \leq 0$ **do** ▷ check for (3.106)
4: $\quad \hat{\mathcal{L}} \leftarrow \hat{\mathcal{L}} \setminus \operatorname{argmin}_{l \in \hat{\mathcal{L}}} \hat{\omega}_l a_l^{-1}$ ▷ drop smallest $\hat{\omega}_l a_l^{-1}$
5: $\quad \mu \leftarrow \left(\sum_{l \in \hat{\mathcal{L}}} \hat{\omega}_l a_l - P\right) \left(\sum_{l \in \hat{\mathcal{L}}} a_l^2\right)^{-1}$ ▷ find water level (3.105)
6: **end while**
7: $\omega_l \leftarrow \left(\hat{\omega}_l - \mu^{\text{lin}} a_l\right)_+$, $\forall l$ ▷ new scaling variables (3.107)
8: $\boldsymbol{\Omega}^{\text{lin}} \leftarrow \sum_l \omega_l \boldsymbol{A}_l$ ▷ new worst case noise (3.107)

3.5.6.4 Scaled Identity and Multiple Linear Constraints

The update equations for $\boldsymbol{\Omega}^{\text{id}}$ and $\boldsymbol{\Omega}_l^{\text{lin}}$ can directly be taken from equation (3.98) and (3.107) for the shaping and the multiple linear constraints, respectively. But, the water level in the equations has to be exchanged with a combined water level μ derived from all constraints. To find the combined water level, the combined trace constraint of the scaled identity constraint and the multiple linear constraints is evaluated:

$$P = \operatorname{tr}\left(\boldsymbol{\Omega}^{\text{id}} \boldsymbol{C}^{\text{id}}\right) + \sum_l \operatorname{tr}\left(\boldsymbol{\Omega}_l^{\text{lin}} \boldsymbol{C}_l^{\text{lin}}\right). \tag{3.108}$$

Based on equation (3.95) and equation (3.104), this can be written as

$$P = \sum_n \left(\hat{\lambda}_n^{\mathrm{id}} c - \mu c^2\right)_+ + \sum_l \left(\hat{\omega}_l a_l - \mu a_l^2\right)_+, \tag{3.109}$$

which can be generalized to

$$P = \sum_t \left(\hat{\xi}_t \psi_t - \mu \psi_t^2\right)_+. \tag{3.110}$$

$\hat{\xi}_t$ is from the union set of all eigenvalues $\hat{\lambda}_n^{\mathrm{id}}$ and scaling variables $\hat{\omega}_l$ of the unconstrained updates, ψ_t is from the union set of all corresponding factors c and a_l, and t is from the union set $\mathcal{T}$ of all corresponding indices n and l. The combined water level is

$$\mu = \frac{\sum_{t \in \hat{\mathcal{T}}} \hat{\xi}_t \phi_t - P}{\sum_{t \in \hat{\mathcal{T}}} \phi_t^2}. \tag{3.111}$$

$\hat{\mathcal{T}}$ is the set of indices corresponding to positive entries of the union set of all eigenvalues λ_n and scaling variables ω_l of the projected update matrices. This set is initialized with $\hat{\mathcal{T}} \leftarrow \mathcal{T}$ and the entry

$$\operatorname*{argmin}_{t \in \hat{\mathcal{T}}} \frac{\hat{\xi}_t}{\phi_t} \tag{3.112}$$

is dropped, until

$$\min_{t \in \hat{\mathcal{T}}} \frac{\hat{\xi}_t}{\phi_t} - \mu > 0 \tag{3.113}$$

holds. See Algorithm 3.8 for the pseudo code.

Algorithm 3.8 Orthogonal projection with scaled identity and multiple linear constraints

Require: unconstrained updates $\hat{\boldsymbol{\Omega}}^{\mathrm{id}}$, $\hat{\omega}_l$, $\forall l$, constraint values c, a_l, $\forall l$, joint limit P

$\hat{\boldsymbol{U}}^{\mathrm{id}} \hat{\boldsymbol{\Lambda}}^{\mathrm{id}} \hat{\boldsymbol{U}}^{\mathrm{id,H}} \leftarrow \mathrm{eig}(\hat{\boldsymbol{\Omega}}^{\mathrm{id}})$ ▷ eigenvalue decomposition

$\hat{\xi}_t \leftarrow \hat{\lambda}_t^{\mathrm{id}}$ and $\phi_t \leftarrow c$ $\forall t \in \{1, \ldots, N\}$

$\hat{\xi}_t \leftarrow \hat{\omega}_{t-N}$ and $\phi_t \leftarrow a_{t-N}$ $\forall t \in \{N+1, \ldots, N+L\}$

$\hat{\mathcal{T}} \leftarrow \left\{t \in \{1, \ldots, N+L\} : \hat{\xi}_t > 0\right\}$ ▷ initialize joint index set

$\mu \leftarrow \left(\sum_{t \in \hat{\mathcal{T}}} \hat{\xi}_t \phi_t - P\right) \left(\sum_{t \in \hat{\mathcal{T}}} \phi_t^2\right)^{-1}$ ▷ find initial water level (3.111)

while $\min_{t \in \hat{\mathcal{T}}} \hat{\xi}_t - \mu \phi_t \leq 0$ **do** ▷ check for (3.113)

  $\hat{\mathcal{T}} \leftarrow \hat{\mathcal{T}} \setminus \operatorname{argmin}_{t \in \hat{\mathcal{T}}} \hat{\xi}_t \phi_t^{-1}$ ▷ drop smallest $\hat{\xi}_t \phi_t^{-1}$

  $\mu \leftarrow \left(\sum_{t \in \hat{\mathcal{T}}} \hat{\xi}_t \phi_t - P\right) \left(\sum_{t \in \hat{\mathcal{T}}} \phi_t^2\right)^{-1}$ ▷ find water level (3.111)

end while

$\omega_l \leftarrow (\hat{\omega}_l - a_l \mu)_+$, $\forall l$ ▷ new scaling variables (3.107)

$\boldsymbol{\Omega} \leftarrow \hat{\boldsymbol{U}}^{\mathrm{id}} \left(\hat{\boldsymbol{\Lambda}}^{\mathrm{id}} - \mu c \mathbf{I}\right)_+ \hat{\boldsymbol{U}}^{\mathrm{id,H}} + \sum_l \omega_l \boldsymbol{A}_l$ ▷ new worst case noise (3.98) and (3.107)

3.5.7 Convergence Behaviour

Comparing the performance of different algorithms requires a detailed analysis of their complexity. In the vicinity of the optimum, the infeasible start Newtons method converges in general with less iterations, while the gradient-projection algorithm gets with less iterations into the vicinity of the optimum. The infeasible start Newtons method converges with less iterations in total at a higher complexity per iteration compared to the gradient-projection algorithm. It has been proven in [74], that the gradient-projection method is very efficient for finding the optimal transmit covariances with fixed noise. As the problem is closely related, it can be assumed that the optimal noise covariance matrices with fixed transmit covariances can be found also very efficiently with the gradient-projection algorithm.

Most existing algorithms for the weighted sum rate maximization with multiple linear constraints alternate between an optimization of the transmit covariance matrices until convergence and an optimization of the worst case noise until convergence [67, 70]. This is just like Algorithm 3.1, but with different methods for finding the optimal transmit covariance matrices and worst case noise covariance matrices, respectively. In the context of these algorithms, an iteration step is a complete run of the outer loop of Algorithm 3.1. As Algorithm 3.2 does not have inner loops, the required iterations of each optimization are counted as the number of times a transmit covariance update step is done. Histograms of the iterations required with Algorithm 3.2 and 3.1 are compared in the following. Figures 3.8 and 3.9 depict two linear constraints and scaled identity shaping, respectively.

In the setup, the base station has $N = 4$ antennas and serves $K = 4$ single antennas users ($M = 1$) with sum power $P = 10\,\mathrm{W}$ and arbitrary weights $[w_1, w_2, w_3, w_4] = [1, 2, 3, 4]$. The simulations are averaged over 10000 i.i.d. channel realizations, where each entry of the channel vector $\boldsymbol{h}_k$ has a zero-mean complex Gaussian distribution with variance one. The first linear constraint is the sum power constraint. The second linear constraint forbids to transmit anything in the direction of an additionally generated channel. The maximum sum-rate was found by running Algorithm 3.2 with 1000 iterations, respectively.

The proof of convergence depends on the selected type of step-size as described in Section 3.5.3 and [28, Section 5.1.3]. The on demand step-size reduction combined with a diminishing step-size assures convergence but not necessarily to the optimal point. Therefore, it has to be checked in the downlink, if all constraints are met after the algorithms converged. In this case, the convergence can easily be checked with a change in the weighted sum rate. The convergence to the optimal point can be checked by looking at the size of the gradient. This requires that the gradient is projected onto the tangent cone. If the norm of all projected gradients is smaller than ϵ, the optimal point is found within the given accuracy. Checking for convergence with this method is more accurate, but requires to project the gradients onto the tangent cone in every step, which non-negligibly increases the complexity. Additionally, the results can be compared to known results. The optimization with a power constraint and one forbidden direction can be compared to results with the transmit covariance matrices selected in the kernel of the forbidden direction as described in Section 2.5.1.2. In the simulations, the algorithms always converged to the optimal rate while fulfilling all constraints with the given accuracy.

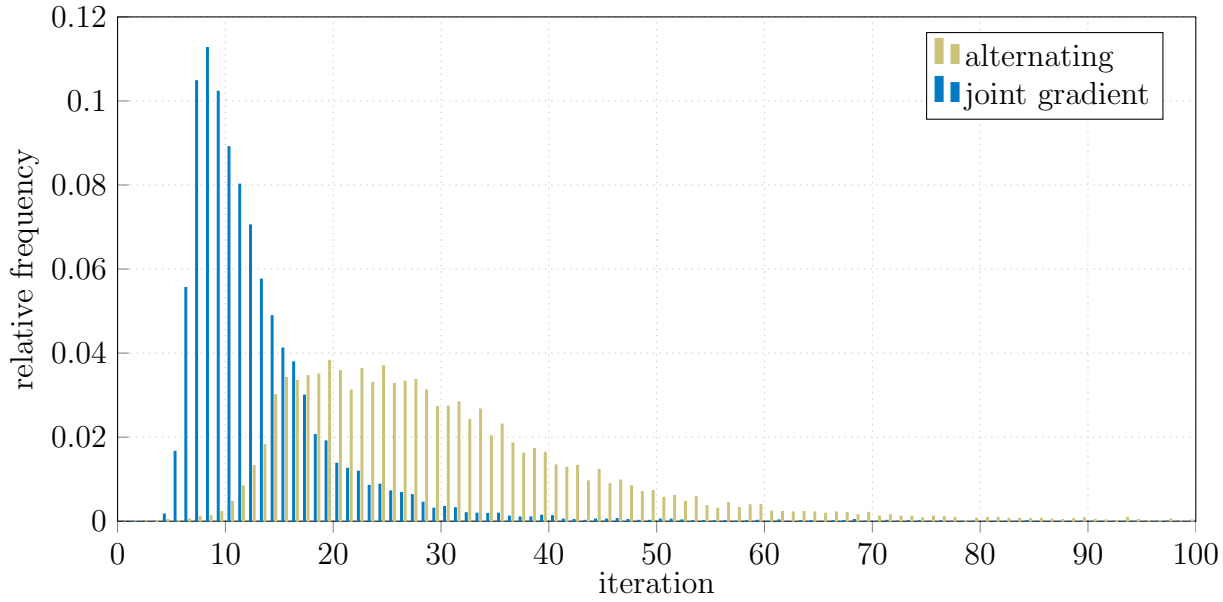


Figure 3.8. Relative frequency of the number of iterations required to find the maximum weighted sum rate with two linear constraints (sum power and one forbidden direction). The algorithms are stopped, if a sum-rate larger than $(1-\varepsilon)$ times the maximum sum-rate was achieved and each constraint was hurt by less than ε ($\varepsilon = 10^{-3}$).

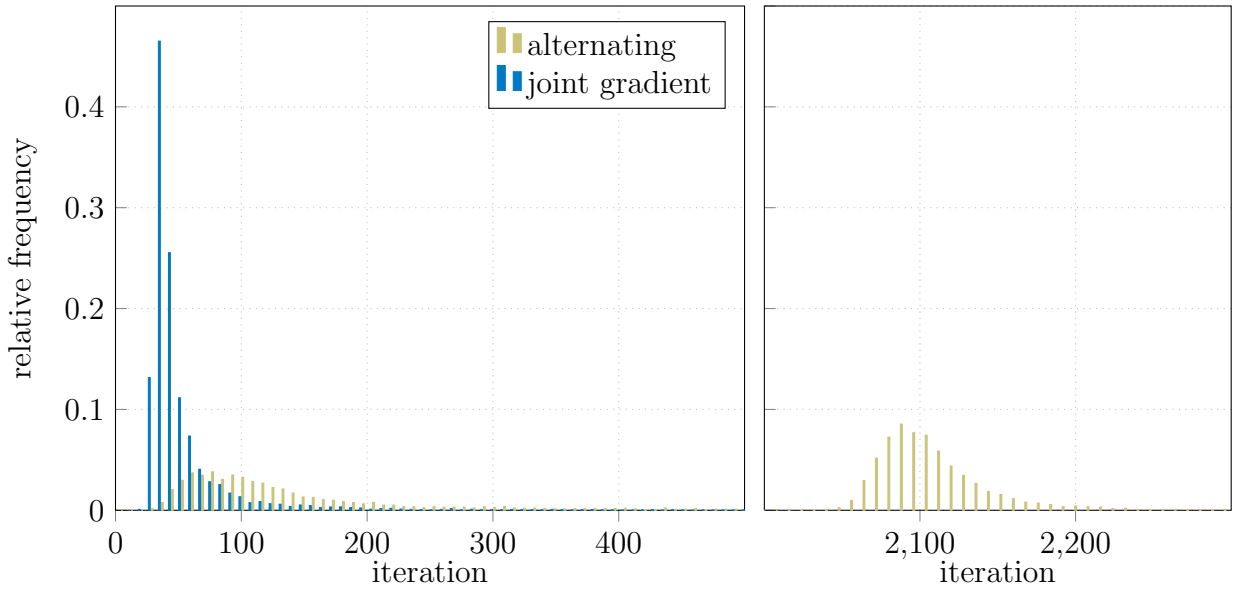


Figure 3.9. Relative frequency of the number of iterations required to find the maximum weighted sum rate with the scaled identity constraint. The algorithms are stopped, if a sum-rate larger than $(1-\varepsilon)$ times the maximum sum-rate was achieved and the squared Frobenius norm of the constraint error matrix was less than ε ($\varepsilon = 10^{-3}$).

3.6 Conclusion

With an adaptation of the minimax uplink-downlink duality, it could be shown that the weighted sum rate maximization with multiple linear and/or linear conic constraints can be solved efficiently with an alternating gradient-projection algorithm. The required orthogonal

projection of the gradients onto the tangent cone of the constraint set and the projection of the updates onto the constraint set itself were derived and described in detail.

Switching after each gradient step between an update of the transmit covariance matrices and the noise covariance matrices has a superior convergence behavior compared to an optimization, which optimizes the transmit and noise covariance matrices until convergence in turns.

4. Intercell Interference Robustness Methods

In state-of-the-art networks, orthogonal frequency bands are assigned to adjacent base stations to separate the transmissions within these cells. The smaller the fraction of the frequency band, the frequency reuse, at each base station is, the less base stations in the near vicinity have to use the same frequency band and the distance to the next base station utilizing the same frequency band can be increased. A smaller frequency reuse reduces the relative influence of the intercell interference compared to the thermal noise at the same transmit power level. But, next generation mobile communication standards long to use the full frequency band at each base station to increase the spectral efficiency. The rate scales linearly with the bandwidth and a complete frequency reuse at each base station has the highest efficiency. The problem of the strong intercell interference is supposed to be encountered with cooperative beamforming in multiple antenna systems.

But, as it could be shown in Chapter 2, only parts of the intercell interference can be mitigated with cooperative strategies. Since measuring all interference channels is unrealistic in a large scale system, the intercell interference is split into two parts—the intercell interference over known channels and over unknown channels. The interference over the known channels can be handled with cooperative strategies. For the unknown interference channels, the interference can only be regarded as noise. What are the properties of this interference which has to be regarded as noise?

In a network with interference coordination, the base stations are assumed to calculate their beamforming in a distributed manner. In order to optimize the beamforming and compute the possible achievable rates for the link rate adaptation, the base stations need information about the intercell interference variances at each mobile device. The received intercell interference variance at a mobile device depends on the interference channel vectors and the sum transmit covariance matrices of the interfering base stations. Even if the channels are assumed to stay constant, the intercell interference at each mobile device will change the moment any base station changes its transmit covariance matrix and cannot be known in advance. Therefore, the base stations compute their beamforming based on an assumed intercell interference. The base stations are blind to the intercell interference change and take the risk, that the actual intercell interference is larger than assumed and the mobile device cannot decode the transmitted symbols or that the actual intercell interference is smaller than assumed and valuable resources are wasted [4]. This so called "flash light effect" was already addressed in [91] and is described in more detail in Section 4.1.

Uncertainty in the intercell interference has a variety of negative effects. The link rate adaption can fail because the SINR during the transmission is unknown, which might lead to unexpected outages. Some operations in higher layers, such as scheduling and resource

allocation, also depend on the SINR and cannot be performed optimally based on assumed intercell interference values. Please note that such an uncertainty in the intercell interference always exists in a realistic cellular system, while in an idealized network MIMO scenario with fully centralized coordination, the problem of intercell interference blindness does not arise. In state-of-the-art systems, the intercell interference instationarity is mitigated with *hybrid automatic repeat request* (HARQ) with soft combining. In this chapter, beamforming strategies handling the intercell interference blindness problem are discussed. The incorporation of HARQ and its effects into these strategies is presented in [29, 33] and Section 4.11.

Most optimizations in the literature ignore the intercell interference blindness problem completely and just assume to know the intercell interference or utilize the expectation of the intercell interference or an intercell interference realization from a previous step as the assumed intercell interference. But, this results in a mismatch between the cost function of the optimization and the actual performance measure and motivates to look at signal processing strategies to fight the intercell interference blindness. In [4] the *conservative gambling* strategy to counteract the intercell interference blindness and the resulting uncertainty about the supported rates is presented, where the data is transmitted at reduced rates with a common back-off factor β. With this strategy, the risk of assuming a rate, which is not supported and would lead to complete outage is reduced, but also the possibly achievable rates are reduced (Section 4.4). Another approach presented in this paper is the *stabilization* method, which simply requires the transmit vectors to stay constant for all time. Dropping the adaptive beamforming removes the uncertainty in the intercell interference variance at the cost of dramatically reduced possible rates. The conservative gambling method was extended by Shirani-Mehr et al. in [29] to conservative gambling with individual back-off factors for each mobile device.

The authors of [4] also define an upper bound for the possible rates in systems with intercell interference instationarity. For the upper bound, the actual intercell interference is assumed to be known and the transmit strategies at all base stations are iteratively updated in parallel. As the problem is not convex, convergence to the optimal value cannot be guaranteed. If the base stations synchronize the update of their beamforming, it will be possible to measure the actual intercell interference with a second pilot, which removes the uncertainty in the intercell interference but increases the overhead [3, 31] (Section 4.7). In [30] and recently also in [92], it was proposed to optimize the transmit covariances at each base station based on the expected rate of the associated mobile devices. With this approach, the system for which the transmit covariances are optimized and the system in which the covariances are utilized become the same (See Section 4.6). By including the risks of a changing intercell interference in the precoder selection, the base stations allocate more resources to mobile devices which are critical for the utility to improve the expected rates.

A different method following the idea of the stabilization approach to handle the intercell interference blindness problem was suggested in [32]. The sum covariance matrices of the transmit symbols at each base station are forced to scaled identity matrices, which still leaves room for an optimization of the individual transmit covariance matrices for the mobile devices. This constraint completely removes the uncertainty in the intercell interference variance and the SINR values of the served mobile devices can be known at the base stations. But, the shaping constraint on the transmitter also reduces the achievable rates (Section 4.9). A combination of the expected rate optimization with a less restrictive shaping constraint was proposed in [80]. With a loosening factor, the strictness of the shaping constraint is

adjusted and the benefits of the stabilization and the expected rate method are combined (Section 4.10).

The problem of the intercell interference blindness is discussed in [31,93,94] for scenarios, where only long term channel state information is available. In [95], Kalman prediction is utilized to guess the intercell interference in the next time slot. Neely et al. [96] investigated the intercell interference blindness from a networking point of view with generalized links between different nodes but regardless of the transmit covariance selection. A subchannel allocation strategy in OFDM, which stabilizes the intercell interference variances at mobile devices was proposed in [5]. Similar problems to the intercell interference blindness were investigated by Huh et al. in [97], where the uncertainty in the SINR is caused by non-perfect channel state knowledge, and by Shirani-Mehr et al. in [98] for varying channels. Huh addressed the problem in the asymptotic region, while Shirani-Mehr utilized an heuristic robustness method.

The adaption of the covariance shaping method to linear precoding with unitary beamforming is discussed in [99] and tested with an LTE system simulator in [100]. In these simulations, the benefits of unitary precoding as interference robustness method are negligible compared to conservative gambling combined with HARQ. This coincidences with the results from Section 4.11, where the performance in the presence of HARQ can only be improved with the expected rate method or loosened covariance shaping. Conservative gambling with individual back-off factors in combination with HARQ was investigated by Shirani-Mehr et al. in [29].

To perform the expected rate method, the cumulative distribution functions of the intercell interference at each associated mobile device need to be available at the serving base station. The cumulative distribution functions can be approximated with long term measurements at the mobile devices. To reduce the feedback, a probability distribution can be matched to the measurements at the mobile devices. Then, only the parameters of the distribution function need to be transmitted to the base station (Section 4.2). It would also be possible to have a rough estimate of the cumulative distribution function directly based on the channel measurements. This would not require any additional measurements and feedback for the cumulative distribution function. The actual intercell interference cannot be known in advance in the regarded scenario, because the transmit covariances at all base station change at the same time, while the channels are assumed to be constant for the block of transmission [55, 62]. The cumulative distribution functions of the intercell interference for this scenario are different from the cumulative distribution functions, where the channels also change over time. Although the later scenario with changing channels is more realistic, especially for cumulative distribution functions, which are approximated with long term measurements, the scenario with fixed channels is assumed to reduce the simulation complexity. The cost of acquiring the cumulative distribution functions is neglected in the following.

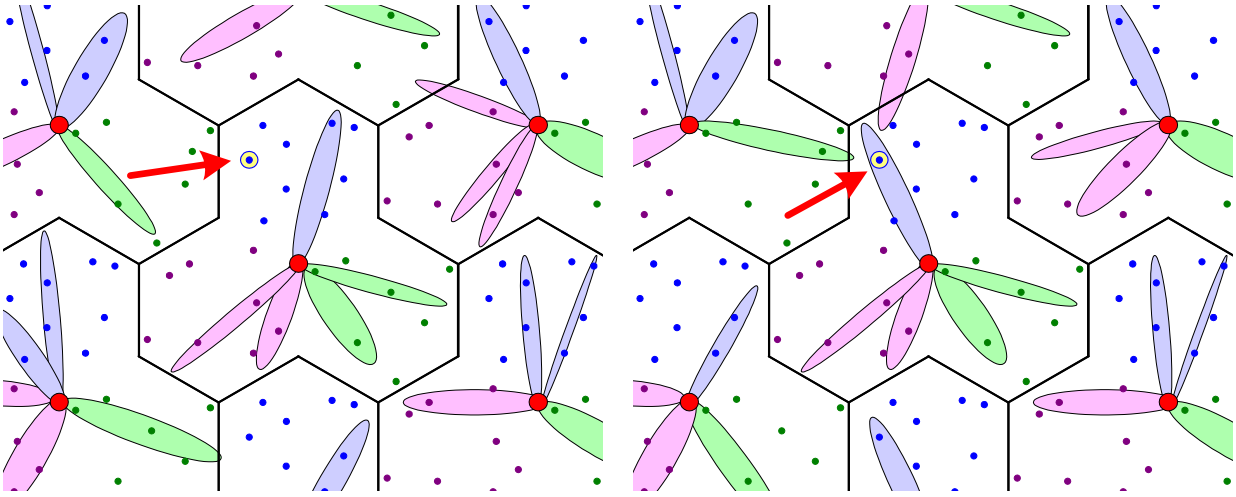
For the description of the different methods handling the intercell interference blindness, we assume a system without cooperation. The complete intercell interference is regarded as noise and its variance is described with a random process with known parameters. Therefore, the base stations do not influence each other anymore and we can limit ourselfs to a system model with a single base station and random intercell interference variances at each mobile device, respectively (Section 4.3).

In this chapter, different methods are presented to make mobile communication systems robust against the intercell interference blindness problem for a general utility function and the weighted sum rate maximization as an example. The adaption to other utility functions and their impact on the intercell interference blindness methods are presented in Chapter 5. Parts of this chapter can be found in [30, 33, 80, 101].

4.1 Intercell Interference Blindness Problem

Interference mitigation can only be performed for known interference channels. The interference over the unknown channels has to be regarded as noise and changes unpredictably the moment any base station employs a newly calculated transmit strategy. This causes the interference blindness problem. The actual intercell interference, SINR, and supported rate are unknown after an update of the transmit covariance matrices. The achievable rates cannot be optimized directly, merely the expectations of these rates.

Even in scenarios with cooperation among the base stations this type of interference cannot be completely eliminated. Cooperation is always limited in realistic systems, because the measurement of all interference channels and a coordination of all beamformers in the network cannot be implemented [1–3] (See Section 2). This interference over the unmeasured channels scales with the common transmit power at the base stations, such systems are always interference limited. Without loss of generality, the base stations do not coordinate their beamforming in the considered scenario. The channels between a mobile device and the interfering base stations are not measured. All intercell interference is regarded as noise.



(a) Time slot one (b) Time slot two

Figure 4.1. Measured and actual system

The intercell interference blindness problem is visualized in Figure 4.1 [4]. Each red dot is a site with three collocated base stations. The beams belonging to the same base station are indicated by the same color and the same origin. It is assumed that the base stations do not coordinate their beamforming and all interference channels remain unmeasured. In time slot one (see Figure 4.1 (a)), the base station in the middle with the blue beam serves a mobile device in the upper right corner of the associated cell. The encircled mobile device has a very good possible SINR. All the other base stations point their beams in other directions

and cause only a small amount of interference at this mobile device. The base station in the center decides to serve this mobile device in the next time slot and selects a beam to concentrate its power at this mobile device. At the same time, all other base stations updated their transmit strategy and, unfortunately, point their beams at the encircled mobile device in the second time slot (see Figure 4.1 (b)). The SINR of the encircled mobile device drops dramatically and makes it unfavorable to serve this mobile device. From the point of view of the mobile device, the system in the first and the second time slot are completely different, the intercell interference, SINR and supported rate changed unpredictably.

The actual interference $\theta_{b,k}^{\text{actual}}$ and the supported rate $r_{b,k}^{\text{actual}}$ at mobile device (b,k) during the transmission cannot be known in advance. Even if all base stations would update their beamforming at the same time, the intercell interference could not be known before the base stations have choosen their transmit covariances. The base stations are blind to the intercell interference change and stand the risk, that the intercell interference increases and the mobile device cannot decode the transmitted symbols or that the intercell interference decreases and valuable resources are wasted. An assumed intercell interference variance $\theta_{b,k}^{\text{assumed}}$ and the corresponding assumed rate $r_{b,k}^{\text{assumed}}$ have to be used for the optimization of the transmit processing. The mismatch between the assumed intercell interference $\theta_{b,k}^{\text{assumed}}$ and the true intercell interference $\theta_{b,k}^{\text{actual}}$ is the intercell interference blindness problem and can be formulated as

$$r_{b,k}^{\text{achieved}} = \begin{cases} r_{b,k}^{\text{assumed}}, & \text{for } \theta_{b,k}^{\text{assumed}} \geq \theta_{b,k}^{\text{actual}}, \\ 0, & \text{for } \theta_{b,k}^{\text{assumed}} < \theta_{b,k}^{\text{actual}}. \end{cases}, \tag{4.1}$$

where the information outage probability is used as error probability [4, 29, 102, 103].

After the optimization of the transmit covariance matrices, the base station assigns the data rate $r_{b,k}^{\text{assumed}}$ with corresponding modulation and coding to user (b,k). If the assumed rate is larger than the supported rate, mobile device (b,k) cannot decode the data and the transmission fails. In other words, when the actual intercell interference is larger than the assumed intercell interference, the channel is worse than assumed and the mobile device cannot decode the signals successfully. If the assumed intercell interference is smaller than the actual intercell interference, mobile device (b,k) can only communicate with the assumed rate $r_{b,k}^{\text{assumed}}$. The achieved rate $r_{b,k}^{\text{achieved}}$ will not be equal to the supported rate $r_{b,k}^{\text{actual}}$ with the actual intercell interference $\theta_{b,k}^{\text{actual}}$ but to the assumed rate $r_{b,k}^{\text{assumed}}$ based on the assumed intercell interference $\theta_{b,k}^{\text{assumed}}$. The same rate could be achieved with less power, if the actual intercell interference is known. The over-assigned transmit power is not only wasted, but produces more interference than necessary in adjacent cells.

If the system is optimized based on some assumed intercell interference, the mismatch between the assumed and the actual intercell interference will lead to a mismatch between the cost function and the performance measure. The system, for which the transmit covariance matrices are optimized, is different to the system, where the transmit covariances are employed. To counteract this problem and match the cost function and the performance measure, the general objective is to maximize the expectation of the sum utility $U(\bullet)$ of the whole system:

$$\max_{\boldsymbol{Q}_{b,k} \succeq \mathbf{0}\ \forall (b,k) \in \mathcal{K}} \ \mathrm{E}\left[\sum_{(b,k)\in\mathcal{K}} U\left(r_{b,k}^{\text{achieved}}\right) \right] \quad \text{s.t. } \mathrm{tr}\left(\boldsymbol{Q}_b\right) \leq P. \tag{4.2}$$

The variance $\theta_{b,k}^{\text{actual}}$ of the actual intercell interference has to be regarded as a random process. Even if the utility is a linear function, the rate is a non linear function of the intercell interference. Therefore, the expectation of the rate cannot be simply exchanged by the rate calculated with the expectation of the intercell interference. Different methods solving (4.2) are presented in the following sections.

4.2 Intercell Interference Statistics

For some of the intercell interference robustness methods and the single cell system model, the cumulative distribution functions of the intercell interference at each associated mobile device need to be available at the serving base station. The cumulative distribution functions can be approximated with long term measurements at the mobile devices. It could also be possible to estimate a rough cumulative distribution function directly based on the channel measurements. This would not require any additional measurements and feedback for the cumulative distribution function.

The intercell interference can be modeled based on interference modeling research and experiments, e.g. [104–106]. But, to include the effects of the different intercell interference robustness techniques and utilities, we derive them from simulations in a system model with many base stations. Although the base stations will not know and use the unmeasured interference channels, they are calculated for the simulations. With the block fading assumption, the intercell interference variance change depends only on the change of the transmit covariances of the interfering base stations. These covariances are optimized without taking the interference they produce over the unknown channels into account. Therefore, the statistics of the intercell interference variance can be found by looking at many independent realizations of these covariances, while the channels are constant.

Looking at the system model from Chapter 2 and assuming that the sum transmit covariance matrix of base station $\hat{b}$ follows a Wishart distribution,

$$\boldsymbol{Q}_{\hat{b}} \sim \mathrm{W}_N \left(\frac{P}{N^2} \mathbf{I}_N, N \right), \tag{4.3}$$

with N degrees of freedom and scale matrix $\frac{P}{N^2}\mathbf{I}_N$, the intercell interference this base station causes at mobile device (b,k) will follow a gamma distribution

$$\theta_{\hat{b},b,k} = \boldsymbol{h}_{\hat{b},b,k}^{\mathrm{H}} \boldsymbol{Q}_{\hat{b}} \boldsymbol{h}_{\hat{b},b,k} \sim \Gamma \left(N, \frac{P}{N^2} \boldsymbol{h}_{\hat{b},b,k}^{\mathrm{H}} \boldsymbol{h}_{\hat{b},b,k} \right) \tag{4.4}$$

with scale parameter $\frac{P}{N^2}\boldsymbol{h}_{\hat{b},b,k}^{\mathrm{H}}\boldsymbol{h}_{\hat{b},b,k}$ and a shape parameter N. The correct calculation of the distribution of the intercell interference is a difficult task, on the one hand the base stations will always transmit with the full transmit power and, therefore, the covariances cannot follow a Wishart distribution. On the other hand the sum of many gamma distributed random variables with different scale parameters is an impasse.

Although the transmit covariances $\boldsymbol{Q}_{\hat{b}}$ result from the optimizations and the sum over all the intercell interference over the unknown channels is taken, it can be shown with simulations, that the intercell interference variance $\theta_{b,k} = \sum_{\hat{b} \in \mathcal{B} \setminus \mathcal{C}_{b,k}} \theta_{\hat{b},b,k}$ can be approximated quite good with a gamma distribution,

$$\theta_{b,k} \sim \Gamma \left(a_{b,k}, b_{b,k} \right), \tag{4.5}$$

where the shape parameter $a_{b,k} = (\mathrm{E}[\theta_{b,k}])^2(\mathrm{var}[\theta_{b,k}])^{-1}$ and the rate parameter $b_{b,k} = \mathrm{E}[\theta_{b,k}](\mathrm{var}[\theta_{b,k}])^{-1}$ can be derived from the mean $\mathrm{E}[\theta_{b,k}]$ and variance $\mathrm{var}[\theta_{b,k}]$ of the intercell interference variance. Figure 4.2 compares the empirical distribution function of the intercell interference with 1000 realizations of the covariance matrices and the approximated cumulative distribution function with a gamma distribution, which takes the mean and variance from the samples. The simulations are done in a system without cooperation, the expected rate method from Section 4.6 to assure intercell interference robustness and the maximum sum rate as cost function.

