Unified approach for optimisation of single-user and multi-user multiple-input multiple-output wireless systems

von Diplom-Ingenieur
Eduard Axel Jorswieck
aus Berlin

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
\- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss

Vorsitzender: Prof. Dr. Gerhard Mönich
Berichter: Prof. Dr. Dr. Holger Boche
Berichter: Prof. Dr. Björn Ottersten (KTH Stockholm)

Tag der wissenschaftlichen Aussprache: 13. September 2004

Berlin 2004
D 83

Zusammenfassung

Mehrantennensysteme werden in zukünftigen Mobilfunksystemen der dritten und vierten Generation eingesetzt werden, um die spektrale Effizienz, die Zuverlässigkeit und die Qualität der drahtlosen Übertragung zu verbessern. In der Theorie wurde bewiesen, dass die Kanalkapazität dieser Mehrantennensysteme linear steigt mit der Anzahl der verwendeten Sende- und Empfangsantennen. Eine andere wichtige Kenngröße neben der Kanalkapazität ist der mittlere quadratische Fehler, wenn der optimale lineare Empfänger eingesetzt wird. Beide Kenngrößen variieren mit den Eigenschaften des Mehrantennen-Kanals und des betrachteten Systems, z.Bsp. mit der Art der Kanalinformation am Sender und Empfänger. Sogar partielle Kanalinformation am Sender erhöht die Leistungsfähigkeit des Mehrantennensystems beträchtlich. In dieser Arbeit werden die Eigenschaften von Mehrantennensystemen mit einem oder mit mehreren Benutzern in einem zellularen Kontext analysiert und neue optimale Sendestrategien entworfen, die die statistischen Eigenschaften des räumlichen Kanals und die Art der Kanalinformation am Sender berücksichtigen.

Im Szenario mit einem Teilnehmer wird die mittlere Leistungsfähigkeit des Mehrantennensystems unter dem Einfluss von einem räumlich korreliertem Schwundkanal und mit verschiedenen Arten von Kanalinformationen am Sender und mit perfekter Kanalkenntnis am Empfänger analysiert. Zuerst wird ein mathematisches Maß für die räumliche Korrelation basierend auf der Majorisierungstheorie definiert. Dadurch wird es möglich, die mittlere Performanz als Funktion der sende- und empfangsseitigen Korrelation im Kontext von Schur-konvexen und Schur-konkaven Funktionen zu beschreiben. Ausserdem stellt man fest, dass die Performanz-Maße zu einer allgemeinen Klasse von Funktionen gehören, die als Spur einer matrixmonotonen Funktion darstellbar sind. Wir verwenden Löwners Darstellung von operator-monotonen Funktionen, um auf einer abstrakten Ebene die optimalen Sendestrategien und den Einfluss der Korrelation auf die Performanz zu charakterisieren. Die optimale Sendestrategie ohne Kanalkenntnis am Sender ist eine Leistungsgleichverteilung in alle Richtungen. Dieses Ergebnis wird für räumlich korrelierte Kanäle bewiesen, indem die robusteste Sendestragie gegen die schlechteste Korrelation berechnet wird. Die mittlere Performanz ohne Kanalinformation am Sender ist eine Schur-konkave Funktione bezüglich Korrelation am Sender oder Empfänger. Desweiteren, wird die optimale Sendestrategie für den Fall hergeleitet, in dem der Sender die Langzeitstatistik des Kanals kennt. Ein iterativer Algorithmus löst das Problem der optimalen Leistungsverteilung. Die sogenannte Beamforming-Region ist der SNR Bereich, in dem ein einziger räumlicher Datenstrom die maximale mittlere Leistung erreicht. Dieser SNR Bereich ist relevant, da hier eine sehr einfache Empfängerstruktur und eine gut verstandene Kanalkodierung eingesetzt werden können. Schließlich leiten wir die generalisierte Waterfilling Lösung als optimale Sendestratgie für perfekte Kanalkenntnis am Sender und Empfänger her und charakterisiern die Eigenschaften dieses Verfahrens.

In einem zellularen Mobilfunksystem greifen mehrere Teilnehmer zur gleichen Zeit auf derselben Frequenz auf eine gemeinsame Basisstation zu oder eine Basisstation sendet gleichzeitig Daten für mehrere Teilnehmer. Die Interzell- und Intrazellinterferenz in einem solchen System erzeugt räumlich gefärbtes Rauschen auf einer einzelnen Mobilfunkstrecke. Daher kann als erster Ansatz ein Mehrantennensys-

tem mit einem Teilnehmer und gefärbtem Rauschen betrachtet werden. Wir leiten die Performanz unter dem schlechtesten möglichen Rauschen und unter verschiedenen Annahmen bezüglich des Rauschens her, um Einsichten in die erreichbare Performanz des Mehrantennensystems im zellularen Kontext mit Inter- und Intrazellinterferenz zu erhalten. Durch bestimmte Rauschfärbung kann sowohl die Kanalkenntnis als auch die Kooperationsfähigkeit an den Sendeantennen verloren gehen. Der nächste Schritt besteht darin, die Sendestrategien aller Teilnehmer einer Zelle zu berücksichtigen. Im letzten Abschnitt der Arbeit wird die augenblickliche Summen-Performanz des Mehrantennen Mehrfachzugriffskanals und des Mehrantennen Broadcast-Kanals unter individuellen oder Summenleistungsbeschränkungen maximiert. Als Summen-Performanz wird entweder die Summenkapazität mit sukzessiver Interferenz-auslöschung im Uplink oder mit Costa-Vorkodierung im Downlink, sowie der normierte mittelere quadratische Summenfehler eingesetzt, falls ein Mehrbenutzer-MMSE Empfänger verwendet wird. Die gemeinsame Kovarianzmatrixoptimierung kann unter Verwendung der Karush-Kuhn-Tucker Optimalitätsbedingungen in eine abwechselnde Leistungsoptimierung und normierte Kovarianzmatrixoptimierung zerlegt werden. Letztere wiederum zerfällt in eine Art modifizierte Einbenutzer Kovarianzmatrixoptimierung mit gefärbtem Rauschen. Die konkrete Struktur dieser Einbenutzer Optimierung hängt von der konkreten Performanz-Metrik ab. Der vorschlagene iterative Algorithmus löst das Summen-Performanz Optimierungsproblem auf effiziente Weise.

Abstract

Multiple-input multiple-output (MIMO) systems will be applied in wireless communications in order to increase the performance, spectral efficiency, and reliability. Theoretically, the channel capacity of those systems grows linearly with the number of transmit and receive antennas. An important performance metric beneath capacity is the normalised mean square error (MSE) under the assumption of optimal linear reception. Clearly, both performance measures depend on the properties of the MIMO channel as well as on the considered system approach, e.g. on the type of channel state information which is available at the transmitter. It has been shown that even partial CSI at the transmitter can increase the performance. In this thesis, we analyse the performance and design optimal transmit strategies of single- and multiuser MIMO systems with respect to the statistical properties of the fading channel and under different types of CSI at the transmit side.

In the single-user scenario, we study the average performance of the system under spatial correlated fading and with different types of CSI at the transmitter and with perfect CSI at the receiver. First, we introduce a measure of correlation which is based on Majorization. As a result, the average performance is analysed as a function of correlation in the context of Schur-convexity and Schur-concavity. Furthermore, we observe that the performance metrics belong to a general class of functions which are the trace of a matrix-monotone function. We use Löwner's representation of operator monotone functions in order to derive the optimum transmission strategies and to characterise the impact of correlation on the average performance. The optimal transmit strategy without CSI at transmitter is equal power allocation. We prove this result for spatial correlated channels by analysing the most robust transmit strategy under worst case correlation. The average performance without CSI is a Schur-concave function with respect to transmit and receive correlation. In addition to this, we derive the optimal transmission strategy with long-term statistics knowledge at the transmitter and propose an iterative algorithm. The beamforming-range is the SNR range in which only one data stream spatially multiplexed achieves the maximum average performance. This range is important, because of its simple receiver structure and well known channel coding. Finally, we derive the generalised water-filling transmit strategy for perfect CSI and characterise its properties.

If the single-user MIMO link is placed into a cellular system in which multiple users at the same time on the same frequency access one common base station or in which one base station transmits to multiple users, the interference colours the noise. This means, we can continue to study a single-user link now with coloured noise as a first approach. In order to gain insights into the performance under interference conditions, we derive the worst case noise performance for three different noise scenarios. We show that the cooperation and the CSI at the transmitter get lost if some type of worst case noise is applied. If all transmit strategies of all participating users are incooperated into the analysis, we arrive at the multi-user MIMO system. Finally, we maximise the instantaneous sum performance of MIMO multiple access channels (MAC) or broadcast channels (BC) under individual or sum power constraints. The sum performance is either the sum capacity if SIC is applied in the uplink or if Costa Precoding is applied in the downlink, or the normalised sum MSE if a multiuser MMSE receiver is applied at the base. Using the Karush-Kuhn-Tucker optimality conditions, we show that the mutual covariance matrix optimisation can be decomposed into power allocation and covariance matrix optimisation under individual power constraints, which can be decomposed into a kind of modified single-user covariance matrix optimisation treating the other users as noise. The concrete structure of the single-user program depends on the performance metric. The proposed algorithms efficiently solve the multi-user MIMO sum performance optimisation problem.

Acknowledgements

During my PhD studies, I had the opportunity to meet many interesting and extraordinary individuals. First of all, I thank my advisor Professor Holger Boche for taking his time to guide and support me throughout the three years. He always made it possible to share and discuss ideas and problems. I am very fortunate to have learned information theory and related mathematics from such a great teacher. His enthusiasm, knowledge, energy, and visions have motivated me to a great extent and will have a lasting effect. I thank Professor Björn Ottersten for serving as second referee and reader of this dissertation.

I want to thank Aydin Sezgin for sharing ideas and room with me during our PhD studies. The comments of Oliver Henkel improved the representation of this thesis. Many thanks go to my project leader Clemens von Helmolt as well as to Volker Jungnickel who have proposed many interesting problems which concerned implementation and realization aspects of multiple-antenna systems.

My gratitude extends to Thomas Haustein, Martin Schubert, Gerhard Wunder, Sławomir Stańczak, Tobias Oechtering, Marcin Wiczanowski, Volker Pohl, and Peter Jung for collaborating and co-authoring papers with me. I would like to thank all colleagues at the Heinrich-Hertz-Institut and the Technical University of Berlin for stimulating discussions and additional support.

Contents

List of Figures

List of Tables

Abbreviations and nomenclature

$\alpha_k(p)$ Optimality condition coefficients, see equation (2.41), page 33

$\mathrm{Ei}(a,b)$ Exponential integral $\mathrm{Ei}(a,b) = \int_1^\infty exp(-\kappa b)\kappa^{-a} d\kappa$, page 32

$\hat{\boldsymbol{w}}_k$ Column of white random matrix $\boldsymbol{W}$ with transmit correlation, see equation (2.24), page 26

$\mathcal{CN}(m,c)$ Complex normal distribution with mean m and covariance c, page 5

$\mathcal{I}$ Set of active directions or users, page 111

$\mathcal{L}$ Lagrangian function, page 105

$\boldsymbol{\lambda}(\boldsymbol{X})$ Eigenvalues of arbitrary Hermitian matrix $\boldsymbol{X}$, page 84

$\boldsymbol{\lambda}^R$ Vector of eigenvalues of receive correlation matrix, page 19

$\boldsymbol{\lambda}^T$ Vector of eigenvalues of transmit correlation matrix, page 19

$\boldsymbol{\Lambda}_Z$ Diagonal matrix with eigenvalues of noise covariance matrix, page 84

$\boldsymbol{\lambda}_{cc}^T, \boldsymbol{\lambda}_{cc}^R$ Vector of eigenvalues of fully correlated transmit and receive correlation matrix, page 40

$\boldsymbol{\lambda}_{nc}^T, \boldsymbol{\lambda}_{nc}^R$ Vector of eigenvalues of completely uncorrelated receive and transmit correlation matrix, page 40

$\boldsymbol{H}$ Flat-fading MIMO channel matrix, page 3

$\boldsymbol{H}_k$ Channel matrix of user k, page 104

$\boldsymbol{h}_{ij}$ Entry in the i-th column and j-th row of the channel matrix $\boldsymbol{H}$, page 19

$\boldsymbol{n}$ Noise vector, page 4

$\boldsymbol{Q}$ Transmit covariance matrix, page 5

$\boldsymbol{Q}_k$ Transmit covariance matrix of user k, page 104

$\boldsymbol{R}_R$ Receive correlation matrix, page 18

$\boldsymbol{R}_T$ Transmit correlation matrix, page 18

$\boldsymbol{W}$ Random matrix with 'white' iid zero-mean complex Gaussian entries, page 18

$\boldsymbol{x}$ Transmitted signal vector, page 4

$\boldsymbol{y}$ Received signal vector, page 4

$\boldsymbol{Z}$ Noise covariance matrix, page 84

Φ (Average) performance metric, page 24

ϕ Inner matrix-monotone performance function, page 24

ρ SNR: ratio of transmit power to noise power at receiver, page 5

σ_N^2 Noise power, page 84

$\tilde{\boldsymbol{w}}_k$ Column of white random matrix $\boldsymbol{W}$ with receive correlation, see equation (2.22), page 26

C Average channel capacity, page 6

c Instantaneous channel capacity, page 6

$F^{[1]}(x)$ First derivative of matrix-monotone function F, first divided difference at x , see equation (2.31), page 28

$h(x)$ Entropy of random variable x, page 3

$I(a; b)$ Mutual information between a and b, page 4

K Number of mobile users, page 7

MSE Average normalized MSE, page 6

mse Instantaneous normalized MSE, page 6

MSE_k Achievable normalised MSE of user k, page 8

n_R Number of receive antennas, page 3

n_T Number of transmit antennas, page 3

P Transmit power constraint, page 5

p_k Individual power constraint of user k or direction k, page 18

R_k Achievable transmission rate of user k, page 8

ARQ Automatic Repeat reQuest, page 1

BC Broadcast Channel, page 2

CSI Channel State Information, page 2

DSL Digital Subscriber Line, page 125

DSM Dynamic Spectrum Management, page 125

iid Identically Independent Distributed, page 13

LOS Line Of Sight, page 13

MAC Multiple Access Channell, page 2

MAP Maximum A-Posteriori, page 6

MIMO Multiple Input Multiple Output, page 1

MISO Multiple Input Single Output, page 1

ML Maximum Likelihood, page 6

MMSE Minimum Mean Square Error, page 6

OFDM Orthogonal Frequency Division Multiplexing, page 128

pdf Probability Density Function, page 13

QoS Quality of Service, page 13

SIC Successive Interference Cancellation, page 2

SIMO Single Input Multiple Output, page 1

SNR Signal to Noise Ratio, page 1

TDMA Time Division Multiple Access, page 2

WLAN Wireless Local Area Network, page 1

1 Introduction

1.1 Motivation

In mobile communication networks, the properties of the underlying physical channel have great impact on the performance and reliability of the system. The fading channel varies in time, frequency, and space. From a traditional point of view, these fluctuations are the limiting factor of wireless communication. From an information theoretic point of view, these fluctuations can be exploited. They provide the possibility to communicate even more reliable and secure. Recently, the spatial dimension was found to increase the performance and spectral efficiency of a wireless system by simply adding more transmit and receive antennas [FG98, Tel99] under idealistic assumptions. Since those papers were published, multiple antenna techniques have developed to the most active area in research in wireless communications.

As a result, the industry and standardisation bodies considered multiple-input multiple-output (MIMO) systems as a promising approach in order to obtain reliable high transmission rates which are required to satisfy the users needs. It is certain that the technology needs to deliver higher data rates to the consumer at a lower cost per data bit. One technology that has been proposed is the use of multiple antennas at both the cellular side and handsets. Other important considered techniques beneath MIMO are adaptive modulation and coding, hybrid ARQ, and fast cell selection [3gp01, 3gp02].

Dual-antenna diversity is already used on WLAN PC cards, access points, and personal digital cellular handsets in Japan. However, the antenna diversity is used only by antenna selection. Most major antenna infrastructure manufacturers offer smart antenna systems.

Theoretically, multiple antennas have been shown to enable major increase in the data capacity, peak data rate, performance, and reliability. However, the actual achievable performance depends on the properties of the radio signal propagation environment. Furthermore, the performance of MIMO systems depend on the type and amount of channel state information (CSI) at the receiver and transmitter side. In contrast to the spectral dimension which comes without the cost of interference between orthogonal subcarriers, the signals in space are multiplexed and interfere with each other. By using sophisticated transmit strategies with perfect or partial CSI at the transmitter, it is possible to fully exploit the spatial dimension. Interestingly, depending on the type of CSI, the properties of the channel have different impact on the optimal transmit algorithms and on the achievable performance. Because of the wide range of possible MIMO techniques, it is important to better understand the theoretical limits of MIMO channels and the right techniques which are able to achieve these limits. One important property of MIMO channels is the correlation at the transmit and receive antenna array [GBGP02, CTK02].

This thesis provides an analysis of the maximal achievable reliable transmission rate of single-user and multi-user MIMO systems. The first half is devoted to single-user MIMO systems and their average performance under different types of CSI and under transmit and receive correlation. The second half analyses multi-user MIMO systems in terms of worst case noise analysis and sum performance optimisation under individual and sum power constraints.

1.2 Notation

Vectors are denoted in bold letters $\boldsymbol{x}$. Matrices are written in bold capital letters $\boldsymbol{H}$. Transpose is $[\cdot]^T$, the conjugate transpose is $[\cdot]^H$. The matrix (pseudo) inverse is denoted by $[\cdot]^{-1}$. $tr(\boldsymbol{A})$ denotes the trace of the matrix $\boldsymbol{A}$, i.e. $tr(\boldsymbol{A}) = \sum_{k=1}^{n} \boldsymbol{A}_{k,k}$. $\lambda_i(\boldsymbol{A})$ is the ith eigenvalue of the matrix $\boldsymbol{A}$. $\lambda_{max}(\boldsymbol{A})$ is the largest eigenvalue of the matrix $\boldsymbol{A}$. $\mathcal{R}(x)$ denotes the real part of the complex variable x. $\mathbb{C}_{+}^{n}$ denotes the set of positive semidefinite matrices. $||\boldsymbol{a}||$ is the l_1-norm, i.e. $||\boldsymbol{a}|| = \sum_{i=1}^{n} |a_i|$. The partial order for vectors is denoted by $\boldsymbol{a} \succ \boldsymbol{b}$ and means vector $\boldsymbol{a}$ majorizes vector $\boldsymbol{b}$. The order for matrices is denoted by $\boldsymbol{A} \succ \boldsymbol{B}$ and means that the difference $\boldsymbol{A} - \boldsymbol{B}$ is positive definite. The expectation operator is $\mathbb{E}_{\boldsymbol{X}}$ and means expectation with respect to the random variable $\boldsymbol{X}$. $\text{diag}(\boldsymbol{A})$ is the vector with diagonal entries of $\boldsymbol{A}$ and $\text{Diag}(\boldsymbol{a})$ is a matrix with entries of the vector $\boldsymbol{a}$ on the diagonal.

1.3 Performance metrics and preliminaries

In this section, we give an informal overview over the performance metrics for single- and multiuser wireless transmission systems which are analysed in the following chapters of this thesis. The complete signal model follows in chapter 2 and chapter 3. All results in this thesis regard the performance metrics introduced in this section.

1.3.1 Single-user systems: Mutual information and related average performance metrics

The first part of this thesis deals with single-user MIMO systems. In order to characterise the general capacity and performance gain of multiple antenna systems the information theoretic *channel capacity* and related metrics are of great interest. They provide upper bounds on the transmission rate for which information can be transmitted over the fading channel with arbitrary small probability of error [CT91]. These useful metrics are the *average mutual information*, the *ergodic channel capacity*, and the *average normalised mean-square error*. We will introduce these metrics on an informal basis and point out the connections and differences between those quantities.

In his seminal work [Sha48], Shannon introduced the notion of mutual information, which measures the amount of information which is contained in some observed variable $\boldsymbol{y}$ about the random variable $\boldsymbol{x}$. The *Mutual information* for the two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$I(\boldsymbol{x}; \boldsymbol{y}) = h(\boldsymbol{y}) - h(\boldsymbol{y}|\boldsymbol{x}) = h(\boldsymbol{x}) - h(\boldsymbol{x}|\boldsymbol{y}) \tag{1.1}$$

with the differential entropy $h(\boldsymbol{y})$ and the conditional differential entropy $h(\boldsymbol{y}|\boldsymbol{x})$. The *differential entropy* of a continuous random variable, e.g. of the random vector $\boldsymbol{y}$ is given by [CT91, Chapter 9]

$$h(\boldsymbol{y}) = -\mathbb{E} \log f_y = -\int_{y \in \mathcal{S}_y} f_y(y) \log f_y(y) dy$$

with the probability density function $f_y(y)$. $\mathcal{S}_y$ is the support of $f_y(y)$. In this thesis, all logarithms $\log(x)$ are binary logarithms. The *conditional differential entropy* of a random variable $\boldsymbol{y}$ under $\boldsymbol{x}$ is given by

$$h(\boldsymbol{y}|\boldsymbol{x}) = -\int_{\substack{x\in\mathcal{S}_x\\ y\in\mathcal{S}_y}} f(x,y)\log f(x|y)dxdy$$

with joint density function $f(x,y)$ and $f(x|y) = \frac{f(x,y)}{f(y)}$ and support $\mathcal{S}_x \times \mathcal{S}_y$ of $f(x,y)$.

The block-flat fading single-user MIMO system is characterised by a flat-fading channel matrix $\boldsymbol{H}$ of size $n_R \times n_T$, which changes independently from block to block and the additive noise vector $\boldsymbol{n}$ at the receiver. Figure (1.1) shows the corresponding block diagram.

Figure 1.1: Block flat-fading MIMO system.

In figure (1.1), the transmitted vector $\boldsymbol{x}$ is multiplied by the flat-fading channel matrix $\boldsymbol{H}$ and then the noise vector is added. The received signal is given by[1]

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}. \tag{1.2}$$

In this model, the channel matrix $\boldsymbol{H}$ as well as the noise $\boldsymbol{n}$ is a random entity. The input dimension of the MIMO system equals the number of transmit antennas and is given by n_T. The number of output dimensions of the MIMO systems equals the number of receive antennas and is given by n_R[2]. The model depends on the propagation scenario, i.e. on the distribution of $\boldsymbol{H}$. Well established models are the Rayleigh channel, i.e. the multipath environment without a line of sight (LOS) component in which $\boldsymbol{H}$ is modelled as zero-mean independent identically distributed (iid) complex Gaussian, i.e. $\boldsymbol{H} \sim \mathcal{CN}(0, \boldsymbol{I})$ or the Rice channel with a LOS component $\boldsymbol{H} \sim \mathcal{CN}(\boldsymbol{K}, \boldsymbol{I})$ with constant LOS matrix $\boldsymbol{K}$. The mutual coupling between the transmit and receive antennas can be modeled by correlation matrices [GBGP02, Mol04]. Possible singularities within the channel, are described as 'key-holes' [CFGV02]. All quantities in (1.2) the transmit vector $\boldsymbol{x}$, the received vector $\boldsymbol{y}$, the noise vector $\boldsymbol{n}$, and the channel matrix $\boldsymbol{H}$ are random variables with proper probability density functions (pdf).

The mutual information between $\boldsymbol{y}$ and $\boldsymbol{x}$ under the assumption that the receiver side knows the channel realization $\boldsymbol{H}$ is given by

$$\begin{aligned} I(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H}) &= h(\boldsymbol{y}|\boldsymbol{H}) - h(\boldsymbol{y}|\boldsymbol{H},\boldsymbol{x}) \\ &= h(\boldsymbol{y}|\boldsymbol{H}) - h(\boldsymbol{n}). \end{aligned} \tag{1.3}$$

Throughout the whole thesis, the receiver is assumed to know the channel state $\boldsymbol{H}$ perfectly. CSI at the receiver is achieved either by channel estimation using pilot signals or by blind channel estimation. The differential entropy of the noise vector $\boldsymbol{n}$ in (1.3) depends on the pdf of the random variable $\boldsymbol{n}$. Thermal receiver

[1]In order to keep notation simple, we omit the block index.

[2]The signal model for flat-fading MIMO channels in (1.2) can be easily extended to frequency-selective channels by increasing the input and output dimensions of the MIMO system.

noise is modelled as a zero-mean complex Gaussian variable, i.e. $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma_n^2 \boldsymbol{I})$. Therefore, the entropy of the random noise vector is given by [CT91, Theorem 9.6.5][3]

$$h(\boldsymbol{n}) = \log\left((2\pi e)^{n_R} \det\left(\sigma_n^2 \boldsymbol{I}\right)\right). \tag{1.4}$$

The entropy of the received vector $\boldsymbol{y}$ in (1.3) depends on the pdf of the input vector $\boldsymbol{x}$. The transmitter chooses the pdf of $\boldsymbol{x}$ in order to maximise the mutual information. The largest entropy is achieved for zero-mean complex Gaussian distributed random variables. Therefore, the transmitter chooses $\boldsymbol{x} \sim \mathcal{CN}(0, \boldsymbol{Q})$ with transmit covariance matrix $\boldsymbol{Q}$. Note that the random variables $\boldsymbol{x}$ and $\boldsymbol{n}$ are independent. As a result, the received vector for fixed channel realization $\boldsymbol{H}$ is zero-mean complex Gaussian distributed with receive covariance matrix $\boldsymbol{W} = \sigma_n^2 \boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H$. This follows from the fact that a linear combination of Gaussian random variables stays Gaussian and that the sum of independent zero-mean Gaussian variables is Gaussian distributed with zero-mean and covariance matrix which is the sum of the individual covariance matrices. Therefore, the conditional entropy of the received vector $\boldsymbol{y}$ is given by

$$\begin{aligned} h(\boldsymbol{y}|\boldsymbol{H}) &= \log\left((2\pi e)^{n_R} \det \boldsymbol{W}\right) \\ &= \log\left((2\pi e)^{n_R} \det\left(\sigma_n^2 \boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)\right). \end{aligned} \tag{1.5}$$

Finally, from the entropy in (1.4) and the conditional entropy in (1.5) follows the *instantaneous mutual information* of the MIMO system in (1.2) with $\rho = \frac{1}{\sigma_n^2}$

$$f(\boldsymbol{Q}, \boldsymbol{H}, \rho) = I(\boldsymbol{y}; \boldsymbol{x}|\boldsymbol{H}) = \log\det\left(\boldsymbol{I} + \rho \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right). \tag{1.6}$$

The instantaneous mutual information is a function of the transmit covariance matrix $\boldsymbol{Q}$, the channel realization $\boldsymbol{H}$, and the inverse noise variance ρ. Often, we will normalise the transmit power to one, i.e. $\operatorname{tr} \boldsymbol{Q} = P = 1$. Then, ρ corresponds directly to the signal-to-noise-ratio (SNR). The mutual information in (1.6) depends on the channel realization $\boldsymbol{H}$ and is therefore a random entity.

The instantaneous mutual information has been derived and analysed in the seminal work of Telatar [Tel99] and Foschini and Gans [FG98]. The equation (1.6) corresponds exactly to the result of the alternative derivation in [Tel99, Section 3.2] and to the derivation in the appendix of [FG98].

Finally, the expression in (1.6) can be maximised with respect to $\boldsymbol{Q}$ under a sum power constraint in order to get the *instantaneous channel capacity* of the MIMO system, i.e.

$$c(\rho, \boldsymbol{H}) = \max_{\boldsymbol{Q}:\, \operatorname{tr} \boldsymbol{Q} \le P} f(\boldsymbol{Q}, \boldsymbol{H}, \rho). \tag{1.7}$$

The channel capacity gives the upper bound of information which can be transmitted over the MIMO channel with arbitrary small probability of error if the channel is in state $\boldsymbol{H}$. The instantaneous mutual information in (1.6) as well as the instantaneous channel capacity in (1.7) are random variables because they depend directly on the channel realization $\boldsymbol{H}$. In order to get an average expression about the mutual information and the channel capacity which can be achieved, the expectation

[3] Note, that the factor $1/2$ in front of the differential entropy in [CT91] vanishes if *complex* random variables are considered.

with respect to $\boldsymbol{H}$ is computed in order to obtain the *average mutual information*

$$F(\boldsymbol{Q},\rho) = \mathbb{E}_{\boldsymbol{H}} \log\det\left(\boldsymbol{I} + \rho\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \tag{1.8}$$

and to obtain the *average channel capacity*

$$C(\rho) = \mathbb{E}_{\boldsymbol{H}} \max_{\boldsymbol{Q}:\ \mathrm{tr}\ \boldsymbol{Q}\leq P} f(\boldsymbol{Q},\boldsymbol{H},\rho). \tag{1.9}$$

If the random fading process $\boldsymbol{H}$ fulfils the ergodic property, the quantity $C(\rho)$ in (1.9) is the *ergodic channel capacity* of the MIMO system.

Interestingly, another performance metric is closely related to the average mutual information and to the channel capacity of the MIMO system: the mean square error (MSE). For computing these upper bounds on the achievable transmission rate with arbitrary small probability of error, the receiver applies the optimum detector and decoder. The optimum receiver algorithm is either the maximum likelihood (ML) or maximum a-posteriori (MAP) algorithm. However, capacity also can be achieved by decision feedback minimum mean square error (MMSE) detection [VG97]. The linear MMSE receiver reduces the computational complexity at the receiver side. If we apply the linear MMSE receiver, the performance metric changes from the average mutual information to the *normalised MSE* [VAT99]. The linear MMSE receiver weights the received signal vector $\boldsymbol{y}$ by the Wiener filter

$$\hat{\boldsymbol{x}} = \rho\boldsymbol{H}^H\left[\boldsymbol{I} + \rho\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right]^{-1}\boldsymbol{y}. \tag{1.10}$$

The covariance matrix of the estimation error $\boldsymbol{K}_\epsilon$ is given by

$$\boldsymbol{K}_\epsilon = \mathbb{E}_{\boldsymbol{H}}\left[(\hat{\boldsymbol{x}} - \boldsymbol{x})(\hat{\boldsymbol{x}} - \boldsymbol{x})^H\right] \tag{1.11}$$

The *normalised MSE* is defined as the trace of the normalised covariance matrix of the estimation error in (1.11) [HB03, JB03d]

$$\begin{aligned} mse(\sigma_n^2,\boldsymbol{Q},\boldsymbol{H}) &= \mathrm{tr}\ \left(\boldsymbol{Q}^{-1/2}\boldsymbol{K}_\epsilon\boldsymbol{Q}^{-1/2}\right) \\ &= n_T - \mathrm{tr}\ \left(\rho\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\left[\boldsymbol{I} + \rho\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right]^{-1}\right) \end{aligned} \tag{1.12}$$

and its average over channel realizations is called *normalised average MSE*. It is given by

$$MSE(\sigma_n^2,\boldsymbol{Q}) = n_T - \mathbb{E}_{\boldsymbol{H}}\ \mathrm{tr}\ \left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\left[\sigma_n^2\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right]^{-1}\right) \tag{1.13}$$

1.3.2 Multiuser systems: Sum performance metrics

The instantaneous performance metrics from the single-user analysis can be directly transfered to the multiuser case. However, in multiuser systems, the transmission from one user is disturbed by interference from the other users. In general, there are two possible multiuser scenarios: In the uplink, multiple users want to transmit their data to one common base station. The uplink is called the multiple access channel (MAC). In the downlink, the base transmits data to multiple users. This is called the broadcast channel (BC). The communication channel between each user and the base is assumed to be block-flat fading.

Multiple access channel

In the MAC, K users transmit their data simultaneously to the base station. All mobiles have n_T transmit antennas[4]. The received vector is given by

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{n}.$$

The mutual information for user k according to the definition in section 1.3.1 under the assumption that the receiver knows the channel realization $\boldsymbol{H}_k$ is given by

$$\begin{aligned} I(\boldsymbol{y}; \boldsymbol{x}_k | \boldsymbol{H}_k) &= \log\det\left(\boldsymbol{I} + \rho \sum_{l=1}^{K} \boldsymbol{H}_l \boldsymbol{Q}_l \boldsymbol{H}_l^H\right) \\ &\quad - \log\det\left(\boldsymbol{I} + \rho \sum_{l=1, l\neq k}^{K} \boldsymbol{H}_l \boldsymbol{Q}_l \boldsymbol{H}_l^H\right) \end{aligned} \tag{1.14}$$

with SNR ρ and transmit covariance matrices $\boldsymbol{Q}_k$. The transmit signals of the users are assumed to be zero-mean independent complex Gaussian distributed with covariance matrix $\boldsymbol{Q}_k$. This pdf maximises the individual mutual information of each user. Obviously, the individual mutual information of user k depends on the multiuser interference and noise, i.e. it is a function of all transmission channels $\boldsymbol{H}_k$ between the users and the base, the SNR ρ and the transmit strategies $\boldsymbol{Q}_k$ of all users

$$R_k(\mathcal{Q}, \mathcal{H}, \rho) = \log\det\left(\frac{\boldsymbol{I} + \rho \sum_{l=1}^{K} \boldsymbol{H}_l \boldsymbol{Q}_l \boldsymbol{H}_l^H}{\boldsymbol{I} + \rho \sum_{l=1, l\neq k}^{K} \boldsymbol{H}_l \boldsymbol{Q}_l \boldsymbol{H}_l^H}\right) \tag{1.15}$$

with the set of covariance matrices $\mathcal{Q}$ and the set of channel realisations $\mathcal{H}$

$$\mathcal{Q} = \{\boldsymbol{Q}_1, \boldsymbol{Q}_2, ..., \boldsymbol{Q}_K\} \quad \text{and} \quad \mathcal{H} = \{\boldsymbol{H}_1, \boldsymbol{H}_2, ..., \boldsymbol{H}_K\}.$$

The achievable rate of user k is denoted by R_k. It is possible, that the receiver first detects the signals of a set of users and subtracts them from the received signal before detecting user k. As long as the users transmit at a rate smaller than or equal to their mutual information, their signals are detected with arbitrary small probability of error and therefore correctly subtracted. Let us assume that the signals of users 1 to $k-1$ are correctly subtracted, then the individual mutual information of user k is given by

$$R_k^{SIC}(\mathcal{Q}, \mathcal{H}, \rho) = \log\det\left(\frac{\boldsymbol{I} + \rho \sum_{l=1}^{K} \boldsymbol{H}_l \boldsymbol{Q}_l \boldsymbol{H}_l^H}{\boldsymbol{I} + \rho \sum_{l=k+1}^{K} \boldsymbol{H}_l \boldsymbol{Q}_l \boldsymbol{H}_l^H}\right). \tag{1.16}$$

The receiver starts with user one, detects his data, and subtracts it from the received signal. The received signal for user one is interfered by all other users. Then the second user is detected and subtracted. The second user gets interference from all but the first user. This procedure continues until the last user is detected without any interference. This approach is called successive interference cancellation (SIC).

[4] It is easy to generalize the results to the general case in which each mobile has a different number of transmit antennas.

Usually, one assumes that the data of all users is detected without errors because the users transmit with rate below their capacity.

If we assume that the receiver detects the user signals in a linear fashion, the optimal choice is the linear multiuser MMSE receiver. The corresponding performance metric is the individual normalised MSE of user k which is given by

$$MSE_k = n_T - \operatorname{tr}\left(\rho \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^H \left(\rho \sum_{l=1}^{K} \boldsymbol{H}_l \boldsymbol{Q}_l \boldsymbol{H}_l^H + \boldsymbol{I}\right)^{-1}\right). \tag{1.17}$$

In contrast to the capacity, it is not possible to perform SIC without error propagation, because the argument with the error free reception is missing. Therefore, each user k experiences interference from all other users. The achievable MSE region is given by all MSE tuples $(m_1, ..., m_K)$ for which $(m_1 \geq MSE_1, ..., m_K \geq MSE_K)$ applies.

Using the individual rate or the individual MSE, each user can require its quality-of-service (QoS) by giving a minimum rate r_k or a maximum MSE m_k which has to be achieved. The problem of the fulfilment of service requirements consists of computing a transmit strategy which ensures for all $1 \leq k \leq K$ that $R_k \geq r_k$ or $MSE_k \leq m_k$ by minimising the individual $p_k \leq P_k$ or sum transmit power $\sum_{k=1}^{K} p_k \leq P$.