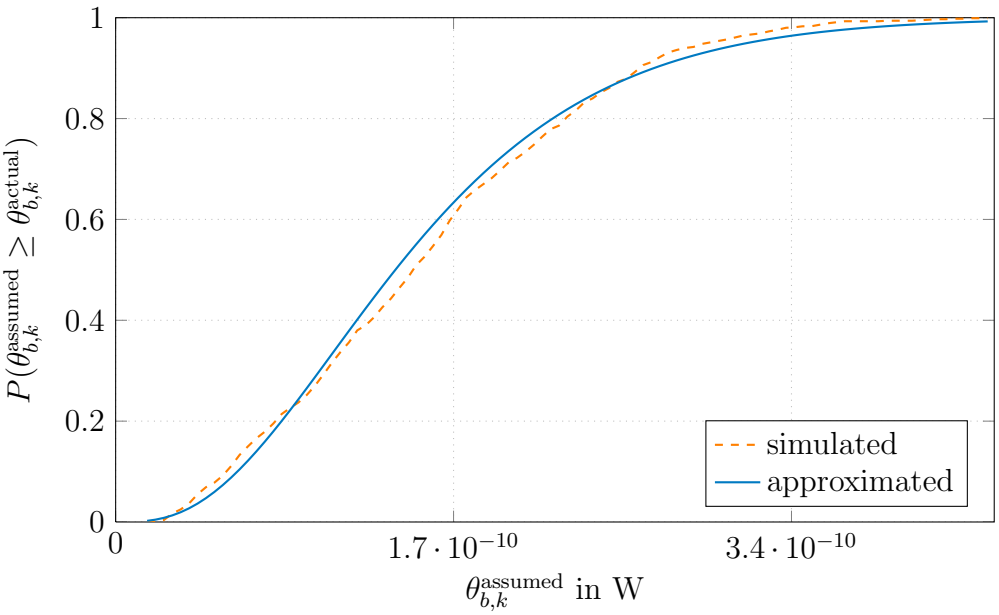


Figure 4.2. Cumulative distribution function of the intercell interference variance

In reality and in our simulations, the assumption of Wishart distributed transmit covariance matrices is incorrect. Letting a base station always transmit with full transmit power already violates the assumption of a Wishart distribution. The simulations operate on an histogram of intercell interference realizations instead of the probability distribution. The first round of intercell interference realizations is generated with scaled identities as transmit covariances for the interfering base stations. New transmit covariances are found with these intercell interference realizations and these new transmit covariances are used for the calculation of new intercell interference realizations. Only with the second set of intercell interference realizations the calculated expectation of the rates and the simulated expectation become equal. The scaled identity matrices are always of full rank, while the second set of covariances is not in general. A further iteration with the intercell interference realizations and covariances does not change the results.

Since the applied cooperative technique, the robustness method, and the utility change the selection of the covariance matrices, a different distribution of the intercell interference variance has to be found for each transmit strategy combination. E.g. with regard to the upper bound from Chapter 2, if some interference channels are measured and set to zero, the statistics of the intercell interference will chance accordingly. Implementing the stabilization or covariance shaping method removes the uncertainty in the intercell interference variance and the cumulative distribution function becomes a step function. Even changing from the

sum rate optimization to a fairness optimization will change the distribution of the intercell interference variance. For each transmit strategy combination an individual distribution approximation has to be made and individual $a_{b,k}$ and $b_{b,k}$ have to be calculated for each mobile device, respectively.

In the following, we assume that the intercell interference variance follows a gamma distribution and that the mean and the variance of this distribution are known through a combination of longterm and instantaneous measurements at the mobile devices, respectively. In Figure 4.3, it can be seen that this approximation of the intercell interference variance distribution has only a negligible error.

4.3 Single Cell System Model

For simplicity but without loss of generality, we limit ourself to a scenario with no cooperation among the base stations an no measured interference channels in this chapter. All methods and results can be transferred to any type of cooperative scenario with realistic assumptions about the cooperation. In realistic scenarios, some interference channels remain unmeasured and cannot be utilized for cooperation. The interference over these channels has to be regarded as noise with a random variance.

In systems, where the intercell interference is regarded as noise, the supported rate of a mobile device does not relate to the interfering base station. The supported rate depends only on the transmit covariances of the serving base station. The network utility optimization, where all rates of all mobile devices are jointly optimized with all transmit covariances in the system, splits into individual cell sum utility optimizations at each base station, respectively:

$$\max_{\boldsymbol{Q}_{b,k}\succeq\mathbf{0}\ \forall(b,k)\in\mathcal{K}_b} \mathrm{E}\left[\sum_{(b,k)\in\mathcal{K}_b} U\left(r_{b,k}^{\text{achieved}}\right)\right] \quad \text{s.t. } \mathrm{tr}\left(\boldsymbol{Q}_b\right) \leq P, \tag{4.6}$$

where only the rates of the mobile devices associated with the base station are optimized regarding the transmit covariances belonging to the same base station. The optimizations for the different intercell interference robustness methods and utilities mostly boil down to weighted sum rate maximizations with additional constraints on the transmit covariance matrices (See the following Sections and Chapter 5), which can be solved with the algorithms presented in Chapter 3.

The intercell interference variance belonging to the interference over the unknown channels can be described as a random process, where the channels are technically fixed, but the transmit covariance matrices at the interfering base stations need to be regarded as random. Therefore and because the cell sum utility optimization splits into individual optimizations, it is sufficient to look at the signal processing of a single cell. The modeling of the interference over the not measured interference channels as noise with a random variance and the handling of this random process with the following methods does not only simplify the system model, it is the only way to make any realistic interference coordination technique applicable in a large scale cellular system.

The base station has N transmit antennas and serves $K = |\mathcal{K}|$ single-antenna mobile devices, where a mobile device is specified by the index k. The vector $\boldsymbol{h}_k \in \mathbb{C}^N$ is the channel between the antennas of the base station and mobile device k. The channel is generated according to the 3GPP MIMO urban macro cell model with the same parameters

described in Section 2.4.1. $\boldsymbol{Q}$ is the sum transmit covariance matrix of the base station, which is the sum of the individual covariance matrices $\boldsymbol{Q}_k$ for the different mobile devices. The rate of user k can be expressed as

$$r_k = \log\left(1 + \frac{\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{Q}_k \boldsymbol{h}_k}{\sum_{\hat{k}>k} \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{Q}_{\hat{k}} \boldsymbol{h}_k + \theta_k + \sigma_\eta^2}\right), \tag{4.7}$$

where σ_η^2 is the variance of the thermal noise. $\sum_{\hat{k}>k} \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{Q}_{\hat{k}} \boldsymbol{h}_k$ is the variance of the intracell interference with dirty paper coding [60]. θ_k is the intercell interference variance at mobile device k. As described in Section 4.2, θ_k can be approximated by a gamma distribution $\theta_k \sim \Gamma(a_k, b_k)$, where a_k and b_k are derived from the 3GPP MIMO urban macro cell model. It is assumed that the statistics of the intercell interference measured in the past are available to the base station. Please note that the intercell interference statistics strongly depend on the selected transmit strategy and have to be adopted for the different methods, which handle the intercell interference blindness problem.

The expected rate cost function of the single cell system can be formulated similarly to (4.6) as

$$\max_{\boldsymbol{Q}_k \succeq \boldsymbol{0}\ \forall k} \ \mathrm{E}\left[\sum_k U\left(r_k^{\text{achieved}}\right)\right] \quad \text{s.t. } \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P. \tag{4.8}$$

Perfect CSI is assumed, which means pilot contamination and other errors during the channel measurements are neglected. Within the block-fading block length T_{block}, the channels stay constant while the intercell interference θ_k and the transmit covariance matrices $\boldsymbol{Q}_k$ can vary at each time slot.

4.4 Conservative Gambling

One method to deal with the intercell interference mismatch problem is conservative gambling [4]. With the gambling method, the intercell interference variance mismatch is simply accepted. In the first step of conservative gambling, the probability of a successful transmission is assumed to be one and the expectation of the achieved rate is set to the assumed rate. The optimization problem reads as

$$\max_{\boldsymbol{Q}_k \succeq \boldsymbol{0}\ \forall k} \ \sum_k U\left(r_k^{\text{assumed}}\right) \quad \text{s.t. } \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P. \tag{4.9}$$

To reduce the risk of an outage, a conservative link rate adaptation is used. This works well, if the actual intercell interference θ_k^{actual} does not differ too much from the assumed intercell interference $\theta_k^{\text{assumed}}$:

$$r_k^{\text{achieved}} = \begin{cases} r_k^{\text{transmitted}} = (1-\beta) r_k^{\text{assumed}}, & \text{for } r_k^{\text{transmitted}} \leq r_k^{\text{actual}}, \\ 0, & \text{for } r_k^{\text{transmitted}} > r_k^{\text{actual}}. \end{cases} \tag{4.10}$$

The assumed rate depends on the selected assumed intercell interference variance $\theta_k^{\text{assumed}}$. The assumed intercell interference variance $\theta_k^{\text{assumed}}$ can be the expected value of the intercell interference, the intercell interference measured in the last time frame or some predefined

value. To lower the risk of a failed transmission, a common backoff β is introduced and the base stations serve the mobile devices with modest rates $r_k^{\text{transmitted}} = (1-\beta) r_k^{\text{assumed}}$. The transmit covariance matrices $\boldsymbol{Q}_k$ are optimized based on the assumed intercell interference variance $\theta_k^{\text{assumed}}$ and the common backoff β is applied after the transmit covariance matrices $\boldsymbol{Q}_k$ are selected.

The backoff β provides a new degree of freedom in the system and should be chosen such that the objective is maximized:

$$\beta^{\text{opt}} = \underset{0 \leq \beta \leq 1}{\text{argmax}} \ \text{E}\left[\sum_k U\left(r_k^{\text{achieved}}\right) \right]. \tag{4.11}$$

The backoff β is a constant value during the operation of a base station and should be optimized during a training phase or initialization of the base station by evaluating the distribution of the intercell interference variances at the mobile devices. In the simulations, β is optimized in the end. After the transmit covariances and assumed rates for many realizations are calculated, the optimal β is found with a line search.

The backoff reduces the risk of an outage, but it also reduces the possible rates. When the actual intercell interference is smaller than the assumed intercell interference, the mobile devices are only served with the transmitted rates based on the backoff and cannot benefit from the extra resources. Besides, occasionally the conservative link rate adaptation fails completely and some users are still in outage.

Algorithm 4.1 concludes the optimization with the gambling method, where the assumed intercell interference variance is set to its expected value. The backoff factor β^{opt} is given and used to calculate the transmitted rates $r_k^{\text{transmitted}}$. Only the calculation of the optimum transmit covariance matrices $\boldsymbol{Q}_k$ in line 2 depends on the utility function.

Algorithm 4.1 Conservative gambling

Require: β^{opt}, $\text{E}[\theta_k^{\text{actual}}]$ $\forall k$

1: $\theta_k^{\text{assumed}} \leftarrow \text{E}[\theta_k^{\text{actual}}]$
2: compute transmit covariance matrices $\boldsymbol{Q}_k$ according to (4.9) with $\theta_k^{\text{assumed}}$
3: compute r_k^{assumed} according to (4.7) with $\boldsymbol{Q}_k$ and $\theta_k^{\text{assumed}}$
4: $r_k^{\text{transmitted}} \leftarrow (1-\beta^{\text{opt}}) r_k^{\text{assumed}}$
5: compute r_k^{achieved} according to (4.10)

4.5 Conservative Gambling with Individual Backoff Factors

The mobile devices are situated in very different interference situations [42]. Some mobile devices are very close to the base station, experience a strong serving channel and suffer from strong interference from the few collocated base stations at the same site. Other mobile devices sit in the center of the cell and see an interference floor with many comparably weak interferers. The mobile devices at the cell edge have the weakest channels and are disturbed by multiple strong interferers, some of those may even be as strong as the serving channel. Using a common backoff factor β for all mobile devices, as it is done with the conservative gambling method (Section 4.4), does not reflect the different interference situations of the individual mobile devices.

To account for the different situations, it is proposed to use an individual scaling factor for each mobile device, respectively [29, 30]. According to equation (4.10), the outage problem can be written as

$$r_k^{\text{achieved}} = \begin{cases} r_k^{\text{transmitted}} = (1-\beta_k) r_k^{\text{assumed}}, & \text{for } r_k^{\text{transmitted}} \leq r_k^{\text{actual}}, \\ 0, & \text{for } r_k^{\text{transmitted}} > r_k^{\text{actual}}. \end{cases}, \tag{4.12}$$

with individual backoff factors. Selecting each β_k freely can also be seen as selecting each $r_k^{\text{transmitted}}$ freely. In the first step, the transmit covariance matrices are selected based on assumed intercell interference variances and the $r_k^{\text{transmitted}}$ are selected in the second step to maximize the sum utility with fixed transmit covariance matrices. The common backoff factor is selected with a line search to maximize the mean expected rate of all mobile devices (See equation (4.11)). The individual transmitted rate has to maximize the utility of the corresponding achieved rate, respectively:

$$r_k^{\text{transmitted, opt}} = \underset{r_k^{\text{transmitted}} \geq 0}{\operatorname{argmax}} \quad \mathrm{E}\left[U\left(r_k^{\text{achieved}} \right) \right]. \tag{4.13}$$

The individual transmitted rates can be optimized separately and the sum in the expectation can be dropped, as a transmitted rate only influences the utility of the associated achieved rate.

If the probability distribution of the intercell interference variance θ_k^{actual} is known, as assumed in the system model, the probability distribution of the supported rate r_k^{actual} can be derived for fixed transmit covariance matrices. Regarding (4.12), the expectation in (4.13) can be formulated as

$$\begin{aligned} \mathrm{E}\left[U\left(r_k^{\text{achieved}}\right)\right] &= U\left(r_k^{\text{transmitted}}\right) \int_0^{r_k^{\text{transmitted}}} f_{r_k^{\text{actual}}}(r_k)\,\mathrm{d}r_k + U(0) \int_{r_k^{\text{transmitted}}}^{\infty} f_{r_k^{\text{actual}}}(r_k)\,\mathrm{d}r_k \\ &= U\left(r_k^{\text{transmitted}}\right) F_{r_k^{\text{actual}}}(r_k^{\text{transmitted}}) + U(0)\left(1 - F_{r_k^{\text{actual}}}(r_k^{\text{transmitted}})\right) \\ &= \left(U\left(r_k^{\text{transmitted}}\right) - U(0)\right) F_{r_k^{\text{actual}}}(r_k^{\text{transmitted}}) + U(0), \end{aligned} \tag{4.14}$$

where $f_{r_k^{\text{actual}}}(r_k)$ is the probability density function of r_k^{actual} and

$$F_{r_k^{\text{actual}}}(r_k) = P\left(r_k^{\text{transmitted}} \leq r_k^{\text{actual}}\right) \tag{4.15}$$

is the probability, that the transmission is successful, or the cumulative distribution function of the random actual rate evaluated at $r_k^{\text{transmitted}}$. The optimal transmitted rate can be calculated by setting the derivative of the objective in (4.14) with respect to the rate to zero. The derivative can be found as

$$\begin{aligned} \frac{\partial\, \mathrm{E}\left[U\left(r_k^{\text{achieved}}\right)\right]}{\partial r_k^{\text{transmitted}}} =& \frac{\partial U\left(r_k^{\text{transmitted}}\right)}{\partial r_k^{\text{transmitted}}} F_{r_k^{\text{actual}}}\left(r_k^{\text{transmitted}}\right) \\ &+ \left(U\left(r_k^{\text{transmitted}}\right) - U(0)\right) f_{r_k^{\text{actual}}}\left(r_k^{\text{transmitted}}\right). \end{aligned} \tag{4.16}$$

With given transmit covariance matrices, setting (4.16) to zero and solving it for the optimal transmitted rate will have exactly one solution, if $U(r_k)$ and $F_{r_k^{\text{actual}}}(r_k)$ are log-convex functions. In this case, the solution can be found numerically with a bisection or Newton–Raphson method.

The procedure of conservative gambling with individual backoff factors can be seen in Algorithm 4.2. In contrast to Algorithm 4.1 the backoff is not an input variable. It is computed inherently in line 3 for each mobile device, respectively. But, the distribution of the intercell interference variance θ_k^{actual} has to be known, which is assumed here to be a gamma distribution with the shaping and rate parameters a_k and b_k, respectively.

Algorithm 4.2 Individual gambling

Require: a_k, b_k $\forall k$

1: $\theta_k^{\text{assumed}} \leftarrow \mathrm{E}[\theta_k^{\text{actual}}] = a_k/b_k$
2: compute transmit covariance matrices $\boldsymbol{Q}_k$ according to (4.9) with $\theta_k^{\text{assumed}}$
3: compute $r_k^{\text{transmitted}}$ according to (4.16) with $\boldsymbol{Q}_k$
4: compute r_k^{achieved} according to (4.12)

4.6 Expected Rate

The individual gambling makes it possible to choose the transmitted rate $r_k^{\text{transmitted}}$ for each mobile device, which optimizes the performance measure according to the outage probability. But, the optimization of the transmit covariance matrices is still based on some assumed intercell interference variance value $\theta_k^{\text{assumed}}$. In the conservative gambling description from [4], the intercell interference variance realization from a previous time slot is used. Algorithm 4.1 in Section 4.4 has the expectation of the intercell interference variance as input variable. With such a selection of the assumed intercell interference, the transmit covariance matrices are not optimized for the system they are used in, which is clearly suboptimal.

The assumed rate is a function of the transmit covariance matrices and the assumed intercell interference variance. The achieved rate is a function of the transmit covariance matrices, the assumed intercell interference and the probability, that the transmission is successful. Similarly to the optimization of the transmitted rate with individual gambling from the previous section, the expected rate optimization can be formulated as

$$\max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \theta_k^{\text{assumed}} \geq 0\ \forall k}} \mathrm{E}\left[\sum_k U\left(r_k^{\text{achieved}}\right)\right] \quad \text{s.t. } \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P, \tag{4.17}$$

where the optimization is not only over the transmit covariance matrices, but also over the assumed intercell interference variance $\theta_k^{\text{assumed}}$ [30]. Looking at

$$r_k^{\text{achieved}} = \begin{cases} r_k^{\text{assumed}}, & \text{for } \theta_k^{\text{assumed}} \geq \theta_k^{\text{actual}}, \\ 0, & \text{for } \theta_k^{\text{assumed}} < \theta_k^{\text{actual}}, \end{cases} \tag{4.18}$$

the expectation in (4.17) can be rewritten, similarly to (4.14), as

$$\begin{aligned} \mathrm{E}\left[U\left(r_k^{\text{achieved}}\right)\right] &= U\left(r_k^{\text{assumed}}\right) \int_0^{\theta_k^{\text{assumed}}} f_{\theta_k^{\text{actual}}}(\theta_k)\,\mathrm{d}\theta_k + U\left(0\right) \int_{\theta_k^{\text{assumed}}}^{\infty} f_{\theta_k^{\text{actual}}}(\theta_k)\,\mathrm{d}\theta_k \\ &= \left(U\left(r_k^{\text{assumed}}\right) - U(0)\right) F_{\theta_k^{\text{actual}}}(\theta_k^{\text{assumed}}) + U\left(0\right). \end{aligned} \tag{4.19}$$

Here $f_{\theta_k^{\text{actual}}}(\theta_k)$ is the probability density function of θ_k^{actual} and $F_{\theta_k^{\text{actual}}}(\theta_k^{\text{assumed}}) = P\left(\theta_k^{\text{assumed}} \geq \theta_k^{\text{actual}}\right)$ is the probability, that the transmission is successful, or the cumulative distribution function of the random intercell interference variance evaluated at $\theta_k^{\text{assumed}}$. As discussed in the single cell system model Section 4.3, it is assumed that these functions describing the distribution of θ_k^{actual} are known. This leads to the joint optimization

$$\max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \theta_k^{\text{assumed}} \geq 0\ \forall k}} \quad \mathrm{E}\left[\sum_k \left(U\left(\theta_k^{\text{assumed}}\right) - U(0)\right) F_{\theta_k^{\text{actual}}}(\theta_k^{\text{assumed}})\right] \quad \text{s.t. } \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P. \tag{4.20}$$

By considering the expectation of the rates with respect to the intercell interference as optimization problem, the cost function of the optimization and the performance measure become the same. Note, that the utility depends only on the assumed intercell interference variance $\theta_k^{\text{assumed}}$ and not on the actual intercell interference realization θ_k^{actual}.

In the case of only one mobile device associated with the base station, the optimal transmit covariance matrix is independent of the distribution of the intercell interference and has to optimize the utility alone. But, if there are multiple mobile devices associated to the base station, the power distribution among the different transmit covariance matrices depends on the assumed SINRs of the mobile devices, which takes the assumed intercell interference variances into account. Similarly to individual gambling, the optimal assumed intercell interference variances can be found by setting the derivative of equation (4.19) with respect to $\theta_k^{\text{assumed}}$ to zero. The derivative is

$$\begin{aligned}\frac{\partial \,\mathrm{E}\left[U\left(r_k^{\text{achieved}}\right)\right]}{\partial \theta_k^{\text{assumed}}} =& \frac{\partial U\left(r_k^{\text{assumed}}\right)}{\partial r_k^{\text{assumed}}} \frac{\partial r_k^{\text{assumed}}}{\partial \theta_k^{\text{assumed}}} F_{\theta_k^{\text{actual}}}\left(\theta_k^{\text{assumed, opt}}\right) \\ &+ \left(U\left(r_k^{\text{assumed}}\big|_{\theta_k \text{assumed, opt}}\right) - U(0)\right) f_{\theta_k^{\text{actual}}}\left(\theta_k^{\text{assumed, opt}}\right)\end{aligned} \tag{4.21}$$

and depends on the transmit covariance matrices. A joint optimization of the transmit covariance matrices and the assumed interference is intractable. Therefore, an alternating optimization is proposed, which optimizes the transmit covariance matrices $\boldsymbol{Q}_k$ and assumed interference variances $\theta_k^{\text{assumed}}$ in problem (4.20) in turns [30]. In every step, the cost function (4.20) will increase and, as it is bounded, it will converge to a locally optimal point.

For a fixed assumed intercell interference, the probability of a successful transmission is fixed and a weighted utility optimization with respect to the transmit covariances remains. For fixed transmit covariances, the assumed intercell interference can be optimized with a root finding algorithm. The optimization of the assumed intercell interference has to be done in the downlink. The weighted sum rate maximization is done with the uplink-downlink duality as discussed in Chapter 3. Note, that the uplink-downlink transformation depends on the selection of the assumed intercell interference as discussed in Section 3.1 and 3.4. Again, there will be only one maximum for fixed transmit covariance matrices, if $U(\theta_k)$ and $F_{\theta_k^{\text{actual}}}(\theta_k)$ are log-convex functions. Algorithm 4.3 summarizes the expected rate method. The individual assumed intercell interference variances are initially set to their expectation, respectively. Then, the transmit covariance matrices and utility maximizing assumed intercell interference variances are found iteratively in a loop, until some convergence criterion is met. The achieved rate is calculated with (4.18). The mean achieved rate over many intercell interference variance realizations converges to the same value as the expected

rate $U\left(\theta_k^{\text{assumed}}\right) F_{\theta_k^{\text{actual}}}(\theta_k^{\text{assumed}})$ with the used assumed intercell interference variances and transmit covariance matrices.

Algorithm 4.3 Expected rate

Require: a_k, b_k $\forall k$

1: $\theta_k^{\text{assumed}} \leftarrow \mathrm{E}[\theta_k^{\text{actual}}] = a_k/b_k$
2: **repeat**
3: compute transmit covariance matrices $\boldsymbol{Q}_k$ according to (4.9) with fixed $\theta_k^{\text{assumed}}$
4: compute $\theta_k^{\text{assumed}}$ according to (4.21) with fixed $\boldsymbol{Q}_k$
5: compute r_k^{assumed} according to (4.7) based on $\boldsymbol{Q}_k$ and $\theta_k^{\text{assumed}}$
6: **until** the objective (4.17) converges
7: compute r_k^{achieved} according to (4.18)

With the sum rate maximization as utility, the mean of the achieved cell sum rate, the expected cell sum rate, and the supported cell sum rate are plotted in Figure 4.3 over the iterations of Algorithm 4.3. At odd iteration steps the transmit beamforming is optimized and at even iterations steps the optimal assumed intercell interference variance is calculated. The low mobility, large scale cellular system model from Chapter 2 with four mobile devices per cell, $K = 4$, four transmit antennas per base station, $N = 4$, no measured interference channels, $L = 0$, a block length of $T_{\text{block}} = 120$, and a common mobile device device speed of $v = 3\,\text{km/h}$ is utilized. The actually supported mean cell sum rate can only be reached, if the actual intercell interference variance values are known. It can be seen that the expected cell sum rate with the gamma distribution approximation and the actually achieved mean cell sum rate calculated with (4.1) converge to almost the same value in only a few iterations.

The expected rate method optimizes the assumed rates based on an approximated distribution of the intercell interference variance, but the actual intercell interference is based on the transmit covariance matrices of the many other base stations in the system. The used distribution of the intercell interference variance is derived from transmit covariance matrices optimized with the expected rate method at a converged state. In contrast to the single cell system model, the distribution of the intercell interference variance changes from iteration to iteration, as the other base stations update their transmit covariance matrices. Therefore, the mean achieved cell sum rate and the mean expected cell sum rate are not equal in the first few iterations. There is a mismatch between the assumed intercell interference variance distribution and the actual distribution. After a few iterations the distributions match and both cell sum rates converge to almost the same value. The remaining error can be explained with the finite resolution and imprecision of the cumulative distribution functions, which are only approximated through Monte Carlo simulations.

Iteration one in Figure 4.3 can be seen as the conservative gambling from Section 4.4 with backoff $\beta = 0$ and iteration two can be seen as the conservative gambling with individual backoff factors. Compared to conservative gambling with an optimized beta, the achieved cell sum rates with the expected rate method are still almost ten percent higher (See Figure 4.6). The disadvantage of this method is that the statistics of the intercell interference have to be known to the base station. With the expected rate method, the base stations allocate more power to the mobile devices, which are critical for the utility. Therefore, the risk of a sudden drop in the sum rate, if a strong interference at the critical mobile device occurs, will be reduced.

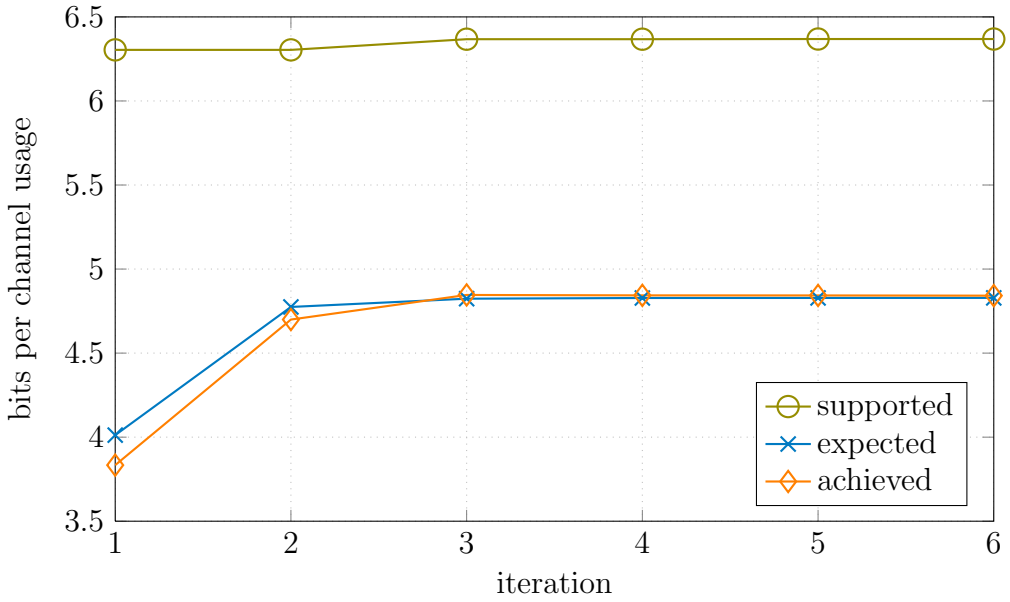


Figure 4.3. Convergence behavior of the cell sum rate with the expected rate mehtod

4.7 Second Pilot

The actual intercell interference variance is unknown when the base stations update their transmit processing. However, the updated supported rates can be made available by the use of a second pilot phase [30, 31]. After the transmit covariances are optimized, based on the measured channels, the base stations transmit a second pilot sequence with the calculated beamforming vectors and the mobile devices can measure and feedback the intercell interference powers. Now, the mobile devices can be served with the supported rates, but the increased overhead decreases the efficiency of the signaling. The discussion in this section assumes a full cellular system as described in Chapter 2, while the problem formulation and algorithm is based on the single cell system model without loss of generality.

The measurement of channel state information for coherent detection differs in *time division duplex* (TDD) and *frequency division duplex* (FDD) [107, 108]. With the second pilot method, it is divided in both systems into two steps. In the first step, the channel vectors and in the second step the intercell interference plus noise are measured. After the first step, the base stations calculate their beamforming according to the measured channels. A measured or assumed intercell interference plus noise variance at this stage can only indicate roughly the supported rates of the mobile devices. The moment the beamforming is applied, the intercell interference plus noise variance and assumed rates are outdated. This effect distorts the transmit covariance matrix optimization and efficient coding can only be applied with a reliable intercell interference plus noise variance and rate.

Therefore, all base stations send orthogonal pilots with the updated transmit covariance matrices in the second step, the second pilot phase. The number of pilot symbols T_{2nd} can be considerably lower compared to the first phase, because the mobile devices have to estimate only a positive, real scalar instead of a complex vector. Now, the mobile devices can measure the actual SINR. The SINR value is subsequently fed back to the associated base station. The associated feedback resources $T_{\text{SINR FB}}$ for the second step are less than for the first step. Keeping the transmit covariance matrices unchanged, the base stations assign proper coding

and modulation schemes and serve the mobile devices with interference aware rates during the data transmission of length T_{data}. T_{block} is the block length defined by the periodicity of channel measurements. In each block the channels are measured and the beamforming is adapted. Under the assumption of block fading, the channels are approximated to be constant during the block length with a coherence time larger or equal to T_{block}. The efficiency of such a scheme is $T_{\text{data}}/T_{\text{block}}$.

In [61], it was shown that the pilot length of an idealized time division duplex system can be used as a lower bound for a frequency division duplex system. In time division duplex systems, the reciprocity of the propagation channels is exploited. The *uplink* (UL) and downlink occupy the same frequency band at orthogonal time slots. In the first step, the channels are measured in the uplink and the gained information is then utilized in the downlink. The mobile devices are split into equally sized subsets $\mathcal{K}_c$. The mobile devices within a subset use pilot sequences, which are orthogonal to each other. But, these pilot sequences are reused in all other subsets. Each base station has to measure at least the channels to its own K mobile devices. The pilot length is at least as large as the number of users in a subset $T_{\text{UL pilots}} \geq |\mathcal{K}_c| \geq K$. For interference coordination it is possible to increase the number of users in a subset and the pilot length to measure additional interference channels. In Figure 4.4 the signaling scheme of time division duplex with the first and second step can be seen.

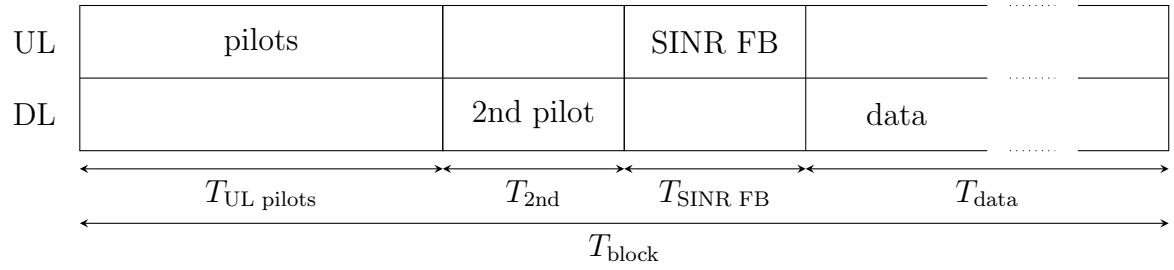


Figure 4.4. TDD signaling with second pilot

The signaling scheme for frequency division duplex can be seen in Figure 4.5. Uplink and downlink are on orthogonal frequency bands and the downlink channel can only be measured at the mobile device, respectively. Then, this information has to be returned to the base station for the optimization of the transmit covariance matrices. In the first step, the base stations are split into subclusters $\mathcal{B}_c$. The base stations within a subcluster use pilots, which are orthogonal to each other, but these pilots are reused in all other subclusters. This is organized in such a way, that each mobile device can measure $C = |\mathcal{B}_c|$ channel vectors. The pilot length has to be at least as large as the number of transmit antennas in a cluster $T_{\text{DL pilots}} \geq NC$. Now, each mobile device could return all these channels, but to keep time division duplex and frequency division duplex comparable, it is assumed that only a limited number of channels is fed back to every base station, respectively. Each mobile device feeds back at least the channel, over which it will be served, to its serving base station. To allow for interference coordination, it is possible that a mobile device feeds also back some interference channels. This additional channel state information is distributed either directly over the air to the interfering base station or via the serving base station and a backhaul network. The limitation to a small number of channels to be fed back reduces the required feedback symbols $T_{\text{chan FB}}$, which are usually the dominant part of this signaling scheme.

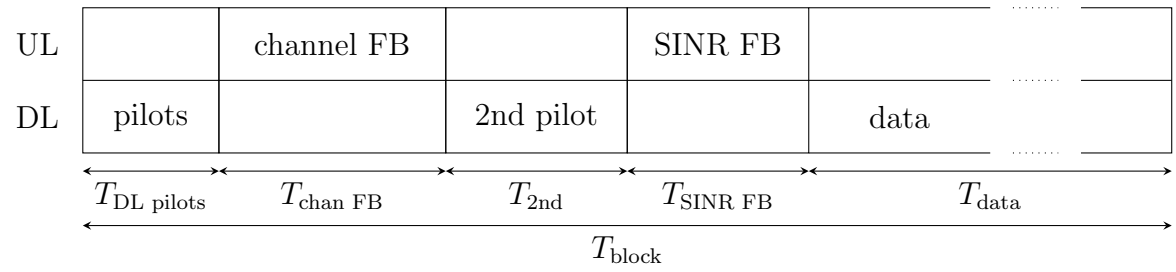


Figure 4.5. FDD signaling with second pilot

The optimization problem of the second pilot method reads the same as the optimization of conservative gambling:

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \sum_k U\left(r_k^{\text{assumed}}\right) \quad \text{s.t. } \operatorname{tr}(\boldsymbol{Q}) \leq P. \tag{4.22}$$

Algorithm 4.4 shows the process of the second pilot method, which starts identically to the conservative gambling method. But, instead of transmitting with reduced rates, the supported rates are measured, or calculated in the simulations, and used for the transmission. Therefore, the achieved rate is equal to the supported rate and there is no risk of an outage.

Algorithm 4.4 Second pilot

Require: $\mathrm{E}[\theta_k^{\text{actual}}]\ \forall k$
1: $\theta_k^{\text{assumed}} \leftarrow \mathrm{E}[\theta_k^{\text{actual}}]$
2: optimize transmit covariance matrices $\boldsymbol{Q}_k$ according to (4.22) with $\theta_k^{\text{assumed}}$
3: measure θ_k^{actual} with second pilot
4: compute $r_k^{\text{transmitted}} = r_k^{\text{actual}}$ according to (4.7) with unchanged $\boldsymbol{Q}_k$ and θ_k^{actual}
5: $r_k^{\text{achieved}} \leftarrow r_k^{\text{transmitted}} = r_k^{\text{actual}}$

The additional second pilot enables the base station to serve the mobile devices with intercell interference-aware rates. However, the increased overhead degrades the signalling efficiency. When there are many interfering base stations, the orthogonal pilot sequences used to distinguish the base stations require a very long pilot phase, which means a large overhead, more communication delay and reduced efficiency. If the pilot sequences are reused in other cells, pilot contamination will reduce the quality of the intercell interference measurements. In this thesis, we do not try to find the exact costs of the second pilot and leave this interesting task for further investigations. But, the second pilot method is used as a reference for performance comparison. In the figures, the performance of this method is plotted, where the negative effect of the second pilot is ignored.

4.8 Genie Assistance

The second pilot is a step towards the upper bound from [4], where the intercell interference variance is simply known through a Genie assistance and not measured. To reach the upper bound, the calculation of the transmit covariance matrices and the measurement of the intercell interference in Algorithm 4.4 should be alternated until convergence is reached. This is done in the loop in Algorithm 4.5, where also the system model from Chapter 2 is

used. For the genie assisted algorithm, the intercell interference variance has to be computed in each iteration and cannot be approximated with a random process. With simulations, it can be shown that the improvement of the utility is negligible after the first iteration.

Algorithm 4.5 Genie assistance

Require:

1: $\boldsymbol{Q}_b \leftarrow \frac{P}{N}\mathbf{I}$
2: get $\theta_{b,k}^{\text{actual}}$ through Genie assistance
3: **repeat**
4: optimize transmit covariance matrices $\boldsymbol{Q}_{b,k}$ according to (4.22) with $\theta_{b,k}^{\text{actual}}$
5: get $\theta_{b,k}^{\text{actual}}$ through Genie assistance
6: **until** convergence is reached
7: compute $r_{b,k}^{\text{actual}}$ according to (4.7) with $\boldsymbol{Q}_{b,k}$ and $\theta_{b,k}^{\text{actual}}$
8: $r_{b,k}^{\text{achieved}} \leftarrow r_{b,k}^{\text{actual}}$

4.9 Covariance Shaping

The uncertainty in the intercell interference can also be eliminated by fixing the transmit covariance matrix as described for the stabilization approach in [4]. If all transmit covariance matrices in the system are fixed, the intercell interference will not change during a block fading block. The mobile devices can be served with interference aware rates, but dropping the adaptive beamforming reduces the possible rates dramatically. The covariance shaping method from Dotzler et al. in [32], is a generalization of the stabilization approach. The robustness to the unpredictable intercell interference is increased by imposing a shaping constraint to the sum transmit covariance, $\sum_k \boldsymbol{Q}_k = \boldsymbol{Q} \preceq \boldsymbol{C}$. If the base station transmits with the maximum power $\text{tr}(\boldsymbol{Q}) = \text{tr}(\boldsymbol{C}) \leq P$, the constraint will be fulfilled with equality. In this case, the sum transmit covariance matrix is fixed to the matrix $\boldsymbol{C}$. But, the individual transmit covariance matrices for the mobile devices can be adapted to maximize the utility. The mobile devices can be served with interference aware rates and adaptive beamforming.

In general, the optimization problem of (4.8) can be reformulated as:

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \ \text{E}\left[\sum_k U\left(r_k^{\text{achieved}}\right)\right] \quad \text{s.t. } \boldsymbol{Q} \preceq \boldsymbol{C},\ \text{tr}\left(\boldsymbol{Q}\right) \leq P. \tag{4.23}$$

With $\boldsymbol{C} = \frac{P}{N}\mathbf{I}$ and at least as many mobile devices as transmit antennas (assuming a full rank joint channel matrix), the base station will always transmit with full power to increase any sum utility $\text{tr}(\boldsymbol{Q}) = \text{tr}(\boldsymbol{C}) = P$ and the shaping constraint has always to be fulfilled with equality $\boldsymbol{Q} = \frac{P}{N}\mathbf{I}$. When the sum transmit covariance matrices of the interfering base stations are fixed in such a way, the intercell interference variance θ_k at mobile device k is known to the base station after the initial measurements of the SINR. The intercell interference variance becomes a fixed value and the expectation operator in the performance measure (4.23) disappears as a result:

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \ \sum_k U\left(r_k^{\text{actual}}\right) \quad \text{s.t. } \boldsymbol{Q} = \frac{P}{N}\mathbf{I}. \tag{4.24}$$

The downlink maximization problem (4.24) can be solved by transforming it to an uplink minimax problem [32]. The uplink minimax optimization is a saddle point problem, which can be solved with an alternating algorithm. The utility maximizing power allocation and the utility minimizing noise realization are found in turns with a joint water spilling algorithm similar to Algorithm 1 in [79] (Chapter 3). The top level Algorithm 4.6 of the covariance shaping is rather simple. But, the optimization of the transmit covariances has become a more complex optimization with the constraint.

Algorithm 4.6 Covariance shaping

Require: $\theta_k^{\text{actual}}\ \forall k$

1: Optimize transmit covariance matrices $\boldsymbol{Q}_k$ according to (4.24) with θ_k^{actual}
2: Compute $r_k^{\text{transmitted}} = r_k^{\text{actual}}$ according to (4.7) with $\boldsymbol{Q}_k$ and θ_k^{actual}
3: $r_k^{\text{achieved}} \leftarrow r_k^{\text{transmitted}} = r_k^{\text{actual}}$

With the covariance shaping method, all base stations restrict their sum transmit covariance matrix to a scaled identity matrix. The uncertainty of the intercell interference variance is eliminated, although the interference channel vectors remain unknown. During T_{block}, the intercell interference variances cannot change when the other base stations update their transmit covariance matrices while fulfilling $\boldsymbol{Q} = \frac{P}{N}\boldsymbol{I}$. The intercell interference is measured before the individual transmit covariance matrices are selected. But, due to the shaping constraint, the intercell interference does not change during the update of these matrices. The rates of disturbed mobile devices do not depend on the optimization of the individual transmit covariance matrices. This motivates again the single cell system model with a random intercell interference variance. Except for the initial SINR measurement, there is no need of measuring the interference channels or any statistics about the intercell interference variance. The cumulative distribution function of the random intercell interference variance becomes a unit step function at the measured intercell interference variance. The problem of interference awareness disappears at the cost of a restriction on the transmit covariances, which reduces the region of achievable date rates.

The individual transmit covariance matrices always have rank one, as they serve single antenna mobile devices. For $\boldsymbol{Q}$ with full rank N, the shaping constraints implies that each eigenvalue should be smaller or equal to $\frac{P}{N}$. If the base station serves only $K < N$ mobile devices, $\boldsymbol{Q}$ will have rank K and only $\frac{K}{N}$ of the transmit power is used. In these cases, $\frac{N}{K}P$ is assigned as the power limit for the covariance shaping method in order to be comparable to the other methods. In the end, the base station uses the transmit power $\frac{K}{N} \cdot \frac{N}{K}P = P$. This also implies, that the constraint on the sum transmit covariance matrix is not fulfilled with equality. The actual intercell interference remains unknown and the expectation operator cannot be dropped. We assume, that the measurements of the SINR are done with the full power sum transmit covariance matrix. Therefore, the maximum intercell interference variance at the mobile devices is known. If the mobile devices are served based on these worst case variances, then there is still no risk of an outage. But, some resources might be wasted, as higher rates could be achieved when a small outage risk is accepted (Section 4.10).

4.10 Loosened Covariance Shaping

By loosening the strict shaping constraint, a controlled instationarity of the intercell interference variance can be introduced. This instationarity can be handled by optimizing the expected rate. With this approach, two different techniques dealing with the intercell interference awareness problem which try to solve the problem with completely different ideas, covariance shaping and the expected rate method, can be combined and a tradeoff can be found.

The combined local optimization reads as

$$\max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \theta_k^{\text{assumed}} \geq 0\ \forall k}} \quad \mathrm{E}\left[\sum_k U\left(r_k^{\text{achieved}}\right)\right] \quad \text{s.t. } \boldsymbol{Q} \preceq \alpha \frac{P}{N}\mathbf{I},\ \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P, \tag{4.25}$$

where $\alpha \geq 1$ loosens the shaping constraint. Note, that the statistics of the random intercell interference variance θ_k^{actual} depend on the shaping constraint. For $\alpha = 1$, the shaping constraint is strict and there is no uncertainty in the intercell interference. For $\alpha \geq N$ the constraint $\sum_k \boldsymbol{Q}_k \succeq \alpha \frac{P}{N}\mathbf{I}$ is not binding and the statistics of the random intercell interference are the same as for the expected rate optimization without any shaping constraint. With the same discussion as in Section 4.6, problem (4.25) can be written as

$$\max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \theta_k^{\text{assumed}} \geq 0\ \forall k}} \quad \mathrm{E}\left[\sum_k \left(U\left(\theta_k^{\text{assumed}}\right) - U(0)\right) F_{\theta_k^{\text{actual}}}(\theta_k^{\text{assumed}})\right] \quad \text{s.t. } \boldsymbol{Q} \preceq \alpha \frac{P}{N}\mathbf{I},\ \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P, \tag{4.26}$$

where the cumulative distribution function $F_{\theta_k^{\text{actual}}}(\theta_k^{\text{assumed}})$ evaluated at $\theta_k^{\text{assumed}}$, i.e., the probability that the transmission is successful for a chosen assumed intercell interference variance, also depends on the loosening factor α all base stations in the system use.

Similarly to the expected rate method, the transmit covariances $\boldsymbol{Q}_k$ and the assumed intercell interference $\theta_{b,k}^{\text{assumed}}$ in problem (4.26) can be optimized with an alternating optimization. For fixed transmit covariances, the derivation of problem (4.26) with respect to the assumed intercell interference variance has to be set to zero, see (4.21). For fixed assumed intercell interference, problem (4.26) is a weighted utility maximization with a loosened shaping. This can be transformed to an uplink problem. The saddle point problem in the uplink can be solved efficiently with a joint scaled gradient descent, which updates the transmit powers q_k and the uplink noise $\boldsymbol{\Omega}$ in parallel. In each step, the updates have to be projected orthogonally onto the constraint set [79] (See Chapter 3).

The complete algorithm, which optimizes the transmit covariances and the assumed intercell interference variances, is listed in Algorithm 4.7. It is essentially the same as the algorithm for the expected rate method, but the transmit covariance matrices are calculated to fulfill the loosened shaping constraint. The algorithm converges, because the cost function improves in every step and the cost function is limited. Convergence is typically reached after three iterations.

For the simulations, the system model from Chapter 2 is used. Every base station has $N = 4$ transmit antennas and serves $K = 4$ mobile devices. For each α, the distributions of the intercell interference variances are approximated based on histograms with many different interfering transmit covariance matrix realizations. The performance of the presented

Algorithm 4.7 Loosened covariance shaping

Require: a_k, b_k $\forall k$, loosening factor α

1: $\theta_k^{\text{assumed}} \leftarrow \mathrm{E}[\theta_k^{\text{actual}}] = a_k/b_k$
2: **repeat**
3: Compute transmit covariance matrices $\boldsymbol{Q}_k$ according to (4.26) with fixed $\theta_k^{\text{assumed}}$
4: Compute $\theta_k^{\text{assumed}}$ according to (4.21) with $\boldsymbol{Q}_k$
5: Compute r_k^{assumed} according to (4.7) with $\boldsymbol{Q}_k$ and $\theta_k^{\text{assumed}}$
6: **until** the objective (4.17) converges
7: Compute r_k^{achieved} according to (4.18)

methods can be seen in Figure 4.6. The normalized average user rate is plotted over the transmit power. The result labeled with "expected" has the rates optimized according to the expected rate algorithm from Section 4.6 without any shaping constraint, while "shaping" is the interference robustness method with the strict shaping constraint from Section 4.9. Both methods yield substantial improvement compared to the conservative link rate adaption algorithm ("gambling") from Section 4.4 with completely different approaches. The loosened covariance shaping presented in this section is labeled "loosened" with a loosening factor of $\alpha = 2.4$. This selection gave the best results at high SNR values. All curves saturate for high power because of the intercell interference. The saturation starts around $P = 1\,\mathrm{W}$. It is assumed that all base stations transmit in the same frequency band, full frequency reuse.

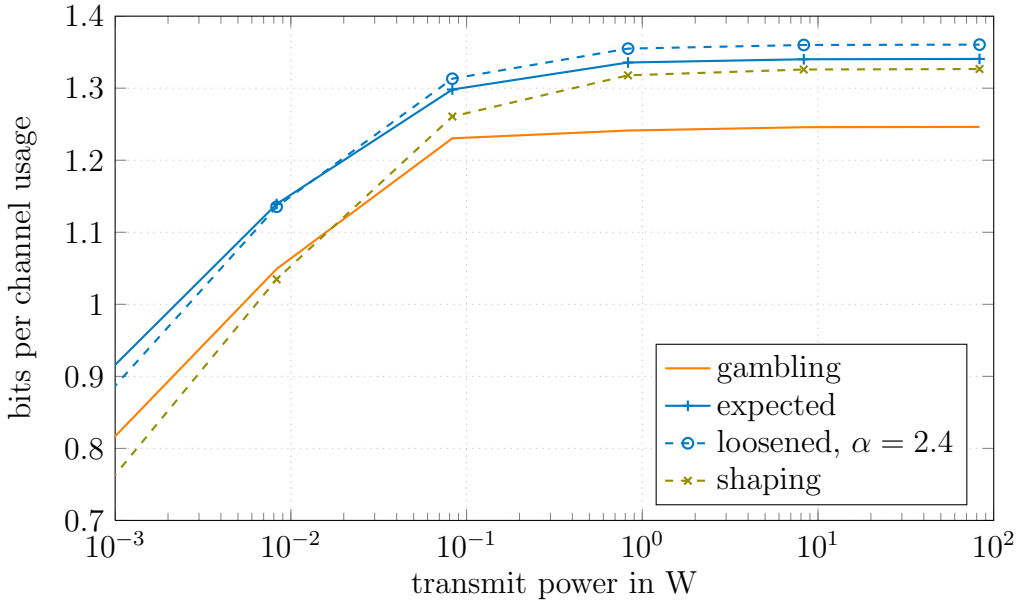


Figure 4.6. Average user rate over transmit power, sum rate maximization

The influence of the loosening on the performance at high SNR values can be seen in Figure 4.7. $\alpha = 1$ and $\alpha = 4$ are the extreme values, where the loosening converges to the scaled identity and the expected rate algorithm, respectively. In low SNR scenarios (Figure 4.8), the effect of the intercell interference vanishes. Therefore, the shaping constraint has no benefit and an unrestricted optimization yields the best results.

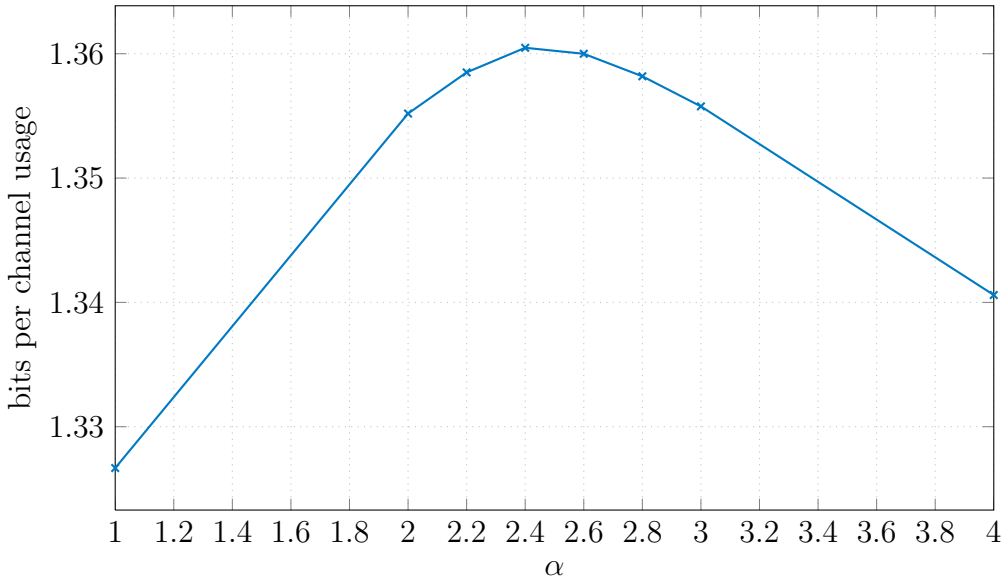


Figure 4.7. Influence ot the loosening at a transmit power of 8.3 W

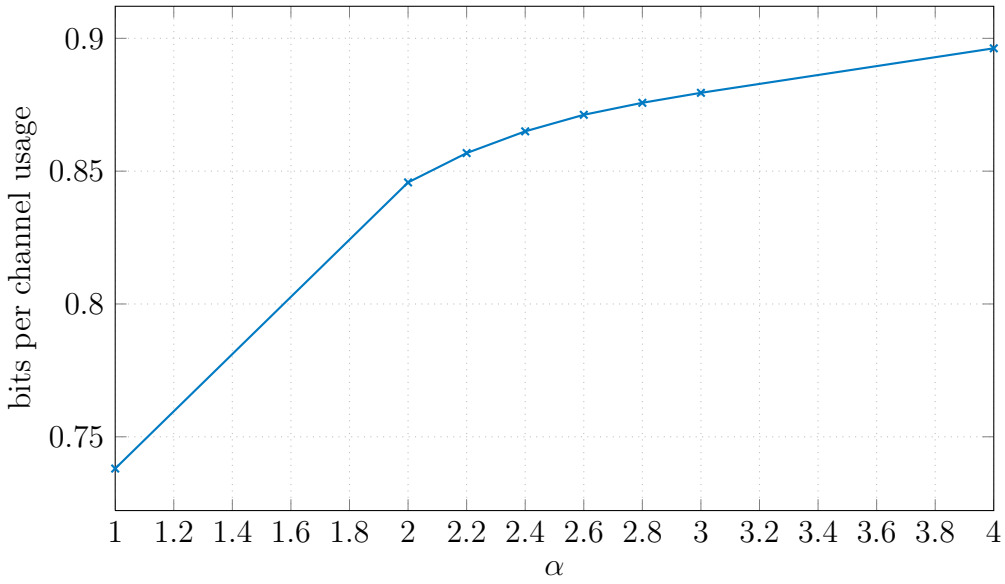


Figure 4.8. Influence ot the loosening at a transmit power of 0.83 mW

The influence of α on the cumulative distribution function of the intercell interference variance is plotted in Figure 4.9 for a single mobile device with a single channel realization. For $\alpha = 1$, the cumulative distribution function is an unit step function at the mean intercell interference as it is expected for the strict covariance shaping. The cumulative distribution functions become flatter for increasing $\alpha \leq 4$. The possible intercell interference values become less predictable.

4.11 Hybrid Automatic Repeat Request

All previously described methods can be combined with *Hybrid Automatic Repeat reQuest* (HARQ) to improve the robustness against the intercell interference blindness prob-

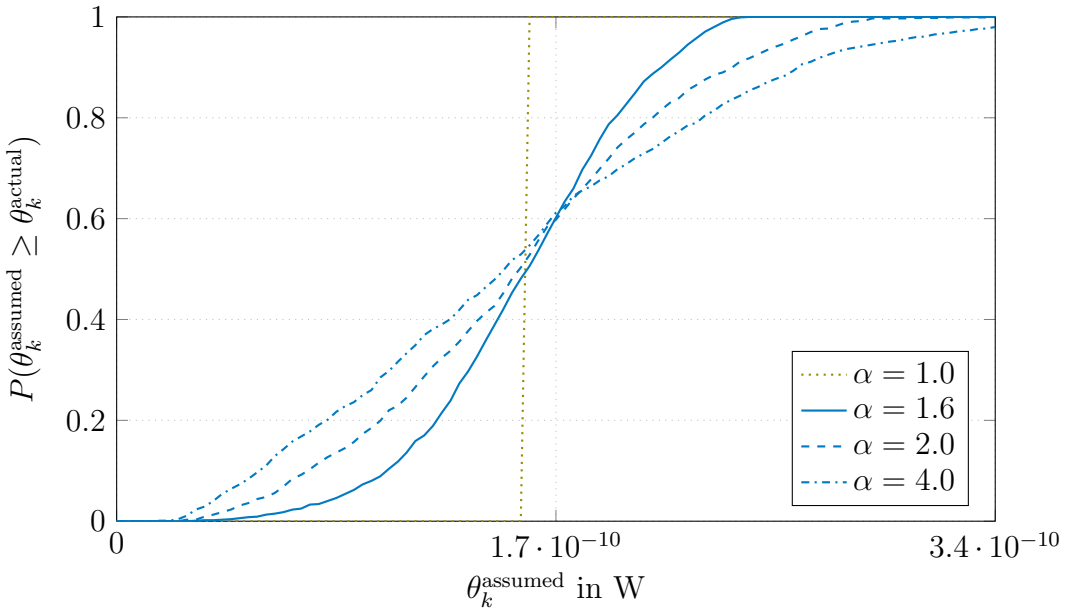


Figure 4.9. Cumulative distribution function of the intercell interference variance

lem [29,33]. The idea of *Automatic Repeat reQuest* (ARQ) is a feedback from the receiver to the transmitter with acknowledgment signals about the success of the decoding. In standard ARQ, redundant bits are added to the original signal using error-detection coding. If the receiver can decode the transmission, it will send an acknowledgment signal in the feedback channel. A receiver, which detects an erroneous message, will request an additional transmission from the sender. Hybrid ARQ is the combination of ARQ with forward error correction coding. Most errors can be corrected with the forward error correction and additional transmissions will only be requested, if unrecoverable errors are detected. In this work, it is still assumed that the link rate adaption is optimal for the assumed SINR and that the forward error correction will correct all errors, if the transmission is coded for a supported rate. The decoding will only fail, if the actual SINR is worse than the assumed SINR. Retransmissions will only be requested, if the actual intercell interference variance is larger than the assumed. The disadvantage of HARQ is the additional overhead and delay caused by the acknowledgment signals and retransmissions.

In typical HARQ implementations, all received transmissions of an HARQ process are stored at the receiver and decoded jointly with code combining. This procedure of soft combining allows the receiver to recover data, which could not be decoded based on any single transmission. Two of the most important combining methods—*Chase Combining* (CC) and *Incremental Redundancy* (IR)—are analyzed in this Section [109–111]. Every retransmission contains the same data bits as the first transmission in a Chase combining-HARQ process. The received transmissions are superimposed with maximum ratio combining. This allows the receiver to decode the data at an SINRs, which is the sum of the SINRs of the individual transmissions [112]. In every transmission in an incremental redundancy-HARQ process, different redundancy versions of the same data bits are transmitted. Every retransmission reduces the initially transmitted rate until the rate matches the combined supported rate of all transmissions [113].

In [114], Mukherjee et al. discuss the problem of intercell interference blindness in LTE and show that it can be mitigated with HARQ. They do not consider the optimization of the transmit processing. In [110,111], the performance of HARQ with Chase Combining and Incremental Redundancy is investigated for wireless standards with link level simulations. Based on such simulations, an abstraction for system level simulations is proposed in [112], where each retransmission is associated with an equivalent SINR gain. The authors of [109] propose several HARQ combining schemes for receivers with multiple antennas in cellular networks.

The maximum number of retransmissions in each HARQ process is set to $T-1$, in order to respect rate or delay requirements. If the receiver cannot decode the data after T transmissions, the data will be discarded and the higher layer will be informed. The utility achieved by mobile device k in an HARQ process is

$$U\left(r_k^{\text{achieved}}\right) = \begin{cases} U\left(r_k^{\text{assumed}}\right), & \text{if decoded within } T \text{ transmissions,} \\ U(0), & \text{else.} \end{cases} \tag{4.27}$$

The intercell interference blindness problem in equation (4.1) is relaxed such, that a transmission will be completed successfully, if the data can be decoded with the combination of all T transmissions. HARQ can be combined with all other methods handling the intercell interference blindness problem. The retransmission process merely increases the probability of a successful transmission at the cost of additional delay.

But, the achieved utility in an HARQ process has to be divided by the number of required transmissions $\hat{T}_k$ of the process to get to the achieved utility per time slot, which is optimized in the following. The according cost function is

$$\Psi^{\text{HARQ}} = \sum_k E_{\boldsymbol{h}_k, \theta_k^{\text{actual}}} \left[\frac{U\left(r_k^{\text{achieved}}\right)}{\hat{T}_k} \right], \tag{4.28}$$

where the expectation is taken over the HARQ processes with respect to the channel realizations $\boldsymbol{h}_k$ and the intercell interference realizations θ_k^{actual}. Based on the renewal-reward theorem [113], the joint expectation can be split in individual expectations of the achieved utility and the required time:

$$\Psi^{\text{HARQ}} = \sum_k \frac{E_{\boldsymbol{h}_k, \theta_k^{\text{actual}}} \left[U\left(r_k^{\text{achieved}}\right) \right]}{E_{\boldsymbol{h}_k, \theta_k^{\text{actual}}} \left[\hat{T}_k \right]}. \tag{4.29}$$

It is assumed that many HARQ processes can be completed during the coherence time T_{block}. It is also assumed that after the T_{block} transmissions of a time frame, all started HARQ processes are completed. This cannot be true in general, but, it only introduces a very small error if the maximum length of an HARQ process will be much smaller than the coherence time, $T \ll T_{\text{block}}$. Under this assumptions, the renewal-reward theorem can be rewritten as the expectation over the block fading blocks of the renewal-reward theorems within one block. With this assumptions, the maximization of the cost function can be done per block:

$$\max \Psi^{\text{HARQ}} = \sum_{k \in \mathcal{K}} \mathrm{E}_{\boldsymbol{h}_k} \left[\max \frac{E_{\theta_k} \left[U\left(r_k^{\text{achieved}}\right) \right]}{E_{\theta_k} \left[\hat{T}_k \right]} \right]. \tag{4.30}$$

All channels stay constant during the coherence time, while the transmit covariance matrices at the interfering base stations vary randomly. Hereby, it is assumed that the HARQ processes for other mobile devices do not introduce correlation to the interference. To support this assumption, it can be argued that multiple HARQ processes are handled in parallel for each mobile device and that the time between retransmissions of the same process varies randomly [29].

4.11.1 Chase Combining

With Chase combining [115], the forward error correction and error detection is wrapped in a repetition code. Every transmission in a Chase combining-HARQ process contains the same bits. With maximum ratio combining at the receiver and if all interference and noise is uncorrelated over different transmit blocks, the effective SINR after t transmissions will be equal to the sum of all individual SINRs of these t transmission [116, Section 14-4].

The HARQ process will be completed successfully, if the sum of all actual SINR values $\gamma_{k,t}^{\text{actual}}$ within this process is not smaller than the assumed SINR $\gamma_k^{\text{assumed}}$:

$$U\left(r_k^{\text{achieved, CC}}\right) = \begin{cases} U\left(r_k^{\text{assumed}}\right), & \text{for } \gamma_k^{\text{assumed}} \leq \sum_{t=1}^{T} \gamma_{k,t}^{\text{actual}} \\ U(0), & \text{for } \gamma_k^{\text{assumed}} > \sum_{t=1}^{T} \gamma_{k,t}^{\text{actual}}. \end{cases} \tag{4.31}$$

Depending on the assumed SINR and the SINR realizations, the number of required transmit blocks for a Chase combining-HARQ process can be found as

$$\hat{T}_k^{\text{CC}} = \begin{cases} 1, & \text{for} \quad \gamma_k^{\text{assumed}} \leq \gamma_{k,1}^{\text{actual}} \\ 2, & \text{for} \quad \gamma_{k,1}^{\text{actual}} < \gamma_k^{\text{assumed}} \leq \gamma_{k,1}^{\text{actual}} + \gamma_{k,2}^{\text{actual}} \\ \vdots & \\ T, & \text{for} \quad \sum_{t=1}^{T-1} \gamma_{k,t}^{\text{actual}} < \gamma_k^{\text{assumed}} \leq \sum_{t=1}^{T} \gamma_{k,t}^{\text{actual}} \\ T, & \text{for} \quad \sum_{t=1}^{T} \gamma_{k,t}^{\text{actual}} < \gamma_k^{\text{assumed}}. \end{cases} \tag{4.32}$$

The probability that the assumed SINR is not larger than the sum over t consecutive SINR values is defined as

$$F_{\gamma_{k,1:t}^{\text{actual}}}(\gamma_k^{\text{assumed}}) = P\left(\gamma_k^{\text{assumed}} \leq \sum_{\hat{t}=1}^{t} \gamma_{k,\hat{t}}^{\text{actual}}\right), \tag{4.33}$$

where $F_{\gamma_{k,0}^{\text{actual}}}(\gamma_k^{\text{assumed}}) = 0$ for any $\gamma_k^{\text{assumed}} > 0$. With equation (4.33), the expectation of (4.31) can be found as

$$E\left[U_k^{\text{CC}}\right] = U\left(r_k^{\text{assumed}}\right) F_{\gamma_{k,1:T}^{\text{actual}}}(\gamma_k^{\text{assumed}}) + U\left(0\right)\left(1 - F_{\gamma_{k,1:T}^{\text{actual}}}(\gamma_k^{\text{assumed}})\right). \tag{4.34}$$

The probability that exactly t transmissions are necessary is

$$F_{\gamma_{k,1:t}^{\text{actual}}}(\gamma_k^{\text{assumed}}) - F_{\gamma_{k,1:t-1}^{\text{actual}}}(\gamma_k^{\text{assumed}}). \tag{4.35}$$

Therefore, the expectation of (4.32) reads as

$$E\left[\hat{T}_k^{\mathrm{CC}}\right] = \sum_{t=1}^{T} t\left(F_{\gamma_{k,1:t}^{\mathrm{actual}}}(\gamma_k^{\mathrm{assumed}}) - F_{\gamma_{k,1:t-1}^{\mathrm{actual}}}(\gamma_k^{\mathrm{assumed}})\right) + T\left(1 - F_{\gamma_{k,1:T}^{\mathrm{actual}}}(\gamma_k^{\mathrm{assumed}})\right)$$
$$= T - \sum_{t=1}^{T-1} F_{\gamma_{k,1:t}^{\mathrm{actual}}}(\gamma_k^{\mathrm{assumed}}). \tag{4.36}$$

The cost function per block fading block with Chase combining-HARQ is

$$\Psi^{\mathrm{CC}} = \sum_{k\in\mathcal{K}} \frac{U\left(r_k^{\mathrm{assumed}}\right) F_{\gamma_{k,1:T}^{\mathrm{actual}}}(\gamma_k^{\mathrm{assumed}}) + U\left(0\right)\left(1 - F_{\gamma_{k,1:T}^{\mathrm{actual}}}(\gamma_k^{\mathrm{assumed}})\right)}{T - \sum_{t=1}^{T-1} F_{\gamma_{k,1:t}^{\mathrm{actual}}}(\gamma_k^{\mathrm{assumed}})}. \tag{4.37}$$

4.11.2 Incremental Redundancy

The data bits are encoded with error detection and a forward error correction code, which adds many redundancy bits. Most of these bits are then punctured to reach the desired code rate. In every transmission of an incremental redundancy-HARQ process different punctured versions are transmitted. Therefore, the code rate changes with every retransmission and the decoding profits from the improved coding gain. If the forward error correction code is infinitely long, infinitely many retransmissions are acceptable, and each transmission contains completely new information, the supported rate of the channel with instationary intercell interference can be reached [29, 114].

The incremental redundancy-HARQ process can be completed successfully in t transmissions, if the assumed rate is not larger than the sum of the individually possible rates within these t transmissions [113]:

$$U\left(r_k^{\mathrm{achieved,\ IR}}\right) = \begin{cases} U\left(r_k^{\mathrm{assumed}}\right), & \text{for } r_k^{\mathrm{assumed}} \leq \sum_{t=1}^{T} r_{k,t}^{\mathrm{actual}} \\ U(0), & \text{for } r_k^{\mathrm{assumed}} > \sum_{t=1}^{T} r_{k,t}^{\mathrm{actual}}. \end{cases} \tag{4.38}$$

Depending on the assumed rate and the actual realizations, the number of required transmit blocks for an incremental redundancy-HARQ process can be found similarly to Chase combining as

$$\hat{T}_k^{\mathrm{IR}} = \begin{cases} 1, & \text{for} \quad r_k^{\mathrm{assumed}} \leq r_{k,1}^{\mathrm{actual}} \\ 2, & \text{for} \quad r_{k,1}^{\mathrm{actual}} < r_k^{\mathrm{assumed}} \leq r_{k,1}^{\mathrm{actual}} + r_{k,2}^{\mathrm{actual}} \\ \vdots & \\ T, & \text{for} \quad \sum_{t=1}^{T-1} r_{k,t}^{\mathrm{actual}} < r_k^{\mathrm{assumed}} \leq \sum_{t=1}^{T} r_{k,t}^{\mathrm{actual}} \\ T, & \text{for} \quad \sum_{t=1}^{T} r_{k,t}^{\mathrm{actual}} < r_k^{\mathrm{assumed}}. \end{cases} \tag{4.39}$$

Following the steps presented for Chase combining ((4.33)–(4.36)), the cost function per block fading block with incremental redundancy-HARQ can be found as

$$\Psi^{\mathrm{IR}} = \sum_{k\in\mathcal{K}} \frac{U\left(r_k^{\mathrm{assumed}}\right) F_{r_{k,1:T}^{\mathrm{actual}}}(r_k^{\mathrm{assumed}}) + U\left(0\right)\left(1 - F_{r_{k,1:T}^{\mathrm{actual}}}(r_k^{\mathrm{assumed}})\right)}{T - \sum_{t=1}^{T-1} F_{r_{k,1:t}^{\mathrm{actual}}}(r_k^{\mathrm{assumed}})}, \tag{4.40}$$

where the probabilities, i.e. cumulative distribution functions, are not with respect to the SINR but the rate itself:

$$F_{r_{k,1:t}^{\text{actual}}}(r_k^{\text{assumed}}) = P\left(r_k^{\text{assumed}} \leq \sum_{\hat{t}=1}^{t} r_{k,\hat{t}}^{\text{actual}}\right). \tag{4.41}$$

In [29, 114], the authors propose to increase the number of allowed transmissions T for HARQ with Incremental Redundancy in LTE. They show that the upperbound rate with known intercell interference values at each time slot, the rate with the Genie or second pilot method, will be reached, if each incremental redundancy-HARQ process runs until successful completion. The argument is as follows: if there is no limit to the number of retransmissions, the supported rate in each time slot will be accumulated until it is larger than or equal to the assumed rate. If the assumed rate is pushed towards infinity and the supported rate at each time slot is bounded, the relative difference between the assumed rate and the accumulated supported rate vanishes asymptotically. This is a loose upper bound, which can only be achieved with an infinite delay and a entirely theoretical forward error correction code.