Another performance metric is the sum of the individual performance metrics. The sum capacity is simply defined as the sum of the individual capacities

$$\sum_{k=1}^{K} R_k^{SIC}(\mathcal{Q}, \mathcal{H}, \rho),$$

i.e. with SIC, we obtain

$$C(\mathcal{Q}, \mathcal{H}, \rho) = \log\det\left(\boldsymbol{I} + \rho \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^H\right). \tag{1.18}$$

The normalised sum MSE is defined in the same manner, i.e. $MSE = \sum\limits_{k=1}^{K} MSE_k$ and it is given by

$$MSE(\mathcal{Q}, \mathcal{H}, \rho) = Kn_T - n_R + \operatorname{tr}\left(\left[\rho \sum_{k=1}^{n_T} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^H + \boldsymbol{I}\right]^{-1}\right). \tag{1.19}$$

The sum capacity and the sum MSE describe the performance of the complete MAC. The system throughput can be measured by the sum capacity in (1.18) or by the sum MSE (1.19).

Broadcast channel

In the MIMO BC scenario, we study the downlink transmission from the base station to the mobiles. The transmission channels between the base and the mobiles are reused from the MAC and by reciprocity[5] we obtain the received vector at mobile

[5]The author in [Tel99] calls reciprocity the fact that the capacity is unchanged if the role of transmitter and receiver is interchanges in a MIMO point-to-point link. In our context, we assume that the uplink and downlink channels are reciprocal because the same frequency is

terminal k as

$$\boldsymbol{y}_k = \sum_{l=1}^{K} \boldsymbol{H}_k^H \boldsymbol{x}_l + \boldsymbol{n}_k. \tag{1.20}$$

The achievable rate for user k under the assumption that the transmit signals intended for all users are zero-mean complex Gaussian distributed with covariance matrix $\boldsymbol{Q}_k$ is given by

$$S_k(\mathcal{Q}, \mathcal{H}, \rho) = \log\det\left(\frac{\boldsymbol{I} + \rho \sum_{l=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_l \boldsymbol{H}_k^H}{\boldsymbol{I} + \rho \sum_{l=1, l\neq k}^{K} \boldsymbol{H}_k \boldsymbol{Q}_l \boldsymbol{H}_k^H}\right). \tag{1.21}$$

The counterpart to SIC at the base station in the MAC, is Costa-Precoding [Cos83]. It is possible, to subtract the already coded data for users 1 to $k-1$ before encoding the data intended for user k without power increase. Therefore, user k receives only interference from users $k+1$ to K. The achievable rate with Costa-Precoding for user k is given by

$$S_k^{CST}(\mathcal{Q}, \mathcal{H}, \rho) = \log\det\left(\frac{\boldsymbol{I} + \rho \sum_{l=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_l \boldsymbol{H}_k^H}{\boldsymbol{I} + \rho \sum_{l=k+1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_l \boldsymbol{H}_k^H}\right). \tag{1.22}$$

Achieving rate S_k^{CST} means that the base station performs Costa-Precoding, the transmit signals are all independent zero-mean complex Gaussian distributed, and the mobile performs optimal detection and decoding.

The achievable sum rate of the MIMO BC with Costa-Precoding equals the sum capacity of the MIMO MAC. Furthermore, it can be shown by Sato's bound [Sat78], that the achievable sum rate of the MIMO BC is the sum capacity itself. Therefore, the sum capacity of the MIMO BC with Costa-Precoding can be achieved by solving the MIMO MAC sum capacity problem and then transforming the transmit covariance matrices as described in [VJG02b]. This simplifies the analysis of the sum performance in terms of the sum capacity, because up- and downlink can be treated together.

The same transformation of transmit covariance matrices from MAC to the BC can be applied for sum MSE minimisation. Therefore, by solving the sum MSE minimisation problem for the uplink, the corresponding downlink problem is solved, too.

This concludes the informal introduction of the performance metrics for the single-user and multi-user MIMO system. In the next section, the contribution of this thesis is summarised and the structure is presented.

1.4 Outline of the thesis

The contributions of this thesis are summarised in the following: The contribution of 2nd chapter consists of the following points: First, we present related work and classify our work. In section 2.2.1 we introduce our channel model. Another ingredient is a rigorous mathematical measure of correlation which is defined in section 2.2.2. We observe that the average performance of the single-user MIMO system can be written as the average trace of an arbitrary matrix-monotone function [Bha97].

used, e.g. in TDD modus.

This observation allows us to analyse different performance metrics as representatives of a much larger class of performance functions. The complete analysis is performed on this meta layer. All results of the average mutual information and the average MSE follow as corollaries. This observation is presented in section 2.3.1. Then, the following results are proven:

- **Optimum transmission strategies:** In section 2.3.2, we justify the intuitive belief that equal power allocation is optimal without CSI at the transmitter even if the transmit antennas and receive antennas are correlated. This is done by considering the worst case transmit and receive correlation and showing that equal power allocation is most robust against that.

  For the case when the transmitter knows the transmit and receive correlation matrix[6], we show that the average performance is optimised by transmitting along the eigenvectors of the transmit correlation matrix. In addition to this, we characterise the optimum power allocation for the covariance feedback case and develop an iterative algorithm which solves the power allocation optimisation problem. At low SNR values only one eigenvalue is supported, i.e. the transmitter performs beamforming. The SNR range in which beamforming is optimal is characterised for arbitrary transmit and receive correlation and arbitrary average performance function.

  With perfect CSI at the transmitter, the transmit strategy is adapted to each channel realization. The MIMO channel is decomposed according to the singular value decomposition first, then optimal power across the parallel Gaussian channels is allocated. We derive the generalised water-filling solution. At low SNR, the beamforming range bases on a instantaneous channel realization is derived.

- **Impact of correlation on the ergodic capacity:** In section 2.3.3, for an arbitrary number of transmit and receive antennas and for arbitrary SNR values we show that uncorrelated MIMO channels yield higher capacities than correlated channels with uninformed transmitter and a receiver which has perfect CSI. We show that the average performance of an open-loop MIMO system is a Schur-concave function with respect to transmit and receive correlation. In addition to this, the difference between the ergodic capacity of fully correlated and completely uncorrelated MIMO channels is studied and a tight lower bound is developed.

  For the case in which the transmitter knows the transmit correlation matrix, we show that the average performance can be either Schur-convex or Schur-concave or nothing with respect to transmit correlation depending on the SNR. For small SNR, beamforming is optimal. If beamforming is optimal, the average performance is Schur-convex with respect to transmit correlation. With respect to receive correlation, we show that the covariance feedback MIMO ergodic channel capacity is a Schur-concave function.

  The closed-loop MIMO average performance is shown to be Schur-convex with respect to transmit and receive correlation for small SNR values and Schur-concave at high SNR values.

- **Comparison to MISO results:** In section 2.3.4, the differences between MIMO and MISO channels are analysed. The general structure of the optimum transmission strategies for the different types of CSI between MISO

[6] This case is called 'covariance feedback', because the knowledge at the transmitter about the two covariance matrices can be achieved by channel estimation at the receiver and slow feedback to the transmitter.

and MIMO systems corresponds well. However, the impact of correlation on MISO and MIMO systems substantially differs. In MISO systems much stronger results can be proven. The average MISO covariance feedback performance is Schur-convex with respect to transmit correlation. The average closed-loop MISO performance is Schur-concave with respect to transmit correlation. As a result, the average performance for the complete set of types of CSI is analysed. In a clearly arranged table, we give a complete summary of all theoretical results.

- **Illustration and discussion of average performance results:** In section 2.3.6, the theoretical results from the last sections, are summarised by illustrative numerical simulations. The properties of the different average performance metrics under correlation and under the different types of CSI can be read off.

In chapter 3, multi-user multiple-antenna systems and their respective performance metrics are studied. We start with a point-to-point link which experiences noise, inter-cell interference and intra-cell interference. We incooperate these three fractions into the noise with specific noise covariance matrix. We compute the worst case noise performance of the MIMO link in three different noise scenarios. The contributions of the worst case noise analysis in section 3.2 are:

- The performance of a MIMO closed-loop system, i.e. perfect CSI at the transmitter, with worst case noise under a trace constraint (or worst case interference) equals the performance of a MIMO open-loop system, i.e. no CSI at the transmitter with white noise, i.e. without interference. The structure of the equivalent system is a single-user MIMO system with uncorrelated noise and without CSI at the transmitter. We do a complete characterisation of the solution of the corresponding minimax expression.
- The worst case noise directions correspond with the left eigenvectors of the channel matrix $\boldsymbol{H}$. The optimal transmit directions correspond with the right eigenvectors of the channel matrix $\boldsymbol{H}$. Both are independent of each other. Therefore, the minimax problem fulfils the saddle point property. The power allocation then is the well-known waterfilling solution.
- The worst case coloured noise decomposes the closed-loop MIMO system into a SIMO MAC with amplified white noise. The transmitter cooperation goes loose and the noise is amplified by a factor equal to the number of receive antennas.

Furthermore, in section 3.3, we explicitly incooperate all transmit strategies into the optimisation of the sum performance of the multiuser MIMO system. The sum performance can be the sum capacity of the MIMO MAC if SIC is applied or the sum capacity of the MIMO BC if Costa Precoding is applied or the sum MSE of the MAC. We assume perfect CSI at transmitter and mobiles and derive the optimal transmit strategies under individual and sum power constraints. The structure of this part of the thesis is summarised in the following:

- At first, the signal model and the associated performance metrics are introduced in section 3.3.1. The problem statements are proposed in section 3.3.2. The sum performance is maximised under a sum power constraint.
- We solve these problems in section 3.3.3 and derive an iterative algorithm which efficiently solves the sum performance optimisation problem of the multiuser MIMO system. We show that the original problem can be decomposed into two parts, namely power allocation and covariance matrix optimisation under individual power constraints. These two parts are alternately

performed and converge to the optimal solution. Furthermore, the covariance matrix optimisation step can be further decomposed into iterative single-user performance optimisation treating the other users as noise.

- In order to get better understanding of the optimal transmit strategy which is in general very complex, we analyse the strategy at low SNR values in section 3.3.4. We completely characterise the SNR range, in which only one user is allowed to transmit at a time.
- This chapter is concluded in section 3.3.5 by a short discussion of the construction of the iterative algorithm. In addition, we provide an illustrative example of average sum capacities for SISO, SIMO, and MIMO MACs with different numbers of users and SNR values.

Finally, in chapter 4, we conclude the thesis and give directions for further research. The list of publications and the bibliography finalise this thesis.

2 Optimal transmission strategies and impact of correlation in single-user multiple-antenna channels

2.1 Related work

The increasing need for fast and reliable wireless communication links has opened the discussion about systems with multiple antennas both located at the transmitter and the receiver - so called multiple-input multiple-output (MIMO) systems [FG98]. Systems with multiple antennas at one side of the link are well known [Jak74] for increasing the capacity and performance. In recent years, it was discovered that MIMO systems have the ability to reach higher transmission rates than one-sided array links [WFGV98] [Tel99].

Many results regarding the capacity of MISO and MIMO systems under different levels of CSI and the corresponding transmission strategies were recently published. First, we review recent results for MISO systems. The MISO case has recently attracted much attention. In [Win98], the potential of multiple antenna systems was pointed out. The capacity of a MISO system with imperfect feedback was first analysed in [VM01] and [NLTW98, NTW99]. In [JG01b, JB02b] the optimum transmission strategy with covariance knowledge at the transmit array with respect to the ergodic capacity was analysed. In [RFLT98, SB01, BS01], the problem of downlink beamforming in MISO systems was solved. In [JG03], the ergodic capacity in the non-coherent transmission scenario with only covariance knowledge at the transmitter and the receiver, is studied.

It has been shown that even partial CSI at the transmitter can increase the capacity of a MISO system. Recently, transmission schemes for optimising capacity in MISO mean-feedback and covariance-feedback systems were derived in [VM01, NLTW98]. The capacity can be achieved by Gaussian distributed transmit signals with a particular covariance matrix. In a block fading model, the general signal processing structure which achieves capacity independent of the type of CSI consists of a Gaussian codebook, a number of beamformers and a power allocation entity [VM01, BCT01]. Additionally, it has been proved that the optimal transmit covariance matrix in the covariance feedback case has the same eigenvectors as the known channel covariance matrix. The complete characterisation of the impact of correlation on the ergodic capacity in MISO systems can be found in [JB04c].

While studying MIMO systems, we can imagine several different scenarios in which the transmitter or the receiver have partial or perfect channel state information (CSI). The capacity of a single-user MIMO system with perfect CSI at the receiver and no CSI at the transmitter was studied in [Tel99]. Equal power allocation was shown to be optimal for uncorrelated Rayleigh fading MIMO channels without CSI at the transmitter. The optimum strategy in order to minimise the outage probability is analysed in [JB03a]. The capacity of a single-user MIMO system with perfect CSI at both the transmitter and the receiver can be derived from the 'water-filling' approach [CT91, Tel99]. If the transmitter has only partial CSI in terms of the channel covariance matrix the optimal transmission strategy is shown in [JG01b] to transmit in the direction of the known channel eigenvectors. The power allocation problem is discussed in [JG01b], [JB02a], [SM02] and [JB02b].

Practical power allocation and rate adaption algorithms are discussed in [MBO04, BMO04]. The case in which both the transmitter and the receiver do not have channel state information remains an open problem. Even for the single-input single-output Rayleigh fading case, the capacity achieving transmission strategy for no CSI has not been solved completely [AFTS01]. For MIMO case, the optimum transmission strategy without CSI at the transmitter and receiver was characterised in [MH99].

Most of the results regarding the ergodic and outage capacity of single user MISO and MIMO systems assume that the transmit and receive antennas are uncorrelated. In reality, there can occur correlation at the transmit antenna array due to the placement of the array and the geometry in the transmission scenario. Especially at the base station which is often un-obstructed, correlation between the antennas can occur.

In literature there are different models which measure the correlation of the transmit and receive antennas in MIMO systems. We use the well established model from [CTK02]. This model is well suited for Rayleigh and Ricean MIMO channels which naturally arise in a rich scattering environment. In [SFGK00] a model for correlation of the multi element antenna was developed and the ergodic and outage capacity was computed under correlation. In [CTK02] the impact of correlation was analysed by studying the asymptotic eigenvalue distribution of the channel matrix for a large number of transmit and receive antennas. The theory of majorization for discrete vectors was extended to continuous probability density functions and it was shown that correlation decreases the ergodic capacity. In [MO02], an approximative expression for the capacity of correlated MIMO channels has been presented. In [MSS01] analytical results for the moments of the mutual information under correlation were derived for large antenna arrays, too. Using the tight capacity bound which was developed in [ONBP02] for high SNR values, the impact of correlation can easily be analysed. All analytical results assume either a large number of antennas or high SNR values. Although the recent results indicate that correlation decreases the ergodic capacity without CSI at the transmitter, the general proof is still missing. For the MISO case, a proof was derived in [BJ04a]: The ergodic capacity of a MISO system with no CSI at the transmitter decreases if the transmit correlation increases. Furthermore, it was shown in [BJ04a], that the capacity loss due to correlation is bounded by some small constant.

The statistics in MIMO channels are entirely characterised by the transmit and receive correlation matrix if the channel between is a rich scattering channel which is assumed Rayleigh distributed. If the channel between the transmit and receive array has singularities or rank deficiencies, the channel statistics are more complex. A method for describing such phenomena was proposed in [Say02]. The analysis in [CFV00, CFGV02] is adapted to several special practical scenarios in which so called keyholes occur. For example, in indoor transmission scenarios in which we have long corridors the channel can be singular. This is not because of correlation at the transmitter or the receiver but because of a keyhole in between. We do not discuss effects like keyholes in this thesis, because we study the impact of correlation of the transmit antenna array separately. Therefore, we assume that the channel between transmitter and receiver does not have keyholes.

2.2 Channel model and basic definitions

2.2.1 Channel model

Consider the standard MIMO block flat-fading channel model

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{2.1}$$

with complex $n_T \times 1$ transmit vector $\boldsymbol{x}$, channel matrix $\boldsymbol{H}$ with $n_R \times n_T$ entries, circularly symmetric complex Gaussian noise $\boldsymbol{n}$ with variance $\frac{\sigma_N^2}{2}\boldsymbol{I}$ per dimension. For convenience, we define the inverse noise variance as $\rho = 1/\sigma_N^2$. We assume that the receiver knows $\boldsymbol{H}$ perfectly.

Let us first describe the signal processing at the transmit antenna array. The transmit covariance matrix is given by

$$\boldsymbol{Q} = \mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}^H\right).$$

Using the eigenvalue decomposition of $\boldsymbol{Q} = \boldsymbol{U}_Q\boldsymbol{\Lambda}_Q\boldsymbol{U}_Q^H$, it becomes obvious how one can construct a particular transmit covariance matrix. The input data stream $d(k)$ is split into m parallel data streams $d_1(k) \dots d_m(k)$. Each parallel data stream is multiplied by a factor $\sqrt{p_1} \dots \sqrt{p_m}$ and then weighted by a beamforming vector $\boldsymbol{u}_1 \dots \boldsymbol{u}_m$. The number of parallel data streams is less than or equal to the number of transmit antennas ($m \leq n_T$). The beamforming vectors $\boldsymbol{u}_i$ have size $n_T \times 1$ with n_T as the number of transmit antennas. The powers $p_1, ..., p_{n_T}$ correspond to the eigenvalues in the diagonal matrix $\boldsymbol{\Lambda}_Q$. The n_T signals of each weighted data stream $\boldsymbol{x}^i(k) = d_i(k) \cdot \sqrt{p_i} \cdot \boldsymbol{u}_i$ are added up $\boldsymbol{x}(k) = \sum_{i=1}^m \boldsymbol{x}^i(k)$ and sent. By omitting the time index k for convenience we obtain in front of the transmit antennas

$$\boldsymbol{x} = \sum_{l=1}^{m} d_l \cdot \sqrt{p_l} \cdot \boldsymbol{u}_l. \tag{2.2}$$

The signal processing structure is shown in figure (2.1). The transmit signal in $\boldsymbol{x}$ has

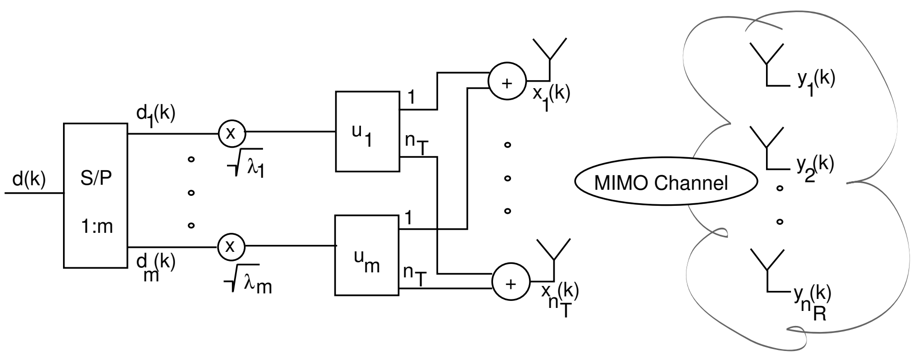

Figure 2.1: Signal processing structure for the MIMO system.

a covariance matrix $\boldsymbol{Q}$ with eigenvalues $p_1, ..., p_m, 0, ..., 0$ and eigenvectors $\boldsymbol{u}_1, ..., \boldsymbol{u}_m$. In order to construct a transmit signal with a given covariance matrix, two signal processing steps are necessary: The first step is the power control $p_1, ..., p_m$ and the second step is multiplying the beamformers $\boldsymbol{u}_1, ..., \boldsymbol{u}_m$. The sum transmit power

$\sum_{k=1}^{n_T} p_k$ is constrained, i.e.

$$\sum_{k=1}^{n_T} p_k = P.$$

Next, we study the stochastic properties of the channel. The correlation of the channel vectors arises in the common downlink transmission scenario in which the base station is un-obstructed [SFGK00]. We follow the model in [GP94] where the subspaces and directions of the paths between the transmit antennas and the receive cluster change more slowly than the actual attenuation of each path.