4.11.3 Cost Function Optimization

The optimization of the cost function (4.37) for Chase combining and (4.40) for incremental redundancy proceeds similarly to the optimization of the cost function as discussed for the expected rate method in Section 4.6. For fixed transmit covariance matrices, the optimal $\gamma_k^{\text{assumed}}$ in (4.37) and the optimal r_k^{assumed} in (4.40) can be found numerically in the downlink. They are associated with an optimal intercell interference, respectively. But, in contrast to the pure expected rate method, here, the weights in the weighted sum utility maximization cannot be fixed for the optimization of the transmit covariance matrices. The transmit covariance matrices are part of the SINR and the rate and, therefore, influence the probability of a successful transmission. The Chase combining cost function for the sum rate utility $U(r) = r$ reads

$$\Psi^{\text{CC}} = \sum_{k\in\mathcal{K}} \frac{F_{\gamma_{k,1:T}^{\text{actual}}}(\gamma_k^{\text{assumed}})}{T - \sum_{t=1}^{T-1} F_{\gamma_{k,1:t}^{\text{actual}}}(\gamma_k^{\text{assumed}})} r_k^{\text{assumed}} = \sum_{k\in\mathcal{K}} w_k r_k^{\text{assumed}}. \tag{4.42}$$

The weighted sum rate maximization from Chapter 3 has to be adapted to solve this problem. With the uplink-downlink duality, the cost function is transformed to the uplink. The weights w_k do not change in this transformation. Algorithm 3.2 contains only updates of the transmit covariance matrices with the sum rate maximization as utility and no further constraints. The transmit covariances in the downlink are found with the primal recovery in equation (3.42) based on the selection of the transmit covariance matrices in the uplink. The transmit covariance matrices in the downlink influence the probability of a successful transmission and, therefore, change the weights. To include this effect in the optimization of the uplink transmit covariance matrices, the gradient in the projected gradient step Algorithm 3.3 is found numerically with respect to the cost function change in the downlink. Also the achieved cost function in each iteration is calculated in the downlink.

The complete alternating optimization, which finds the optimal transmit covariance matrices and the optimal assumed SINR or rate in turns is depicted in Algorithm 4.8. As single antenna mobile devices are assumed, the transmit covariance matrices simplify to

transmit variances q_k, $\forall k$, in the uplink. The update of the downlink transmit covariance matrices is described in lines 7–21. The BC2MAC downlink-uplink transformation is similar to the primal recovery described in Section 3.4.5. In line 9, the numerical gradient with respect to q_k is computed for all k with a very small ϵ. Although, the downlink utility $\Psi(w_{1:K},\ \boldsymbol{Q}_{1:K},\ \theta_{1:K}^{\text{assumed}})$ depends on the weights, downlink transmit covariance matrices, and assumed intercell interferences, the influence of a single changed uplink transmit variance is here denoted by $\Psi(q_k)$ for notation convenience, while all other parameters are fixed. To get to the actual utility with a changed q_k, the downlink covariance matrices have to be calculated with the uplink-downlink transformation (3.42). The assumed SINR and the assumed rate have to be updated and depending on these assumed values, the weights have to be computed as described in equation (4.42). The update of the assumed intercell interference, SINR, and rate and the according weight is described in the lines 22 and 23. The algorithm converges in less than five iterations of the outer loop.

Algorithm 4.8 Expected rate method with HARQ

Require: Intercell interference distribution parameters a_k, b_k $\forall k$, acceptable error ε

1: $d_\mathrm{t} \leftarrow 1$ ▷ initialize inverse step-size
2: $\boldsymbol{Q}_k \leftarrow \frac{P}{NK}\mathbf{I}$, $\forall k$ ▷ initialize DL transmit covariance matrices
3: find optimal $\theta_k^{\text{assumed}}$, $\gamma_k^{\text{assumed}}$, and r_k^{assumed} ▷ maximize (4.37) or (4.40)
4: compute w_k, $\forall k$, with $\gamma_k^{\text{assumed}}$ or r_k^{assumed}, $\forall k$ ▷ (4.42)
5: cost_old $\leftarrow \Psi(w_{1:K},\ \boldsymbol{Q}_{1:K},\ \theta_{1:K}^{\text{assumed}})$ ▷ initial downlink objective (4.37) or (4.40)
6: **repeat**
7: $q_{1:K} \leftarrow \mathrm{BC2MAC}(\boldsymbol{Q}_{1:K},\ \theta_{1:K}^{\text{assumed}})$ ▷ uplink transmit variances
8: **repeat**
9: $g_{\mathrm{t},k} \leftarrow (\Psi(q_k + \epsilon/2) - \Psi(q_k - \epsilon/2))/\epsilon$, $\forall k$ ▷ numerical gradient computation
10: $p_\mathrm{t} \leftarrow P/(\sum_k g_{\mathrm{t},k})$ ▷ preconditioning (3.54)
11: **repeat**
12: $\hat{q}_k \leftarrow q_k + p_\mathrm{t}/d_\mathrm{t} g_{\mathrm{t},k}$, $\forall k$ ▷ unconstr. update (3.46)
13: $\tilde{q}_{1:K} \leftarrow (\hat{q}_{1:K})_\perp$ ▷ joint projection (3.47)
14: cost_new $\leftarrow \Psi(\tilde{q}_{1:K})$ ▷ evaluate objective (4.37) or (4.40)
15: cost_increase $\leftarrow$ cost_new $-$ cost_old
16: $d_\mathrm{t} \leftarrow d_\mathrm{t} + 1$ ▷ decrease step-size
17: **until** cost_increase $\geq -\varepsilon$
18: $q_k \leftarrow \tilde{q}_k$, $\forall k$
19: cost_old $\leftarrow$ cost_new
20: **until** $|\text{cost\_increase}| \leq \varepsilon$
21: $\boldsymbol{Q}_{1:K} \leftarrow \mathrm{MAC2BC}\big(q_{1:K},\ \theta_{1:K}^{\text{assumed}}\big)$ ▷ downlink recovery (3.42)
22: find optimal $\theta_k^{\text{assumed}}$, $\gamma_k^{\text{assumed}}$, and r_k^{assumed} ▷ maximize (4.37) or (4.40)
23: compute w_k, $\forall k$, with $\gamma_k^{\text{assumed}}$ or r_k^{assumed}, $\forall k$ ▷ (4.42)
24: cost_old $\leftarrow \Psi(w_{1:K},\ \boldsymbol{Q}_{1:K},\ \theta_{1:K}^{\text{assumed}})$ ▷ evaluate objective (4.37) or (4.40)
25: **until** objective (4.37) or (4.40) converges
26: **return** $r_k^{\text{achieved}}/\hat{T}_k$ ▷ (4.31) and (4.32) or (4.38) and (4.39)

4.11.4 Simulations

The system model from Chapter 2 is used for the simulations. Every base station has $N = 4$ transmit antennas, serves $K = 4$ single antenna mobile devices, and $T = 4$ transmissions are allowed for each HARQ process. It is assumed that the acknowledgment or negative acknowledgment feedback is delay-free and error-free. Instead of a probability distribution for the intercell interference variance, SINR, or rate, a histogram of realizations is utilized for the expected rate method in the optimizations. The normalized average mobile device rate is plotted over the transmit power in Figure 4.10for the maximum sum rate utility. All curves saturate for high power because of the intercell interference limitation. The saturation starts already at 1 W as it is assumed that all base stations transmit in the same frequency band. The "2nd pilot" rate is the upper bound, which can only be achieved, if the intercell interference is known at the transmitter. The rates with the expected rate method are denoted "expected" without ARQ, as "expected-CC" with Chase combining and "expected-IR" with incremental redundancy. It can be seen that incremental redundancy performs better than Chase combining and no ARQ. The covariance shaping method with a scaled identity from [32] (Section 4.9) is plotted as "shaping" and the conservative gambling algorithm from [4] (Section 4.4) with an optimized backoff factor as "gambling". The gambling algorithm can also be improved with HARQ. The according results are omitted to improve the readability.

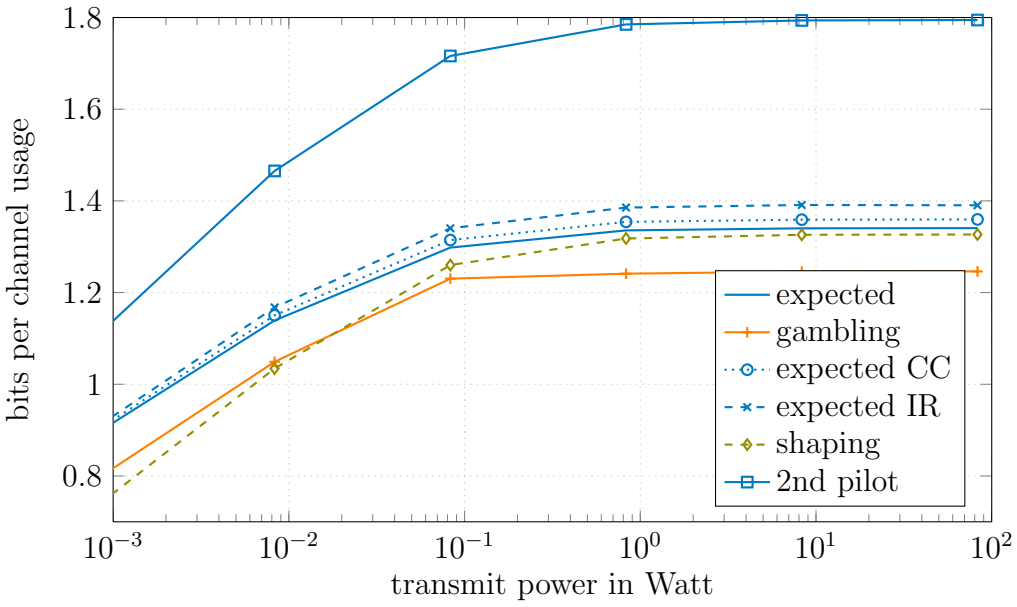


Figure 4.10. Combination of the expected rate method with HARQ, average mobile device rate

4.12 Conclusion

The intercell interference blindness problem arises in all cellular systems for the part of the intercell interference, which has to be regarded as noise. Even if all channels in the system are constant, this part of the intercell interference variance at a mobile devices changes whenever a base station in the system changes its transmit strategy. A simple single cell system model,

where the intercell interference variances are drawn from a gamma distribution, could be derived to simplify the analysis.

The intercell interference blindness problem can be tackled with many different approaches. The simplest approach and baseline for this work is to transmit at lower rates with the conservative gambling method. It is possible to win knowledge about the actual intercell interference variance by measuring it with a second pilot. This is an approach that gets very close to the upper bound where knowledge about the intercell interference variance is gained through an omniscient genie. Nevertheless, due to the increased signaling overhead, the second pilot is not a desirable technique.

To make the system robust against unpredictable changes in the intercell interference variance, the expected rate method takes the actual performance measure, the expectation of the rate with respect to the instationary intercell interference, as cost function. The covariance shaping method tackles the intercell interference blindness problem from a completely different direction by stabilizing the intercell interference variances. These two approaches notably outperform conservative gambling and can be combined to the loosened covariance shaping method, which finds the best compromise between stabilizing the intercell interference variance at mobile devices in other cells and optimizing the expected rates of the own mobile devices.

HARQ with soft combining can also be used to mitigate the disadvantage of intercell interference instationarity. This effect of HARQ could be included into the optimization of the transmit covariance matrices leading to even better results with the expected rate method.

5. Intercell Interference Robustness in Fairness Optimizations

In the previous chapter, the intercell interference blindness problem has been addressed for a general utility U or the *sum rate* (SR) maximization. Maximizing the sum throughput of the system usually ends up with serving only the mobile devices with high SINR. Provider and users of a cellular network are typically interested in a fair resource distribution. The focus of this chapter is on the intercell interference blindness problem in fairness optimizations.

A contemporary overview of different fairness utilities in general and their treatment can be found in [117, Chapter 5] and [118–122]. In [96], Neely et al. investigated the fairness optimization problem with intercell interference robustness from a networking point of view with generalized links between different nodes but regardless of the transmit covariance selection. In the asymptotic region, the transmit covariance matrix selection with fairness utilities was addressed by Huh et al. in [97] with fixed intercell interference variances but varying channel coefficients. A similar problem, where the uncertainty in the SINR is caused by non-perfect channel state knownledge, was approached by Shirani-Mehr et al. in [98] with fairness scheduling and an heuristic approach for the robustness. The problem of intercell interference blindness in systems with fairness scheduling in combination with HARQ was investigated by Shirani-Mehr et al. in [29], where conservative gambling with individual back-off factors was employed as intercell interference robustness method. Recently, the expected rate method was used by Fritzsche et al. in [92] to solve the intercell interference blindness problem with proportional fairness as utility. In the light of intercell interference blindness, Ellenbeck et al. addressed the problem of selecting unitary precoding matrices in a system with proportional-fair scheduling and HARQ [100]. Unitary precoding matrices can be seen as a form of covariance shaping.

The fairness among the mobile devices can be realized with scheduling. With fairness scheduling, the transmit covariance matrices are still selected to maximize a weighted sum rate, but the scheduler changes the weights of the mobile devices from time slot to time slot according to a fairness criterion. In Section 5.2, *Round Robin* (RR), *Throughput-Fair* (TF) scheduling, and *Proportional-Fair* (PF) scheduling are implemented with a fairness scheduler. With round robin, the mobile devices are served in turn. Each mobile device gets the same number of time slots assigned. Under the throughput fairness criterion, the mobile devices with the smallest historical throughput are scheduled, which leads to equal throughput for each mobile device in the end. The goal of proportional fairness is to find a balance point between the sum throughput maximization and equal throughput among all mobile devices [123].

Fairness beamforming does not rely by design on a scheduler. The idea is to find covariance matrices, which serve the mobile devices inherently with rates that optimize the utility. This is done by splitting the utility optimization into a weighted sum rate optimization and an optimization of the weights [119]. In the optimal point, the weights are such that the weighted sum rate optimization selects utility maximizing rates. As the rate region has to be convexified with time sharing in general, multiple different transmit covariance matrix selections at different time slots might be necessary to implement the optimal rates. But, the transmit covariance matrices have to be switched very rarely compared to fairness scheduling. In Section 5.3, *max-min* fairness (MM), *log fairness* (LF), and *proportional fairness* (PF) are discussed for fairness beamforming. The discussion is similarly to [97, 124–126], but for the intercell interference robustness methods from Chapter 4. Parts of this chapter can be found in [101, 127].

5.1 System Model

Without loss of generality, the MU-MISO single cell system model with an intercell interference variance following a Gamma distribution from Section 4.3 is employed for the introduction of the fairness algorithms. The simulations in Section 5.2.1 – 5.2.3 are done with an MU-MIMO system model with four mobile devices and four transmit antennas and without intercell interference to give better insights to the described algorithms. In the simulations Section 5.4, the simulations are done with both, the single cell system model from Section 4.3 and the full cellular system model from Section 2.1. With the single cell system model, the behavior of the instationary intercell interference can be simulated with low complexity.

For fairness scheduling as well as for fairness beamforming, it is assumed that there are two different schedulers. The mobile device set scheduler and the transmit covariance matrix scheduler. The mobile device set scheduler assigns a set of mobile devices $\mathcal{K}$ and an utility function $U(\bullet)$ to a frame of resources. This frame of resources can be split into many subframes of resources. The transmit covariance matrix scheduler is responsible for selecting transmit covariance matrices for each of these subframes with the goal of optimizing the imposed utility for the given mobile device set. With the block fading assumption, the channel can be seen as constant during a resource block, which is defined by the coherence time and frequency. The number of symbols, which can be transmitted during such a resource block is denoted by T_{block}. For simplicity, the time and frequency dimensions are unified to a single dimension, which is denoted as time in the following. For the discussion of the algorithms, it is assumed that T_{block} is very large, which is valid in scenarios with low mobile device mobility. With this assumption, a frame can be set to a single resource block and the mobile device set scheduler updates its selection for each resource block. The resource block is then split into many slots, which still consist of many symbols. Therefore, the transmit covariance matrix scheduler works on many slots with equal channels. But, the intercell interference variance can still change unpredictably from slot to slot. In systems with shorter resource block lengths T_{block}, a frame has to consist of many resource blocks and the transmit covariance matrix scheduler has to work with channels that change from block to block.

5.2 Fairness Scheduling

Fairness scheduling is implemented with the maximum weighted sum rate framework described in Chapter 3. For fairness scheduling, the weights for the mobile devices are called priority coefficients. The selection of the mobile device priority is based on a compromise between the maximum possible sum rate and a fair resource distribution. Mobile devices with a higher weight are preferred in the resource distribution compared to mobile devices with smaller weights. By setting its weight to zero, a mobile devices is removed from the active mobile device set. The transmit covariance matrices are still selected to maximize a weighted sum rate, but the scheduler changes the weights of the mobile devices from slot to slot according to a fairness criterion.

The general objective function of a fairness scheduler, which selects the set $\mathcal{K}^{(t)}$ of mobile devices that are scheduled by the base station in slot t is [123]

$$\max_{\mathcal{K}^{(t)}} \; \mathrm{E}\left[\sum_{k\in\mathcal{K}^{(t)}} \frac{\left(r_k^{(t)}\right)^{\alpha}}{\left(R_k^{(t)}\right)^{\beta}}\right] \qquad \text{s.t.} \quad |\mathcal{K}^{(t)}| = \hat{K}. \tag{5.1}$$

Here, $r_k^{(t)}$ is the data rate potentially achievable by mobile device k in slot t and

$$R_k^{(t)} = \frac{1}{t-1}\sum_{\tau=1}^{t-1} r_k^{(\tau)} \tag{5.2}$$

is the historical average rate of mobile device k in the current block. The design parameter $\hat{K}$ is the cardinality of $\mathcal{K}^{(t)}$. α and β are the factors that tune the fairness of the scheduler.

In general, the parameters α and β can be set to any value between zero and one to design a meaningful cost function. Here, only the values of α and β with the setup from [123] as shown in Table 5.1 are investigated. The scheduler does not care about the channel quality of the mobile devices and serves them in turn with round robin. For the sum rate utility, the scheduler has to look only at the instantaneous possible rates and for throughput fairness, the scheduler ignores the channel quality and serves the mobile devices with the smallest historical throughput. Proportional fairness finds a compromise between maximizing the sum rate and serving the mobile device with the smallest historical throughput. Typically, only one mobile device is served per time slot and $\hat{K}$ is set to one. Here, multiple mobile devices are allowed per time slot, $\hat{K} > 1$, and the selected mobile devices are served with a weighted rum rate maximization. Note, that this can change the utility.

After the scheduler selected the active mobile device set $\mathcal{K}^{(t)}$, the optimization of the transmit covariance matrices can be derived from equation (5.1) for the different fairness utilities with the selection of the coefficients α, β, and the active mobile device set $\mathcal{K}^{(t)}$ as described in Table 5.2. The possible data rates $r_k^{(t)}$ always have to be included in the objective function for the selection of the transmit covariance matrices ($\alpha = 1$). The resulting weighted sum rate maximization for all fairness utilities has the form

$$\max_{\boldsymbol{Q}_k\succeq\mathbf{0}\;\forall k\in\mathcal{K}^{(t)}} \; \mathrm{E}\left[\sum_{k\in\mathcal{K}^{(t)}} w_k^{(t)} r_k^{(t)}\right] \qquad \text{s.t.} \quad \boldsymbol{Q}\in\mathcal{C}, \tag{5.3}$$

where the weights are $w_k^{(t)} = \left(R_k^{(t)}\right)^{-\beta}$, $\forall k, t$. Instead of selecting the active device set $\mathcal{K}^{(t)}$, the scheduler can also select the weights in such a way, that active mobile devices get the weight assigned as described and the weight of not active mobile devices is set to zero. The expectation operator in equation (5.3) takes the instationary intercell interference into account and the problem can be solved with the different strategies for intercell interference robustness described in Chapter 4. The constraint set $\mathcal{C}$ describes in most cases the sum power constraint, the covariance shaping constraint or a combination of both. It is also possible to use the constraint set for interference temperatures or per antenna power constraints. Problem (5.1) and (5.3) can jointly be written for the sum rate maximization as

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \in \mathcal{K}} \ \mathrm{E}\left[\sum_{k \in \mathcal{K}} r_k^{(t)}\right] \qquad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{C}, \tag{5.4}$$

where all weights are set to one and all mobile devices are active all the time.

	RR	SR	TF	PF
α	0	1	0	1
β	0	0	1	1

Table 5.1. α and β with single mobile device per slot

	RR	SR	TF	PF
α	1	1	1	1
β	0	0	1	1
$\mathcal{K}^{(t)}$	$\mathcal{K}_{\mathrm{RR}}^{(t)}$	$\mathcal{K}$	$\mathcal{K}_{\mathrm{TF}}^{(t)}$	$\mathcal{K}$

Table 5.2. α, β, and $\mathcal{K}^{(t)}$ with multiple mobile devices per slot

The general idea is shown in Algorithm 5.1, where the transmit covariance matrix optimization is always done with respect to all mobile devices and the active mobile devices are selected through non-zero weights. The weights assigned by the fairness scheduler $w_k^{(t)}$ can be $w_{\mathrm{RR},k}^{(t)}$, $w_{\mathrm{TF},k}^{(t)}$ or $w_{\mathrm{PF},k}^{(t)}$ according to which fairness criterion is adopted. After computing the weights, the rates $r_k^{(t)}$, $\forall k$, are calculated by solving the maximum weighted sum rate problem (5.3) with different intercell interference robustness methods. The time-varying property of fairness scheduling is shown by the time counter t and the update of the historical average rate $R_k^{(t)}$, $\forall k$. Fairness scheduling can only realize fairness over many time slots.

Algorithm 5.1 Fairness scheduling and intercell interference robustness

Require: assigned user set $\mathcal{K}$, T_{block}

1: $R_k^{(1)} \leftarrow 1\ \forall k$ ▷ initialize historical throughput
2: **for** $t = 1$ **to** T_{block} **do**
3: $\quad w_{1:K}^{(t)} \leftarrow \mathrm{Scheduler}\left(R_{1:K}^{(t)}\right)$ ▷ select weights (5.6), (5.8), (5.10)
4: $\quad r_{1:K}^{(t)} \leftarrow \mathrm{RobustSR}\left(w_{1:K}^{(t)}\right)$ ▷ find robust sum rate (5.3)
5: $\quad R_k^{(t+1)} \leftarrow \frac{1}{t-1} \sum_{\tau=1}^{t-1} r_k^{(\tau)}$, $\forall k$ ▷ update historical throughput (5.2)
6: **end for**

If only a subset of mobile devices are in the active set, it might happen that less mobile devices will be served than the base station has transmit antennas. In this case, the sum

transmit covariance matrix will be rank deficient in scenarios with single antenna receivers. Covariance shaping, where the sum transmit covariance matrix is restricted to a scaled identity matrix, cannot transmit with full power. Each eigenvalue is limited to a fraction of the sum transmit power. If only a few eigenvalues are positive, the used power will still be only a fraction of the total allowed power. To keep the covariance shaping method comparable to the other methods, the scaling of the shaping matrix is increased such, that the full transmit power can be used with the active mobile device set. Only the worst intercell interference variance at the mobile devices can be known in this case. The actual intercell interference will be smaller, which motivated the use of loosened covariance shaping.

5.2.1 Round Robin

With round robin, the scheduler assigns the same amount of resources to all mobile devices. This resource fairness can be achieved by serving the mobile devices in turn. Typically, only a single mobile device is served in each slot with round robin. Here, round robin is relaxed in order to allow serving multiple mobile devices in each slot. In the simulations, two mobile devices, rather than one, are served in each slot, so the multi-user diversity can still be exploited. The two mobile devices are served in such a way, that the sum rate is maximized. An example of this scheduling technique can be seen in Figure 5.1. Four mobile devices are assigned to the block and exactly two mobile devices have non-zero rates at each time instance. Mobile devices with a bad channel will always suffer, if they are paired with a mobile device with a good channel.

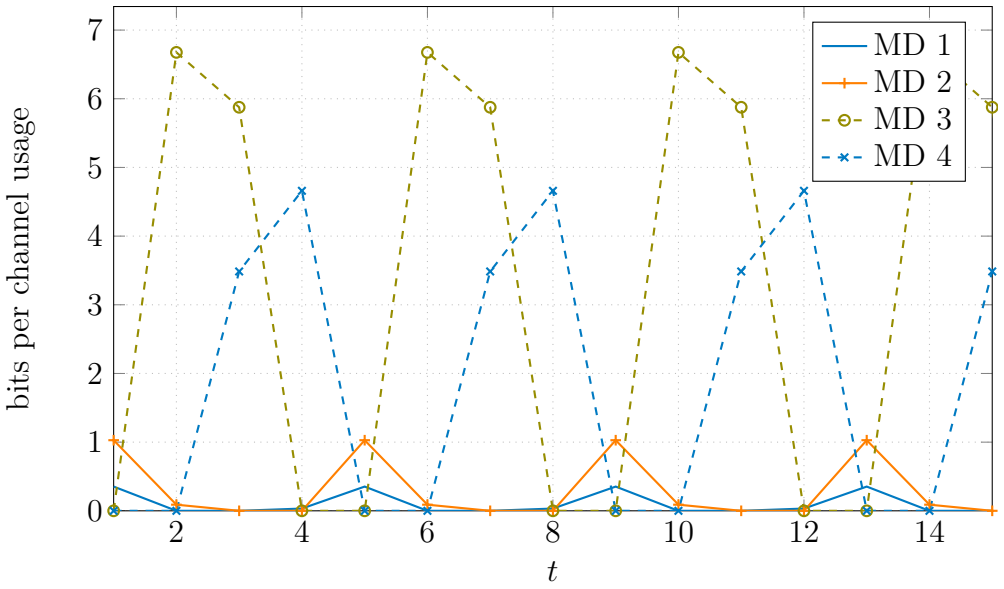


Figure 5.1. Instantaneous mobile device rates with round robin

The objective function,

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \in \mathcal{K}_{\mathrm{RR}}^{(t)}} \mathrm{E}\left[\sum_{k \in \mathcal{K}_{\mathrm{RR}}^{(t)}} r_k^{(t)}\right] \quad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{C}, \tag{5.5}$$

is similar to that of the sum rate maximization (5.4), except that the active set $\mathcal{K}_{\mathrm{RR}}^{(t)}$ is changing from slot to slot as shown in Table 5.3 for two active mobile devices per slot. This can also be reached by plugging the weights,

$$w_{\mathrm{RR},k}^{(t)} = \begin{cases} 1 & \text{for } k \in \mathcal{K}_{\mathrm{RR}}^{(t)}, \\ 0 & \text{else}, \end{cases} \tag{5.6}$$

into optimization (5.3) and maximizing over the complete mobile device set $\mathcal{K}$.

slot number (t)	(1)	(2)	(3)	(4)	(5)	$\cdots$
active MD set $\mathcal{K}_{\mathrm{RR}}^{(t)}$	$\{1,\ 2\}$	$\{2,\ 3\}$	$\{3,\ 4\}$	$\{4,\ 1\}$	$\{1,\ 2\}$	$\cdots$
weight of MD 1, $w_{\mathrm{RR},1}^{(t)}$	1	0	0	1	1	$\cdots$
weight of MD 2, $w_{\mathrm{RR},2}^{(t)}$	1	1	0	0	1	$\cdots$
weight of MD 3, $w_{\mathrm{RR},3}^{(t)}$	0	1	1	0	0	$\cdots$
weight of MD 4, $w_{\mathrm{RR},4}^{(t)}$	0	0	1	1	0	$\cdots$

Table 5.3. Active mobile device set with round robin $\mathcal{K}_{\mathrm{RR}}^{(t)}$

5.2.2 Throughput Fairness

The throughput-fair scheduler always serves the mobile device(s) with the smallest historical throughput. With a long enough block length T_{block}, the rates of all mobile devices converge to the same throughput with this strategy. To achieve this goal, mobile devices with a poor channel quality are scheduled more often compared to mobile devices with strong channels. Again, the original throughput-fair scheduler only serves a single mobile device per slot and, here, multiple mobile devices are served in each slot to exploit the multi-user diversity. The mobile devices, which are scheduled, are served according to the proportional-fair optimization (see Section 5.2.3). In slot (t), $\mathcal{K}_{\mathrm{TF}}^{(t)}$ is the set of scheduled mobile devices according to their relatively small historical throughput $R_k^{(t)}$. With two active mobile devices per slot, the set can be found as

$$\mathcal{K}_{\mathrm{TF}}^{(t)} = \left\{ k_1 = \underset{k \in \mathcal{K}}{\operatorname{argmin}} \left(R_k^{(t)} \right),\ k_2 = \underset{k \in \mathcal{K} \setminus k_1}{\operatorname{argmin}} \left(R_k^{(t)} \right) \right\}. \tag{5.7}$$

The weights for the maximum weighted sum rate problem (5.3) are

$$w_{\mathrm{TF},k}^{(t)} = \begin{cases} \frac{1}{R_k^{(t)}} & \text{if } k \in \mathcal{K}_{\mathrm{TF}}^{(t)}, \\ 0 & \text{else}, \end{cases} \tag{5.8}$$

resulting in the objective function

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \in \mathcal{K}_{\mathrm{TF}}^{(t)}} \mathrm{E}\left[\sum_{k \in \mathcal{K}_{\mathrm{TF}}^{(t)}} \frac{1}{R_k^{(t)}} r_k^{(t)} \right] \qquad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{C}. \tag{5.9}$$

The example rates in Figure 5.2 and 5.3 are qualitative results to give an insight to throughput fairness. Four mobile devices are assigned and two mobile devices are active per slot. With single-user throughput fairness, only one mobile device is scheduled in each slot and the rates of the mobile devices converge to exactly the same throughput. However, with multiple mobile devices per slot, the mobile device with the weakest channel is always paired with at least one other mobile device. Therefore, the mobile device with the worst channel will always achieve a smaller throughput than the rest of the mobile devices. Serving multiple mobile devices per slot violates the targeted fairness criterion. But, the multi-user diversity increases the possible rates and makes the system more robust to the case that all mobile devices are starving because one mobile device is not able to communicate at all. Additionally, serving multiple mobile devices per slot increases the rank of the sum transmit covariance matrix, which is important for optimizations with a shaping constraint.

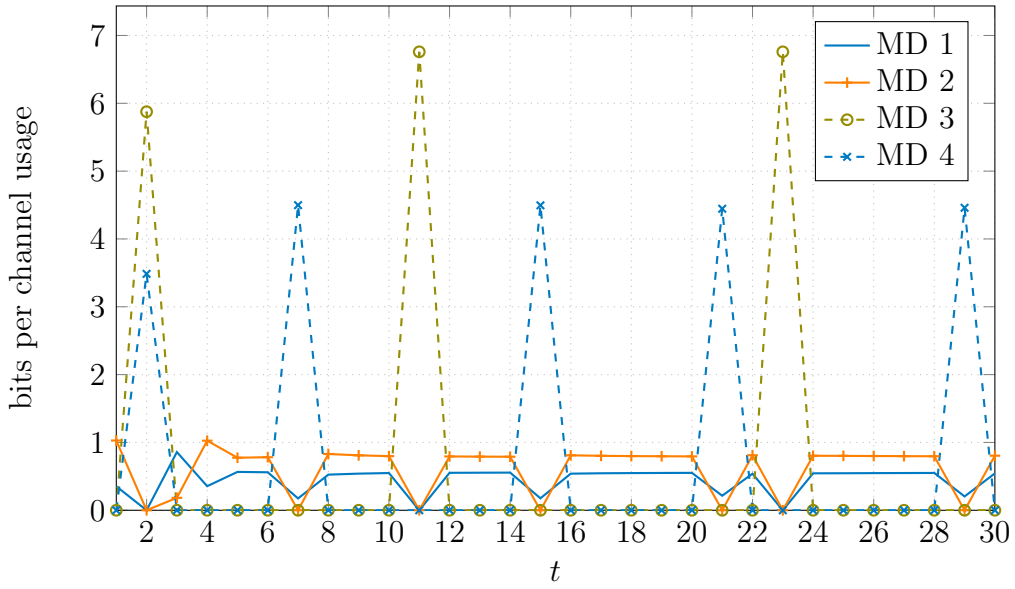


Figure 5.2. Instantaneous mobile device rates with throughput fairness

5.2.3 Proportional Fairness

The goal of proportional fairness is to find a balance between maximum system throughput and equal throughput for all mobile devices. With the weights,

$$w_{\mathrm{PF},k}^{(t)} = \frac{1}{R_k^{(t)}} \quad \forall k, t, \tag{5.10}$$

the cost function is

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \in \mathcal{K}} \ \mathrm{E}\left[\sum_{k \in \mathcal{K}^{(t)}} \frac{1}{R_k^{(t)}} r_k^{(t)}\right] \qquad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{C}. \tag{5.11}$$

The whole set $\mathcal{K}$ of the mobile devices is scheduled in all slots. In the end of T_{block}, mobile devices with poor channels will have a smaller throughput compared to the throughput they

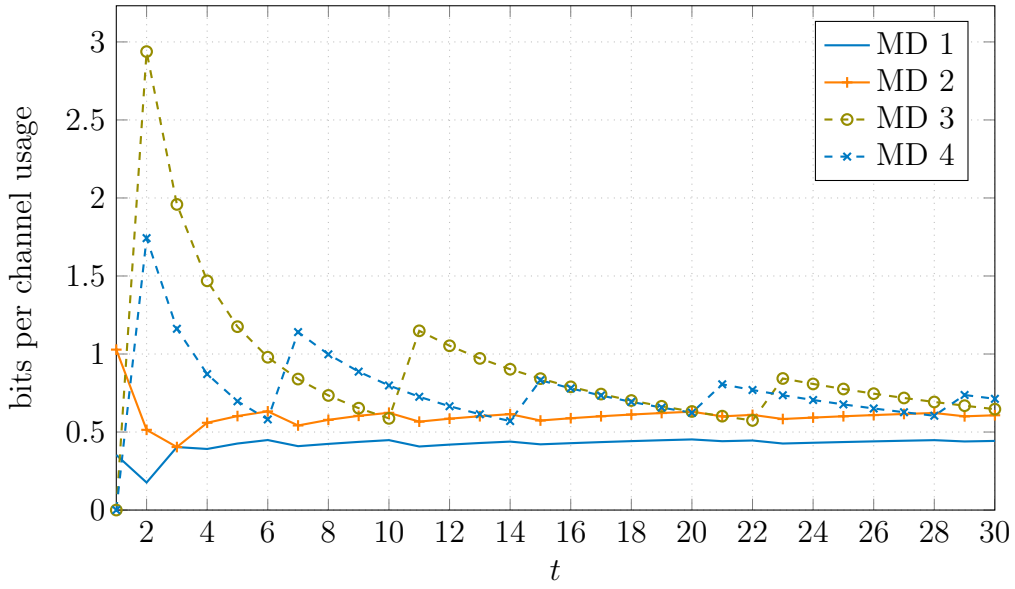


Figure 5.3. Time averaged mobile device rates with throughput fairness

would have with throughput fairness, but they are served more frequently than in the case of the maximum sum rate optimization.

The performance of proportional fairness is shown in Figures 5.4 and 5.5. Several time slots after the initialization, the system converges to the stable state where the mobile devices in the cell are served with the optimum weights $w_k = (r_k)^{-1}$. The instantaneous and time averaged mobile device rates converge to the same value. In general, the instantaneous mobile device rates do not converge and may jump between different realizations. Time-sharing between different covariance matrix realizations might be necessary to achieve the goal of the proportional-fair scheduler.

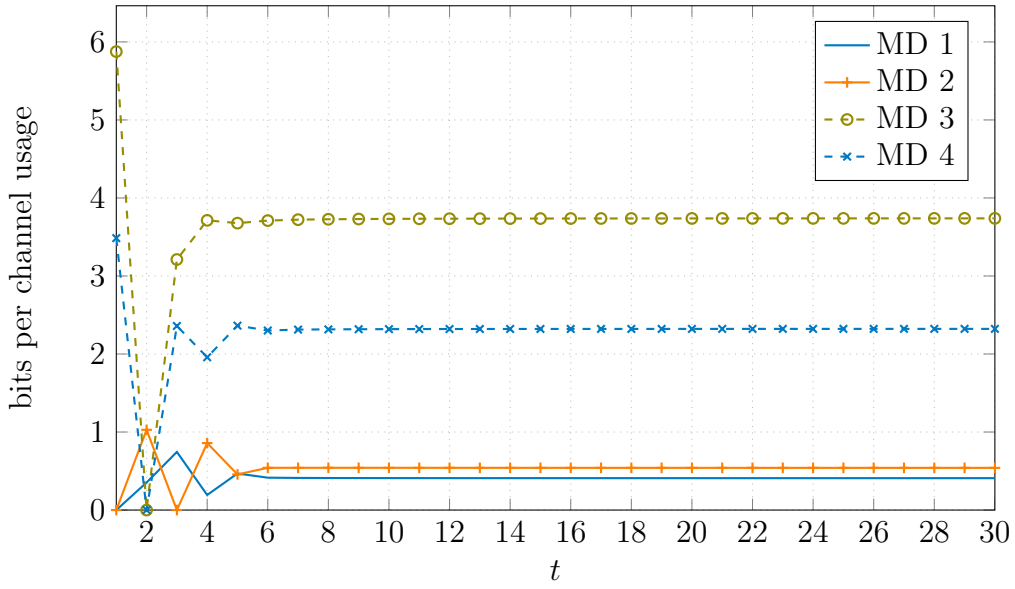


Figure 5.4. Instantaneous mobile device rates with proportional fairness

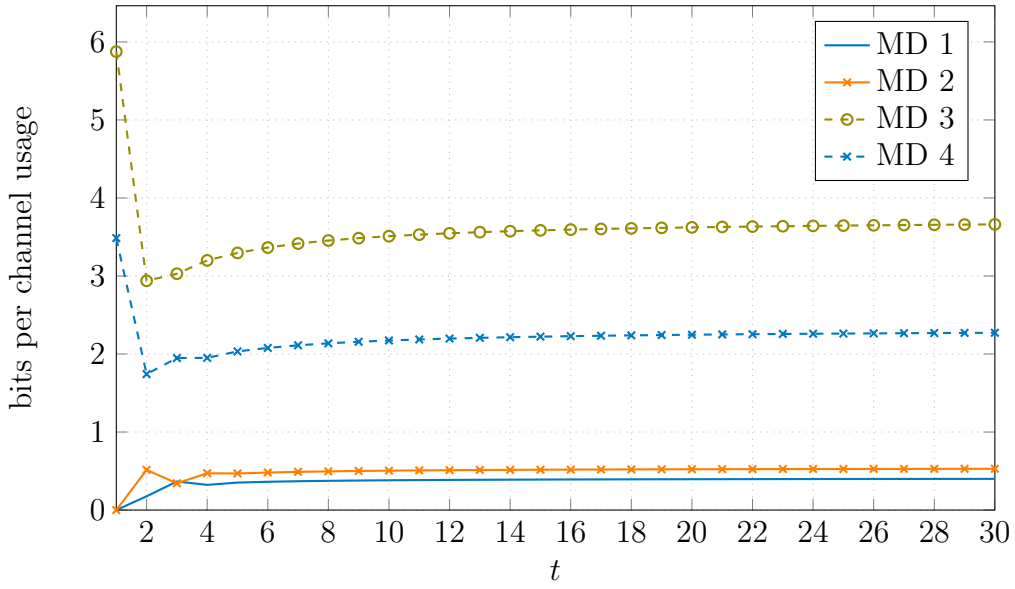


Figure 5.5. Time averaged mobile device rates with proportional fairness

5.3 Fairness Beamforming

The goal of fairness beamforming is to achieve fairness in every time slot. But, the rate region of the downlink will only be convex, if time sharing is employed. One has to alternate between multiple different transmit covariance matrix settings to optimize the desired utility. Therefore, fairness beamforming requires also multiple time slots in general. The algorithm for finding the optimal points is an iterative algorithm, where the different required transmit covariance matrices are the solution at different iterations. The evolution of the iterations can be seen as virtual queue of a scheduler [98, 128]. Fairness beamforming and fairness scheduling achieve the same rates for the same utility. The general objective function,

$$\max_{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \sum_k \mathrm{E}\left[U(r_k)\right] \qquad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{C}, \tag{5.12}$$

can be parameterized in such a way, that each rate $r_k \in \mathbb{R}_0^+$, $\forall k$ has to be smaller or equal than the feasible rate $c_k \in \mathbb{R}_0^+$ [97, 119]:

$$\max_{r_k \geq 0\ \forall k,\ \boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \sum_k \mathrm{E}\left[U(r_k)\right] \qquad \text{s.t.} \quad r_k \leq c_k\ \forall k, \quad \boldsymbol{Q} \in \mathcal{C}. \tag{5.13}$$

The optimization is done with respect to the transmit covariance matrices, which find the optimal c_k, $\forall k$, in the set of all feasible rates, and with respect to the rates r_k, $\forall k$, which are in the end set equal to c_k, $\forall k$. With the Lagrangian multiplier $\lambda_k \in \mathbb{R}_0^+$ for the constraints $r_k \leq c_k$, problem (5.13) can be rearranged as

$$\begin{aligned} &\max_{\substack{r_k \geq 0\ \forall k \\ \boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k,\ \boldsymbol{Q} \in \mathcal{C}}} \sum_k \mathrm{E}\left[U(r_k)\right] - \lambda_k (r_k - c_k) \\ &= \max_{r_k \geq 0\ \forall k} \sum_k \mathrm{E}\left[U(r_k)\right] - \lambda_k r_k + \max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \boldsymbol{Q} \in \mathcal{C}}} \sum_k \lambda_k c_k. \end{aligned} \tag{5.14}$$

The first maximization is solved by setting the derivations with respect to r_k to zero, respectively, while the second part is the maximum weighted sum rate problem discussed in Chapter 3, which is always transformed into an uplink problem. As the rates in uplink and downlink are equal, also the first maximization can be solved in the uplink. The solution can be found by an alternating optimization of r_k, $\forall k$, and c_k, $\forall k$ as described in Section 5.3.3. The expectation operator can be handled by the different methods for intercell interference robustness described in Chapter 4. With the expected rate method, the assumed intercell interference variance can be optimized together with r_k, $\forall k$, or in additional outer loop (See Section 5.3.3).

5.3.1 Max-Min

With the *max-min* (MM) criterion, the smallest weighted rate among the mobile devices is maximized by setting the utility function in (5.13) to [129]:

$$U(r_k) = \begin{cases} w_k r_k & \text{for } k = \underset{k}{\operatorname{argmin}}\,(w_k r_k), \\ 0 & \text{else,} \end{cases} \tag{5.15}$$

$$\sum_k U(r_k) = \min_k (w_k r_k). \tag{5.16}$$

The objective function now becomes

$$\max_{r_k \geq 0\ \forall k,\ \boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \ \mathrm{E}\left[\min_k (w_k r_k)\right] \qquad \text{s.t.} \quad r_k \leq c_k\ \forall k, \quad \boldsymbol{Q} \in \mathcal{C}. \tag{5.17}$$

where the optimization leads to equal weighted rates $w_1 r_1 = w_2 r_2 = \cdots = w_K r_K$.

The idea of the weighted max-min optimization can be seen in Figure 5.6 for the two mobile device case with equal weights $w_1 = w_2 = 1$ and an example rate region $\mathcal{R}$. The function $\min(r_1, r_2)$ has the same value for all rate tuples (r_1, r_2) on the contour line $\sum_k U(r_k) = \hat{U}$, where $\hat{U}$ is a constant. The solution to problem 5.17 can be found by moving the contour line from a large enough $\hat{U}$ to smaller values along the bisecting line of the positive quadrant until the contour line hits the border of the rate region from the upper right side. The optimal point is denoted by $\boldsymbol{r}^{\star}_{\mathrm{MM}}$.

The parameterized problem according to (5.14) is

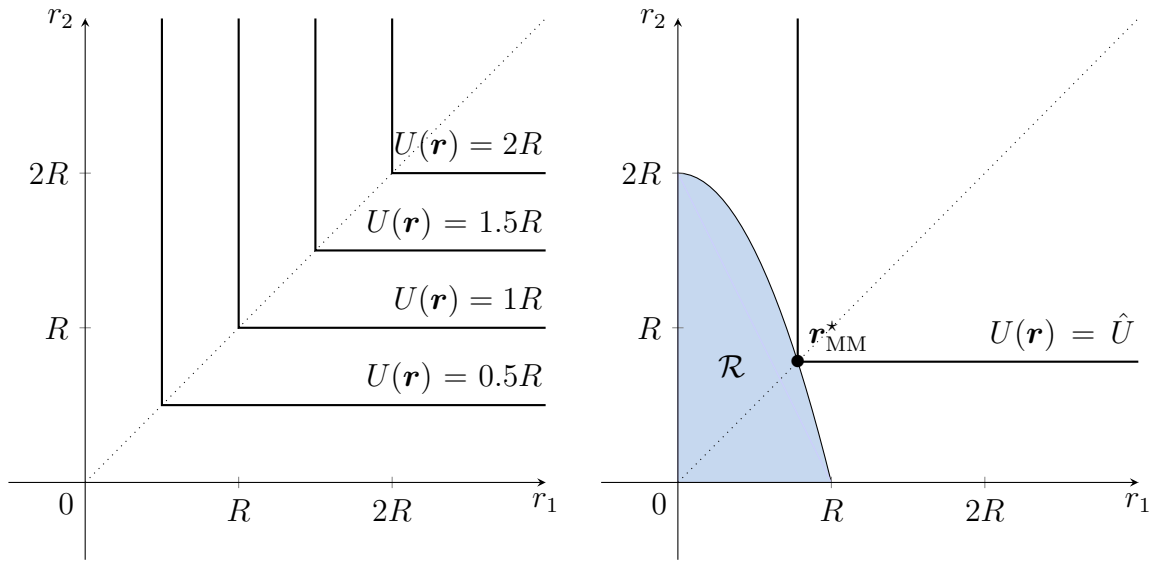
$$\Psi = \max_{r_k \geq 0\ \forall k} \left(\mathrm{E}\left[\min_k (w_k r_k)\right] - \sum_k \lambda_k r_k \right) + \max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \boldsymbol{Q} \in \mathcal{C}}} \ \sum_k \lambda_k c_k. \tag{5.18}$$

The minimization operator picks only the smallest weighted rate, which can be fulfilled by multiple weighted rates. Therefore, the derivative of (5.18) with respect to r_k is

$$\frac{\partial \Psi}{\partial r_k} = \begin{cases} (\partial\, \mathrm{E}\,[w_k r_k])(\partial r_k)^{-1} - \lambda_k & \text{if } k = \operatorname{argmin}_k (w_k r_k), \\ \lambda_k & \text{else.} \end{cases} \tag{5.19}$$

Setting the derivative to zero gives the optimal λ_k, $\forall k$,

$$\lambda_k = \begin{cases} (\partial\, \mathrm{E}\,[w_k r_k])(\partial r_k)^{-1} & \text{if } k = \operatorname{argmin}_k (w_k r_k), \\ 0 & \text{else.} \end{cases} \tag{5.20}$$



(a) Contour lines (b) Graphical solution

Figure 5.6. Max-min fairness with two MDs with equal weights, $U(\boldsymbol{r}) = \sum_k U(r_k) = \min(r_1, r_2)$

For the interference robustness methods, where the expectation operator is exchanged with the identity mapping, the Lagrangian multipliers are

$$\lambda_k = \begin{cases} w_k & \text{if } k = \operatorname{argmin}_k(w_k r_k), \\ 0 & \text{else.} \end{cases} \tag{5.21}$$

For the expected rate method, the Lagrangian multiplier is

$$\lambda_k = \begin{cases} F(\theta_k^{\text{assumed}}) w_k & \text{if } k = \operatorname{argmin}_k(w_k r_k), \\ 0 & \text{else,} \end{cases} \tag{5.22}$$

where $F(\theta^{\text{assumed}})$ is the probability, that all assumed intercell interference variances are not smaller than the actual intercell interference variances. This is slightly different than the previous definition in Section 4.6, which is used in the next section. Setting λ_k to zero is a rather extreme selection. In the optimal point, all λ_k are positive. The convergence behavior of the alternating optimization can be improved by gradually updating all λ_k with a gradient method, which is explained in more detail in Section 5.3.3.

Throughput fairness from Section 5.2.2 and the max-min fairness are closely related and aim for the same goal. Nevertheless, it is possible with max-min fairness to serve all mobile devices at the same slot with the same weighted rate, while this is never possible with scheduling. Additionally, the relaxation in the scheduling for multiple mobile devices in the same time slot alters the utility and not all mobile devices get the same weighted rate in the end.

5.3.2 Log Fairness

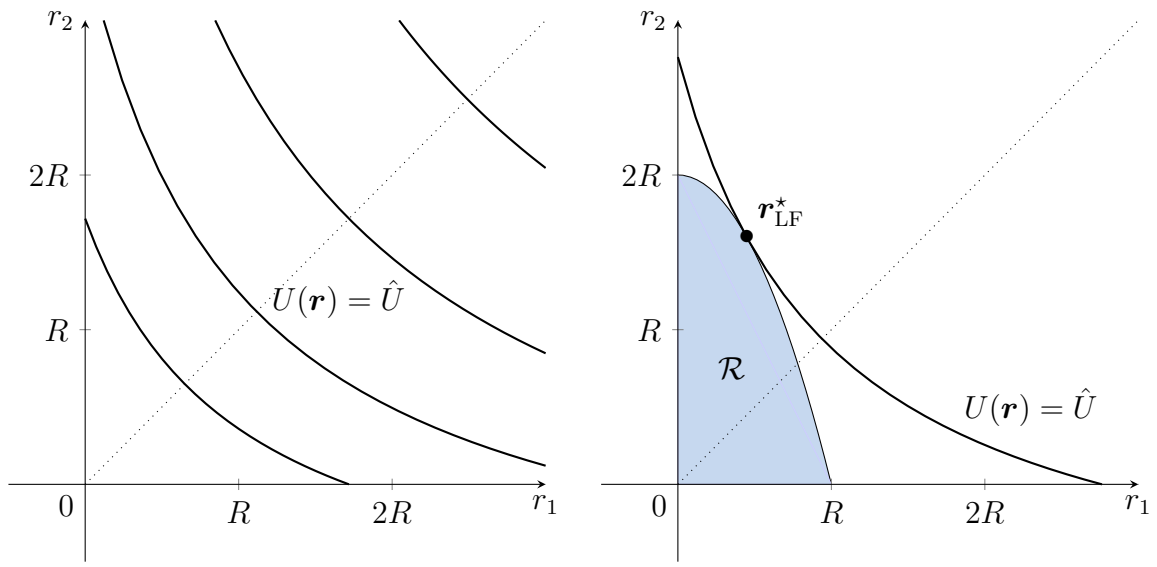
For the *log fairness* (LF) criterion, the utility function becomes the logarithm of the weighted rate added to a design factor γ [119]:

$$\sum_k U(r_k) = \sum_k \log(\gamma + w_k r_k). \tag{5.23}$$

The objective function of log fairness is

$$\max_{r_k \geq 0\ \forall k,\ \boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k} \sum_k \mathrm{E}\left[\log(\gamma + w_k r_k)\right] \qquad \text{s.t.} \quad r_k \leq c_k\ \forall k, \quad \boldsymbol{Q} \in \mathcal{C}. \tag{5.24}$$

The idea for solving problem (5.24) is depicted in Figure 5.7 for the two mobile device case with equal weights $w_1 = w_2 = 1$, an example rate region $\mathcal{R}$, and $\gamma = 1$. The utility function $\sum_k U(r_k) = \log(1+r_1) + \log(1+r_2)$ has the same value for all the rate tuples (r_1, r_2) on the contour line $\sum_k U(r_k) = \hat{U}$. Again, the solution to problem (5.24) can be found by moving the contour line down the bisection line until it touches the border of the rate region in the optimal point $\boldsymbol{r}^\star_{\mathrm{LF}}$ from the upper right side.



(a) Contour lines (b) Graphical solution

Figure 5.7. Log fairness ($\gamma = 1$) with two MDs with equal weights, $U(\boldsymbol{r}) = \log(1+r_1) + \log(1+r_2)$

According to (5.14), the objective is rearranged as

$$\Psi = \max_{r_k \geq 0\ \forall k} \left(\sum_k \mathrm{E}\left[\log(\gamma + w_k r_k)\right] - \lambda_k r_k \right) + \max_{\substack{\boldsymbol{Q}_k \succeq \mathbf{0}\ \forall k \\ \boldsymbol{Q} \in \mathcal{C}}} \sum_k \lambda_k c_k. \tag{5.25}$$

The second maximization of (5.25) is a maximum weighted sum rate problem as described in detail in Chapter 3, where the Lagrangian multipliers λ_k are used as weights. By plugging the probability of successful decoding $F(\theta_k^{\text{assumed}}) = P(\theta_k^{\text{assumed}} \leq \theta_k^{\text{actual}})$, $(1 - F(\theta_k^{\text{asssumed}})) =$

$P(\theta_k^{\text{assumed}} > \theta^{\text{actual}})$ (See equation (4.19)) into equation (5.25), the first maximization can be expressed as

$$\max_{r_k \geq 0\ \forall k} \sum_k F(\theta_k^{\text{assumed}}) \log(\gamma + w_k r_k) + (1 - F(\theta_k^{\text{assumed}})) \log(\gamma) - \lambda_k r_k \tag{5.26}$$

and simplified to

$$\max_{r_k \geq 0\ \forall k} \sum_k F(\theta_k^{\text{assumed}}) \log\left(1 + \frac{w_k}{\gamma} r_k\right) - \lambda_k r_k. \tag{5.27}$$

Setting the derivative of (5.27) with respect to r_k to zero,

$$\frac{\partial \Psi}{\partial r_k} = \frac{w_k}{\gamma + w_k r_k} F(\theta_k^{\text{assumed}}) - \lambda_k = 0, \tag{5.28}$$

allows to compute the optimal λ_k, $\forall k$,

$$\lambda_k = F(\theta_k^{\text{assumed}}) \frac{w_k}{\gamma + w_k r_k}. \tag{5.29}$$

For all intercell interference robustness methods, which use $F(\theta_k^{\text{assumed}}) = 1$, the optimal λ_k, $\forall k$, can be computed as

$$\lambda_k = \frac{w_k}{\gamma + w_k r_k}. \tag{5.30}$$

For the expected rate method, the optimization of $\theta_k^{\text{assumed}}$ can be a part of the first maximization or it can be optimized in an outer loop.

The design parameter γ influences the shape of the contour lines. In the extreme case of $\gamma \to -\min_k w_k r_k$, the contour lines of log fairness become the contour lines of the max-min optimization. In the other extreme case of $\gamma \to \infty$, the contour lines become the contour lines of the *sum rate* (SR) maximization, which are lines with a slope of minus one. The contour lines of the three different criteria and the respective graphical solutions and optimal points can be compared in Figure 5.8. A joint scaling of all λ_k, e.g. with γ, does not change the second optimization,

$$\lambda_k = \frac{\gamma w_k}{\gamma + w_k r_k}. \tag{5.31}$$

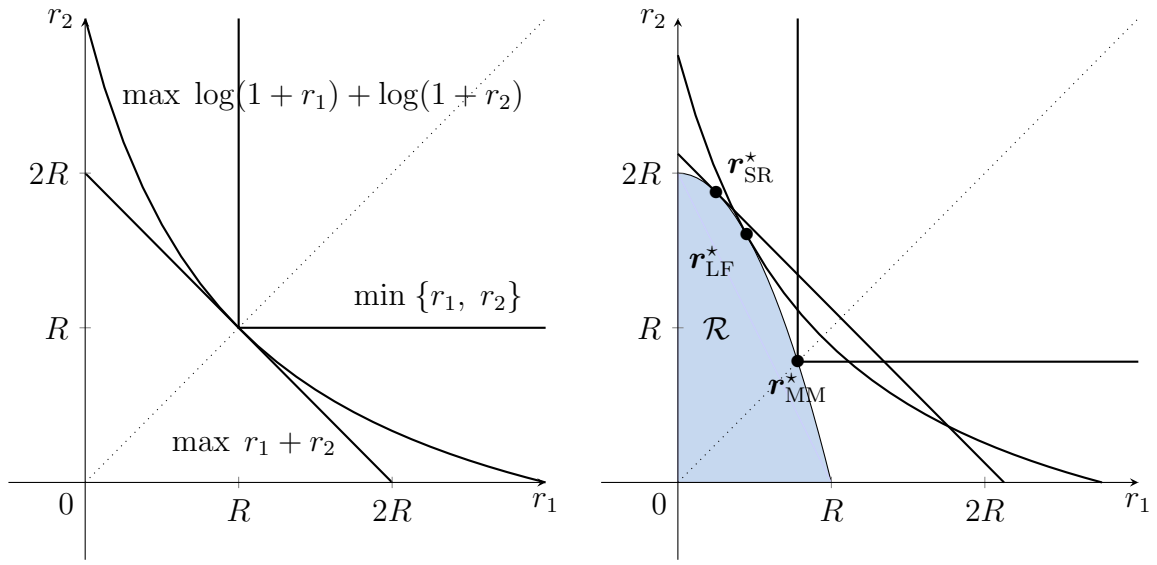
Here it can be easily seen, that $\gamma \to \infty$ yields the maximum sum rate weights,

$$\lambda_k = \frac{w_k}{1 + \frac{w_k}{\gamma} r_k} \overset{\gamma \to \infty}{=} w_k, \tag{5.32}$$

and $\gamma = \epsilon - \min_k(w_k r_k)$, $\epsilon \to 0$, yields the max-min weights,

$$\lambda_k = \frac{\epsilon w_k}{\gamma + w_k r_k} = \frac{w_k}{1 + \epsilon^{-1}(w_k r_k - \min_k(w_k r_k))} \overset{\epsilon \to 0}{=} \begin{cases} w_k & \text{if } k = \operatorname{argmin}_k(w_k r_k), \\ 0 & \text{else.} \end{cases} \tag{5.33}$$

With $\gamma = 0$, the Lagrangian multiplier in the iteration at hand is simply set to the rate from the previous iteration. In the optimal point, the Lagrangian multiplier and the rate converge jointly. Revisiting proportional-fair scheduling from Section 5.2.3, the instantaneous rates converge to their average rate, respectively. In the optimal point, the instantaneous rate is found with a weight that is one over the average rate, which is the rate from the previous time slot. Therefore, log fairness with $\gamma = 0$ is the same as proportional fairness [130].



(a) Contour lines (b) Graphical solution

Figure 5.8. Sum rate, log fairness ($\gamma = 1$), and max-min with two MDs with equal weights

5.3.3 Fairness Beamforming Algorithm

The fairness beamforming optimization has to be embedded into the previously described weighted sum rate algorithms for the different intercell interference robustness methods. The optimal transmit covariance matrices in the downlink and the optimal weights, which assure fairness in the weighted sum rate maximization, can either be found in an alternating optimization or in a joint optimization. The alternating optimization finds the optimal transmit covariance matrices in the downlink according to the algorithms presented in Chapter 3 and the optimal weights in turn. Each optimization is performed until convergence and the alternation is done until everything converges.

The optimal weights can be found more efficiently in a joint optimization with the transmit covariance matrices. To accomplish this, the optimization has to be done in the uplink. The weights and rates in the uplink are the same as in the downlink. Therefore, the update of the weights can also be done in the uplink without any adaption. The weighted sum rate algorithm with multiple constraints 3.2 is extended in Algorithm 5.2 to include the update of the weights, which assure fairness. In contrast to Algorithm 3.2, the utility is used as measure for convergence in Algorithm 5.2. This can directly be used as an implementation for covariance shaping with fairness beamforming. Conservative gambling can be implemented by dropping the optimization of the noise covariance matrices, i.e., setting the sum noise covariance matrix to the identity matrix from the beginning on. The optimal back-off factor β is applied after the optimization of the transmit covariance matrices 4.1 and 4.2. The optimal β has to be found for each utility individually.

As described in [97, 125], the weights or Lagrangian multipliers are not directly updated according to equation (5.20) or (5.29). The weights for the next iteration are found with a gradient step:

$$\lambda_k^{(i+1)} = \lambda_k^{(i)} - s_\lambda^{(i+1)}(c_k^{(i)} - r_k^{(i+1)}). \tag{5.34}$$

Algorithm 5.2 Fairness Beamforming

Require: Accuracy ε, outage probabilities $F(\theta_k^{\text{assumed}})$, $\forall k$, constraints $(\boldsymbol{C}_l, \mathcal{Z}_l)$, $\forall l$

$\boldsymbol{\Xi}_k^{(0)} \leftarrow \frac{P}{MK}\mathbf{I}$, $\forall k$ ▷ initial transmit covariances

$\boldsymbol{\Omega}_{1:L}^{(0)} \leftarrow \left(\frac{P}{L\,\mathrm{tr}(\boldsymbol{C}_{1:L})}\mathbf{I}\right)_{\perp}$ ▷ initial noise covariances (3.51)

$\lambda_k^{(0)} \leftarrow \frac{1}{K}$, $\forall k$ ▷ initial Lagrangian multipliers

$c_{\mathrm{n},1:K}^{(0)} \leftarrow \mathrm{ULRates}\left(\boldsymbol{\Xi}_{1:K}^{(0)},\ \sum_l \boldsymbol{\Omega}^{(0)}\right)$ ▷ initial rates (3.27)

$\hat{U}_{\mathrm{n}}^{(0)} \leftarrow \mathrm{E}\left[\sum_k U(c_{\mathrm{n},k}^{(0)})\right]$ ▷ initial utility (5.16) or (5.23)

$d_{\mathrm{t}} \leftarrow 1$, $d_{\mathrm{n}} \leftarrow 1$, $d_\lambda \leftarrow 1$, $i \leftarrow 0$ ▷ initialize step-size

repeat

  $r_{1:K}^{(i+1)} \leftarrow \mathrm{OptRates}$ ▷ according to (5.36) or (5.35)

  $\lambda_k^{(i+1)} \leftarrow \lambda_k^{(i)} - s_\lambda^{(i+1)}\left(c_{\mathrm{n},k}^{(i)} - r_k^{(i+1)}\right)$, $\forall k$ ▷ see (5.34)

  $\Pi^{(i+1)} \leftarrow \mathrm{sort}\left(\lambda_{1:K}^{(i+1)}\right)$ ▷ find non-increasing order of λ_k

  reorder $\lambda_k^{(i+1)}$, $F(\theta_k^{\text{assumed}})$, $\boldsymbol{H}_k$, and $\boldsymbol{\Xi}_k^{(i+1)}$, $\forall k$ according to $\Pi^{(i+1)}$

  $\boldsymbol{\Xi}_{1:K}^{(i+1)} \leftarrow \mathrm{TransCovStep}\left(\boldsymbol{\Xi}_{1:K}^{(i)}\right)$ ▷ Algorithm 3.3

  $c_{\mathrm{t},1:L}^{(i+1)} \leftarrow \mathrm{ULRates}\left(\boldsymbol{\Xi}_{1:K}^{(i+1)},\ \sum_l \boldsymbol{\Omega}_l^{(i)}\right)$ ▷ rates (3.27)

  $\hat{U}_{\mathrm{t}}^{(i+1)} \leftarrow \mathrm{E}\left[\sum_k U\left(c_{\mathrm{t},k}^{(i+1)}\right)\right]$ ▷ evaluate utility (5.16) or (5.23)

  $\mathrm{cost\_increase} \leftarrow \hat{U}_{\mathrm{t}}^{(i+1)} - \hat{U}_{\mathrm{n}}^{(i)}$

  $\boldsymbol{\Omega}_{1:L}^{(i+1)} \leftarrow \mathrm{NoiseCovStep}\left(\boldsymbol{\Omega}_{1:L}^{(i)}\right)$ ▷ Algorithm 3.4

  $c_{\mathrm{n},1:K}^{(i+1)} \leftarrow \mathrm{ULRates}\left(\boldsymbol{\Xi}_{1:K}^{(i+1)},\ \sum_l \boldsymbol{\Omega}_l^{(i+1)}\right)$ ▷ rates (3.27)

  $\hat{U}_{\mathrm{n}}^{(i+1)} \leftarrow \mathrm{E}\left[\sum_k U\left(c_{\mathrm{n},k}^{(i+1)}\right)\right]$ ▷ evaluate utility (5.16) or (5.23)

  $\mathrm{cost\_decrease} \leftarrow \hat{U}_{\mathrm{n}}^{(i+1)} - \hat{U}_{\mathrm{t}}^{(i+1)}$

  $i \leftarrow i + 1$ ▷ iteration counter

until $\mathrm{cost\_increase} \leq \varepsilon$ **and** $\mathrm{cost\_decrease} \geq -\varepsilon$

$\boldsymbol{Q}_{1:K} \leftarrow \mathrm{MAC2BC}(\boldsymbol{\Xi}_{1:K},\ \mu,\ \sum_l \boldsymbol{\Omega}_l)$ ▷ downlink recovery (3.42)

The gradient of the Lagrangian function of equation (5.14) with respect to λ_k is $c_k - r_k$, $\forall k$. c_k, $\forall k$, are the feasible rates according to the transmit covariance matrices in the recent iteration and r_k, $\forall k$, are the rates maximizing problem (5.18) or (5.26) for fixed λ_k and c_k, respectively. With log fairness, the optimal $r_k^{(i+1)}$, $\forall k$, is (See equation (5.29)) [97]

$$r_K^{(i+1)} = \frac{F(\theta_k^{\text{assumed}})}{\lambda_k^{(i)}} - \frac{\gamma}{w_k}. \tag{5.35}$$

The optimal $r_k^{(i+1)}$, $\forall k$, under max-min fairness can be computed under the necessary condition $\sum_k \lambda_k = 1$ [97] as

$$r_k^{(i+1)} = \left(\sum_k \frac{1}{w_k}\right)^{-1} \sum_k c_k. \tag{5.36}$$

With the weighted sum rate maximization as utility, the optimal λ_k is obviously equal to the weights. The gradient has to be subtracted from the previous solution, because the dual optimization with respect to λ_k, $\forall k$, is a minimization. $s_\lambda^{(i+1)} > 0$ is the step-size, which can be selected according to the methods described in Section 3.5.3. In Algorithm 5.2, an on demand step size reduction of d_λ^{-1} like in Algorithm 3.3 is assumed. The optimal non-increasing order of the weights λ_k has to be maintained (See Section 3.4.2). As soon as the weights are updated, it might be necessary to update the order as well. The weights, transmit covariance matrices, and channels have to be reordered to assure

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \tag{5.37}$$

in every iteration.

The probabilities of a successful transmission are already included in the cost function (5.14) with the expectation operator. For conservative gambling and covariance shaping, the success probability is always one during the optimization. The expected rate method or loosened covariance shaping additionally need to optimize the success probabilities, with respect to the assumed intercell interference variance. It is possible to integrate the update of the intercell interference into the optimization at hand, finding all variable in a single loop. But, the assumed intercell interference has to be optimized in the downlink and changes the uplink problem. Switching between the uplink and the downlink is computationally cost intense. Therefore, the optimization of the assumed intercell interference variances is done in an outer loop, as described in Algorithm 4.3 and 4.7. These algorithms typically converge in very few iterations with respect to the assumed intercell interference, which reduces the necessary uplink-downlink transformations drastically.

Compared to fairness scheduling, the computational complexity of fairness beamforming is higher. But, fairness beamforming can assure a fair allocation in less time slots. In general, the fair allocation can only be assured with time sharing and, therefore, the fairness can only be assured over multiple time slots. This is also the reason, why the algorithm does not converge necessarily. But, the time average of the utility converges to the optimal point.

5.4 Simulation Results

In the following simulations, the single cell system model with random intercell interference variance from Section 4.3 is utilized. The base station serves $K = 4$ single-antenna mobile devices with $N = 4$ transmit antennas. In each realization, $T_{\text{block}} = 50$ time slots are simulated, during which the channel stays constant while the actual intercell interference variance θ_k^{actual}, $\forall k$, can be different in each time slot. The transmit power limit of the base station is $P = 14\,\text{W}$ and the noise variance at each mobile device is $\sigma_\eta^2 = 1.4 \times 10^{-12}\,\text{W}$. The channels are generated according to the same parameters as described in Section 2.4.1.

The random actual intercell interference variances θ_k^{actual}, $\forall k$, are generated with a gamma distribution $\theta_k \sim \Gamma(a_k, b_k)$, $\forall k$, where a_k and b_k are derived from the used channel model (see Section 4.3) [30]. The statistics are individually derived for each mobile device, utility, and intercell interference robustness method and they are assumed to be known to the base station. Perfect channel estimation is assumed and the measurement of the signal to interference plus noise ratio in the piloting phase is also assumed to be perfect. The optimal back-off factor β in conservative gambling is optimized individually for each utility function and it has always a value between 0.12 and 0.2.

The cumulative distribution function of the mobile device rates are compared for different combinations of intercell interference robustness methods and fairness scheduling criteria. The expected rate, conservative gambling and covariance shaping methods are combined with *round robin* (RR), *sum rate* (SR) maximization, *throughput-fair* (TF), and *proportional-fair* (PF) scheduling. The shown rates $\bar{r}_k$, $\forall k$, are the achievable rates averaged over the time block T_{block} and come from all the mobile devices over multiple realizations. Curves with a flat slope have a wider range of rates. A steeper slope can be regarded as more fair.

The highest rates can be reached with the combination of sum rate maximization and the expected rate method. In the outage region $r \leq 1$, covariance shaping has always a much steeper slope than the other robustness methods. With covariance shaping, the intercell interference is always known and no mobile device drops to a rate of zero. The outage rates are in general very small for all utilities and robustness methods because the plots are done without HARQ.