The channel matrix $\boldsymbol{H}$ for the case in which we have *correlated transmit and correlated receive antennas* is modelled as

$$\boldsymbol{H} = \boldsymbol{R}_R^{\frac{1}{2}} \cdot \boldsymbol{W} \cdot \boldsymbol{R}_T^{\frac{1}{2}} \tag{2.3}$$

with transmit correlation matrix $\boldsymbol{R}_T = \boldsymbol{U}_T \boldsymbol{D}_T \boldsymbol{U}_T^H$ and receive correlation matrix $\boldsymbol{R}_R = \boldsymbol{U}_R \boldsymbol{D}_R \boldsymbol{U}_R^H$. $\boldsymbol{U}_T$ and $\boldsymbol{U}_R$ are the matrices with the eigenvectors of $\boldsymbol{R}_T$ and $\boldsymbol{R}_R$ respectively, and $\boldsymbol{D}_T$, $\boldsymbol{D}_R$ are diagonal matrices with the eigenvalues of the matrix $\boldsymbol{R}_T$ and $\boldsymbol{R}_R$, respectively. The random matrix $\boldsymbol{W}$ has zero-mean independent complex Gaussian identically distributed entries, i.e. $\boldsymbol{W} \sim \mathcal{CN}(0, \boldsymbol{I})$.

The most general form of the correlation model consists of a very large correlation matrix of size $(n_T \cdot n_R \times n_T \cdot n_R)$ which incooperates the transmit and receive correlation, i.e. it is the expectation of the outer product of the vectorised channel matrix

$$\boldsymbol{\kappa} = \mathbb{E}\left[\mathrm{vec}(\boldsymbol{H}) \cdot \mathrm{vec}(\boldsymbol{H})^H\right]. \tag{2.4}$$

The correlation matrix $\boldsymbol{\kappa}$ in (2.4) expresses the correlation between each transmit or receive element to every other transmit or receive element. Often, the transmit and the receive antenna array are spatially divided. Then, the following simplification is allowed [CTK02]. In the case in which each receive antenna observes the same correlation between the transmit antennas, i.e. the transmit correlation is independent of the receive antenna and vice versa the receive correlation is independent of the transmit antenna, the correlation model in (2.4) simplifies in comparison to the model in (2.3).

We assume that the entries in the channel matrix are complex Gaussian distributed. Therefore, we can describe the impact of the correlation at the transmitter and the receiver by considering the second moment, i.e. the covariance matrix. The analysis in [CFV00, CFGV02] can not be applied to our scenario because the entries of the channel matrix in [CFV00, CFGV02] are products of complex Gaussian distributed entries. It does not suffice to consider only the second moment in order to analyse the impact of correlation or keyholes. However, the authors in [CFV00, CFGV02] argue that a MIMO system with 'many' keyholes converges to the common MIMO model by application of the central limit theorem.

The model of correlation which we introduced in (2.3) allows to analyse the different performance metrics from chapter 1 with respect to correlation at the transmit and the receive side. In order to provide a stable mathematical theory for studying correlation, we give a mathematical measure of correlation in the next section.

2.2.2 A mathematical measure of correlation

In order to provide a measure of correlation, we take two arbitrarily chosen transmit correlation matrices $\boldsymbol{R}_T^1$ and $\boldsymbol{R}_T^2$ with the constraint that trace($\boldsymbol{R}_T^1$) = trace($\boldsymbol{R}_T^2$) = n_T which is equivalent to

$$\sum_{l=1}^{n_T} \lambda_l^{T,1} = \sum_{l=1}^{n_T} \lambda_l^{T,2}, \tag{2.5}$$

with $\lambda_l^{T,1}$, $1 \leq l \leq n_T$, and $\lambda_l^{T,2}$, $1 \leq l \leq n_T$, are the eigenvalues of the covariance matrix $\boldsymbol{R}_T^1$ and $\boldsymbol{R}_T^2$, respectively.

This constraint regarding the trace of the correlation matrix $\boldsymbol{R}_T$ is necessary because the comparison of two transmission scenarios is only valid if the average path loss is equal. Without receive correlation, the trace of the correlation matrix can be written as

$$\text{trace}(\boldsymbol{R}_T) = \sum_{i=1}^{n_T} \left(\mathbb{E}\left[\boldsymbol{H}\boldsymbol{H}^H\right]\right)_{ii} = \sum_{i=1}^{n_T} \mathbb{E}\left[|\boldsymbol{h}_i|^2\right]. \tag{2.6}$$

However, the RHS of (2.6) is the sum of the average path loss from the transmit antenna $i = 1...n_T$. In order to study the impact of correlation on the achievable capacity separately, the average path loss is kept fixed by applying the trace constraint on the correlation matrices $\boldsymbol{R}_T^1$ and $\boldsymbol{R}_T^2$.

We will say that a correlation matrix $\boldsymbol{R}_T^1$ is more correlated than $\boldsymbol{R}_T^2$ with descending ordered eigenvalues $\lambda_1^{T,1} \geq \lambda_2^{T,1} \geq ... \geq \lambda_{n_T}^{T,1} \geq 0$ and $\lambda_1^{T,2} \geq \lambda_2^{T,2} \geq ... \geq \lambda_{n_T}^{T,2} \geq 0$ if

$$\sum_{k=1}^{m} \lambda_k^{T,1} \geq \sum_{k=1}^{m} \lambda_k^{T,2} \quad 1 \leq m \leq n_T - 1. \tag{2.7}$$

The measure of correlation which we will introduce is defined in a natural way: the larger the first m eigenvalues of the correlation matrices are (with the trace constraint in (2.6)), the more correlated is the MIMO channel. As a result, the most uncorrelated MIMO channel has equal eigenvalues, whereas the most correlated MIMO channel has only one non-zero eigenvalue which is given by $\lambda_1 = n_T$.

Before proceeding with our definition of 'more correlated' in terms of the eigenvalue distribution of the channel covariance matrix, we give the necessary definitions we will need in the following.

Definition 1: For two vectors $\boldsymbol{x}, \boldsymbol{y} \in R^n$ with descending ordered components $x_1 \geq x_2 \geq ... \geq x_n \geq 0$ and $y_1 \geq y_2 \geq ... \geq y_n \geq 0$ one says that the vector $\boldsymbol{x}$ majorizes the vector $\boldsymbol{y}$ and writes

$$\boldsymbol{x} \succ \boldsymbol{y} \ \text{ if } \ \sum_{k=1}^{m} x_k \geq \sum_{k=1}^{m} y_k \ \ , m = 1, ..., n-1. \ \text{ and } \ \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k.$$

The next definition describes a function Φ which is applied to the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ with $\boldsymbol{x} \succ \boldsymbol{y}$:

Definition 2: A real-valued function Φ defined on $\mathcal{A} \subset R^n$ is said to be *Schur-*

convex on $\mathcal{A}$ if

$$\boldsymbol{x} \succ \boldsymbol{y} \text{ on } \mathcal{A} \Rightarrow \Phi(\boldsymbol{x}) \geq \Phi(\boldsymbol{y}).$$

Similarly, Φ is said to be *Schur-concave* on $\mathcal{A}$ if

$$\boldsymbol{x} \succ \boldsymbol{y} \text{ on } \mathcal{A} \Rightarrow \Phi(\boldsymbol{x}) \leq \Phi(\boldsymbol{y}).$$

Example: Suppose that $\boldsymbol{x}, \boldsymbol{y} \in R_+^n$ are positive real numbers and the function Φ is defined as the sum of the squared components of the vectors, i.e. $\Phi_2(\boldsymbol{x}) = \sum_{k=1}^n |x_k|^2$. Then, it is easy to show that the function Φ_2 is Schur-concave on R_+^n, i.e. if $\boldsymbol{x} \succ \boldsymbol{y} \Rightarrow \Phi_2(\boldsymbol{x}) \leq \Phi_2(\boldsymbol{y})$.

We will need the following lemma (see [MO79, Theorem3.A.4]) which is sometimes called Schur's condition. It provides an approach for testing whether some vector valued function is Schur-convex or not.
Lemma 1: Let $\mathcal{I} \subset \mathbb{R}$ be an open interval and let $f : \mathcal{I}^n \rightarrow \mathbb{R}$ be continuously differentiable. Necessary and sufficient conditions for f to be Schur-convex on $\mathcal{I}^n$ are

$$f \text{ is symmetric on } \mathcal{I}^n \tag{2.8}$$

and

$$(x_i - x_j)\left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}\right) \geq 0 \quad \text{forall } 1 \leq i, j \leq n. \tag{2.9}$$

Since $f(\boldsymbol{x})$ is symmetric, Schur's condition can be reduced as in [MO79, p. 57]

$$(x_1 - x_2)\left(\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2}\right) \geq 0. \tag{2.10}$$

From Lemma 1 follows that $f(\boldsymbol{x})$ is a Schur-concave function on $\mathcal{I}^n$ if $f(\boldsymbol{x})$ is symmetric and

$$(x_1 - x_2)\left(\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2}\right) \leq 0. \tag{2.11}$$

The definition of Schur-convexity and Schur-concavity can be extended if another function $\Psi : R \rightarrow R$ is applied to $\Phi(\boldsymbol{x})$. Assume that Φ is Schur-concave, if the function Ψ is monotonically increasing then the expression $\Psi(\Phi(\boldsymbol{x}))$ is Schur-concave, too. If we take for example the function $\Psi(n) = \log(n)$ for $n \in R_+$ and the function Φ_p from the example above, we can state that the composition of the two functions $\Psi(\Phi_p(\boldsymbol{x}))$ is Schur-concave on R_+^n. This result can be generalised for all possible compositions of monotonically increasing as well as decreasing functions, and Schur-convex as well as Schur-concave functions. For further information about majorization theory see [MO79].

The following definition provides a measure for comparison of two correlation matrices.
Definition 3: The transmit correlation matrix $\boldsymbol{R}_T^1$ is more correlated than $\boldsymbol{R}_T^2$ if

and only if

$$\sum_{l=1}^{m} \lambda_l^{T,1} \geq \sum_{l=1}^{m} \lambda_l^{T,2} \quad \text{for} \quad m = 1...n_T, \quad \text{and} \quad \sum_{l=1}^{n_T} \lambda_1^{T,1} = \sum_{l=1}^{n_T} \lambda_2^{T,2}. \tag{2.12}$$

One says that the vector consisting of the ordered eigenvalues $\boldsymbol{\lambda}_1^T$ majorizes $\boldsymbol{\lambda}_2^T$, and this relationship can be written as $\boldsymbol{\lambda}_1^T \succ \boldsymbol{\lambda}_2^T$ like in Definition 1.

Remark I: It can be shown that vectors with more than two components cannot be totally ordered. So there are examples of correlation vectors that cannot compared using our Definition 3, e.g. $\boldsymbol{\eta}^1 = [0.6, 0.25, 0.15]$ and $\boldsymbol{\eta}^2 = [0.55, 0.4, 0.05]$. This is a problem of all possible orders for comparing correlation vectors. Majorization induces only a partial order.

Note that our definition of correlation in Definition 3 differs from the usual definition in statistics. In statistics a diagonal covariance matrix indicates that the random variables are uncorrelated. This is independent of the auto-covariances on the diagonal. In our definition, we say that the antennas are uncorrelated if in addition to statistical independence, the auto-covariances of all entries are equal. This difference to statistics occurs because the direction, i.e. the unitary matrices of the correlation have no impact on our measure of correlation. Imagine the scenario in which all transmit antennas are uncorrelated, but have different average transmit powers because of their amplifiers. In a statistical sense, one would say the antennas are uncorrelated. Our measure of correlation says that the antennas are correlated, because they have different transmit powers. The measure of correlation in Definition 3 is more suitable for the analysis of the performance of multiple antenna systems, because different transmit powers at the antennas obviously have a strong impact on the performance. In this thesis, these effects are considered.

This measure of correlation allows us to analyse the impact of correlation on the various performance metrics introduced in chapter 1 in single-user MIMO systems under different types of CSI. In the following, the measure of correlation is applied to transmit correlation matrices $\boldsymbol{R}_T$ and to receive correlation matrices $\boldsymbol{R}_R$ as well.

Remark II: As mentioned above, the case in which the transmit antennas are fully correlated corresponds to $\lambda_1^T = n_T$, $\lambda_2^T = ... = \lambda_{n_T}^T = 0$. The case in which the transmit antennas are fully uncorrelated corresponds to $\lambda_1^T = \lambda_2^T = ... = \lambda_{n_T}^T = 1$. This illustrates that the expression in (2.12) can be used as a measure for correlation.

Example: At this point, we give another example for the measure of correlation. Assume the situation in figure (2.2). We have two different correlation scenarios. In scenario A and B the largest two eigenvalues ($\lambda_1^A = \lambda_1^B$ and $\lambda_2^A = \lambda_2^B$) are equal. The smallest three eigenvalues in scenario B are equal ($\lambda_3^B = \lambda_4^B = \lambda_5^B$) but in scenario A the smallest three eigenvalues are unequal ($\lambda_3^A > \lambda_4^A > \lambda_5^A$). In addition to this, the sum of all eigenvalues in scenario A and B is equal. Applying the order which is introduced in Definition 3, eigenvalue vector A majorizes eigenvector B ($\boldsymbol{\lambda}^A \succ \boldsymbol{\lambda}^B$).

Scenario B applies for all eigenvalue distributions $\boldsymbol{\lambda}$ with fixed λ_1 and λ_2 and equal trace the 'smallest' eigenvalue distribution, i.e. $\boldsymbol{\lambda}^B \prec \boldsymbol{\lambda}$ for all $\boldsymbol{\lambda}$ with

$$\lambda_1 + \lambda_2 + \sum_{k=3}^{n_T} \lambda_k = 1$$

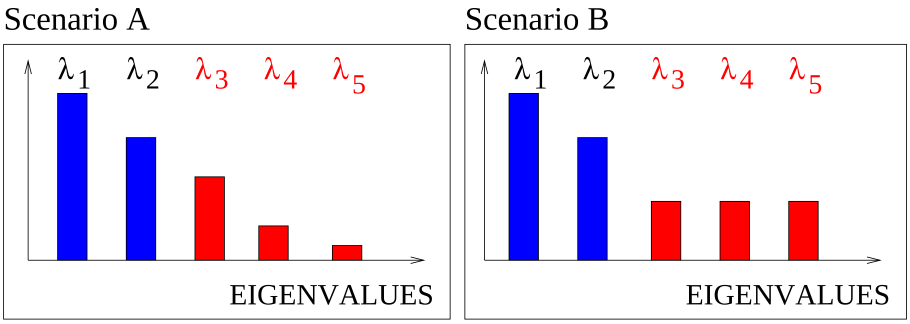

Figure 2.2: Example correlation matrix eigenvalue distribution.

and for $\boldsymbol{\lambda}^B$ with

$$\frac{1-\lambda_1^B-\lambda_2^B}{n_T-2}=\lambda_3^B=\ldots=\lambda_{n_T}^B.$$

Entropy of Gaussian vector As another example of the measure of correlation, consider a random Gaussian distributed vector $\boldsymbol{z}$ of dimension n with $\boldsymbol{z}\sim CN(0,\boldsymbol{R})$ and with covariance matrix $\boldsymbol{R}$. Denote the eigenvalues of the covariance matrix $\boldsymbol{R}$ as $\boldsymbol{r}=[r_1,...,r_n]$. In the following, we show that the entropy $h_r(\boldsymbol{z})$ is a Schur-concave function with respect to the correlation eigenvalues $\boldsymbol{r}$. The entropy of $\boldsymbol{z}$ is given by [CT91, Theorem 9.6.5]

$$h_r(\boldsymbol{z})=\log\left[(2\pi e)^n\det(\boldsymbol{R})\right]=\log\left[(2\pi e)^n\prod_{i=1}^n r_i\right]. \tag{2.13}$$

Further on, we have the following result by [MO79, Theorem 3.F.1.a]

$$\boldsymbol{\lambda}\succ\boldsymbol{\mu}\rightarrow\prod_{i=1}^n\lambda_i\le\prod_{i=1}^n\mu_i. \tag{2.14}$$

Therefore, from (2.13) and (2.14) follows

$$\boldsymbol{\lambda}\succ\boldsymbol{\mu}\rightarrow h_\lambda(\boldsymbol{z})\le h_\mu(\boldsymbol{z}).$$

This can be shown alternatively, by

$$h_r(\boldsymbol{z})=\log\left[(2\pi e)^n\prod_{i=1}^n r_i\right]=\log(2\pi e)^n+\sum_{i=1}^n\log r_i. \tag{2.15}$$

According to [MO79, Proposition 3.C.1],

$$\boldsymbol{\lambda}\succ\boldsymbol{\mu}\rightarrow\sum_{i=1}^n\log\lambda_i\le\sum_{i=1}^n\log\mu_i.$$

2.3 Average performance metrics

We assume that the receiver has perfect CSI. The transmitter has either no CSI, perfect CSI, or knowledge of the transmit correlation matrix $\boldsymbol{R}_T$ and receive correlation matrix $\boldsymbol{R}_R$. Let us study the optimisation problems regarding the average performance metrics average mutual information (1.8) and average normalised MSE (1.13) from chapter 1.