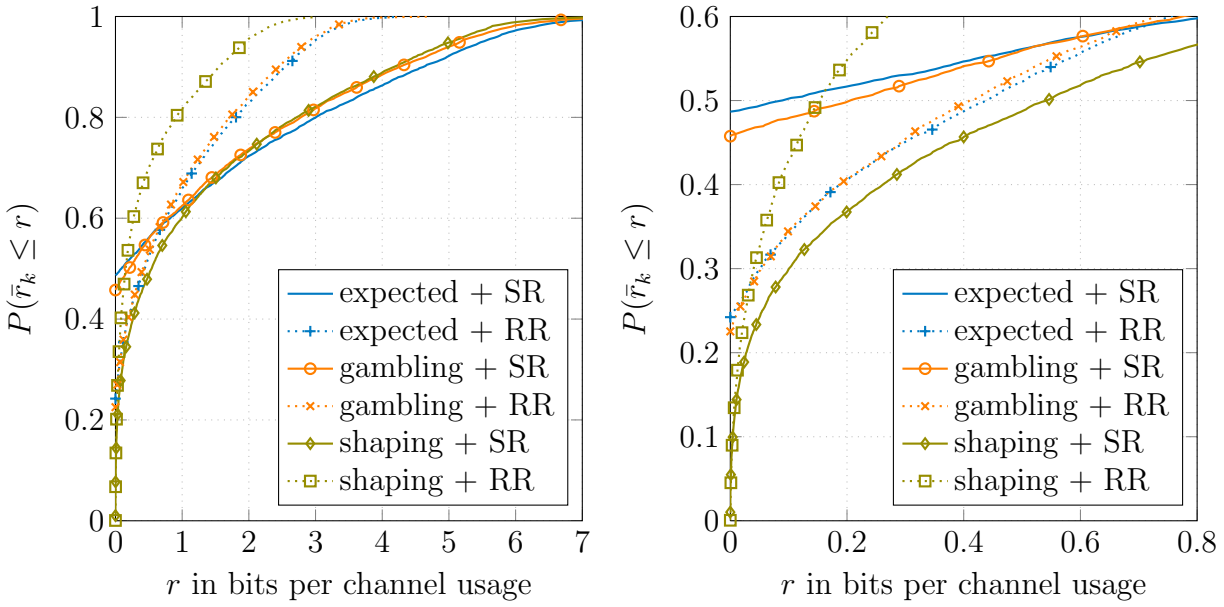
In the simulations, the round robin scheduler picks two mobile devices for each slot. The expected rate method has a higher throughput with round robin and the sum rate maximization than conservative gambling and covariance shaping (See Figure 5.9). With the expected rate method and conservative gambling, the probability that the rate of a mobile device drops to zero is around 25 % with round robin and approaches almost 50 % under the sum rate maximization. In comparison, the curve of covariance shaping is very steep in the outage region and there is zero probability that a mobile device rate is zero, which is the most fair.

The probability is zero because the rate with the worst case intercell interference variance is used and the mobile devices can always decode the data. The intercell interference variance used for covariance shaping with round robin is only a worst case assumption, because the transmit covariance matrix is rank deficient. Only two mobile devices a served, but the base station has four antennas. This is also the reason, why covariance shaping performs badly with round robin. The actual intercell interference variance is in most cases smaller than the worst case intercell interference variance, but the data is always transmitted with the worst case rate.

The same argument is true for throughput fairness, where also only two mobile devices are scheduled per time slot. The expected rate method performs the best with the throughput-fair scheduler. Since the mobile devices are scheduled to have the same throughput in the end and the mobile devices with bad channels are served very frequently, the three curves in Figure 5.10 have very steep slopes. All three methods have probabilities close to zero for an mobile device rate of zero. But, the curves of throughput fairness have a probability of one for rates smaller than two bits per channel usage because the mobile devices with smallest throughput are scheduled all the time.

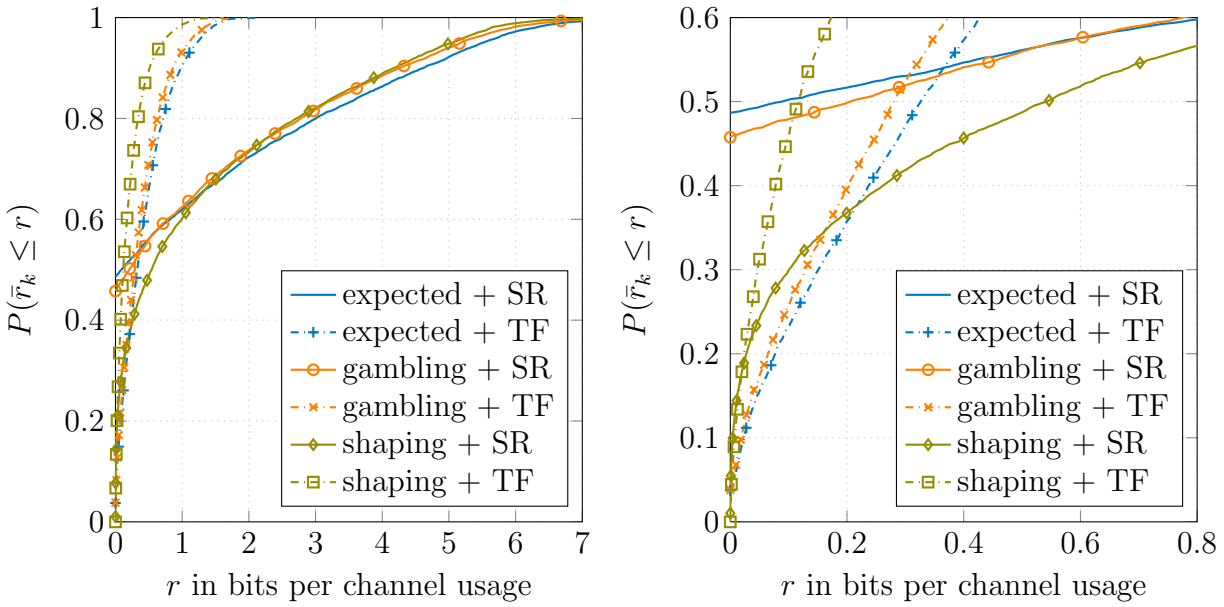
Under the proportional fairness criterion, all mobile devices in the cell are served all the time. Covariance shaping outperforms the other methods in achieving higher rates in Figure 5.11. Proportional fairness finds a balance between fairness and sum throughput. On the one hand, the figures show that almost all mobile devices can always receive some data. With all three methods the probability for an mobile device rate of zero is very small. On the other hand, the probability for high rates is still relatively high.

Figure 5.12 illustrates the effect of parameter γ in *log-fair* (LF) beamforming combined with conservative gambling as robustness method. γ can take any value in the open range $(-\min(w_k r_k), \infty)$. A smaller γ results in a fairer utility, while a larger alpha gives more



(a) Complete region (b) Outage region

Figure 5.9. CDF of achieved rates with round robin

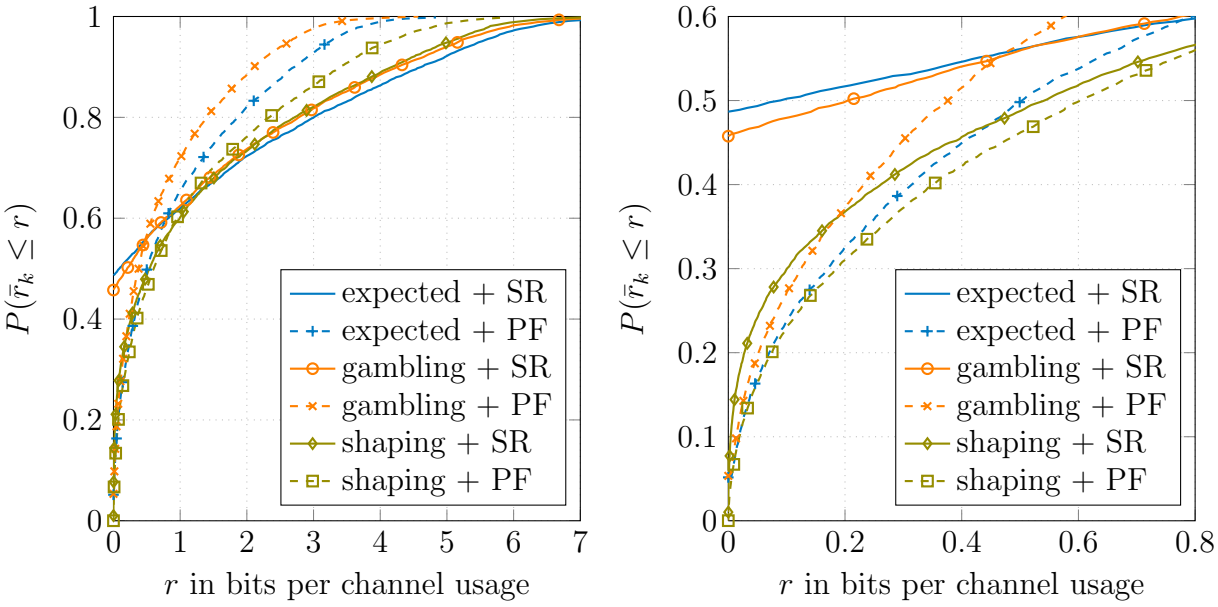


(a) Complete region (b) Outage region

Figure 5.10. CDF of achieved rates with throughput fairness

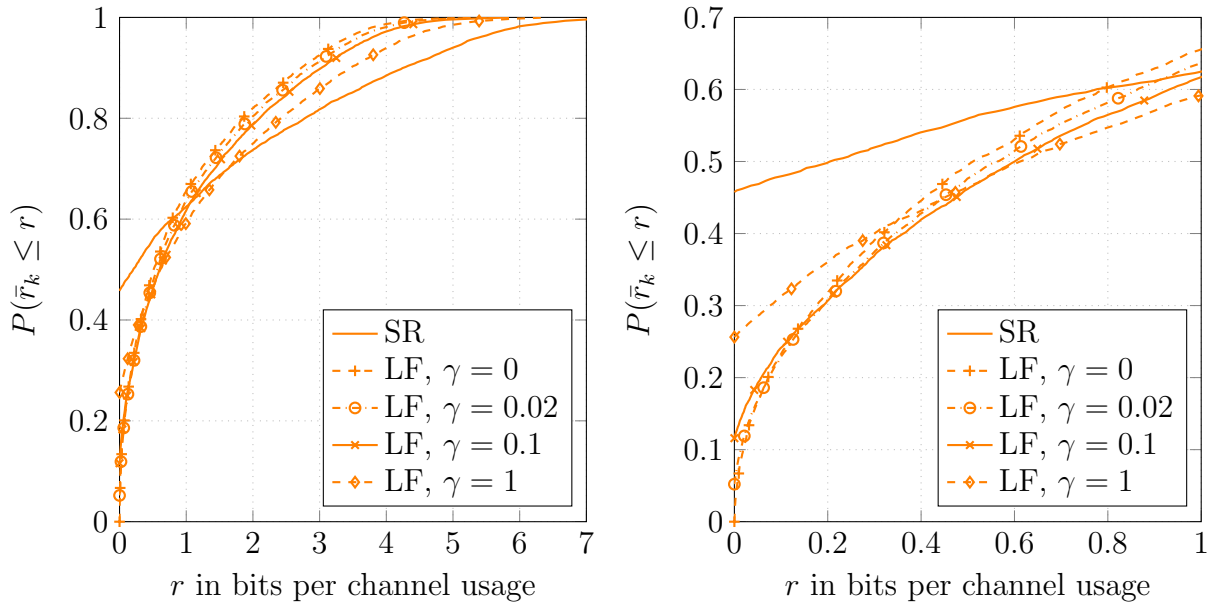
resources to the mobile devices with good channels. In the extreme case, $\gamma = -\min(w_k r_k)$, the log fairness converges to max-min and in the other extreme case, $\gamma = \infty$, it converges to the sum rate maximization. The selection $\gamma = 0$ leads to the proportional-fair allocation, which is also achieved with the proportional-fair scheduler [130].

All different fairness scheduler are depicted in one plot with the expected rate method in Figure 5.13. Tables 5.4 and 5.5 show the mean and median of the time averaged mobile



(a) Complete region (b) Outage region

Figure 5.11. CDF of achieved rates with proportional fairness

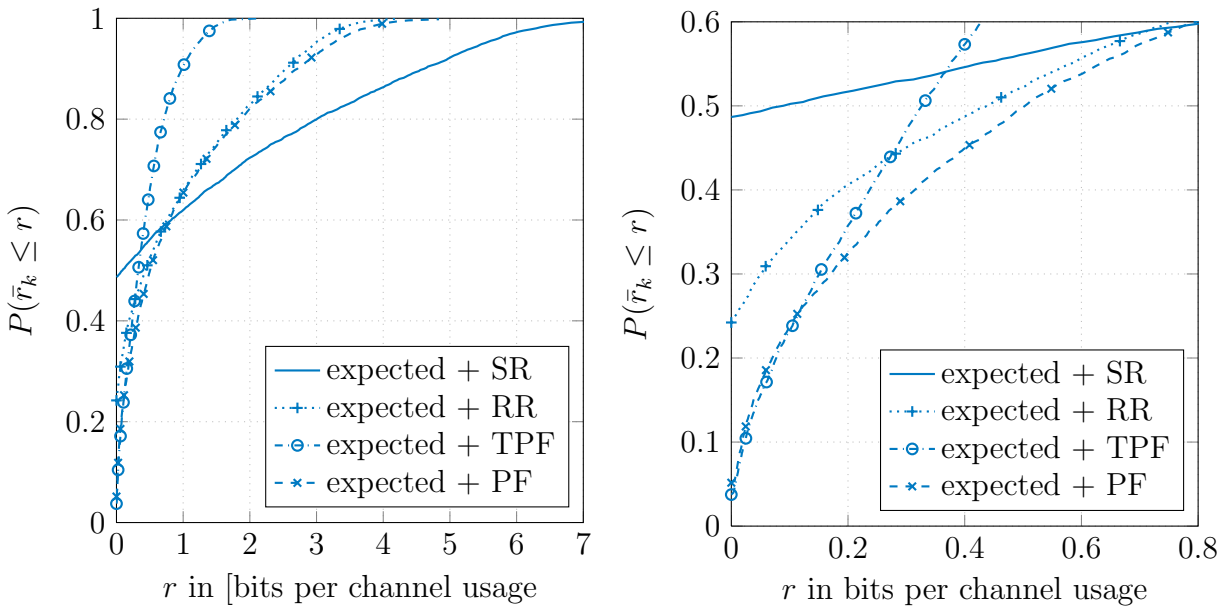


(a) Complete region (b) Outage region

Figure 5.12. CDF of conservative gambling and log fairness with different γ

device rates for all fairness scheduler and robustness method combinations. The third line in the tables is the gain of the expected rate method compared to conservative gambling and the fifth line is the gain of covariance shaping. The mean is optimized with the sum rate maximization, while the median gives a better picture of the fairness among the mobile device rates. The median can be found in the plots as the 50 % outage rate.

The different cell sum utilities are compared for the different robustness approaches in Table 5.6. The expectation operator in the utilities are approximated with the mean, respectively. The maximum sum rate utility is K-times the mean achieved mobile device rate (here four times) from Table 5.4. The max-min or throughput-fair utility is much smaller than the rates in Tables 5.4 and 5.5. Although, the expected utility should always be larger than zero, it might be zero in the simulations due to the limited scheduling slots, which can also be seen in the figures with the cumulative distribution functions. As soon as one of the mobile device rates in a cell drops to zero, the complete max-min utility drops to zero, which cannot be seen in the mean or median. The proportional-fair utility is transformed, it is used as exponent for the base of the natural logarithm, to improve the readability of the table. Again, the expected rate method always outperforms conservative gambling. Covariance shaping combined with proportional fairness has the best performance in the complete picture.



(a) Complete region (b) Outage region

Figure 5.13. CDF of achieved rates with the expected rate method

scheduler	sum rate	round robin	throughput-fair	proportional-fair
gambling	1.269	0.837	0.373	0.741
expected	1.353	0.870	0.423	0.958
% gain	6.6	4.0	13.5	29.3
shaping	1.339	0.473	0.202	1.190
% gain	5.5	-43.5	-45.8	60.6

Table 5.4. Mean achieved mobile device rate in bits per channel usage

scheduler	sum rate	round robin	throughput-fair	proportional-fair
gambling	0.205	0.410	0.281	0.377
expected	0.089	0.438	0.326	0.504
% gain	-56.4	6.9	16.3	33.7
shaping	0.536	0.153	0.117	0.603
% gain	161	-62.7	-58.4	59.9

Table 5.5. Median achieved mobile device rate in bits per channel usage

utility	$\mathrm{E}\left[\max \sum_k r_k\right]$	$\mathrm{E}\left[\max \min_k r_k\right]$	$\exp\left(\frac{1}{K}\mathrm{E}\left[\max \sum_k \log(r_k)\right]\right)$
gambling	5.075	0.086	0.239
expected	5.412	0.100	0.315
% gain	6.6	15.9	31.7
shaping	5.357	0.035	0.384
% gain	5.5	-59.3	60.4

Table 5.6. Mean achieved cell sum utility

5.5 Conclusion

In this chapter, the implementation of the previously presented intercell interference robustness methods are discussed for different fairness utilities. In the first part, the fairness utilities are approximated with a time slot scheduler. In the second part they are handled with fairness beamforming. The time slot scheduler selects weights for a weighted sum rate maximization in each slot, while fairness beamforming tries to find the utility maximizing transmit covariance matrices in each time slot, respectively. Both approaches are strongly related and lead to the same solution for the same utility. Throughput-fair scheduling and max-min beamforming target the same utility, but the throughput-fair scheduler has to limit the number of active mobile devices per slot and therefore deviates slightly from the utility. Proportional-fair scheduling has the same utility as log-fair beamforming with $\alpha = 0$. Detailed algorithms are presented for implementing the considered problems.

The performance of the different utilities and robustness methods are compared with plots of the cumulative distribution function of the mobile device rates. Additionally, the expected utility and the mean and median of the mobile device rates are listed. It could be seen that the transmit covariance matrices for max-min, round robin, and throughput-fair scheduling can be rank deficient in many cases. If this is the case, the covariance shaping method cannot use the complete transmit power and the intercell interference variance at the mobile devices is only a worst case assumption. The performance of covariance shaping suffers dramatically under these circumstances. In general, the performance under fairness utilities greatly increases with the intercell interference robustness methods. Especially the combination of covariance shaping with proportional fairness shows promising results.

Appendix

A1 Interference Visualization

The system utilized for simulations in a full cellular network can be seen in Figure A1. There are 19 sites with 3 base stations each placed homogeneously in the network. Each base station serves the mobile devices of the hexagonal shaped cell it covers. The wrap-around method is used in order to have each cell experiencing interference from two rings of sites around itself. The 57 base stations are copied, including their beamforming, and placed six times around the central cluster. Each mobile device only sees the 57 base stations, which are closest by Euclidean distance. In Figure A1, the central cluster is inked slightly darker than the wrap-around clusters. The placement and orientation of the base stations is indicated by small arrows. The numbers in the cells identify the cell and associated base station uniquely. It can be seen, that the numbers are reused in the copied clusters.

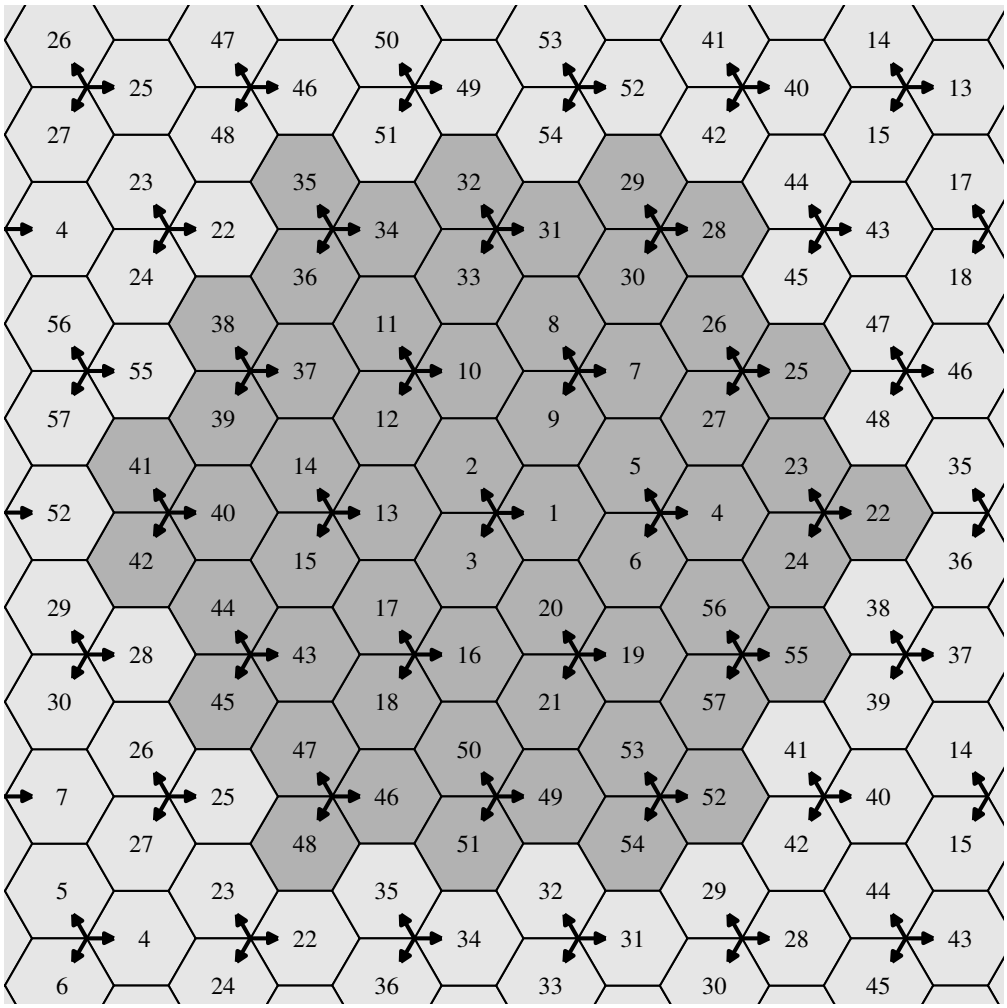


Figure A1. Cellular layout and cell numbering

The simulations are based on the 3GPP MIMO urban macro cell channel model at a center frequency of 2 GHz. The distance between to neighboring sites is 500 m and the minimum distance between a mobile device and a site is 25 m. Each base station is equipped with $N = 4$ antennas and can serve multiple single antenna receivers at the same time. The antennas are assumed as hypothetical isotropic radiators with a spacing of half a wavelength (0.5λ) between the antennas of the same array. The fading realizations of antennas of the

same array are correlated in general. The channels can consist of line-of-sight and non-line-of-sight paths with random angular spread and a delay spread of $\sigma_{\mathrm{DS}} = 0.5\,\mu\mathrm{s}$. Orthogonal frequency-division multiplexing is employed to acquire frequency-flat small-band channels. The simulations are done in the downlink with a transmit power of $P = 14\,\mathrm{W}$ and a thermal noise variance of $\sigma_\eta^2 = 1.4 \cdot 10^{-13}\,\mathrm{W}$ at the receivers, which gives a transmit power over noise power ratio of 140 dB. The blue colored region in the center of Figure A2 is the region, in which a mobile device has in average the strongest channel to base station 1 according to the channel model. The shape of the region is slightly different compared to the hexagonal layout. In the simulations, the user association is based on the hexagonal layout nevertheless. This can be justified with the hysteresis of the hand-over process between base station associations.

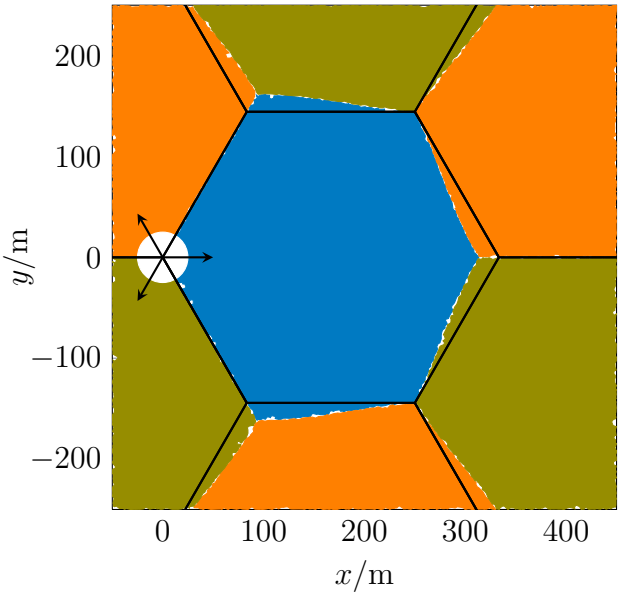


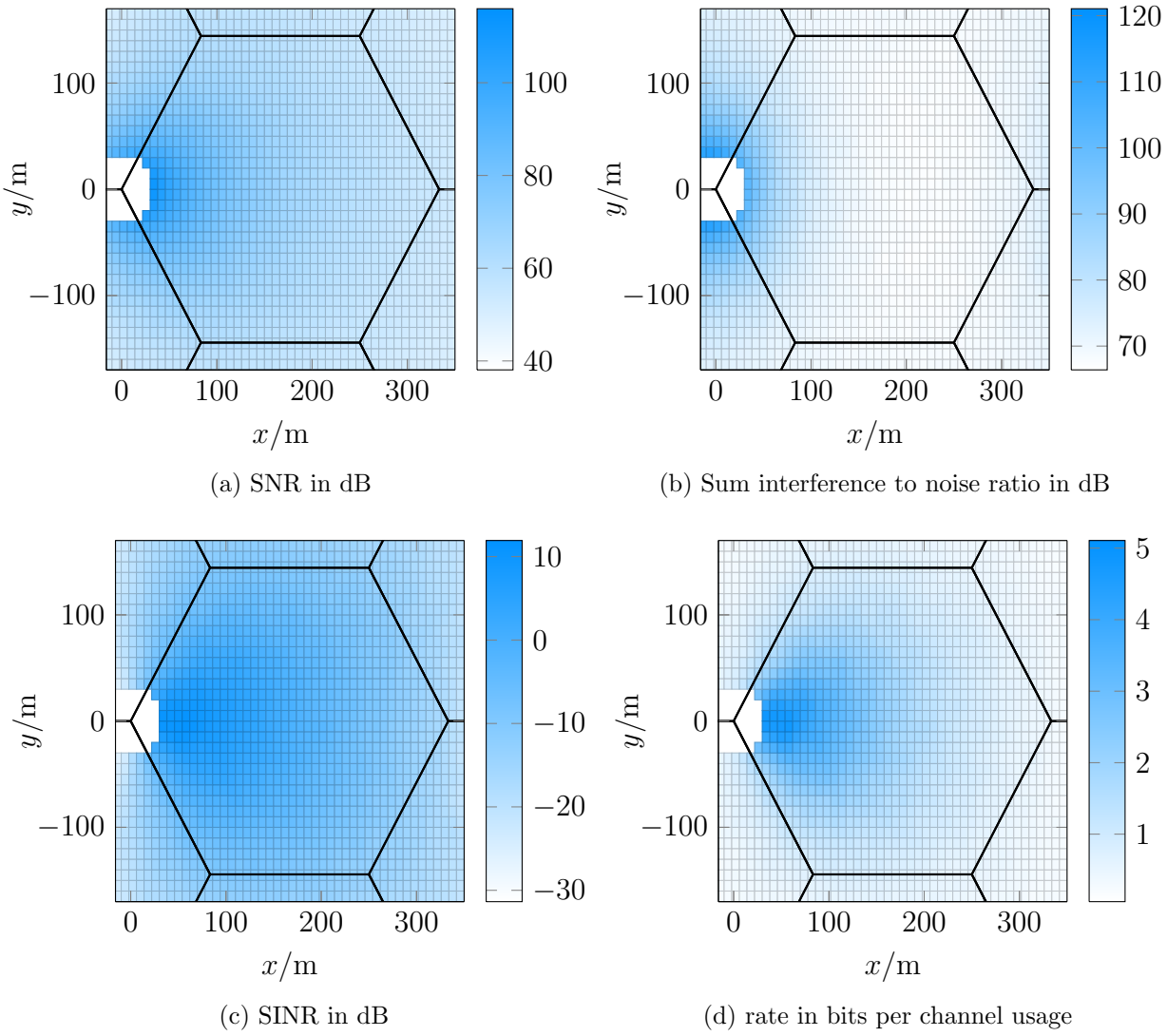
Figure A2. Cell boundaries

The dependency of the average channel quality on the position within the cell can be seen in Figure A3. The SNR in the covered area is rather large on average, as can be seen in Figure A3 (a). In this appendix, the SNR is calculated by multiplying the average channel norm with the transmit power over receiver noise ratio. Mobile devices further away from the base station experience a much smaller SNR than mobile devices close to the base station due to the larger path-loss. The influence of the antenna pattern can also be seen. More energy is directed into the cell than in other direction. This increases the SNR of the mobile devices in the cell while reducing the interference caused in other cells.

The sum interference power of all base stations in the network over the thermal noise in dB can be seen in Figure A3 (b) for the area of cell one. This sum interference is calculated by summing up all the SNRs of the not serving base stations, which is the case in systems with full frequency reuse. The interference is very strong in general. Mobile devices close to the base station suffer from a very strong interference caused by the collocated base stations. The interference is the smallest in the center of the cell and increases towards the far end of the cell slightly.

The SINR and rate according to the combination of the previously described SNR and sum interference power can be seen in Figure A3 (c) and A3 (d), respectively. Site vicinity users have the best SINR and rate in general. The SINR and rate decrease dramatically

over the covered area but much less than the SNR decreases solely due to path-loss. The achievable rates at the far end of the cell are rather small.



(a) SNR in dB

(b) Sum interference to noise ratio in dB

(c) SINR in dB

(d) rate in bits per channel usage

Figure A3. Position dependency of channel quality in cell one

A1.1 Mobile Device Classification

Based on the observations in the previous section, the mobile devices can be classified into

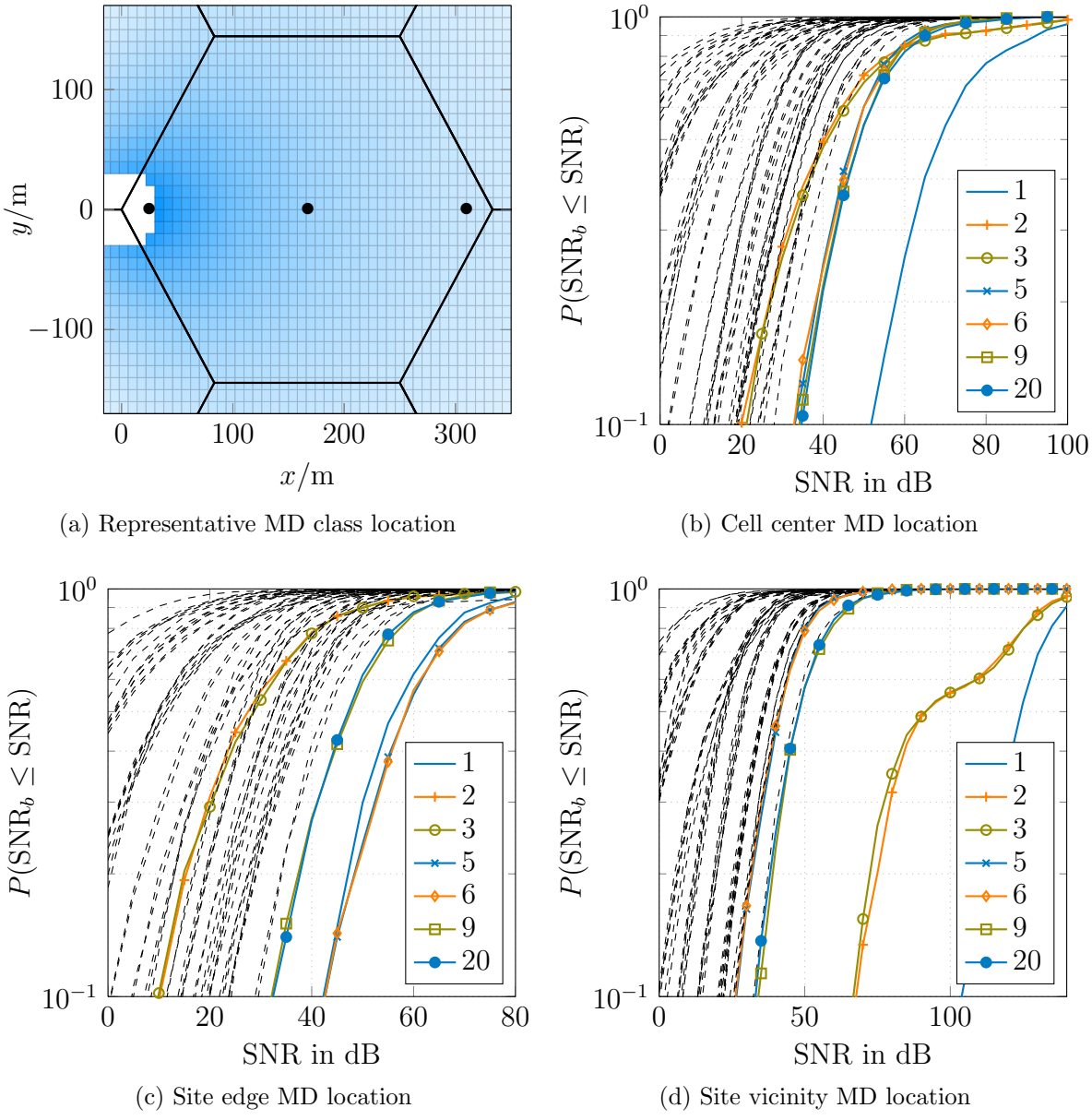
- cell center mobile devices,
- site edge mobile devices, and
- site vicinity mobile devices.

Different locations in the cell, which represent the different classes and are used in the following simulations, are depicted in Figure A4 (a).

The remaining plots in Figure A4 show cumulative distribution functions of the SNR of all base stations in the network to a mobile device positioned at the three different locations, respectively. The here used SNR is calculated by multiplying the channel norm of the channel between the base station and the mobile device with the transmit power to thermal noise

power ratio. Therefore, this SNR is not necessarily the SNR of a serving base station but any base station in the network. The curves associated with different base stations can be distinguished with the numbers given in the legend of the plots. These numbers correspond to the base station numbering in Figure A1. The SNRs of base station one and the six surrounding base stations are colorized and marked, while the rest of the SNRs remain unmarked.

The cumulative distribution functions for a mobile device located in the cell center can be found in Figure A4 (b). The SNR of base station 1 is very strong and has a large gap to the SNRs of the interfering base stations. The SNRs of the closest neighbors are almost equally strong, there is no single strong interfering base station.



(a) Representative MD class location

(b) Cell center MD location

(c) Site edge MD location

(d) Site vicinity MD location

Figure A4. Mobile device classification and SNR CDFs of all base stations

The plot for the mobile device positioned at the site edge can be seen in Figure A4 (c). The SNR of base station one is very strong, but the interfering base stations five and six, which are located vis-à-vis of base station one, have a similar SNR. Therefore, these base stations cause very strong interference. The gap to the rest of the interfering base stations is small compared to the gap in the cell center location. The interference from the base stations of the same site is negligible compared to the other interference.