All average performance metrics introduced in chapter 1 can be written in the following generalised form

$$\Phi(\rho, \boldsymbol{Q}, \boldsymbol{R}_T, \boldsymbol{R}_R) = \mathbb{E}_{\boldsymbol{W}} \text{ tr } \phi(\rho \boldsymbol{R}_R^{1/2} \boldsymbol{W} \boldsymbol{R}_T^{1/2} \boldsymbol{Q} \boldsymbol{R}_T^{1/2} \boldsymbol{W}^H \boldsymbol{R}_R^{1/2}) \tag{2.16}$$

with zero-mean complex Gaussian iid random matrix $\boldsymbol{W}$ and with the matrix-valued function ϕ defined on the set of positive semidefinite matrices. Using the following matrix-valued function $\phi_1(\boldsymbol{X})$ in the RHS of (2.16)

$$\phi_1(\boldsymbol{X}) = \log(\boldsymbol{I} + \boldsymbol{X}) \tag{2.17}$$

the generalised average performance function Φ in (2.16) corresponds in every detail to the average mutual information in (1.8). Using another matrix-valued function $\phi_2(\boldsymbol{X})$ in (2.16)

$$\phi_2(\boldsymbol{X}) = \frac{n_T}{n_R} \boldsymbol{I} - \boldsymbol{X} [\boldsymbol{I} + \boldsymbol{X}]^{-1} \tag{2.18}$$

the function Φ in (2.16) directly corresponds to the average normalised MSE in (1.13). The average normalised MSE is to be minimised. The first term in the RHS of (2.18) does neither depend on the transmission strategy nor on the transmit or receive correlation. It depends only on the number of transmit and receive antennas. Therefore, the minimisation of the normalised average MSE can be expressed as a maximisation of the function

$$\tilde{\phi}_2(\boldsymbol{X}) = \boldsymbol{X} [\boldsymbol{I} + \boldsymbol{X}]^{-1} \tag{2.19}$$

In (2.16) it is assumed that the transmit signals are complex Gaussian distributed with transmit covariance matrix $\boldsymbol{Q}$ with power constraint tr $(\boldsymbol{Q}) = P$. The transmit covariance matrix can be understood as an operator that maps from the set of all channel matrices $\mathcal{H} = \{\boldsymbol{H}\}$ to the set of transmit covariance matrices under the trace constraint $\mathcal{Q} = \{\boldsymbol{Q} : \boldsymbol{Q} \succeq \boldsymbol{0}, \text{ tr } \boldsymbol{Q} = P\}$. The maximum of (2.16) with respect to transmit strategy $\boldsymbol{Q}$ depends on the CSI at the transmitter. In general, the generalised performance function Φ in (2.16) is optimised with respect to transmit policy $\boldsymbol{Q}$ under the trace constraint. As a result, for fixed SNR ρ, transmit $\boldsymbol{R}_T$ and receive correlation matrices $\boldsymbol{R}_R$, we obtain the following class of optimisation problems

$$\text{max or min} \quad \Phi(\boldsymbol{Q}, \rho, \boldsymbol{R}_T, \boldsymbol{R}_R) \text{ s.t. } \boldsymbol{Q} \succeq \boldsymbol{0} \text{ and } \text{tr } \boldsymbol{Q} = P. \tag{2.20}$$

If the transmitter has perfect CSI, the transmitter can adapt its transmit strategy $\boldsymbol{Q}$ to every channel matrix realization $\boldsymbol{H}$. The transmit strategy changes with every channel realization. In case of average mutual information maximisation, the optimal transmission strategy is the well known 'water-filling' [Tel99], [CT91]. Then, the maximum mutual information corresponds to the *ergodic channel capacity* of a closed-loop MIMO system. In case of average normalised MSE minimisation, the optimal transmission strategy is some kind of altered 'water-filling' [HB03]. In section 2.3.2, we derive the general structure of the optimal transmit strategy $\boldsymbol{Q}$ for arbitrary inner performance function $\boldsymbol{\phi}$.

If the transmitter has knowledge about the transmit and receive correlation, the transmit strategy does not depend on each instantaneous channel realization. The transmit covariance matrix is kept fixed for a pair of transmit and receive correlation matrices. In order to describe the optimisation problem which leads to the optimal transmission strategy with covariance feedback, we discuss the expression in (2.16).

The main problem in computing the average generalised performance metric Φ in (2.16) is the expectation operator. We obtain for the expectation of the inner matrix-valued function

$$\begin{aligned} & \mathbb{E}_{\boldsymbol{W}} \operatorname{tr} \phi(\rho \boldsymbol{R}_R^{1/2} \boldsymbol{W} \boldsymbol{R}_T^{1/2} \boldsymbol{Q} \boldsymbol{R}_T^{1/2} \boldsymbol{W}^H \boldsymbol{R}_R^{1/2}) \\ = \; & \mathbb{E}_{\boldsymbol{W}} \operatorname{tr} \phi(\rho \boldsymbol{\Lambda}_R^{1/2} \boldsymbol{W} \boldsymbol{\Lambda}_T^{1/2} \boldsymbol{Q} \boldsymbol{\Lambda}_T^{1/2} \boldsymbol{W}^H \boldsymbol{\Lambda}_R^{1/2}) \\ = \; & \mathbb{E}_{\boldsymbol{w}_{1,\dots,n_T}} \operatorname{tr} \phi\left(\rho \sum_{k=1}^{n_T} \lambda_k^T p_k \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H\right). \end{aligned} \tag{2.21}$$

The first identity in (2.21) follows from the fact that the distribution of the random variable $\boldsymbol{W}$ and $\boldsymbol{WU}$ and $\boldsymbol{UW}$ for unitary $\boldsymbol{U}$ are equal. This fact has been extensively used in [MH99]. Furthermore, note that the trace of $\boldsymbol{UWU}^H$ is equal to the trace of $\boldsymbol{W}$ for unitary $\boldsymbol{U}$. The random vectors $\tilde{\boldsymbol{w}}_k$ for $1 \leq k \leq n_T$ are given by

$$\tilde{\boldsymbol{w}}_k = \begin{pmatrix} \sqrt{\lambda_1^R} w_{1,k} \\ \sqrt{\lambda_2^R} w_{2,k} \\ \vdots \\ \sqrt{\lambda_{n_R}^R} w_{n_R,k} \end{pmatrix} \tag{2.22}$$

with zero-mean complex Gaussian iid random variables $w_{i,j}$ with $1 \leq i \leq n_R$ and $1 \leq j \leq n_T$. Note that the LHS in (2.21) can be further rewritten as

$$\operatorname{tr} \phi(\rho \boldsymbol{R}_R^{1/2} \boldsymbol{W} \boldsymbol{R}_T^{1/2} \boldsymbol{Q} \boldsymbol{R}_T^{1/2} \boldsymbol{W}^H \boldsymbol{R}_R^{1/2}) = \operatorname{tr} \phi\left(\rho \sum_{k=1}^{n_R} \lambda_k^R \hat{\boldsymbol{w}}_k \hat{\boldsymbol{w}}_k^H\right) \tag{2.23}$$

with $\hat{\boldsymbol{w}}_k$ for $1 \leq k \leq n_R$ as

$$\hat{\boldsymbol{w}}_k = \begin{pmatrix} \sqrt{p_1 \lambda_1^T} w_{k,1} \\ \sqrt{p_2 \lambda_2^T} w_{k,2} \\ \vdots \\ \sqrt{p_{n_T} \lambda_{n_T}^T} w_{k,n_T} \end{pmatrix}. \tag{2.24}$$

The equations (2.21) and (2.23) express the symmetry of the average performance metric with respect to the correlation properties at the transmitter and the receiver. For the equal power allocation transmit strategy, one can observe that the impact of correlation at the transmit antennas or at the receive antennas is equal. This intuitive explanation will be affirmed in Corollary 7 and Theorem 3.

In the case in which the transmitter knows the transmit correlation matrix and receive correlation matrix, the optimum eigenvectors of the transmit covariance matrix $\boldsymbol{Q}$ correspond to the eigenvectors of the transmit correlation matrix. This will be proven in section 2.3.2. For the average mutual information, this was shown in [JVG01] for completely uncorrelated receive antennas and in [JB03b] for receive correlation. The optimal power allocation is given by

$$\boldsymbol{\Lambda}_Q^{opt} = \arg_{\boldsymbol{p}} \left[\max_{\sum_{k=1}^{n_T} p_k = P} \mathbb{E} \operatorname{tr} \left(\phi\left(\rho \sum_{i=1}^{n_T} \lambda_i^T p_i \tilde{\boldsymbol{w}}_i \tilde{\boldsymbol{w}}_i^H\right)\right)\right] \tag{2.25}$$

with $\tilde{\boldsymbol{w}}_i$ given in (2.22). Using the optimum transmit strategy characterised so far, the maximum of the average mutual information (for ϕ_1 in (2.25)) corresponds to the *ergodic channel capacity* of a MIMO system with long-term knowledge about the transmit and receive correlation.

If the transmitter has no CSI, it can neither adapt the transmit strategy to the instantaneous channel realization nor to the long-term channel information. The transmit strategy is fixed for all channel realizations. Furthermore, it is difficult to define a channel capacity, because the average mutual information itself is a variable which depends on the transmit and receive correlation. In addition to this, it is not obvious how to define the *best* transmit strategy. What correlation should be assumed? In the next subsection, we propose an approach in order to solve these problems. We show that the optimal transmit covariance matrix which solves (2.20) under the worst case channel correlation (see section 2.3.2) is given by

$$\boldsymbol{Q}^{opt} = \frac{P}{n_T}\boldsymbol{I}. \tag{2.26}$$

In the next section, we prove that equal power allocation as assumed in (2.26) is most robust against worst case correlation without CSI at the transmitter.

Additionally, the solution to the optimisation problem in (2.20) is characterised and the well-known waterfilling solution for perfect CSI at the transmitter is briefly reviewed in the next section.

2.3.1 Properties of the inner matrix-valued performance function: Matrix-concavity

Before proceeding to the next section, it is important to know the properties of the function ϕ in the optimisation problem (2.20). The following definitions will prove useful. Matrices and functions are discussed in [Bha97, HJ91].

The function $\boldsymbol{F} : \mathbb{C}^n_+ \rightarrow \mathbb{C}^n_+$ is a matrix-valued function. The function $\boldsymbol{F}$ maps from the set of positive semidefinite matrices to the set of positive semidefinite matrices. We consider only matrix-valued functions which act by a scalar function $F(x)$ on each eigenvalue of the matrix $\boldsymbol{A}$ motivated by the spectral theorem in linear algebra. Therefore, the function $\boldsymbol{F}$ affects only the eigenvalues of the matrix $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$, i.e. $\boldsymbol{F}(\boldsymbol{A}) = \boldsymbol{U}\boldsymbol{F}(\boldsymbol{\Lambda})\boldsymbol{U}^H$ and $\boldsymbol{F}(\boldsymbol{\Lambda}) = \operatorname{diag}(F(\lambda_1), ..., F(\lambda_n))$.

We will assume that F is *matrix-monotone*, i.e. $\boldsymbol{F}(\boldsymbol{A}) \succeq \boldsymbol{F}(\boldsymbol{B})$ holds if $\boldsymbol{A} - \boldsymbol{B} \succeq \boldsymbol{0}$ [Bha97, Section V]. In our case the function $\boldsymbol{\phi}_1(\boldsymbol{A})$ for the mutual information and the function $\tilde{\boldsymbol{\phi}}_2(\boldsymbol{A})$ for the normalised MSE are matrix-monotone.

A matrix-valued function is *matrix-concave* if

$$\boldsymbol{F}\left[(1-\lambda)\boldsymbol{A} + \lambda\boldsymbol{B}\right] \succeq (1-\lambda)\boldsymbol{F}(\boldsymbol{A}) + \lambda\boldsymbol{F}(\boldsymbol{B}). \tag{2.27}$$

A function $\boldsymbol{F}$ is called *matrix-convex* if $-\boldsymbol{F}$ is matrix-convex. Both performance functions $\boldsymbol{\phi}_1(\boldsymbol{X})$ and $\boldsymbol{\phi}_2(\boldsymbol{X})$ are matrix-concave and matrix-convex, respectively.

Therefore, we restrict our analysis to the cases in which the inner function $\boldsymbol{\phi}$ in the optimisation problem (2.20) are matrix-concave or matrix-monotone functions[1].

Next, we define according to Löwner's Theorem [Lö34] [Bha97, Section V.4] the

[1]In [HP82, Theorem 2.5] it is shown that all operator monotone functions on $(0, \infty)$ are operator concave.

following class of matrix valued functions

$$\tilde{\boldsymbol{F}}(\boldsymbol{A}) = \int_0^\infty s\boldsymbol{A}\left(s\boldsymbol{I}+\boldsymbol{A}\right)^{-1} d\mu(s) \tag{2.28}$$

with a positive measure $\mu \in [0,\infty)$, i.e. μ is monotonically increasing, positive, and continuous from the left [KF70], with entries

$$\left[\tilde{\boldsymbol{F}}(\boldsymbol{A})\right]_{kl} = \int_0^\infty s\left[\boldsymbol{A}\left(s\boldsymbol{I}+\boldsymbol{A}\right)^{-1}\right]_{kl} d\mu(s). \tag{2.29}$$

Every matrix-monotone function $\boldsymbol{F}$ can be expressed by a unique positive measure μ defined on $[0,\infty]$ as the function $\tilde{\boldsymbol{F}}$ in (2.28) and (2.29) plus a constant term and a linear term. In this thesis, we consider only the class of matrix-monotone functions which can be expressed with the representation in (2.28).

As a result, the optimisation problem in (2.20) can be expressed using Löwner's representation of matrix-convex functions as

$$\begin{aligned}\Phi(\rho,\boldsymbol{Q},\boldsymbol{R}_T,\boldsymbol{R}_R) \quad &= \quad \mathbb{E}\ \mathrm{tr} \int_0^\infty s\left(\rho\boldsymbol{R}_R^{1/2}\boldsymbol{W}\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}\boldsymbol{W}^H\boldsymbol{R}_R^{1/2}\right) \\ &\quad \cdot\left[s\boldsymbol{I}+\left(\rho\boldsymbol{R}_R^{1/2}\boldsymbol{W}\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}\boldsymbol{W}^H\boldsymbol{R}_R^{1/2}\right)\right]^{-1} d\mu(s).\end{aligned} \tag{2.30}$$

Note that the measure $d\mu(s)$ in (2.30) depends on the inner performance function ϕ. Often, the representation in (2.30) is useful for solving the optimisation problem (2.20) in the forthcoming sections. Note that the term inside the integral equals the second summand of the matrix-valued function ϕ_2. The integral and the trace operator can be interchanged because the trace is a finite sum.

Often, we will need the first derivative of the matrix valued function F at $\boldsymbol{A}$

$$DF(\boldsymbol{A})(\boldsymbol{H}) = \left.\frac{d}{dt}\right|_{t=0} F(\boldsymbol{A}+t\boldsymbol{H}).$$

There is an interesting relationship between the derivative $DF(\boldsymbol{A})$ and the matrix $F^{[1]}(\boldsymbol{A})$. According to the approach in this section, it is defined in the following way: $F^{[1]}(\boldsymbol{A})$ acts on the eigenvalues of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}_A\boldsymbol{U}^H$ with

$$F^{[1]}(\boldsymbol{A}) = \boldsymbol{U}\mathrm{diag}(f'(\lambda_1), f'(\lambda_2), ..., f'(\lambda_n))\boldsymbol{U}_H \tag{2.31}$$

Since the matrix $F^{[1]}(\boldsymbol{A})$ can be conveniently expressed using the positive measure μ, we will often use this matrix in the following. Let us call the matrix $F^{[1]}(\boldsymbol{A})$ the 'outer derivative' of F.

Examples of performance metrics of the matrix-monotone class

In communications systems analysis, the performance of the system is often measured by a function which belongs to the aforementioned class, i.e. it can be represented as the trace of a matrix-monotone function with channel $\boldsymbol{H}$ and transmit strategy $\boldsymbol{Q}$ as an argument $\mathbb{E}$ tr $\phi(\rho\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H)$.

The average mutual information can be expressed as this type of function as shown above $\phi(\boldsymbol{X}) = \log\det\left(\boldsymbol{I}+\rho\boldsymbol{X}\right)$. The distribution $\mu(s)$ in the Löwner representation is then given by $\mu(s) = \frac{u(s-1)}{s^2}$ with the step function $u(t) = 0$ for $t < 0$ and $u(t) = 1$ for $t \geq 0$.

For the normalized MSE, the function ϕ is $\phi(\boldsymbol{X}) = \boldsymbol{X}\left[\boldsymbol{I} + \boldsymbol{X}\right]^{-1}$. In this case, the distribution $\mu(s)$ is given by $\mu(s) = \delta(s-1)$ with the dirac impulse $\delta(0) = 1$ and otherwise $\delta(t) = 0$ for $t \neq 0$.

The Chernoff bound of orthogonal space-time block codes can be represented [NBP02] by taking the natural logarithm and $\phi(\boldsymbol{X}) = \log(n_\alpha)(-\alpha \boldsymbol{X})$ with symbol constellation parameters n_α as the number of signal point with minimum distance and $\alpha = \rho \cdot d_{min}$ with SNR ρ and minimum distance d_{min}.

2.3.2 Optimum transmission strategies

In this section, we derive the transmission strategies which maximise the average performance metrics for the single-user MIMO system in (2.20) with no CSI, covariance knowledge, and perfect CSI at the transmitter.

Optimum transmission strategy without CSI at the transmitter

In this section, we derive the optimum transmission strategy without CSI at the transmitter and with correlation at the transmit antenna array. In order to do so we have to introduce additional constraints on the transmit correlation matrix. In the following, we consider the least favourable transmit correlation matrix and search for the transmit strategy which maximises the average performance metric under this correlation matrix. This kind of min-max expression for the mutual information in the case in which the transmitter has no a priori information about the type of channel class, was defined as the channel capacity of the compound channel in [Wol78]. A similar approach in [PCL03b] studies the maximum with respect to the transmit covariance matrix and the minimum with respect to the channel realization of the instantaneous capacity in a flat-fading MIMO channel. In addition to this, the worst case capacity of a MIMO system is studied in [GHIM01]. In our case, the generalised performance function Φ is to be maximised by transmit covariance matrix and to be minimised by transmit correlation matrix.