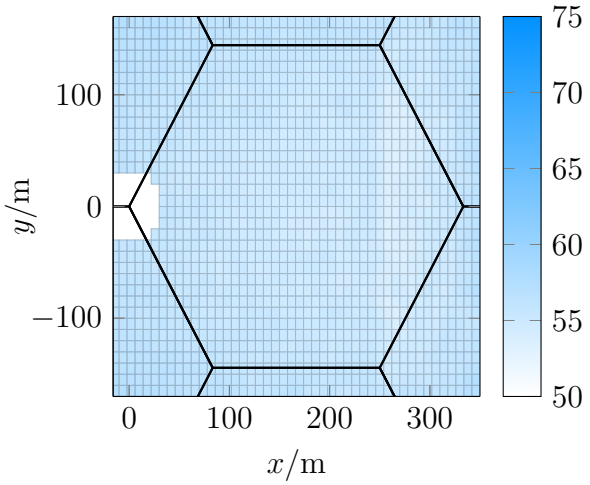
A very different picture can be seen for the mobile device placed very close to the base station (see Figure A4 (d)). The SNR of base station one is extremely strong and is only followed by the also strong SNR of the two collocated base stations. These two base stations will cause the largest potion of the interference at mobile devices located in this position. The gap to the rest of the interference is extremely large.

A1.2 Frequency Reuse and Site Cooperation

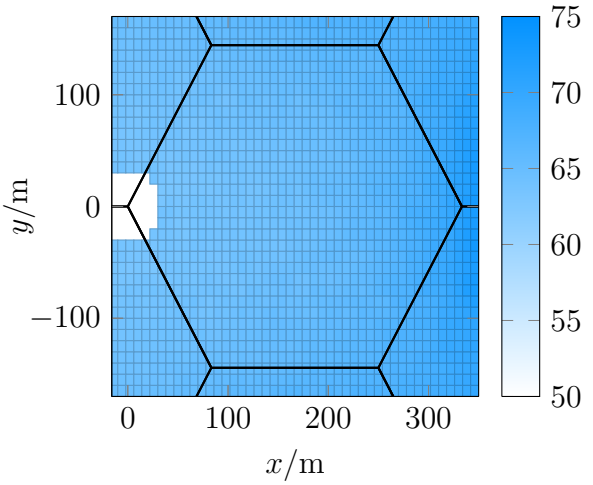
The influence of frequency reuse with reuse factor three, one ring of neighboring base stations transmits in a different frequency band, can be seen for the sum interference to noise ratio, SINR, and rate in Figure A5 (a), (c), and (e), respectively. The rates are calculated with a pre-log factor of a third, to include the effect of frequency reuse. The influence of network MIMO site cooperation, where each site forms a cooperative cluster, is shown for the sum interference to noise ratio, SINR, and rate in Figure A5 (b), (d), and (f), respectively. In this scenario, the three base stations collocated at each site can be seen as a single base station with three times the antennas and three times the transmit power, respectively. To calculate the rates, the received power was scaled down by a third to respect the fact that three times more mobile devices have to be served by the combined base station.

The sum intercell interference to noise ration is considerably smaller with both types of interference mitigation compared to no interference mitigation in Figure A3 (b). Both methods remove the interference from the collocated base stations completely. The interference is in general smaller with frequency reuse, as it removes the interference from all neighboring base stations. Additionally, the intercell interference to noise ratio is almost evenly distributed over the cell with frequency reuse. With site cooperation the sum interference to noise ratio increases towards the cell edge. The SINR is drastically larger with both methods compared to no interference mitigation in Figure A3 (c). The SINR is more evenly distributed over the cell with frequency reuse than with site cooperation.

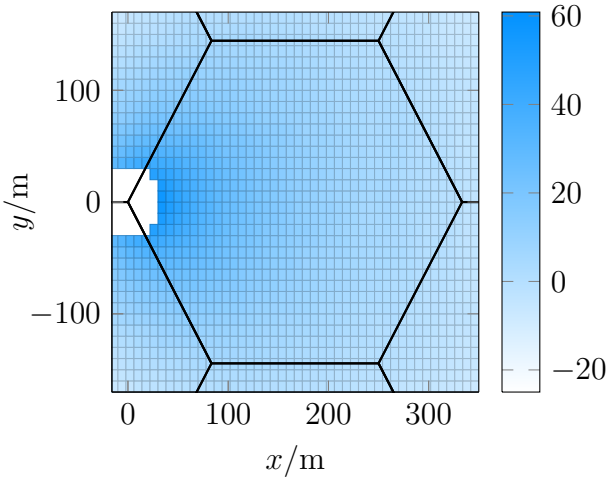
Although the SINR is similar with both methods, there is a big difference in the rates. The rates with frequency reuse are scaled down with a pre-log factor, while the SINR is scaled down in the case of site cooperation. The rates with site cooperation are astonishingly high, but neglect the cost for acquiring the additional channel state information. Frequency reuse has no further implementation issue than splitting and separating the frequency bands between the base stations. On closer inspection, it can also be seen that only the rates close to the base station are high, while the rates towards the cell center and close to the cell edge are rather small with site cooperation. Calculating the rates for site cooperation by simply scaling down the transmit power can also be challenged. In both cases, the rates are higher than with no interference mitigation in Figure A3 (d).



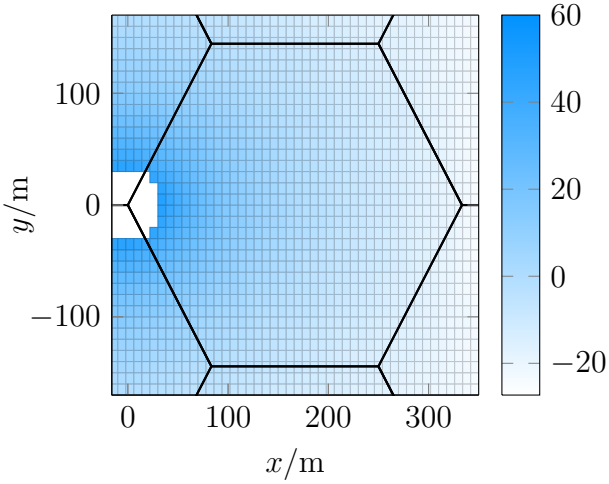
(a) Sum interference to noise ratio in dB, reuse 3



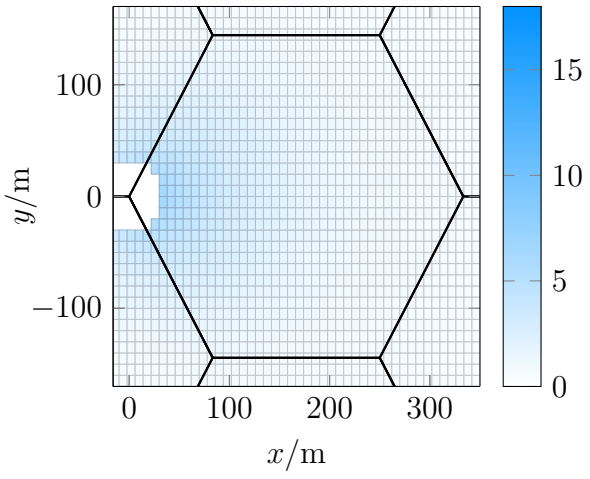
(b) Sum interference to noise ratio in dB, cooperation



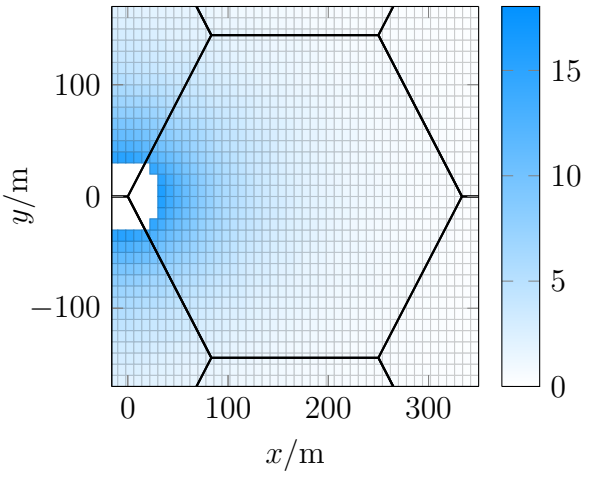
(c) SINR in dB, reuse 3



(d) SINR in dB, cooperation



(e) Rate in bits per channel usage, reuse 3



(f) Rate in bits per channel usage, cooperation

Figure A5. Frequency reuse and site cooperation

A2 Pilot Contamination

In a cellular system, where pilot sequences are used to measure the channels, these pilot sequences have to be reused by many devices. The interference caused by the simultaneous transmission of the same pilot sequence from different antennas is called pilot contamination [56]. The different transmissions cannot be separated and a linear combination of the different channels is measured. In a system with time division duplex, the channels are measured in the uplink. Mobile devices transmit pilot sequences and the base stations estimate the channels. The gained information is then used by the base stations in the downlink to optimize the transmit strategy. The transmission is optimized for the estimated channel and, as this is a linear combination of a serving channel and multiple interference channels, will produce more interference at mobile devices, which used the same pilot sequence.

In the following, it is assumed that all pilot sequences are transmitted in the same time and frequency resources. The K mobile devices within a cell have to use orthogonal pilot sequences, in order to have the associated base stations obtain meaningful estimates. A pilot length of $T_{\text{pilot}} = K+L$ allows up to $K+L$ orthogonal pilot sequences. Each base station can measure the channels to its own K mobile devices and L interference channels, additionally. Using all possible pilot sequences is always beneficial, as it allows to learn L interference channels at each base station, respectively, and minimizes the pilot contamination.

With a common mobile device transmit power P_{MD} and the disjoint sets $\mathcal{P}_p$ of all mobile devices sharing the same pilot sequence with index p, the pilot contamination at base station $\check{b}$ for the received transmission from mobile device (b,k) can be found as

$$\boldsymbol{\Theta}_{\text{pc},\check{b},b,k} = P_{\text{MD}} T_{\text{pilot}} \sum_{(\hat{b},\hat{k})\in\mathcal{P}_p\setminus(b,k)} \boldsymbol{R}_{\boldsymbol{h},\check{b},\hat{b},\hat{k}}, \tag{A1}$$

$$= P_{\text{MD}} T_{\text{pilot}} \left(\boldsymbol{R}_{\boldsymbol{h},\check{b},p} - \boldsymbol{R}_{\boldsymbol{h},\check{b},b,k}\right), \tag{A2}$$

where $\boldsymbol{R}_{\boldsymbol{h},\check{b},b,k} = \mathrm{E}\left[\boldsymbol{h}_{\check{b},b,k}\boldsymbol{h}_{\check{b},b,k}^{\mathrm{H}}\right]$ is the covariance matrix of channel $\boldsymbol{h}_{\check{b},b,k}$ and $\boldsymbol{R}_{\boldsymbol{h},\check{b},p} = \sum_{(\hat{b},\hat{k})\in\mathcal{P}_p} \boldsymbol{R}_{\boldsymbol{h},\check{b},\hat{b},\hat{k}}$ is the sum of the covariance matrices of all channels from mobile devices using the pilot sequence with index p to base station $\check{b}$. As described in [107] and [131] for the point-to-point and the broadcast channel, respectively, the measurement error due to noise and interference during the piloting can be modeled as Gaussian noise for the data transmission. With an MMSE estimator, the error covariance matrix due to contamination is

$$\boldsymbol{R}_{\text{pc},\check{b},b,k} = \left(\boldsymbol{R}_{\boldsymbol{h},\check{b},b,k}^{-1} + P_{\text{MD}} T_{\text{pilot}} \left(\sigma_\eta^2 \mathbf{I} + \boldsymbol{\Theta}_{\text{pc},\check{b},b,k}\right)^{-1}\right)^{-1} \tag{A3}$$

$$= \boldsymbol{R}_{\boldsymbol{h},\check{b},b,k} + \boldsymbol{R}_{\boldsymbol{h},\check{b},b,k}\left(\frac{\sigma_\eta^2}{P_{\text{MD}} T_{\text{pilot}}}\mathbf{I} + \boldsymbol{R}_{\boldsymbol{h},\check{b},p}\right)^{-1} \boldsymbol{R}_{\boldsymbol{h},\check{b},b,k}, \tag{A4}$$

where σ_η^2 is the thermal noise.

Combined with the thermal noise, the channel state information outdating, and the background interference, the noise at mobile device (b,k) described in Section 2.1 can be extended to

$$\sigma_{b,k}^2 = \sigma_\eta^2 + \sigma_{\text{od},b,k}^2 + \frac{P}{N}\operatorname{tr}(\boldsymbol{R}_{\text{pc},b,b,k}) + \frac{P}{N}\sum_{\check{b}\in\mathcal{C}_{b,k}\setminus b} \operatorname{tr}(\boldsymbol{R}_{\text{pc},\check{b},b,k}) + \theta_{\text{bg}}. \tag{A5}$$

This still neglects the fact that the time division duplex measurements in the uplink and used for the downlink will always contain a calibration error because the reciprocity is only an idealized assumption.

In the following, different strategies for allocating pilot sequences to mobile devices with respect to pilot contamination are discussed.

A2.1 Random Allocation

For the random pilot sequence allocation strategy, the pilot sequences are assigned to the mobile devices based on their index. The base stations are randomly labeled with an index b and the mobile devices associated to base station b are also randomly labeled with the index k. The set of all mobile devices, which use the pilot sequence with index p is

$$\mathcal{P}_p^{\text{rand}} = \{(b,k) : (b,k) \in \mathcal{K}, p = \mod((b-1)K + k, T_{\text{pilot}})\}, \tag{A6}$$

where $\mod(a,b) = a - b\left\lfloor \frac{a}{b} \right\rfloor + 1$ is a slightly modified division algorithm. This assures that all mobile devices within one cell use a different pilot sequence, as long as $T_{\text{pilot}} \geq K$. Therefore, each base station can acquire a contaminated estimate of the channels of the associated mobile devices utilizing K of the pilot sequences. For each of the remaining L pilot sequences, each base station can get an estimate of a linear combination of the interference channels linking to mobile devices with the specific pilot sequence, respectively.

A2.2 Strongest Interferer Allocation

The upper bound in Section 2.4 reaches the highest rates, if the L known interference channels at each base station are the channels over which the base stations would cause the strongest decrease in rate, respectively. Finding these optimal interference channels is a problem, which is very hard to tackle. Therefore, it is proposed to suboptimally choose the L interference channels at each base station, which have the largest Euclidean norm, respectively.

To identify the channels a base station is supposed to measure, the set $\mathcal{L}_b$ is created for each base station, which contains the K mobile devices associated to the base station b and the L mobile devices with the strongest interference channels linked to base station b, respectively. The mobile devices in a set $\mathcal{L}_b$ should use pilot sequences, which are orthogonal to each other. Otherwise, the base stations could not get meaningful estimates of the channels. The goal is now to find the sets $\mathcal{P}_p$, which minimize the trace of the the sum of all pilot contaminations, whereas each pilot sequence has to appear in each set $\mathcal{L}_b$:

$$\begin{aligned}\left\{\mathcal{P}_1^{\text{opt}}, \ldots, \mathcal{P}_{T_{\text{pilot}}}^{\text{opt}}\right\} = \underset{\left\{\mathcal{P}_1,\ldots,\mathcal{P}_{T_{\text{pilot}}}\right\}}{\operatorname{argmin}} & \sum_{\check{b}\in\mathcal{B}} \sum_{(b,k)\in\mathcal{L}_{\check{b}}} \operatorname{tr}\left(\boldsymbol{\Theta}_{\text{pc},\check{b},b,k}\right) \\ & \text{s.t. } \mathcal{P}_p \cap \mathcal{L}_b \neq \emptyset \ \forall p, b.\end{aligned} \tag{A7}$$

The different sets $\mathcal{L}_b$ for the different base stations are overlapping. Therefore, it is not always feasible to allocate the pilot sequences to the mobile devices in such a way, that in every set $\mathcal{L}_b$ the mobile devices use all different pilot sequences. To get an always feasible optimization, the minimization (A7) is relaxed and different sets $\mathcal{P}_p^{\check{b},b,k}$ of mobile devices using the same pilot sequence are allowed for each channel estimation, respectively. Therefore,

there are no unique sets $\mathcal{P}_p$ any more and they have to be exchanged in equation (A1) with $\mathcal{P}_p^{\check{b},b,k}$ for each channel measurement.

$$\mathcal{P}_p^{\check{b},b,k} = \underset{\mathcal{P}_p}{\operatorname{argmin}} \operatorname{tr}\left(\boldsymbol{\Theta}_{\mathrm{pc},\check{b},b,k}\right) \quad \text{s.t. } \mathcal{P}_p \cap \mathcal{L}_{\tilde{b}} \neq \emptyset \ \forall \tilde{b}. \tag{A8}$$

For each channel measurement, the optimization will generate a different pilot sequence allocation. This is clearly not implementable, as the different pilot allocations may conflict with each other. An mobile device may be assigned to multiple pilot sequences, which is not possible.

A successive allocation algorithm is used to solve problem (A8) suboptimally. The pseudocode of the algorithm can be seen in Table A.3. The algorithm is initialized by setting $\mathcal{P}_p^{\check{b},b,k} = \{(b,k)\}$. In a random order, every set $\mathcal{L}_{\tilde{b}}$ is visited and $\mathcal{P}_p^{\check{b},b,k} \cap \mathcal{L}_{\tilde{b}}$ is checked. If the sets have a non empty intersection, i.e., an mobile device in the set $\mathcal{L}_{\tilde{b}}$ is already assigned to the same pilot as mobile device (b,k), this set is skipped. Otherwise, all mobile devices in the set $\mathcal{L}_{\check{b}}$ are removed from the set $\mathcal{L}_{\tilde{b}}$ because they have to use a different pilot sequence than mobile device (b,k). In the remaining set $\mathcal{L}_{\tilde{b}} \setminus \mathcal{L}_{\check{b}}$, the mobile device with the weakest channel linked to base station $\check{b}$ is picked. If all mobile devices would be removed from previously visited sets, the constraint $\mathcal{P}_p^{\check{b},b,k} \cap \mathcal{L}_{\tilde{b}} \neq \emptyset$, $\forall \tilde{b}$, might not be fulfilled.

Algorithm A.3 Strongest Interferer Allocation

Require:

for $\check{b} \in \mathcal{B}$ **do**
  for $(b,k) \in \mathcal{L}_{\check{b}}$ **do**
    $\mathcal{P}_p^{\check{b},b,k} \leftarrow \{(b,k)\}$
    for $\tilde{b} \in \mathcal{B} \setminus \check{b}$ **do**
      if $\mathcal{P}_p^{\check{b},b,k} \cap \mathcal{L}_{\tilde{b}} = \emptyset$ **then**
        $\mathcal{P}_p^{\check{b},b,k} \leftarrow \mathcal{P}_p^{\check{b},b,k} \cup \underset{(\hat{b},\hat{k}) \in \mathcal{L}_{\tilde{b}} \setminus \mathcal{L}_{\check{b}}}{\operatorname{argmin}} \operatorname{tr}(\boldsymbol{R}_{\boldsymbol{h},\check{b},\hat{b},\hat{k}})$
      end if
    end for
  end for
end for

On the one hand, because the sets $\mathcal{L}_b$ overlap, it is not always optimal to chose the mobile devices with the weakest channel in the interfering sets. It might be better to chose a mobile device with a stronger channel, which appears in multiple sets and, therefore, reduces the number of mobile devices in the set $\mathcal{P}_p^{\check{b},b,k}$. On the other hand, an implementable solution would force to allocate the pilot sequences to the mobile devices without any conflicts and, therefore, reduce the degrees of freedom. This approach is used to get an insight on the benefit of an optimized pilot allocation, where each base station can measure the mobile devices of its cell and the L strongest interferers.

A3 List of Used Symbols

Symbol	Element of	Description
$(\bullet)^{\text{achieved}}$		achieved value
$(\bullet)^{\text{actual}}$		actual value during the transmission
$(\bullet)^{\text{assumed}}$		assumed value during the optimization
$(\bullet)^{\text{CC}}$		Chase combining
$(\bullet)_{\text{coop}}$	$\mathbb{R}_0^+$	system with cooperation
$(\bullet)^{\text{HARQ}}$		hybrid automatic repeat request
$(\bullet)^{\text{id}}$		scaled identity shaping constraint
$(\bullet)^{\text{IR}}$		incremental redundancy
$(\bullet)_{\text{LF}}$		log fairness
$(\bullet)^{\text{lin}}$		linear constraint
$(\bullet)_{\lambda}$		associated with update of λ_k
$(\bullet)_{\text{MM}}$		max-min fairness
$(\bullet)_{\text{n}}$		associated with update of the noise covariance matrix
$(\bullet)_{\text{noco}}$	$\mathbb{R}_0^+$	system without cooperation
$(\bullet)^{\text{per ant}}$		per antenna power constraint
$(\bullet)_{\text{PF}}$		proportional fairness
$(\bullet)_{\text{RR}}$		round robin
$(\bullet)^{\text{shape}}$		shaping constraint
$(\bullet)_{\text{SR}}$		sum rate maximization
$(\bullet)_{\text{t}}$		associated with update of the transmit covariance matrices
$(\bullet)_{\text{TF}}$		throughput fairness
$(\bullet)^{\text{transmitted}}$		utilized value for the transmission
$(\bullet)_{\text{upper}}$	$\mathbb{R}_0^+$	upper bound system, measured interference channels disappear
$(\bullet)_{\text{zfici}}$	$\mathbb{R}_0^+$	all base stations transmit in the kernel of the measured interference channels
$\mathbf{0}$		all zeros matrix of appropriate size
$\emptyset$		empty set
∞		infinity
∂		partial derivative
a, a_l	$\mathbb{R}_0^+$	downlink linear constraint value
$\boldsymbol{a}^{\text{H}}$, $\boldsymbol{a}_l^{\text{H}}$	$\mathbb{C}^N$	downlink linear constraint direction vector
$\boldsymbol{A}$, $\boldsymbol{A}_l$	$\mathbb{C}^{N\times N}$	downlink linear constraint matrix
b, $\hat{b}$, $\check{b}$, $\tilde{b}$	$\mathcal{B}$	index of base station
(b,k), $(\hat{b},\hat{k})$	$\mathcal{K}$	index of mobile device in set of all mobile devices
B	$\mathbb{N}$	number of base stations
$\mathcal{B}$		set of base stations
$\mathcal{B}_c$	$\mathcal{B}$	set of base stations with orthogonal pilots
BS_b	$\mathcal{B}$	serving base station b

Symbol	Element of	Description
$\mathrm{BS}_{\hat{b}}$	$\mathcal{B}$	interfering base station $\hat{b}$ with measured channel
$\mathrm{BS}_{\check{b}}$	$\mathcal{B}$	interfering base station $\check{b}$ with **not** measured channel
c	$3 \cdot 10^8\,\mathrm{m/s}$	speed of light in air (Chapter 2)
c	$\mathbb{R}_0^+$	scaling factor of scaled identity matrix (Chapter 3)
c_k	$\mathbb{R}_0^+$	feasible rate of mobile device k
c_{ij}	$\mathbb{C}$	entry on the ith row and jth column of $\boldsymbol{C}$
C	$\mathbb{R}$	cardinality of $\mathcal{B}_c$
$\boldsymbol{C}$, $\boldsymbol{C}_l$	$\mathbb{C}^{N\times N}$	downlink shaping constraint matrix
$\mathcal{C}$		constraint set for the downlink transmit covariance matrix
$\mathcal{C}_{b,k}$	$\mathcal{B}$	set of base stations that measured the channel to mobile device (b,k)
$\mathcal{C}_\mathrm{n}$		constraint set for the uplink noise covariance matrices
$\mathbb{C}$		set of complex numbers
d	$\mathbb{N}$	inverse projected gradient step-size
d		differential operator
$\boldsymbol{e}_l$	$\{0,\ 1\}^N$	lth column of the identity matrix of appropriate size
E	$\mathbb{R}_0^+$	sum outdating error variance in each block
E[•]		expectation
exp(•)		$\mathrm{e}^{(\bullet)}$, where e is Euler's number, base of the natural logarithm
f	$\mathbb{R}$	frequency
$\hat{f}$	$\mathbb{R}$	frequency on the boundary of region with maximal correlation for given number of symbols
f_0	$\mathbb{R}_0^+$	reference frequency
$\hat{f}_0$	$\mathbb{R}_0^+$	maximum frequency distance to reference frequency
f_C	$\mathbb{R}_0^+$	carrier frequency
f_D	$\mathbb{R}_0^+$	maximum Doppler frequency
$f_X(x)$		probability density function of X at x
$F(x)$, $F_X(x)$		cumulative distribution function of X at x
g_l	$\mathbb{R}$	gradient of the uplink cost function with respect to the Lagrangian multiplier of linear constraint l projected onto the tangent cone of the constraint set
$\hat{g}_l$	$\mathbb{R}$	associated unconstrained gradient
$g_{\mathrm{t},k}$	$\mathbb{R}$	gradient of the uplink cost function with respect to the transmit variance of mobile device k (MU-SIMO)
$\hat{\boldsymbol{G}}_\mathrm{n}$	$\mathbb{C}^{N\times N}$	unconstrained gradient of the uplink cost function with respect to the noise covariance matrix
$\boldsymbol{G}_{\mathrm{n},l}$	$\mathbb{C}^{N\times N}$	gradient of the uplink cost function with respect to the Lagrangian multiplier of constraint l projected onto the tangent cone of the constraint set
$\hat{\boldsymbol{G}}_{\mathrm{n},l}$	$\mathbb{C}^{N\times N}$	associated unconstrained gradient

Symbol	Element of	Description
$\boldsymbol{G}_{\mathrm{t},k}$	$\mathbb{C}^{M\times M}$	gradient of the uplink cost function with respect to the transmit covariance matrix of mobile device k projected onto the tangent cone of the constraint set (MU-MIMO)
$\hat{\boldsymbol{G}}_{\mathrm{t},k}$	$\mathbb{C}^{M\times M}$	associated unconstrained gradient (MU-MIMO)
h_i	$\mathbb{R}$	ith entry of $\boldsymbol{h}^{\mathrm{T}}$
$\boldsymbol{h}^{\mathrm{T}}$	$\mathbb{R}^{1\times N}$	general, real valued MISO downlink channel
$\boldsymbol{h}_k$	$\mathbb{C}^{N\times 1}$	SIMO uplink channel of mobile device k
$\boldsymbol{h}_k^{\mathrm{H}}$	$\mathbb{C}^{1\times N}$	MISO downlink channel of mobile device k
$\boldsymbol{h}_{\hat{b},b,k}$	$\mathbb{C}^{N}$	channel between base station $\hat{b}$ and mobile device (b,k)
$\hat{\boldsymbol{h}}_{\hat{b},b,k}$	$\mathbb{C}^{N}$	associated channel estimate
$\boldsymbol{H}_k^{\mathrm{H}}$, $\boldsymbol{H}_k^{\mathrm{downlink}}$	$\mathbb{C}^{M\times N}$	downlink channel of mobile device k
$\boldsymbol{H}_k$, $\boldsymbol{H}_k^{\mathrm{uplink}}$	$\mathbb{C}^{N\times M}$	uplink channel of mobile device k
$\tilde{\boldsymbol{H}}_k$	$\mathbb{C}^{N\times M}$	substituted uplink channel of mobile device k
i	$\mathbb{N}$	iteration counter
$\mathbf{I}$		identity matrix of appropriate size
$\mathrm{J}_0(\bullet)$		first kind and zeroth order Bessel function
k, $\hat{k}$, k_1, k_2	$\mathcal{K}_b$, $\mathcal{K}$	index of mobile device associated with base station b or in single cell network
(k,m)	$\{(1,1),\ldots,(K,M)\}$	index of eigenvalue of all $\boldsymbol{\Xi}_k$
K	$\mathbb{N}$	number of mobile devices per base station
$\hat{K}$	$\mathbb{N}$	number of mobile devices scheduled per time slot
$\mathcal{K}$		set of all mobile devices in the network
$\mathcal{K}_b$	$\mathcal{K}$	set of mobile devices associated with base station b
$\mathcal{K}_c$	$\mathcal{K}$	set of mobile devices with orthogonal pilots
$\mathcal{K}^{(t)}$	$\mathcal{K}$	set of mobile devices scheduled in slot (t)
l	$\{1,\ldots,L\}$	index of constraint
L	$\mathbb{N}$	number of measured interference channels per base station (Chapter 2)
L	$\mathbb{N}$	number of downlink constraints (Chapter 3 and 5)
$\hat{L}$	$\mathbb{N}$	number of positive ω_l
$\mathcal{L}$, $\mathcal{L}(\bullet)$		Lagrangian function
$\hat{\mathcal{L}}$		set of indices associated with positive ω_l, ω_l that are zero
$\tilde{\mathcal{L}}$		set of indices associated with ω_l that are zero
$\mathcal{L}_b$		set of mobile devices to which base station b measured the channel
m	$\{1,\ldots,M\}$	index of eigenvalue of $\boldsymbol{\Xi}_k$, $\hat{\boldsymbol{\Xi}}_k$, $\forall k$
M	$\mathbb{N}$	number of antennas at each mobile device
$\hat{M}$	$\mathbb{N}$	number of positive eigenvalues of all $\boldsymbol{\Xi}_k$

Symbol	Element of	Description
$\mathrm{MD}_{b,k}$	$\mathcal{K}$	mobile device (b,k)
$\hat{\mathcal{M}}$		set of indices associated with positive eigenvalues of all $\boldsymbol{\Xi}_k$
$\mathrm{mod}\ (a,b)$		modified division algorithm $\mathrm{mod}\ (a,b) = a - b\left\lfloor \frac{a}{b} \right\rfloor + 1$
n	$\{1,\ldots,N\}$	index of eigenvalue of $\boldsymbol{\Omega}$ (Chapter 3)
n	$\mathbb{N}$	HARQ process index (Chapter 4)
$\boldsymbol{n}_k$	$\mathbb{C}^M$	downlink noise at mobile device k
N	$\mathbb{N}$	number of antennas at each base station
$\hat{N}$	$\mathbb{N}$	number of positive eigenvalues of $\boldsymbol{\Omega}$
$\hat{\mathcal{N}}$		set of indices associated with positive eigenvalues of $\boldsymbol{\Omega}$
$\mathbb{N}$		set of natural numbers including zero
p	$\mathbb{R}_0^+$	preconditioning factor of the projected gradient update (Chapter 3)
p	$\mathbb{N}$	index of pilot sequence (Appendix A2)
P	$\mathbb{R}_0^+$	sum transmit power
P_l	$\mathbb{R}_0^+$	transmit power of antenna l
P_{MD}	$\mathbb{R}_0^+$	mobile device transmit power
$P(\bullet)$		probability mass function
$\mathcal{P}_p$, $\mathcal{P}_p^{\check{b},b,k}$		set of mobile devices using the same pilot sequence p
q_k	$\mathbb{R}_0^+$	SIMO uplink transmit variance of mobile device k
$\hat{q}_k$	$\mathbb{R}_0^+$	unconstrained update of transmit variance
$\tilde{q}_k$	$\mathbb{R}_0^+$	temporary projected update of transmit variance
q_{ij}	$\mathbb{C}$	entry on the ith row and jth column of $\boldsymbol{Q}$
$\boldsymbol{Q}$, $\boldsymbol{Q}_b$	$\mathbb{C}^{N\times N}$	downlink sum transmit covariance matrix
$\boldsymbol{Q}_k$, $\boldsymbol{Q}_{b,k}$	$\mathbb{C}^{N\times N}$	downlink transmit covariance matrix for a mobile device
$\hat{\boldsymbol{Q}}_b$, $\hat{\boldsymbol{Q}}_{b,k}$	$\mathbb{C}^{N-L\times N-L}$	transmit covariance matrices in kernel of measured interference channels
$\tilde{\boldsymbol{Q}}$, $\tilde{\boldsymbol{Q}}_k$	$\mathbb{C}^{N\times N}$	substituted downlink transmit covariance matrices
$\boldsymbol{Q}^{\mathrm{diag}}$	$\mathbb{R}^{2\times 2}$	diagonal downlink sum transmit covariance matrix
$\boldsymbol{Q}^{\mathrm{Re}}$	$\mathbb{R}^{2\times 2}$	real valued downlink sum transmit covariance matrix
r, r_k, $r_{b,k}$	$\mathbb{R}_0^+$	rate of mobile device in bits per channel usage
$\bar{r}_k$	$\mathbb{R}_0^+$	average rate in one block in bits per channel usage
r_k^{downlink}	$\mathbb{R}_0^+$	downlink rate of mobile device k in bits per channel usage
r_k^{uplink}	$\mathbb{R}_0^+$	uplink rate of mobile device k in bits per channel usage
$\boldsymbol{r}$	$\mathbb{R}_0^{+,K}$	rate vector in bits per channel usage
$\boldsymbol{r}^\star$	$\mathbb{R}_0^{+,K}$	optimal rate vector
R_k	$\mathbb{R}_0^+$	historical average rate in bits per channel
$\boldsymbol{R}_{\boldsymbol{h},\check{b},b,k}$	$\mathbb{C}^{N\times N}$	covariance matrix of channel $\boldsymbol{h}_{\check{b},b,k}$
$\boldsymbol{R}_{\boldsymbol{h},\check{b},p}$	$\mathbb{C}^{N\times N}$	sum covariance matrix of channels from mobile devices with pilot p to base station $\check{b}$
$\boldsymbol{R}_{\mathrm{pc},\check{b},b,k}$	$\mathbb{C}^{N\times N}$	contamination error covariance matrix of channel $\boldsymbol{h}_{\check{b},b,k}$
$\mathcal{R}$		rate region