The line of argument is the following: We show that the transmission strategy which is most robust against the worst case transmit correlation is equal power allocation. This is done by an upper and a lower bound which coincide with another. In order to derive the lower bound, it is necessary to know the worst case correlation which minimises the performance function if equal power allocation is applied. This result is derived in section 2.3.3 where the impact of correlation on the average performance function is analysed. Therefore, our line of argument is not a circular proof. In the derivation of the optimum strategy we merely use the fact that the worst case transmit correlation for equal power allocation is the completely correlated case. The receive correlation is kept fixed during the derivation.

Let us define

$$F(\boldsymbol{Z}) = \mathbb{E} \operatorname{tr} \phi(\rho \tilde{\boldsymbol{W}} \boldsymbol{Z} \tilde{\boldsymbol{W}}^H) \tag{2.32}$$

with matrix $\tilde{\boldsymbol{W}} = \boldsymbol{R}_R^{1/2} \cdot \boldsymbol{W}$. Then the average mutual information or the normalised average MSE with correlation at the transmit antenna array can be written as $F(\boldsymbol{R}_T^{1/2} \boldsymbol{Q} \boldsymbol{R}_T^{1/2})$ with ϕ_1 or $\tilde{\phi}_2$ as inner performance function. Obviously, the transmitter is going to maximise the function F, but it has no a priori information about the transmit correlation, i.e. the channel belongs to some class but the transmitter does not know which. Following the approach of the compound channel in [Wol78], the transmitter is pessimistic and it assumes that the channel belongs to the worst case class of channels [2]. The worst case correlation is obtained if we assume that the correlation player knows the strategy of the transmit player, i.e. the correlation is a function of the transmit strategy $\boldsymbol{R}_T(\boldsymbol{Q})$. The performance function is minimised with respect to this correlation and we have the following max-min problem

$$\begin{aligned}\underline{\phi} &= \sup_{\operatorname{tr}(\boldsymbol{Q})\le P} \inf_{\operatorname{tr}(\boldsymbol{R_T})=n_T} F(\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}) \\ &= \sup_{\sum_{k=1}^{n_T} p_k)\le P} \inf_{\operatorname{tr}(\boldsymbol{R_T})=n_T} \mathbb{E}\operatorname{tr}\phi\left(\rho\sum_{k=1}^{n_T} p_k\lambda_k^T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right). \end{aligned} \tag{2.33}$$

For the inner infimum with fixed transmit covariance matrix $\boldsymbol{Q}$ we have

$$\inf_{\operatorname{tr}(\boldsymbol{R}_T)=n_T} F(\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}) \le \min_{1\le k\le n_T} \mathbb{E}\operatorname{tr}\phi\left(\rho p_k n_T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right) \tag{2.34}$$

where we have used the rank one diagonal correlation matrix which has zeros on the diagonal but on the k-th position, i.e.

$$\boldsymbol{R}_T = \operatorname{diag}[0,...,0,n_T,0,...,0].$$

The RHS of (2.34) can be written as

$$\begin{aligned}&\min_{1\le k\le n_T} \mathbb{E}\operatorname{tr}\phi\left(\rho p_k n_T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right) =^{(a)} \mathbb{E}\operatorname{tr}\phi\left(\rho p_{\min} n_T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right) \\ &\le^{(b)} \mathbb{E}\operatorname{tr}\phi\left(\rho\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right) =^{(c)} \min_{\operatorname{tr}(\boldsymbol{R_T})=n_T} F(\boldsymbol{R}_T^{1/2}\hat{\boldsymbol{Q}}\boldsymbol{R}_T^{1/2}).\end{aligned} \tag{2.35}$$

In (2.35) we have used in (a) that the minimum is achieved if the k-th position with a nonzero entry in the correlation matrix corresponds to the smallest transmit covariance matrix eigenvalue. Step (b) follows from the observation that for the smallest transmit covariance matrix eigenvalue only P/n_T power can be allocated. Finally, in step (c) we used the equal power allocation transmit covariance matrix, i.e. $\hat{\boldsymbol{Q}} = \frac{P}{n_T}\boldsymbol{I}$. The RHS in (2.35) does not depend on the transmit covariance matrix. If we take the supremum of (2.35) over all transmit covariance matrices with power constraint P, it holds

$$\sup_{\operatorname{tr}(\boldsymbol{Q})\le P} \inf_{\operatorname{tr}(\boldsymbol{R_T})=n_T} F(\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}) \le \mathbb{E}\operatorname{tr}\phi\left(\rho\tilde{\boldsymbol{w}}_1\tilde{\boldsymbol{w}}_1^H\right). \tag{2.36}$$

Furthermore, the following lower bound on the max-min capacity holds

$$\begin{aligned}\sup_{\operatorname{tr}(\boldsymbol{Q})\le P} \inf_{\operatorname{tr}(\boldsymbol{R_T})=n_T} F(\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}) &\ge \inf_{\operatorname{tr}(\boldsymbol{R_T})=n_T} F(\frac{\rho}{n_T}\boldsymbol{R}_T) \\ &= \mathbb{E}\operatorname{tr}\phi(\rho\tilde{\boldsymbol{w}}_1\tilde{\boldsymbol{w}}_1^H).\end{aligned} \tag{2.37}$$

The infimum in (2.37) is characterised in Theorem 1. The result in Theorem 3 states that if equal power allocation is used, the average performance function is lowest for completely correlated transmit antennas. From (2.36) and (2.37) it follows that

[2] An optimistic transmitter would choose to assume the best case class and a realistic transmitter would choose a kind of average. However, we follow the pessimistic transmitter.

the max-min mutual information in (2.32) is given by

$$\sup_{\text{tr }(\boldsymbol{Q})\leq P} \inf_{\text{tr }(\boldsymbol{R_T})=n_T} F(\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}) = \mathbb{E}\text{ tr }\phi(\rho\tilde{\boldsymbol{w}}_1\tilde{\boldsymbol{w}}_1^H) = \mathbb{E}\text{ tr }\phi(\rho\sum_{k=1}^{n_R}\lambda_k^R w_k) \tag{2.38}$$

and can be achieved by equal power allocation, i.e. $\hat{\boldsymbol{Q}} = \frac{P}{n_T}\boldsymbol{I}$. Note, that the argument of the function ϕ in (2.38) is a scalar.

The max-min performance metric in (2.38) can be further analysed. In the RHS of (2.38), the random variables $\tilde{\boldsymbol{w}}_1$ include the receive correlation. The result in Theorem 3 is applied for receive correlation in Corollary 7 and it states that if equal power allocation is used, the average performance function is lowest for completely correlated receive antennas. Therefore, the minimum of the max-min mutual information in (2.38) with respect to receive correlation $\boldsymbol{R}_R$ is given by

$$\inf_{\text{tr }(\boldsymbol{R_R})=n_R} \mathbb{E}\text{ tr }\phi(\rho\tilde{\boldsymbol{w}}_1\tilde{\boldsymbol{w}}_1^H) = \mathbb{E}\text{ tr }\phi(\rho n_R w_1)$$

In the case in which the inner function ϕ is the mutual information, we obtain

$$\inf_{\text{tr }(\boldsymbol{R_R})=n_R} \mathbb{E}\text{ tr }\phi_1(\rho\tilde{\boldsymbol{w}}_1\tilde{\boldsymbol{w}}_1^H) = \mathbb{E}\log(1+\rho n_R w_1) = \exp(1/(\rho n_R))\text{Ei}(1,1/(\rho n_R)).$$

with the exponential integral $\text{Ei}(a,b) = \int_1^\infty exp(-\kappa b)\kappa^{-a}d\kappa$. In the case in which the inner function ϕ is the modified average normalised MSE, we obtain

$$\inf_{\text{tr }(\boldsymbol{R_R})=n_R} \mathbb{E}\text{ tr }\tilde{\phi}_2(\rho\tilde{\boldsymbol{w}}_1\tilde{\boldsymbol{w}}_1^H) = \mathbb{E}\frac{\rho n_R w_1}{1+\rho n_R w_1} = 1 - \frac{\exp(1/(\rho n_R))\text{Ei}(1,1/(\rho n_R))}{\rho n_R}.$$

The derivation in this section has shown that equal power allocation is the most robust transmit strategy if the transmitter has no CSI but the channel is correlated. The max-min average mutual information could be called *compound channel capacity* according to the notion introduced in [Wol78]. The transmitter knows that the channel belongs to the class of Rayleigh fading channels with transmit and receive correlation, but does neither know the concrete instantaneous channel realization nor the correlation matrices a priori.

Optimum transmission strategy with covariance matrix knowledge at the transmitter

In this section, we derive the optimum transmit strategy if the transmitter knows the transmit and receive correlation matrix. At first, we show that the optimum transmit directions, i.e. the optimal eigenvectors of the transmit covariance matrix correspond to the eigenvectors of the transmit correlation matrix. Furthermore, we characterise the optimum power allocation in terms of the necessary and sufficient Karush-Kuhn-Tucker conditions. Finally, we derive an iterative algorithm which solves the power allocation problem.

Optimal eigenvectors of transmit covariance matrix

The following lemma provides the optimal transmit directions. It shows that the

optimal transmit eigenvectors correspond to the eigenvectors of the known transmit correlation matrix.
Lemma 2: The optimal transmit covariance matrix eigenvectors correspond with the eigenvectors of the transmit correlation matrix which are known at the transmitter, i.e.

$$\max_{\substack{\boldsymbol{Q}\succeq 0\\ \mathrm{tr}\,\boldsymbol{Q}\leq P}} \Phi(\rho, \boldsymbol{Q}, \boldsymbol{R}_T, \boldsymbol{R}_R) = \max_{\sum_{k=1}^{n_T} p_k = P} \mathbb{E}\ \mathrm{tr}\ \left(\phi\left(\rho \sum_{i=1}^{n_T} \lambda_i^T p_i \tilde{\boldsymbol{w}}_i \tilde{\boldsymbol{w}}_i^H\right)\right)$$

as in (2.25).

The proof can be found in section 2.4.1. As a result, only the power allocation has to be characterised. The *beamformers* at the transmitter equal the eigenvectors of the transmit correlation matrix.

Characterisation of the optimum power allocation

In order to characterise the optimum power allocation we define the power vector $\boldsymbol{p} = [p_1, ..., p_{n_T}]$ with the sum power constraint $||\boldsymbol{p}|| \leq P$. For fixed transmit and receive correlation matrix eigenvalues λ_k^T and λ_k^R, the average performance metric is a function of the power allocation, which follows from (2.25)

$$\Phi(\boldsymbol{p}, \boldsymbol{\lambda}^R, \boldsymbol{\lambda}^T) = \mathbb{E}\ \mathrm{tr}\ \phi\left(\rho \sum_{k=1}^{n_T} p_k \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H\right). \tag{2.39}$$

The maximum of the average performance metric in (2.39) with respect to the power allocation is given by

$$\hat{\Phi}(\boldsymbol{\lambda}^R, \boldsymbol{\lambda}^T) = \max_{\sum_{k=1}^{n_T} p_k = P} \mathbb{E}\ \mathrm{tr}\ \phi\left(\rho \sum_{k=1}^{n_T} p_k \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H\right). \tag{2.40}$$

Furthermore, we define the following coefficients

$$\alpha_k(\hat{\boldsymbol{p}}) = \rho \lambda_k^T \mathbb{E}\left(\tilde{\boldsymbol{w}}_k^H \phi^{[1]}\left(\rho \sum_{l=1}^{n_T} \hat{p}_l \lambda_l^T \tilde{\boldsymbol{w}}_l \tilde{\boldsymbol{w}}_l^H\right) \tilde{\boldsymbol{w}}_k\right). \tag{2.41}$$

In (2.41), the term $F^{[1]}(\boldsymbol{X})$ is the 'first derivative' of the matrix-valued function $F(\boldsymbol{X})$ as defined in section 2.3.1. Finally, we define the set of indices for which a given power allocation has entries greater than zero

$$\mathcal{I}(\hat{\boldsymbol{p}}) = \{k \in \{1, ..., n_T\} : \hat{p}_k > 0\}. \tag{2.42}$$

The following theorem provides a characterisation of the power allocation $\hat{\boldsymbol{p}}$ which maximises the expression in (2.39).
Theorem 1: A necessary and sufficient condition for the optimality of a power allocation $\hat{\boldsymbol{p}}$ is

$$\{k_1, k_2 \in \mathcal{I}(\hat{\boldsymbol{p}}) \implies \alpha_{k_1} = \alpha_{k_2} \text{ and} \tag{2.43}$$

$$k \notin \mathcal{I}(\hat{\boldsymbol{p}}) \implies \alpha_k \leq \max_{l \in \mathcal{I}(\hat{\boldsymbol{p}})} \alpha_l\}. \tag{2.44}$$

This means that all indices l which obtain some power p_l greater than zero have the same $\alpha_l = \max_{k\in\{1,...,n_T\}} \alpha_k$. Furthermore, all other α_i are less or equal to α_l.

The proof can be found in section 2.4.2.

In case, ϕ is the mutual information the coefficients $\alpha_k(\hat{\boldsymbol{p}})$ are given by

$$\alpha_k^1(\hat{\boldsymbol{p}}) = \rho\lambda_k^T \mathbb{E}\left(\tilde{\boldsymbol{w}}_k^H \left(\boldsymbol{I} + \rho\sum_{l=1}^{n_T}\hat{p}_l\lambda_l^T\tilde{\boldsymbol{w}}_l\tilde{\boldsymbol{w}}_l^H\right)^{-1}\tilde{\boldsymbol{w}}_k\right).$$

In case, ϕ is the modified normalised average MSE the coefficients $\alpha_k(\hat{\boldsymbol{p}})$ are given by

$$\alpha_k^2(\hat{\boldsymbol{p}}) = \rho\lambda_k^T \mathbb{E}\left(\tilde{\boldsymbol{w}}_k^H \left(\boldsymbol{I} + \rho\sum_{l=1}^{n_T}\hat{p}_l\lambda_l^T\tilde{\boldsymbol{w}}_l\tilde{\boldsymbol{w}}_l^H\right)^{-2}\tilde{\boldsymbol{w}}_k\right).$$

Algorithm for optimum power allocation

We use Theorem 1 from the last section to provide the following algorithm which computes the optimum power allocation for the MIMO system with covariance feedback (algorithm 1).

Algorithm 1 Optimum power allocation

Require: given $\boldsymbol{\mu}$ and SNR ρ
1: $\boldsymbol{p}^1 = [1, 0, ..., 0]$
2: **for** $i = 1$ to $n_T - 1$ **do**
3:   **if** $\alpha_i(\boldsymbol{p}^i) \geq \alpha_{i+1}(\boldsymbol{p}^i)$ **then**
4:     optimum solution is given in $\boldsymbol{p}^i$
5:   **else**
6:     find $\boldsymbol{p}^{i+1}$ with $\alpha_1(\boldsymbol{p}^{i+1}) = ... = \alpha_{i+1}(\boldsymbol{p}^{i+1})$
7:   **end if**
8: **end for**
Ensure: $\sum_{k=1}^{n_T} p_k = 1$

We start with the beamforming solution in $\boldsymbol{p}^1 = [1, 0, ..., 0]$ and check whether the condition in (2.44) is fulfilled. If it is not fulfilled we split the transmission power to direction one and two ($\boldsymbol{p}^2 = [p_1, p_2, 0, ..., 0]$) in such a way that $\alpha_1(\boldsymbol{p}^2) = \alpha_2(\boldsymbol{p}^2)$. Next, we recheck the condition in (2.44) for $\boldsymbol{p}^2$ again and so on.

In the following, we derive a simple approach to compute the step in line 6 in algorithm 1. In order to find a $\boldsymbol{p}^{i+1}$ which solves

$$\alpha_1(\boldsymbol{p}^{i+1}) = \alpha_2(\boldsymbol{p}^{i+1}) = ... = \alpha_{i+1}(\boldsymbol{p}^{i+1}) \tag{2.45}$$

we propose the following approach. At first, let us define i functions $f_1(\boldsymbol{p}), ..., f_i(\boldsymbol{p})$ with

$$f_k(\boldsymbol{p}) = (\alpha_1(\boldsymbol{p}) - \alpha_k(\boldsymbol{p}))^2 \quad \forall k = 1...i.$$

Next, we define the objective function $g(\boldsymbol{p})$ with

$$g(\boldsymbol{p}) = \sum_{k=1}^{i} f_k(\boldsymbol{p}). \tag{2.46}$$

Finally, the $\boldsymbol{p}^{i+1}$ which solves (2.45) is the null of $g(\boldsymbol{p})$, i.e. it numerically solves the optimisation problem

$$\boldsymbol{p}^{i+1} = \arg \min_{\sum\limits_{k=1}^{i+1} p_k^{i+1}=1} g(\boldsymbol{p}). \tag{2.47}$$

Observe that there exists only one $\boldsymbol{p}^{i+1}$ which solves (2.45). As a result, algorithm 1 provides the optimum power allocation for given SNR, transmit, and receive correlation matrix. However, the coefficients $\alpha_k(\hat{\boldsymbol{p}})$ are difficult to compute. Therefore, we further characterise the optimum power allocation strategy for low and high SNR values to gain more insights into the analytical structure of the optimal power allocation.

Optimum strategy for low SNR:

The SNR range in which only one direction is supported, i.e. $p_1 = P$ and $p_2 = p_3 = ... = p_{n_T} = 0$ is named the *beamforming range.* The following theorem provides the necessary and sufficient condition for *beamforming* to be optimal with respect to optimisation in (2.25).
Theorem 2: The solution to the optimisation in (2.25) is given by *beamforming* $p_1 = P$, $p_2 = p_3 = ... = p_{n_T} = 0$ if and only if the following condition is fulfilled

$$\rho\lambda_2^T \mathbb{E}\ \mathrm{tr}\ \left[\boldsymbol{R}_R \phi^{[1]}(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H)\right] \leq \rho\lambda_1^T \mathbb{E}\left[\tilde{\boldsymbol{w}}_1^H \phi^{[1]}(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H)\tilde{\boldsymbol{w}}_1\right]. \tag{2.48}$$

The proof can be found in appendix 2.4.4.