Symbol	Element of	Description
$\mathbb{R}$		set of rational numbers
$\mathbb{R}_0^+$		set of non-negative rational numbers
$\mathbb{R}^+$		set of positive rational numbers
s	$\mathbb{R}_0^+$	projected gradient step-size
$\boldsymbol{S}_k$	$\mathbb{C}^{M\times M}$	Lagrangian multiplier for the positive-semidefinite constraint on $\boldsymbol{\Xi}_k$
$\boldsymbol{S}_l$	$\mathbb{C}^{N\times N}$	Lagrangian multiplier for the positive-semidefinite constraint on $\boldsymbol{\Omega}_l$
t	$\mathbb{R}$	time (Capter 2)
t	$\mathcal{T}$	index of union of all eigenvalues λ_n^{id} and ω_l (Chapter 3)
t	$\{1,\ldots,T\}$	HARQ transmission index (Chapter 4)
t, τ	$\{1,\ldots,$ $T_{\mathrm{block}}\}$	slot index (Chapter 5)
$\hat{t}$	$\mathbb{R}$	time on the boundary of region with maximal correlation for given number of symbols
t_0	$\mathbb{R}$	reference time
$\hat{t}_0$	$\mathbb{R}$	maximum time distance to reference time
T	$\mathbb{N}$	maximal number of transmissions in an HARQ process
$\hat{T}_k$	$\{1,\ldots,T\}$	actual number of transmissions required in an HARQ process
T_{2nd}	$\mathbb{N}$	length of second pilot
T_{block}	$\mathbb{N}$	total number of symbols available in each block
$T_{\mathrm{chan\ FB}}$	$\mathbb{N}$	number of symbols for channel feedback in each block
T_{data}	$\mathbb{N}$	number of symbols for data transmission in each block
$T_{\mathrm{DL\ pilots}}$	$\mathbb{N}$	length of downlink pilot
T_{pilot}	$\mathbb{N}$	length of piloting sequence
$T_{\mathrm{SINR\ FB}}$	$\mathbb{N}$	number of symbols for SINR feedback in each block
$T_{\mathrm{UL\ pilots}}$	$\mathbb{N}$	length of uplink pilot
$\boldsymbol{T}_l$	$\mathbb{C}^{N\times N}$, $\mathcal{Z}_l$	Lagrangian multiplier for the subspace constraint on $\boldsymbol{\Omega}_l$
$\mathcal{T}$		set of indices unifying all eigenvalues $\hat{\lambda}_n^{\mathrm{id}}$ and $\hat{\omega}_l$
$\hat{\mathcal{T}}$		union of $\hat{\mathcal{N}}$ and $\hat{\mathcal{L}}$
U, $U(\bullet)$		monotonic utility function
$\hat{U}$	$\mathbb{R}_0^+$	specific utility value
$\hat{\boldsymbol{U}}_k$	$\mathbb{C}^{M\times M}$	modal matrix of $\boldsymbol{\Xi}_k$, $\hat{\boldsymbol{\Xi}}_k$
v	$\mathbb{R}_0^+$	common mobile device speed in km/h
$\boldsymbol{v}$	$\mathbb{C}^N$	uplink sum receive signal
$\boldsymbol{V}_b$	$\mathbb{C}^{N\times N-L}$	kernel of measured interference channels at base station b
w_k	$\mathbb{R}_0^+$	weighted sum rate weight
$\mathrm{W}_p(\boldsymbol{V},n)$		Wishart distribution of size p with n degrees of freedom and shaping matrix $\boldsymbol{V}$
x_i, x_j	$\mathbb{C}$	ith and jth element of $\boldsymbol{x}$
$\boldsymbol{x}$	$\mathbb{C}^N$	downlink sum transmit signal

Symbol	Element of	Description
$\boldsymbol{x}_k$	$\mathbb{C}^N$	downlink transmit signal for mobile device k
$\boldsymbol{y}_k$	$\mathbb{C}^M$	downlink receive signal at mobile device k
$\boldsymbol{Z}$, $\boldsymbol{Z}_l$	$\mathcal{Z}$, $\mathcal{Z}_l$	downlink conic constraint apex shifting matrix
$\mathcal{Z}$, $\mathcal{Z}_l$		downlink conic constraint subspace of feasible apex shifting
$\mathcal{Z}^\perp$, $\mathcal{Z}_l^\perp$		orthogonal subspace to $\mathcal{Z}$, $\mathcal{Z}_l$
α	$\mathbb{R}_0^+$	scaling variable for interference temperatures (Chapter 2)
α	$[1, \infty[$	loosened covariance shaping loosening factor (Chapter 4)
α_k	$\mathbb{R}_0^+$	difference weight in the weighted sum rate of summand k
α, β	$\mathbb{R}_0^+$	variables selecting type of fairness in fairness scheduling (Chapter 5)
β	$[0, 1]$	backoff factor for conservative gambling (Chapter 4)
β_k	$[0, 1]$	backoff factor for mobile device k of conservative gambling with individual backoff factors
γ	$\mathbb{R}_0^+$	intercell interference limit / interference temperature (Chapter 2)
γ_k	$\mathbb{R}_0^+$	SINR of mobile device k (Chapter 4)
γ	$\mathbb{R}$	variable selecting type of fairness with log fairness (Chapter 5)
$\gamma_{b,\hat{b},\hat{k}}$	$\mathbb{R}_0^+$	interference temperature for base station b at mobile device $(\hat{b}, \hat{k})$
$\Gamma(a_k, b_k)$, $\Gamma(a_{b,k}, b_{b,k})$		gamma distribution with shape parameter a_k, $a_{b,k}$ and scale parameter b_k, $b_{b,k}$
ε	$\mathbb{R}_0^+$	sufficient accuracy in the algorithms
ϵ	$\mathbb{R}_0^+$	feasible step size in the tangent cone projection (Chapter 3)
ϵ	$\mathbb{R}_0^+$	vanishing variable of a numerical gradient (Chapter 4)
ϵ	$\mathbb{R}_0^+$	vanishing variable used in log fairness (Chapter 5)
θ_k, $\theta_{b,k}$	$\mathbb{R}_0^+$	intercell interference variance
$\theta_{\hat{b},b,k}$	$\mathbb{R}_0^+$	intercell interference variance summand at mobile device (b, k) caused by base station $\hat{b}$
θ_{bg}	$\mathbb{R}_0^+$	out-of-cluster intercell interference variance
$\theta_{\mathrm{blind},b,k}$	$\mathbb{R}_0^+$	intercell interference variance over not measured channels at mobile device (b, k)
$\theta_{\max}$	$\mathbb{R}_0^+$	maximal intercell interference variance in the system
$\boldsymbol{\Theta}_{\mathrm{pc},\check{b},b,k}$	$\mathbb{C}^{N\times N}$	pilot contamination of channel $\boldsymbol{h}_{\check{b},b,k}$
λ	$\mathbb{R}_0^+$	wave length
λ_k	$\mathbb{R}_0^+$	Lagrangian multiplier for the constraint $r_k \leq c_k$
$\lambda_{k,m}$, $\hat{\lambda}_{k,m}$	$\mathbb{R}_0^+$	mth eigenvalues of $\boldsymbol{\Xi}_k$, $\hat{\boldsymbol{\Xi}}_k$
$\boldsymbol{\Lambda}$, $\hat{\boldsymbol{\Lambda}}$	$\mathbb{R}^{M\times M}$	diagonal matrix with eigenvalues of $\boldsymbol{\Xi}_k$, $\hat{\boldsymbol{\Xi}}_k$
μ	$\mathbb{R}_0^+$	Lagrangian multiplier of the sum power/trace constraint
ν	$\mathbb{R}$	Lagrangian multiplier of sum constraint in tangent cone projection

Symbol	Element of	Description
$\boldsymbol{\nu}$	$\mathbb{C}^N$	uplink noise
ξ	$\mathbb{R}_0^+$	signaling efficiency
$\hat{\xi}_t$	$\mathbb{R}_0^+$	union of all eigenvalues $\hat{\lambda}_n^{\mathrm{id}}$ and Lagrangian multipliers $\hat{\omega}_l$
$\boldsymbol{\xi}_k$	$\mathbb{C}^M$	uplink transmit signal of mobile device k
$\boldsymbol{\Xi}_k$	$\mathbb{C}^{M\times M}$	uplink transmit covariance matrix of mobile device k
$\hat{\boldsymbol{\Xi}}_k$	$\mathbb{C}^{M\times M}$	unconstrained update of transmit covariance matrix
$\tilde{\boldsymbol{\Xi}}_k$	$\mathbb{C}^{M\times M}$	temporary projected update of transmit covariance matrix
Π		mobile device order in the weighted sum rate
$\rho(t,f)$	$\mathbb{R}$	correlation coefficient between channel at (t_0, f_0) and outdated channel at (t_0+t, f_0+f)
$\hat{\rho}(\bullet)$		boundary of region with maximal correlation for given number of symbols
$\sigma_{b,k}^2$	$\mathbb{R}_0^+$	sum noise variance at mobile device (b,k)
σ_{DS}	$\mathbb{R}_0^+$	root mean square delay spread
σ_e^2(t,f)	$\mathbb{R}_0^+$	normalized channel state outdating error variance between channel at (t_0, f_0) and outdated channel at (t_0+t, f_0+f)
$\bar{\sigma}_e^2$	$\mathbb{R}_0^+$	mean normalized channel state outdating error variance
σ_η^2	$\mathbb{R}_0^+$	thermal noise variance
$\sigma_{\mathrm{od},b,k}^2$	$\mathbb{R}_0^+$	sum channel state information outdating error variance at mobile device (b,k)
Φ, $\Phi(\bullet)$		uplink cost function
ω_l	$\mathbb{R}_0^+$	Lagrangian multiplier of downlink linear constraint
$\hat{\omega}_l$	$\mathbb{R}_0^+$	unconstrained update of the Lagrangian multiplier
$\boldsymbol{\Omega}$	$\mathbb{C}^{N\times N}$, $\mathcal{Z}^\perp$	uplink noise covariance matrix
$\boldsymbol{\Omega}_l$	$\mathbb{C}^{N\times N}$, $\mathcal{Z}_l^\perp$	Lagrangian multiplier of the lth downlink conic constraint
$\hat{\boldsymbol{\Omega}}_l$	$\mathbb{C}^{N\times N}$	unconstrained update of Lagrangian multiplier
$\tilde{\boldsymbol{\Omega}}_l$	$\mathbb{C}^{N\times N}$	temporary projected update of Lagrangian multiplier
ψ_t	$\mathbb{R}_0^+$	union of c and all a_l
Ψ, $\Psi(\bullet)$		downlink cost function

A4 List of Used Acronyms

Acronym	Description
3GPP	3rd Generation Partnership Project telecommunications associations collaboration
ARQ	automatic repeat request
BC2MAC	downlink-uplink transformation of the transmit covariance matrices (Section 3.4.5)
BS	base station
CC	Chase combining
CDF	cumulative distribution function

Acronym	Description
CoMP	cooperative multi-point
CSI	channel state information
DL	downlink
DS	delay spread
FB	feedback
FDD	frequency division duplex
HARQ	hybrid automatic repeat request
HARQ-CC	hybrid automatic repeat request with Chase combining
HARQ-IR	hybrid automatic repeat request with incremental redundancy
IR	incremental redundancy
LB NC	lower bound without cooperation
LB ZF	lower bound with zero forcing of the intercell interference
LF	log fairness
LTE	Long-Term Evolution wireless communication standard
MAC2BC	recovery of the downlink transmit covariance matrices (Section 3.4.5)
Max-SINR	maximal signal-to-interference-plus-noise-ratio as cost function
MD	mobile device
MIMO	multiple-input and multiple-output
MM	max-min fairness
MMSE	minimum mean square error
MU-MIMO	multi-user multiple-input and multiple-output
NoiseCovStep	projected gradient update of the noise covariance matrix (Section 3.5.2)
OptRates	select rates in fairness beamforming algorithm (Section 5.3.3)
PF	proportional fairness
RobustSR	intercell interference robust sum rate maximization (Chapter 4)
RR	round robin
SINR	signal-to-interference-plus-noise-ratio
SNR	signal-to-noise-ratio
SR	sum rate maximization
TDD	time division duplex
TF	throughput fairness
TransCovStep	projected gradient update of the transmit covariance matrices (Section 3.5.1)
UL	uplink
ULRates	calculate uplink rates (Section 3.4.1)

Bibliography

[1] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation, and Reduced-Complexity Scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[2] A. Lozano, R. W. Heath, Jr., and J. G. Andrews, "Fundamental Limits of Cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.

[3] H. H. Brunner, M. T. Ivrlač, and J. A. Nossek, "Upper Bound to Interference Coordination with Channel State Information Outdating," in *2011 11th European Wireless Conf.*, Vienna, Austria, Apr. 27–29, 2011.

[4] M. T. Ivrlač and J. A. Nossek, "Intercell-Interference in the Gaussian MISO Broadcast Channel," in *2007 IEEE Global Telecommun. Conf. (GLOBECOM)*, Washington, D.C., USA, Nov. 26–30, 2007, pp. 3195–3199.

[5] G. Fodor and C. Koutsimanis, "A Low Intercell Interference Variation Scheduler for OFDMA Networks," in *2008 IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 19–23, 2008, pp. 3078–3084.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity. Part I. System Description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[7] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[8] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded Cooperation in Wireless Communications: Space-Time Transmission and Iterative Decoding," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 362–371, Feb. 2004.

[9] G. J. Foschini, K. Karakayali, and R. A. Valenzuela, "Coordinating Multiple Antenna Cellular Networks to Achieve Enormous Spectral Efficiency," *IEE Proc. Commun.*, vol. 153, no. 4, pp. 548–555, Aug. 2006.

[10] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming Intercell Interference in Indoor Wireless Systems," in *2007 Conf. Rec. 41st Asilomar Conf. Signals, Syst. and Comput. (ACSSC)*, Pacific Grove, CA, USA, Nov. 4–7, 2007, pp. 83–87.

[11] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, "Increasing Downlink Cellular Throughput with Limited Network MIMO Coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983–2989, Jun. 2009.

[12] S. W. Peters and R. W. Heath, Jr., "User Partitioning for Less Overhead in MIMO Interference Channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 592–603, Feb. 2012.

[13] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[14] P. Wang, H. Wang, L. Ping, and X. Lin, "On the Capacity of MIMO Cellular Systems with Base Station Cooperation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3720–3731, Nov. 2011.

[15] O. Simeone *et al.*, "Cooperative Wireless Cellular Systems: An Information-Theoretic View," *Found. and Trends® Commun. and Inf. Theory*, vol. 8, no. 1-2, pp. 1–177, Aug. 2011.

[16] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Local Base Station Cooperation Via Finite-Capacity Links for the Uplink of Linear Cellular Networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 190–204, Jan. 2009.

[17] R. Irmer *et al.*, "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[18] A. Barbieri *et al.*, "Coordinated Downlink Multi-Point Communications in Heterogeneous Cellular Networks," in *2012 Inf. Theory and Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 5–10, 2012, pp. 7–16.

[19] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[20] J. G. Andrews, W. Choi, and R. W. Heath, Jr., "Overcoming Interference in Spatial Multiplexing MIMO Cellular Networks," *IEEE Wireless Commun.*, vol. 14, no. 6, pp. 95–104, Dec. 2007.

[21] V. H. Mac Donald, "The Cellular Concept," *Bell System Tech. J.*, vol. 58, no. 1, pp. 15–41, Jan. 1979.

[22] W. Choi and J. G. Andrews, "The Capacity Gain from Intercell Scheduling in Multi-Antenna Systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 714–725, Feb. 2008.

[23] A. Dotzler, W. Utschick, and G. Dietl, "Fractional Reuse Partitioning for MIMO Networks," in *2010 IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010.

[24] A. Saleh, A. Rustako, and R. Roman, "Distributed Antennas for Indoor Radio Communications," *IEEE Trans. Commun.*, vol. 35, no. 12, pp. 1245–1251, Dec. 1987.

[25] H. Dai, A. F. Molisch, and H. V. Poor, "Downlink Capacity of Interference-Limited MIMO Systems with Joint Detection," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 442–453, Mar. 2004.

[26] J. G. Andrews, "Interference Cancellation for Cellular Systems: a Contemporary Overview," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 19–29, Apr. 2005.

[27] A. Dotzler, M. Riemensberger, and W. Utschick, "Minimax Duality for MIMO Interference Networks," *MDPI Inform.*, vol. 7, no. 2, pp. 1–42, Mar. 2016. [Online]. Available: http://www.mdpi.com/2078-2489/7/2/19

[28] R. Hunger, *Analysis and Transceiver Design for the MIMO Broadcast Channel*, 1st ed., ser. Foundations in Signal Processing, Communications and Networking, H. Boche, R. Mathar, and W. Utschick, Eds. Heidelberg, Germany: Springer, 2013.

[29] H. Shirani-Mehr, H. Papadopoulos, S. A. Ramprashad, and G. Caire, "Joint Scheduling and ARQ for MU-MIMO Downlink in the Presence of Inter-Cell Interference," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 578–589, Feb. 2011.

[30] H. H. Brunner and J. A. Nossek, "Handling Unknown Interference in Cellular Networks with Interference Coordination," in *2012 16th Int. ITG Workshop Smart Antennas (WSA)*, Dresden, Germany, Mar. 7–8, 2012, pp. 133–139.

[31] R. Bendlin, H. H. Brunner, M. T. Ivrlač, J. A. Nossek, and Y.-F. Huang, "Two-Phase Scheduling and Leakage-Based Precoding in Wireless Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3148–3157, Jul. 2013.

[32] A. Dotzler, M. Riemensberger, W. Utschick, and G. Dietl, "Interference Robustness for Cellular MIMO Networks," in *2012 IEEE 13th Int. Workshop Signal Process. Advances Wireless Commmun. (SPAWC)*, Cesme, Turkey, Jun. 17–20, 2012, pp. 229–233.

[33] H. H. Brunner, J. Braun, A. Mezghani, and J. A. Nossek, "Precoding for Systems with Soft Combining to Counteract Instationary Intercell Interference," in *2014 IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Florence, Italy, May 4–9, 2014, pp. 1150–1154.

[34] M. K. Karakayali, G. J. Foschini, R. A. Valenzuela, and R. D. Yates, "On the Maximum Common Rate Achievable in a Coordinated Network," in *2006 IEEE Int. Conf. Commun. (ICC)*, Istanbul, Turkey, Jun. 11–15, 2006, pp. 4333–4338.

[35] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving "Massive MIMO" Spectral Efficiency with a Not-so-Large Number of Antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.

[36] G. Caire, S. A. Ramprashad, and H. C. Papadopoulos, "Rethinking Network MIMO: Cost of CSIT, Performance Analysis, and Architecture Comparisons," in *2010 Inf. Theory and Appl. Workshop (ITA)*, San Diego, CA, USA, Jan. 31–Feb. 5, 2010.

[37] S. A. Ramprashad and G. Caire, "Cellular vs. Network MIMO: A Comparison Including the Channel State Information Overhead," in *2009 IEEE 20th Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Tokyo, Japan, Sep. 13–16, 2009, pp. 878–884.

[38] H. H. Brunner, M. T. Ivrlač, and J. A. Nossek, "Hard Limit to Interference Coordination," to be submitted for publication.

[39] 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.

[40] 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.

[41] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the Capacity of Wireless Networks through Distributed Interference Alignment," in *2008 IEEE Global Telecommun. Conf. (GLOBECOM)*, New Orleans, LO, USA, Nov. 30–Dec. 4 2008.

[42] H. H. Brunner and J. A. Nossek, "Mitigation of Intercell Interference without Base Station Cooperation," in *2010 14th Int. ITG Workshop Smart Antennas (WSA)*, Bremen, Germany, Feb. 23–24, 2010, pp. 1–7.

[43] A. Dotzler, W. Utschick, and G. Dietl, "Efficient Zero-Forcing Based Interference Coordination for MISO Networks," in *2011 IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, Yokohama, Japan, May 15–18, 2011.

[44] H. Huh, H. C. Papadopoulos, and G. Caire, "Multiuser MISO Transmitter Optimization for Intercell Interference Mitigation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4272–4285, Aug. 2010.

[45] J. Huang, R. A. Berry, and M. L. Honig, "Distributed Interference Compensation for Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[46] 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.

[47] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A Distributed Numerical Approach to Interference Alignment and Applications to Wireless Interference Networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[48] S. W. Peters and R. W. Heath, Jr., "Cooperative Algorithms for MIMO Interference Channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.

[49] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[50] R. Tresch and M. Guillaud, "Clustered Interference Alignment in Large Cellular Networks," in *2009 IEEE 20th Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Tokyo, Japan, Sep. 13–16, 2009, pp. 1024–1028.

[51] W. Mennerich and W. Zirwas, "Reporting Effort for Cooperative Systems Applying Interference Floor Shaping," in *2011 IEEE 22nd Int. Symp. Personal Indoor and Mobile Radio Commun. (PIMRC)*, Toronto, ON, USA, Sep. 11–14, 2011, pp. 541–545.

[52] R. K. Mungara, D. Morales-Jimenez, and A. Lozano, "System-Level Performance of Interference Alignment," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1060–1070, Feb. 2015.

[53] D. Aziz, F. Boccardi, and A. Weber, "System-Level Performance Study of Interference Alignment in Cellular Systems with Base-Station Coordination," in *2012 IEEE 23rd Int. Symp. Personal Indoor and Mobile Radio Commun. (PIMRC)*, Sydney, Australia, Sep. 9–12, 2012, pp. 1155–1160.

[54] C. Suh, M. Ho, and D. N. C. Tse, "Downlink Interference Alignment," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2616–2626, Sep. 2011.

[55] N. Jindal and A. Lozano, "A Unified Treatment of Optimum Pilot Overhead in Multipath Fading Channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[56] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot Contamination Problem in Multi-Cell TDD Systems," in *2009 IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, Jun. 28–Jul. 3, 2009, pp. 2184–2188.

[57] M. H. Castañeda, I. Slim, A. Mezghani, and J. A. Nossek, "Transceiver Design in Multiuser MISO Systems with Imperfect Transmit CSI," in *2010 14th Int. ITG Workshop Smart Antennas (WSA)*, Bremen, Germany, Feb. 23–24, 2010, pp. 442–449.

[58] H. H. Brunner, M. H. Castañeda, and J. A. Nossek, "How Much Training is Needed for Interference Coordination in Cellular Networks?" in *2011 15th Int. ITG Workshop Smart Antennas (WSA)*, Aachen, Germany, Feb. 24–25, 2011.

[59] H. H. Brunner and J. A. Nossek, "Interference Temperature Analysis in a Large Scale Cellular System," in *2013 17th Int. ITG Workshop Smart Antennas (WSA)*, Stuttgart, Germany, Mar. 13–14, 2013.

[60] M. Costa, "Writing on Dirty Paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[61] T. L. Marzetta and B. M. Hochwald, "Fast Transfer of Channel State Information in Wireless Systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.

[62] A. Lozano and D. Porrat, "Non-Peaky Signals in Wideband Fading Channels: Achievable Bit Rates and Optimal Bandwidth," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 246–257, Jan. 2012.

[63] S. O. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ, USA: Prentice Hall, 2001.

[64] W. C. Jakes, *Microwave Mobile Communications*, 2nd ed. New York, NY, USA: Wiley-IEEE-Press, 1994.

[65] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[66] S. A. Jafar and S. Vishwanath, "Generalized Degrees of Freedom of the Symmetric Gaussian K User Interference Channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3297–3303, Jul. 2010.

[67] H. Huh, H. Papadopoulos, and G. Caire, "MIMO Broadcast Channel Optimization under General Linear Constraints," in *2009 IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, Jun. 28–Jul. 3, 2009, pp. 2664–2668.

[68] "Spacial Channel Model for Multiple Input Multiple Output (MIMO) Simulations," ETSI 3GPP, Sophia Antipolis Cedex, France, Tech. Rep. 3GPP TR 25.996 version 9.0.0 Release 9, Dec. 2009.

[69] W. Yu, "Uplink-Downlink Duality via Minimax Duality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 361–374, Feb. 2006.

[70] L. Zhang, R. Zhang, Y. Liang, Y. Xin, and H. V. Poor, "On Gaussian MIMO BC-MAC Duality With Multiple Transmit Covariance Constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.

[71] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[72] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Oct. 2003.

[73] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum Power Iterative Water-Filling for Multi-Antenna Gaussian Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[74] R. Hunger, D. A. Schmidt, M. Joham, and W. Utschick, "A General Covariance-Based Optimization Framework Using Orthogonal Projections," in *2008 IEEE 9th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Recife, Brasil, Jul. 6–9, 2008, pp. 76–80.

[75] A. Liu, Y. Liu, H. Xiang, and W. Luo, "Iterative Polite Water-Filling for Weighted Sum-Rate Maximization in iTree Networks," in *2010 IEEE Global Telecommun. Conf. (GLOBECOM)*, Miami, FL, USA, Dec. 6–10, 2010.

[76] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative Water-Filling for Gaussian Vector Multiple-Access Channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[77] W. Yu and T. Lan, "Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[78] A. Liu, Y. Liu, H. Xiang, and W. Luo, "Polite Water-Filling for Weighted Sum-Rate Maximization in MIMO B-MAC Networks Under Multiple Linear Constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 834–847, Feb. 2012.

[79] H. H. Brunner, A. Dotzler, W. Utschick, and J. A. Nossek, "Weighted Sum Rate Maximization with Multiple Linear Conic Constraints," in *2015 IEEE Int. Conf. Commun. (ICC)*, London, United Kingdom, Jun. 8–12, 2015, pp. 4635–4640.

[80] ——, "Intercell Interference Robustness Tradeoff with Loosened Covariance Shaping," in *2014 18th Int. ITG Workshop Smart Antennas (WSA)*, Erlangen, Germany, Mar. 12–13, 2014.

[81] I. E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Trans. Telecommun.*, vol. 10, pp. 585–596, Nov. 1999.

[82] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge University Press, 2004.

[83] T. Wiegart, C. Hellings, and W. Utschick, "Close-to-Optimal Partial Decode-and-Forward Rate in the MIMO Relay Channel via Convex Programming," in *2016 20th Int. ITG Workshop Smart Antennas (WSA)*, Munich, Germany, Mar. 9–11, 2016.

[84] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Stanford, CA, USA: Stanford University Press, 1958.

[85] A. A. Goldstein, "Convex Programming in Hilbert Space," *Bull. Am. Math. Soc.*, vol. 70, no. 5, pp. 709–710, May 1964.

[86] E. S. Levitin and B. T. Polyak, "Constrained Minimization Methods," *USSR Comput. Math. Math. Phys.*, vol. 6, no. 5, pp. 1–50, Dec. 1966.

[87] G. P. McCormick and R. A. Tapia, "The Gradient Projection Method under Mild Differentiability Conditions," *SIAM J. Control*, vol. 10, no. 1, pp. 93–98, 1972.

[88] D. P. Bertsekas, "On the Goldstein-Levitin-Polyak Gradient Projection Method," *IEEE Trans. Autom. Control*, vol. 21, no. 2, pp. 174–184, 1976.

[89] J. C. Dunn, "Global and Asymptotic Convergence Rate Estimates for a Class of Projected Gradient Processes," *SIAM J. Control Optim.*, vol. 19, no. 3, pp. 368–400, 1981.

[90] M. Kallio and C. H. Rosa, "Large-Scale Convex Optimization via Saddle-Point Computation," *Operations Research*, vol. 47, no. 1, pp. 93–101, Feb. 1999.

[91] A. Osseiran and A. Logothetis, "Closed Loop Transmit Diversity in WCDMA HS-DSCH," in *2005 IEEE 61st Veh. Technol. Conf. (VTC Spring)*, vol. 1, Stockholm, Sweden, May 30–Jun. 1, 2005, pp. 349–353.

[92] R. Fritzsche, P. Rost, and G. P. Fettweis, "Robust Rate Adaptation and Proportional Fair Scheduling With Imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4417–4427, Aug. 2015.

[93] R. Bendlin, Y.-F. Huang, M. T. Ivrlač, and J. A. Nossek, "Cost-Constrained Transmit Processing in Wireless Cellular Networks with Universal Frequency Reuse," in *2010 44th Ann. Conf. Inf. Sci. and Syst. (CISS)*, Princeton, NJ, USA, Mar. 17–19, 2010.

[94] ——, "Enhancing the Network Sum-Rate without Sharing Channel State Information between Base Stations," in *2010 IEEE Global Telecommun. Conf. (GLOBECOM)*, Miami, FL, USA, Dec. 6–10, 2010.

[95] R. Bendlin, Y.-F. Huang, J. A. Nossek, and M. T. Ivrlač, "Circumventing Base Station Cooperation through Kalman Prediction of Intercell Interference," in *2008 Conf. Rec. 42nd Asilomar Conf. Signals, Syst. and Comput. (ACSSC)*, Pacific Grove, CA, USA, Oct. 26–29, 2008, pp. 1462–1466.

[96] M. J. Neely, E. Modiano, and C. P. Li, "Fairness and Optimal Stochastic Control for Heterogeneous Networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 396–409, Apr. 2008.

[97] H. Huh, S. H. Moon, Y. T. Kim, I. Lee, and G. Caire, "Multi-Cell MIMO Downlink With Cell Cooperation and Fair Scheduling: A Large-System Limit Analysis," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.

[98] H. Shirani-Mehr, G. Caire, and M. J. Neely, "MIMO Downlink Scheduling with Non-Perfect Channel State Knowledge," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2055–2066, Jul. 2010.

[99] A. Dotzler, G. Dietl, and W. Utschick, "Unitary Precoding for MIMO Interference Networks," in *2012 IEEE Global Telecommun. Conf. (GLOBECOM)*, 2012.

[100] J. Ellenbeck, A. Dotzler, and W. Utschick, "Interference Robustness for MIMO Networks — System-Level Performance Evaluation for LTE-Advanced," in *2012 IEEE Globecom Workshops*, Anaheim, CA, USA, Dec. 3–7, 2012, pp. 658–663.

[101] H. H. Brunner, Y. Song, A. Dotzler, W. Utschick, and J. A. Nossek, "Robustness against Instationary Intercell Interference," to be submitted for publication.

[102] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information Theoretic Considerations for Cellular Mobile Radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[103] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[104] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.

[105] E. Salbaroli and A. Zanella, "Interference Characterization in a Finite Poisson Field of Nodes with Shadowing," in *2008 IEEE 19th Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Cannes, France, Sep. 15–18, 2008.

[106] X. Yang and A. P. Petropulu, "Co-Channel Interference Modeling and Analysis in a Poisson Field of Interferers in Wireless Communications," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 64–76, Jan. 2003.

[107] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[108] T. L. Marzetta, "BLAST Training: Estimating Channel Characteristics for High Capacity Space-Time Wireless," in *Proc. 37th Ann. Allerton Conf. Commun., Control, and Computing*, Monticello, IL, USA, Sep. 22–24, 1999, pp. 958–966.

[109] E. W. Jang, J. Lee, H.-L. Lou, and J. M. Cioffi, "On the Combining Schemes for MIMO Systems with Hybrid ARQ," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 836–842, Feb. 2009.

[110] P. Frenger, S. Parkvall, and E. Dahlman, "Performance Comparison of HARQ with Chase Combining and Incremental Redundancy for HSDPA," in *2001 IEEE 54th Veh. Technol. Conf. (VTC Fall)*, vol. 3, Atlantic City, NJ, USA, Oct. 7–11 2001, pp. 1829–1833.

[111] A. Das, F. Khan, A. Sampath, and H.-J. Su, "Performance of Hybrid ARQ for High Speed Downlink Packet Access in UMTS," in *2001 IEEE 54th Veh. Technol. Conf. (VTC Fall)*, vol. 4, Atlantic City, NJ, USA, Oct. 7–11 2001, pp. 2133–2137.

[112] J. C. Ikuno, M. Wrulich, and M. Rupp, "Performance and Modeling of LTE H-ARQ," in *2009 13th Int. ITG Workshop Smart Antennas (WSA)*, Berlin, Germany, Feb. 16–18, 2009.

[113] G. Caire and D. Tuninetti, "The Throughput of Hybrid-ARQ Protocols for the Gaussian Collision Channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001.

[114] S. Mukherjee, S. A. Ramprashad, and Y. Li, "Performance Bounds and Improvements to HARQ in LTE under Unknown and Highly Variable Interference Due to MIMO," in *2011 IEEE Veh. Technol. Conf. (VTC Fall)*, San Francisco, CA, USA, Sep. 5–8, 2011.

[115] D. Chase, "A Combined Coding and Modulation Approach for Communication over Dispersive Channels," *IEEE Trans. Commun.*, vol. 21, no. 3, pp. 159–174, Mar. 1973.

[116] J. G. Proakis, *Digital Communications*, 3rd ed. New York, NY, USA: McGraw-Hill, 1995.

[117] J. Brehmer, *Utility Maximization in Nonconvex Wireless Systems*, 1st ed., ser. Foundations in Signal Processing, Communications and Networking, H. Boche, R. Mathar, and W. Utschick, Eds. Heidelberg, Germany: Springer, 2012.

[118] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An Axiomatic Theory of Fairness in Network Resource Allocation," in *2010, IEEE 29th Conf. Comput. Commun. (INFOCOM)*, San Diego, CA, USA, Mar. 15–19, 2010, pp. 1–9.

[119] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures," *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[120] S. S., M. Wiczanowski, and H. Boche, *Fundamentals of Resource Allocation in Wireless Networks: Theory and Algortihms*, 2nd ed., ser. Foundations in Signal Processing, Communications and Networking, H. Boche, R. Mathar, and W. Utschick, Eds. Heidelberg, Germany: Springer, 2008.

[121] S. Stanczak, M. Wiczanowski, and H. Boche, "Distributed Utility-Based Power Control: Objectives and Algorithms," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 5058–5068, Oct. 2007.

[122] J. Mo and J. Walrand, "Fair End-to-End Window-Based Congestion Control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[123] H. J. Kushner and P. A. Whiting, "Convergence of Proportional-Fair Sharing Algorithms under General Conditions," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1250–1259, Jul. 2004.

[124] D. Neumann, A. Dotzler, W. Utschick, and G. Dietl, "Optimal Utility-Based Multi-User Scheduling and Low-Complexity Alternatives," in *2013, 9th Int. ITG Conf. Syst., Commun. and Coding (SCC)*, Munich, Germany, Jan. 21–24, 2013.

[125] H. Huh, G. Caire, S.-H. Moon, and I. Lee, "Multi-Cell MIMO Downlink with Fairness Criteria: the Large System Limit," in *2010 IEEE Int. Symp. Inform. Theory (ISIT)*, Austin, TX, USA, Jun. 13–18 2010, pp. 2058–2062.

[126] J. Liu and Y. Thomas Hou, "Weighted Proportional Fairness Capacity of Gaussian MIMO Broadcast Channels," in *2008 IEEE 27th Conf. Comput. Commun. (INFOCOM)*, Phoenix, AZ, USA, Apr. 15–17, 2008, pp. 385–393.

[127] Y. Song, H. H. Brunner, and J. A. Nossek, "Intercell Interference Blindness in Fairness Optimizations," in *2015 19th Int. ITG Workshop Smart Antennas (WSA)*, Ilmenau, Germany, Mar. 3–5, 2015.

[128] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource Allocation and Cross-Layer Control in Wireless Networks," *Found. Trends Network.*, vol. 1, no. 1, pp. 1–144, Apr. 2006.

[129] B. Song, Y. H. Lin, and R. L. Cruz, "Weighted Max-Min Fair Beamforming, Power Control, and Scheduling for a MISO Downlink," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 464–469, Feb. 2008.

[130] F. Kelly, "Charging and Rate Control for Elastic Traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, Jan. 1997.

[131] M. Kobayashi, G. Caire, and N. Jindal, "How Much Training and Feedback are Needed in MIMO Broadcast Channels?" in *ISIT 2008, IEEE International Symposium on Information Theory*, Jul. 2008, pp. 2663–2667.