Theorem 2 can be specialised for specific choices of ϕ and for uncorrelated receive antennas. We have the following corollaries from Theorem 2:
Corollary 1: The solution to the optimisation in (2.25) is given by *beamforming* $p_1 = P$, $p_2 = p_3 = ... = p_{n_T} = 0$ if and only if the receive antennas are uncorrelated and the following condition is satisfied so.

$$\rho\lambda_2^T \leq \frac{\mathbb{E}\left[\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H \phi^{[1]}(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H)\right]}{\mathbb{E}\left[\phi^{[1]}(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H)\right]}$$

For the special case, in which the mutual information is maximised, we have the following corollary.
Corollary 2: Capacity can be achieved by beamforming for given eigenvalues of transmit and receive correlation matrix and given SNR if and only if the inequality

$$\rho\lambda_2^T \leq \frac{1 - \mathbb{E}\left(\frac{1}{1+\rho\lambda_1^T \sum\limits_{k=1}^{n_R} \lambda_k^R \omega_k}\right)}{\sum\limits_{k=1}^{n_R} \lambda_k^R - \rho\lambda_1^T \sum\limits_{k=1}^{n_R} \lambda_k^R \tau_k} \tag{2.49}$$

with $\tau_k = \mathbb{E}\left(\frac{\lambda_k^R \omega_k}{1+\rho\lambda_1^T \sum_{k=1}^{n_R} \lambda_k^R \omega_k}\right)$ is fulfilled. The ω_k are independent identically distributed following a standard exponential distribution.

This result corresponds to the result in [JB03b].

If we optimise the mutual information, it follows the beamforming range with uncorrelated receive antennas, i.e. $\boldsymbol{R}_R = \boldsymbol{I}$ in the following corollary.

Corollary 3: Capacity can be achieved by beamforming for uncorrelated receive antennas for given eigenvalues of transmit correlation matrix and given SNR if and only if the inequality

$$\rho\lambda_2 \leq \frac{1 - \mathbb{E}\left(\frac{1}{1+\rho\lambda_1||\boldsymbol{w}_1||^2}\right)}{\mathbb{E}\left(\frac{1}{1+\rho\lambda_1||\boldsymbol{w}_1||^2}\right) + n_R - 1} \tag{2.50}$$

is fulfilled.

This corresponds to the result in [JB03b, JG02].

Remark: The inequality in (2.50) can be written with

$$\mathbb{E}\left(\frac{1}{1+\rho\lambda_1||\boldsymbol{w}_1||^2}\right) = \left(\frac{1}{\xi}\right)^{n_R} e^{\frac{1}{\xi}}\Gamma(1-n_R, \frac{1}{\xi}),$$

with $\xi = \rho\mu_1$, the incomplete gamma function defined as $\Gamma(a,z) = \Gamma(a) - \frac{z^a}{a}$ ${}_1F_1(a, 1+a, -z)$, and with the confluent hyper-geometric function ${}_1F_1(e, f, g)$ as

$$\rho\lambda_2 \leq \frac{1 - \left(\frac{1}{\rho\lambda_1}\right)^{n_R} e^{\frac{1}{\rho\lambda_1}}\Gamma(1-n_R, \frac{1}{\rho\lambda_1})}{\left(\frac{1}{\rho\lambda_1}\right)^{n_R} e^{\frac{1}{\rho\lambda_1}}\Gamma(1-n_R, \frac{1}{\rho\lambda_1}) + n_R - 1}. \tag{2.51}$$

For the special case in which the modified average normalised MSE is maximised, we obtain the following necessary and sufficient condition for optimality of beamforming for uncorrelated receive antennas:
Corollary 4: The minimum normalised average MSE can be achieved by beamforming for uncorrelated receive antennas, given eigenvalues of channel correlation matrix, and given SNR if and only if the inequality

$$\rho\lambda_2 \leq \frac{\mathbb{E}\left(\frac{\rho\lambda_1||\boldsymbol{w}_1||^2}{(1+\rho\lambda_1||\boldsymbol{w}_1||^2)^2}\right)}{\mathbb{E}\left(\frac{1}{(1+\rho\lambda_1||\boldsymbol{w}_1||^2)^2}\right) + n_R - 1} \tag{2.52}$$

is fulfilled.

The behaviour at high SNR values cannot be described in closed form. The optimal power allocation strongly depends on the system parameter, i.e. the number of transmit and receive antennas. For MISO systems, the optimal power allocation for SNR approaching infinity is in general not equal power allocation [JB03f].

Optimum transmission strategy with perfect CSI at the transmitter

If the transmitter has perfect CSI, the optimal transmit strategy $\boldsymbol{Q}$ is adapted to every instantaneous channel realization $\boldsymbol{H}$. This results in the following optimisation problem

$$\max_{\substack{\boldsymbol{Q}\succeq 0 \\ \text{tr}\,\boldsymbol{Q}\leq P}} \text{tr}\,\phi\left(\rho\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right). \tag{2.53}$$

The optimum solution to (2.53) is characterised by the following Lemma 3.
Lemma 3: The optimisation problem in (2.53) is solved by transmit strategy $\boldsymbol{Q} = \boldsymbol{U}\text{diag}(p_1, ..., p_{n_T})\boldsymbol{U}^H$ with eigenvectors $\boldsymbol{U}$ that correspond to the eigenvectors of the channel realization $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}_H\boldsymbol{U}^H$. Assume that $\boldsymbol{H}$ has full rank. The optimal

power allocation $p_1^*, ..., p_{n_T}^*$ fulfils the following necessary and sufficient optimality condition for all $1 \leq k \leq n_T$:

$$p_k^* = \left(\frac{1}{\rho \lambda_k^H} \tilde{\phi}^{[1]} \left(\frac{\nu}{\rho \lambda_k^H} \right) \right)^+ \tag{2.54}$$

with $\nu > 0$ such that $\sum_{k=1}^{n_T} p_k^* = P$ and with $a^+ = \max(a, 0)$. The function $\tilde{\phi}^{[1]}$ is the inverse function of the first derivative of the function ϕ.

The Lemma 3 directly follows from the KKT optimality conditions and the proof can be found in appendix 2.4.6.

Remark: The characterisation in (2.54) admits a 'water-filling' power allocation algorithm [PF04]. Note, the function $\phi^{[1]}(x)$ is monotonically decreasing because $\phi(x)$ is (matrix)-concave. As a result, the inverse function $\tilde{\phi}^{[1]}(x)$ is monotonically decreasing with x, too. The algorithm works as follows: Sort the eigenvalues λ_k in increasing order; start with $j = 2$ and choose ν such that $\frac{1}{\lambda_k^H} \tilde{\phi}^{[1]} \left(\frac{\nu}{\lambda_k^H} \right) > 0$. If $\sum_{k=1}^{j} p_k \leq P$ increase j. Otherwise, $p_1, ..., p_{j-1}$ are active and they share the power P such that their KKT conditions are equal.

The optimal power allocation and the Lagrangian multiplier in the optimisation in (2.53) can be rewritten using the alternative representation of the (matrix)-monotone function. The optimal power allocation is given by

$$\begin{aligned} \rho \lambda_k^H \int_0^\infty \frac{s^2}{(s + \rho \lambda_k^H p_k)^2} d\mu(s) &= \nu \qquad \text{if } p_k > 0 \\ \rho \lambda_k^H \int_0^\infty d\mu(s) &\leq \nu \qquad \text{otherwise.} \end{aligned} \tag{2.55}$$

An interesting question concerns the uniqueness of the optimal power allocation. The following Lemma 4 proves that the optimal power allocation in (2.54) is unique.

Lemma 4: The optimal power allocation in (2.54) and (2.55) is unique.

The proof by contradiction can be found in appendix 2.4.7.

For the two special cases, mutual information and MSE optimisation, we obtain the following two corollaries. Define $\xi = \min(n_T, n_R)$.
Corollary 5: The transmit strategy for maximising the mutual information with perfect CSI at the transmitter is spatial 'water-filling' in direction of the eigenvectors of the known channel matrix $\boldsymbol{H}$. The optimum power allocation for each channel realization $\boldsymbol{H}$ with eigenvalues $\lambda_1^H, ..., \lambda_\xi^H$ is characterised for $1 \leq i \leq \xi$ by

$$p_i = \left(\nu - \frac{1}{\rho \lambda_i^H} \right)^+ \tag{2.56}$$

with power constraint

$$\sum_{i=1}^{\xi} p_i = \sum_{i=1}^{\xi} \left(\nu - \frac{1}{\rho \lambda_i^H} \right)^+ = P.$$

Remark: This corollary follows from Lemma 3 by inserting the performance function $\phi_1(x) = log(1 + x)$. The inverse function of the first derivative is then given by

$$\tilde{\phi}_1^{[1]}(y) = \frac{1}{y} - 1.$$

The water-filling power allocation for maximisation of mutual information in parallel Gaussian channels can be found in [CT91], too.
Corollary 6: The transmit strategy for minimising the MSE with perfect CSI at the transmitter is spatial 'water-filling' in direction of the eigenvectors of the known channel matrix $\boldsymbol{H}$. The optimum power allocation for each channel realization $\boldsymbol{H}$ with eigenvalues $\lambda_1^H, ..., \lambda_\xi^H$ is characterised for $1 \leq i \leq \xi$ by

$$p_i = \left(\frac{1}{\sqrt{\nu \rho \lambda_i^H}} - \frac{1}{\rho \lambda_i^H} \right)^+ \tag{2.57}$$

with power constraint $\sum_{i=1}^{\xi} p_i = P$.

Remark: This corollary follows from Lemma 3 by inserting the performance function $\phi_2(x) = \frac{x}{1+x}$. The inverse function of the first derivative is then given by

$$\tilde{\phi}_2^{[1]}(y) = \frac{1}{\sqrt{y}} - 1.$$

The water-filling power allocation for minimisation of the average MSE can be found in [SSB$^+$02, Table 1] and [HB03].

Optimum strategy for low SNR values:

The beamforming range depends on the instantaneous channel realization by its largest and second largest eigenvalue λ_1^H and λ_2^H. From the representation in (2.55) follows the following condition for optimality of beamforming

$$\frac{\lambda_2^H}{\lambda_1^H} \leq \int_0^\infty \frac{s^2}{(s + \rho \lambda_1^H)^2} d\mu(s). \tag{2.58}$$

For mutual information maximisation, beamforming is optimal if and only if

$$\rho \leq \frac{\lambda_1^H - \lambda_2^H}{\lambda_1^H \lambda_2^H}. \tag{2.59}$$

The inequality (2.59) contains the random variables λ_1^H and λ_2^H.

For MSE minimisation, beamforming is optimal if and only if

$$\rho \leq \sqrt{\frac{1}{\lambda_1^{H,2}} + \frac{\lambda_1^H}{\lambda_2^H} - 1} - \frac{1}{\lambda_1^H}. \tag{2.60}$$

The closed-loop MIMO performance function serves as an upper bound for the achievable open-loop and covariance feedback MIMO performance. In the next section, the impact of correlation on the average performance metrics of the MIMO system is analysed.

2.3.3 Impact of correlation on the average performance in MIMO channels

In this section, we study the impact of correlation on the average performance metrics in MIMO system with no CSI, perfect CSI, and covariance matrix knowledge at the transmitter and perfect CSI at the receiver. We will show that the average performance without CSI at the transmitter decreases with increasing transmit correlation, i.e. the average performance in (2.16) as a function of correlation properties is Schur-concave.

For the case in which the transmitter has covariance matrix knowledge or perfect CSI, we will show that in MIMO systems sometimes correlation increases the average performance, and sometimes it decreases it, depending on the SNR range.

The special case in which the receiver's side has only one receive antenna is discussed in section 2.3.4. Much stronger results can be derived for the MISO system. The results are compared in section 2.3.5.

The completely uncorrelated transmit and receive correlation vector is

$$\boldsymbol{\lambda}_{nc}^T = [1, ..., 1] \qquad \boldsymbol{\lambda}_{nc}^R = [1, ..., 1] \tag{2.61}$$

and the completely correlated transmit and receive correlation vector is

$$\boldsymbol{\lambda}_{cc}^T = [n_T, 0, ..., 0] \qquad \boldsymbol{\lambda}_{cc}^R = [n_R, 0, ..., 0]\,. \tag{2.62}$$

Average performance of open-loop MIMO channel

At first, we fix the receive correlation vector $\bar{\boldsymbol{\lambda}}^R$ and prove the following theorem which states that the average performance for the open-loop MIMO channel is a Schur-concave function with respect to the transmit correlation matrix eigenvalues $\boldsymbol{\lambda}^T$. The average performance metric as a function of transmit and receive correlation can be written as

$$\Phi_{MIMO}^{noCSI}(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R) = \mathbb{E}\ \mathrm{tr}\ \phi\left(\rho \sum_{k=1}^{n_T} \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H\right) \tag{2.63}$$

Note that the receive correlation vector is implicitely contained in the random vectors $\tilde{\boldsymbol{w}}_k$ defined in (2.22).

Theorem 3: For fixed receive correlation vector $\bar{\boldsymbol{\lambda}}^R$ and vector $\boldsymbol{\lambda}_0^T$ and for an arbitrary vector $\boldsymbol{\lambda}_1^T$ which majorizes vector $\boldsymbol{\lambda}_0^T$, i.e.

$$\boldsymbol{\lambda}_1^T \succeq \boldsymbol{\lambda}_0^T,$$

it follows that

$$\Phi_{MIMO}^{noCSI}(\boldsymbol{\lambda}_0^T, \bar{\boldsymbol{\lambda}}^R) \geq \Phi_{MIMO}^{noCSI}(\boldsymbol{\lambda}_1^T, \bar{\boldsymbol{\lambda}}^R).$$

The proof can be found in appendix 2.4.8.

Remark: Theorem 3 states that transmit correlation decreases the average performance of a MIMO system with perfect CSI at the receiver and no CSI at the transmitter. In terms of average performance, the completely correlated case performs the worst, and the completely uncorrelated performs the best. Note that the significance of Theorem 3 is much greater. On the one hand, the optimal transmit strategy without CSI at the transmitter with respect to the worst case correlation was derived using this result. On the other hand, the result in Theorem 3 can be further applied to receive correlation as well.

Next, we fix the transmit correlation vector $\bar{\boldsymbol{\lambda}}^T$ and give the following corollary which states that the average performance is a Schur-concave function with respect to the receive correlation matrix eigenvalues $\boldsymbol{\lambda}^R$.
Corollary 7: For fixed transmit correlation vector $\bar{\boldsymbol{\lambda}}^T$, fixed vector $\boldsymbol{\lambda}_0^R$, and for arbitrary vector $\boldsymbol{\lambda}_1^R$ which majorizes vector $\boldsymbol{\lambda}_0^R$, i.e.

$$\boldsymbol{\lambda}_1^R \succeq \boldsymbol{\lambda}_0^R,$$

it follows that

$$\Phi_{MIMO}^{noCSI}(\bar{\boldsymbol{\lambda}}^T, \boldsymbol{\lambda}_0^R) \geq \Phi_{MIMO}^{noCSI}(\bar{\boldsymbol{\lambda}}^T, \boldsymbol{\lambda}_1^R).$$

The proof of Theorem 3 can be used together with the application of the alternative representation in (2.23) to prove Corollary 7.

The next corollary emphasizes the behaviour at the both extreme correlation scenarios, complete correlation as well as no correlation at all.
Corollary 8: The average performance of the single-user MIMO system with uninformed transmitter and perfect CSI at the receiver is highest with uncorrelated transmit antennas:

$$\Phi_{MIMO}^{noCSI}(\boldsymbol{\lambda}_{nc}^T, \bar{\boldsymbol{\lambda}}_R) \geq \Phi_{MIMO}^{noCSI}(\boldsymbol{\lambda}^T, \bar{\boldsymbol{\lambda}}_R)$$

for all $\boldsymbol{\lambda}^T$ which fulfill $\operatorname{tr}(\boldsymbol{\lambda}^T) = n_T$. The average performance of the single-user MIMO system with uninformed transmitter and perfect CSI at the receiver is lowest for correlated transmit antennas:

$$\Phi_{MIMO}^{noCSI}(\boldsymbol{\lambda}_{cc}^T, \bar{\boldsymbol{\lambda}}_R) \leq \Phi_{MIMO}^{noCSI}(\boldsymbol{\lambda}^T, \bar{\boldsymbol{\lambda}}_R)$$

for all $\boldsymbol{\lambda}^T$ which fulfill $\operatorname{tr}(\boldsymbol{\lambda}^T) = n_R$.

The corollary follows direclty from Theorem 1 because the vector of eigenvalues of the correlation matrix for the completely uncorrelated case $\boldsymbol{\lambda}_{nc}$ is majorized by all other eigenvalue vectors. The other way around, the completely correlated vector $\boldsymbol{\lambda}_{cc}$ majorizes all other vectors.

From Theorem 3 the next two corollaries follows for the average mutual information and the average MSE. The average mutual information as a function of transmit and receive correlation eigenvalues is described by

$$C(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R) = \mathbb{E}\operatorname{tr}\log\left(\boldsymbol{I} + \rho\sum_{k=1}^{n_T}\lambda_k^T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right) = \mathbb{E}\operatorname{tr}\log\left(\boldsymbol{I} + \rho\sum_{l=1}^{n_R}\lambda_l^R\hat{\boldsymbol{w}}_l\hat{\boldsymbol{w}}_l^H\right)$$

The average sum MSE as a function of transmit and receive correlation is given by

$$\begin{aligned} MSE(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R) &= \mathbb{E}\operatorname{tr}\left(\rho\sum_{k=1}^{n_T}\lambda_k^T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right)\left[\boldsymbol{I} + \rho\sum_{k=1}^{n_T}\lambda_k^T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right]^{-1} \\ &= \mathbb{E}\operatorname{tr}\left(\rho\sum_{l=1}^{n_R}\lambda_k^R\hat{\boldsymbol{w}}_k\hat{\boldsymbol{w}}_k^H\right)\left[\boldsymbol{I} + \rho\sum_{l=1}^{n_R}\lambda_k^R\hat{\boldsymbol{w}}_k\hat{\boldsymbol{w}}_k^H\right]^{-1}. \end{aligned}$$

Corollary 9: The average mutual information in single-user flat-fading MIMO channels is a Schur-concave function with respect to the transmit correlation eigenvalues for fixed receive correlation, i.e.

$$\boldsymbol{\lambda}_1^T \succeq \boldsymbol{\lambda}_2^T \qquad \Longrightarrow \qquad C(\boldsymbol{\lambda}_1^T, \bar{\boldsymbol{\lambda}}^R) \le C(\boldsymbol{\lambda}_2^T, \bar{\boldsymbol{\lambda}}^R).$$

Furthermore, the average mutual information is Schur-concave with respect to the receive correlation eigenvalues for fixed transmit correlation eigenvalues, too.

$$\boldsymbol{\lambda}_1^R \succeq \boldsymbol{\lambda}_2^R \qquad \Longrightarrow \qquad C(\bar{\boldsymbol{\lambda}}_1^T, \boldsymbol{\lambda}_1^R) \le C(\bar{\boldsymbol{\lambda}}_2^T, \boldsymbol{\lambda}_2^R)$$

Remark: Note that the Schur-concavity of the average mutual information holds for all numbers of transmit and receive antennas and for all SNR values. As a result, the impact of correlation on the average mutual information is described.
Corollary 10: The average MSE in single-user flat-fading MIMO channels in which the receiver performs MMSE detection is a Schur-convex function with respect to the transmit correlation eigenvalues for fixed receive correlation, i.e.

$$\boldsymbol{\lambda}_1^T \succeq \boldsymbol{\lambda}_2^T \qquad \Longrightarrow \qquad MSE(\boldsymbol{\lambda}_1^T, \bar{\boldsymbol{\lambda}}^R) \ge MSE(\boldsymbol{\lambda}_2^T, \bar{\boldsymbol{\lambda}}^R).$$

Furthermore, the average MSE is Schur-convex with respect to the receive correlation eigenvalues for fixed transmit correlation eigenvalues, too.

$$\boldsymbol{\lambda}_1^R \succeq \boldsymbol{\lambda}_2^R \qquad \Longrightarrow \qquad MSE(\bar{\boldsymbol{\lambda}}_1^T, \boldsymbol{\lambda}_1^R) \ge MSE(\bar{\boldsymbol{\lambda}}_2^T, \boldsymbol{\lambda}_2^R)$$

Performance degradation due to correlation

Next, we analyse the capacity loss due to transmit and receive correlation. The difference between the completely uncorrelated and completely correlated scenario is a function of the SNR and the number of transmit and receive antennas. In general, it is given by

$$\Delta(\rho, n_T, n_R) = \mathbb{E}\ \mathrm{tr}\ \left[\phi(\rho \boldsymbol{W}\boldsymbol{W}^H) - \phi(\rho n_T n_R w_1)\right] \tag{2.64}$$

with w_1 standard exponentially distributed, i.e. pdf $p(w)_{w_1} = \exp(-w)$ and $\boldsymbol{W}$ has zero-mean iid complex Gaussian entries. We observe that in the first term of the RHS in (2.64) $\min(n_T, n_R)$ degrees of freedom are available, whereas in the second term only one degree of freedom is available. The following Lemma shows that the difference in (2.64) is monotonically increasing with the SNR.
Lemma 5: The performance loss due to correlation $\Delta(\rho, n_T, n_R)$ in (2.64) increases with the SNR ρ.

The proof can be found in appendix 2.4.9.

The exact analysis is given in the next subsection for the average mutual information.

Average mutual information degradation due to correlation

The average mutual information in the case of uncorrelated transmit and receive antennas with uninformed transmitter and perfect CSI at the receiver is given by

$$C(\boldsymbol{\lambda}_{nc}^T, \boldsymbol{\lambda}_{nc}^R) \quad = \quad \mathbb{E} \log\det\left(\frac{1}{\rho}\boldsymbol{I} + \sum_{k=1}^{n_T} \boldsymbol{w}_k \boldsymbol{w}_k^H\right) - \log\det\left(\frac{1}{\rho}\boldsymbol{I}\right). \tag{2.65}$$

The average mutual information in the case of fully correlated receive and transmit antennas in given by

$$C(\boldsymbol{\lambda}_{cc}^T, \boldsymbol{\lambda}_{cc}^R) = \mathbb{E} \log \left(\frac{1}{\rho} + n_T n_R w_1 \right) - \log \left(\frac{1}{\rho} \right). \tag{2.66}$$

The difference between the uncorrelated $C(\boldsymbol{\lambda}_{nc}^T, \boldsymbol{\lambda}_{nc}^R)$ from (2.65) and correlated $C(\boldsymbol{\lambda}_{cc}^T, \boldsymbol{\lambda}_{cc}^R)$ from (2.66) case is a function of the SNR, the number of transmit, and the number of receive antennas:

$$\Delta_{nc}^{cc}(\rho, n_T, n_R) \quad = \quad C(\boldsymbol{\lambda}_{nc}^T, \boldsymbol{\lambda}_{nc}^R) - C(\boldsymbol{\lambda}_{cc}^T, \boldsymbol{\lambda}_{cc}^R) \tag{2.67}$$

Theorem 4: The capacity loss due to correlation of the single user MIMO system with an uninformed transmitter and perfect CSI with n_T transmit antennas, n_R receive antennas, and at a SNR of ρ is lower bounded by

$$\Delta_{nc}^{cc}(\rho, n_T, n_R) \geq \sum_{k=1}^{n_T} \Psi(n_R - k + 1) + \gamma - \log(n_R n_T) + \log \left((P\rho)^{n_T - 1} \right) \tag{2.68}$$

with the Psi-function. The bound in (2.68) becomes tight for SNR approaching infinity.

The proof can be found in 2.4.10.

Remark I: The result in (2.68) can be alternatively approximated for SNR approaching infinity. In [SM00, section 2.1, (11)], for SNR approaching infinity, the ergodic capacity of a flat-fading completely uncorrelated MIMO system is approximated as

$$C_{\rho \to \infty}^{uc} \approx \log\left((\rho)^{n_T}\right) + \sum_{j=1}^{n_T} \Psi(n_R - n_T + j) - n_T \log(n_T). \tag{2.69}$$

Remark II: If the transmitter has more than one antenna, the difference in (2.68) grows without bound with SNR and with the number of transmit antennas. This means that the information loss due to correlation can be serious high. In the case in which we have one receive antenna, the loss in average mutual information due to correlation is bounded by some small constant.

Impact of correlation with covariance feedback

For the case in which the transmitter knows the transmit correlation matrix, the average performance is neither Schur-convex nor Schur-concave. First we present necessary and sufficient condition which has to be fulfilled so that the average performance is Schur-convex. It is immediately observed that this condition is not fulfilled for high SNR values but for low SNR values. In addition to this, the following derivation provides a characterisation of the optimal power allocation scheme related to transmit correlation. This characterisation reduces the search space for power allocation vectors and decreases the computational complexity of the iterative algorithm in section 2.3.2.

The average performance with covariance knowledge at the transmitter as a function of transmit and receive correlation is given as

$$\Phi_{MIMO}^{cov}(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R) = \max_{\sum_{k=1}^{n_T} p_k = P} \mathbb{E}\, \mathrm{tr} \left(\phi \left(\rho \sum_{k=1}^{n_T} p_k \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H \right) \right). \tag{2.70}$$

In the following, we assume that the transmit correlation matrix has full rank and keep the receive correlation matrix $\boldsymbol{R}_R$ fixed.
Lemma 6: The average performance of the MIMO system with correlation knowledge at the transmitter is a Schur-convex function with respect to the transmit correlation matrix eigenvalues $\boldsymbol{\lambda}^T$ if and only if the following inequality is fulfilled by the optimum power allocation $\boldsymbol{p} = [p_1, ..., p_{n_T}]$ and the transmit correlation matrix eigenvalues $\boldsymbol{\lambda}^T = [\lambda_1^T, ..., \lambda_{n_T}^T]$ for all $1 \leq l \leq n_T - 1$

$$\frac{p_l}{\lambda_l^T} \geq \frac{p_{l+1}}{\lambda_{l+1}^T} \tag{2.71}$$

The proof can be found in the section 2.4.11.

Remark: Using the condition in (2.71) it is simple to determine whether transmit correlation increases or decreases the average performance. Note that the transmit correlation as well as the SNR and the optimal power allocation are known at the transmitter.

In order to apply the result to the theory of majorization which we use to compare different correlation scenarios, we need a connection between the condition in Lemma 6 and majorization. With this connection, we have statements about Schur-convexity or Schur-concavity of the ergodic channel capacity in the covariance feedback scenario.

The condition (2.71) in Lemma 6 is stronger than the majorization between the power allocation vector $\boldsymbol{p}$ and the transmit correlation vector $\boldsymbol{\lambda}^T$ [MO79, Proposition 5.B.1]. From (2.71) follows (with normalised transmit power and transmit correlation $\sum p_k = \sum \lambda_k^T$)

$$\boldsymbol{p} \succeq \boldsymbol{\lambda}^T \tag{2.72}$$

From (2.72) does not follow the condition (2.71) in Lemma 6. However, we can use (2.72) to check whether (2.71) can be fulfilled and whether the ergodic capacity is Schur-convex. This approach provides the following results:

- *For low SNR in the beamforming range:* If only one direction is supported, i.e. $p_1 = P$ and $p_2 = p_3 = ... = p_{n_T} = 0$ the condition from Lemma (6) is fulfilled, because the RHS of (2.71) are always zero.

- *For high SNR:* In general, more than one data stream is multiplexed. If $n_R \geq n_T$, equal power allocation turns out to be optimal. In those cases, the average performance is Schur-concave (Theorem 3) and the performance loss due to correlation is given by Theorem 4.

The ergodic channel capacity for covariance knowledge at the transmitter as a function of transmit and receive correlation is given by

$$C_{MIMO}^{cov}(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R) = \max_{\sum_{k=1}^{n_T} p_k = P} \mathbb{E} \log\det \left(\boldsymbol{I} + \rho \sum_{k=1}^{n_T} p_k \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H \right). \tag{2.73}$$

Hence, the ergodic capacity in MIMO systems with covariance feedback is Schur-convex at low SNR values and Schur-concave in the high SNR range (if equal power allocation is optimal, i.e. $n_R \geq n_T$).

In order to characterise the impact of receiver correlation on the average performance of the covariance feedback MIMO system, we prove the following theorem.

Theorem 5: The ergodic capacity of the MIMO system with covariance knowledge at the transmit antenna array and with fixed transmit correlation $\bar{\boldsymbol{\lambda}}^T$ is Schur-concave with respect to the receive correlation $\boldsymbol{\lambda}^R$, i.e. from $\boldsymbol{\lambda}_0^R \succeq \boldsymbol{\lambda}_1^R$ follows for the ergodic capacity in (2.73) that $C(\bar{\boldsymbol{\lambda}}^T, \boldsymbol{\lambda}_0^R) \leq C(\bar{\boldsymbol{\lambda}}^T, \boldsymbol{\lambda}_1^R)$.

The sketch of the proof of Theorem 5 can be found in section 2.4.12.

We state the following theorem that analyses the impact of receive correlation on the achievable capacity. It is a corollary of the Theorem 5. However, we present it together with its proof because the proof itself is interesting.
Lemma 7: The achievable capacity of a MIMO system with covariance feedback is biggest in case of uncorrelated receive antennas.

The proof can be found in section 2.4.13.

The specialisation of the results from this section for the mutual information and average MSE, result in the following corollaries. The average mutual information corresponds with the ergodic channel capacity, because a coding theorem and its converse can be proved [GV97] for covariance feedback.
Corollary 11: The ergodic capacity as well as the average MSE for a MIMO system with perfect CSI at the receiver and covariance knowledge at the transmitter is Schur-convex with respect to transmit correlation $\boldsymbol{\lambda}^T$ for fixed receive correlation $\boldsymbol{\lambda}^R$ if and only if the following condition with the optimal power allocation $\boldsymbol{p}^*$ is fulfilled for all $1 \leq l \leq n_T - 1$

$$\frac{p_l}{\lambda_l^T} \geq \frac{p_{l+1}}{\lambda_{l+1}^T}.$$

The discussion from above holds for the ergodic capacity and the average MSE, too. For small SNR values, correlation is helpful, since it increases the ergodic capacity and decreases the average MSE.

Impact of correlation with perfect CSI

The average performance with perfect CSI at the transmitter as a function of transmit and receive correlation is given by

$$\Phi_{MIMO}^{pCSI}(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R) = \mathbb{E} \max_{\sum_{k=1}^{n_T} p_k = P} \operatorname{tr} \phi \left(\rho \sum_{k=1}^{n_T} p_k \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H \right). \tag{2.74}$$

In general, it is complicated to analyse the average performance of a system in which the transmission strategy depends on each random channel realization and therefore is a random variable, too. Hence, we discuss the impact of correlation for low and high SNR values, only. For low SNR in average only one direction is supported by the waterfilling power allocation, i.e. $\mathbb{E}p_1 = P$ and $\mathbb{E}p_2 = \mathbb{E}p_3 = ... = \mathbb{E}p_{n_T} = 0$. Therefore, the average performance is for small SNR values Schur-convex with respect to transmit correlation. Analogue, for high SNR, we know that equal power allocation is optimal if the number of receive antennas is larger than or equal to the number of transmit antennas. Therefore, the average performance is for high SNR values Schur-concave with respect to transmit correlation. The impact of transmit correlation depends on the distribution of the eigenvalues of the channel realizations. The probability density function of the eigenvalues of the correlated Wishart matrix is in general not known in closed form. However, for asymptotic many transmit and receive antennas there exist formulas [SB95] for the empirical distribution function and an analysis in [CTK02, MO02] which indicates that the ergodic capacity is Schur-concave with respect to receive correlation for asymptotic high SNR.

We conjecture that the closed-loop MIMO ergodic channel capacity is Schur-convex for small SNR values and Schur-concave for high SNR values with respect to transmit *and* receive correlation, because in the low SNR regime, only the largest instantaneous channel eigenvalue is supported and the more correlated the transmit or the receive antennas are, the larger is the largest eigenvalue in average.

The impact of correlation for all three types of CSI at the transmitter is illustrated in section 2.3.6. The more CSI is available, the higher is the average performance. In addition to this, the three types of CSI are related by the following identities:
Lemma 8: The average performance of the MIMO system with the three different types of CSI, namely perfect, covariance, and no CSI are related by their extreme cases, i.e. for completely correlated transmit antennas $\boldsymbol{\lambda}^T_{cc}$ and completely uncorrelated transmit antennas $\boldsymbol{\lambda}^T_{nc}$ and fixed receive correlation $\boldsymbol{\lambda}^R$. It holds

$$\Phi^{noCSI}_{MIMO}(\boldsymbol{\lambda}^T_{nc}, \boldsymbol{\lambda}^R) = \Phi^{cov}_{MIMO}(\boldsymbol{\lambda}^T_{nc}, \boldsymbol{\lambda}^R) \text{ and } \Phi^{cov}_{MIMO}(\boldsymbol{\lambda}^T_{cc}, \boldsymbol{\lambda}^R) = \Phi^{pCSI}_{MIMO}(\boldsymbol{\lambda}^T_{cc}, \boldsymbol{\lambda}^R).$$

This lemma follows from the fact that the transmission strategies for completely uncorrelated transmitters for no CSI and covariance knowledge and for completely correlated transmitters for covariance knowledge and perfect CSI are equal. In the MISO case, this lemma can be specialised to provide a full inequality chain between all types of CSI and all comparable transmit correlation vectors.

These reflections complete the analysis of the impact of correlation on the average performance of MIMO systems under different types of CSI at the transmitter and with perfect CSI at the receiver. Logically, all results stay valid if we consider MISO systems in which the receiver has only one receive antenna. However, much stronger results can be derived for the MISO case as in the MIMO case. Especially, for covariance feedback and perfect CSI at the transmitter. Therefore, the impact of correlation on the average performance in MISO systems is analysed in the next section.

2.3.4 Impact of correlation on the average performance in MISO systems

In this section, results for the average performance in MISO channels are given which cannot be derived from the general MIMO case. For the open-loop MISO performance follows from Theorem 3 that the average performance is Schur-concave with respect to the transmit correlation eigenvalues. The average performance is given with $\boldsymbol{Q} = \frac{1}{n_T}\boldsymbol{I}$ by

$$\Phi^{noCSI}_{MISO}(\rho, \boldsymbol{\lambda}^T) = \mathbb{E} \operatorname{tr} \phi \left(1 + \frac{\rho}{n_T} \sum_{k=1}^{n_T} \lambda^T_k w_k \right) \tag{2.75}$$

with iid standard exponentially distributed w_k.

The optimal transmit strategy with perfect CSI at the transmitter and one receive antenna is beamforming in direction of the largest channel matrix eigenvalue, because $\lambda_2(\boldsymbol{H}) = \lambda_3(\boldsymbol{H}) = ... = \lambda_\xi(\boldsymbol{H}) = 0$. Therefore, the average performance for perfect CSI is given by

$$\Phi^{pCSI}_{MISO}(\rho, \boldsymbol{\lambda}^T) = \mathbb{E} \operatorname{tr} \phi \left(\rho \sum_{k=1}^{n_T} \lambda_k^T w_k \right). \tag{2.76}$$

Remark: With the substitution $\rho = \tilde{\rho}/n_T$ we obtain for the performance with perfect CSI (2.76) the same expression as for the performance without CSI (2.75). Therefore, the average performance with perfect CSI is Schur-concave with respect to transmit correlation. This is in contrast to the general MIMO case in which the performance can be Schur-concave if and only if the condition from Lemma 6 is fulfilled.

Finally, the average performance of the MISO system with covariance feedback is given by

$$\Phi^{cov}_{MISO}(\rho, \boldsymbol{\lambda}^T) = \max_{\sum p_k = P} \mathbb{E} \operatorname{tr} \phi \left(\rho \sum_{k=1}^{n_T} p_k \lambda_k^T w_k \right). \tag{2.77}$$

The next Theorem is proven for the ergodic channel capacity only. We consider the ergodic channel capacity of the MISO system for the covariance feedback case which is given by

$$C^{cov}_{MISO}(\rho, \boldsymbol{\lambda}^T) = \max_{\sum p_k = P} \mathbb{E} \log \det \left(1 + \rho \sum_{k=1}^{n_T} p_k \lambda_k^T w_k \right). \tag{2.78}$$

The next theorem states that the ergodic capacity in (2.78) is Schur-convex with respect to the correlation vector $\boldsymbol{\lambda}^T$.
Theorem 6: For arbitrary channel correlation vectors $\boldsymbol{\lambda}_1^T$ and $\boldsymbol{\lambda}_2^T$ we have the following implication

$$\boldsymbol{\lambda}_1^T \succeq \boldsymbol{\lambda}_2^T \Longrightarrow C^{cov}_{MISO}(\rho, \boldsymbol{\lambda}_1^T) \geq C^{cov}_{MISO}(\rho, \boldsymbol{\lambda}_2^T). \tag{2.79}$$

i.e. the capacity of the single user MISO system with covariance feedback is Schur-convex.

The proof can be found in appendix 2.4.14.

Average mutual information in MISO systems

In this section, the average mutual information in MISO systems for different types of CSI under transmit correlation is derived from the general results in the last sections.

The inequality chain in the next corollary shows the relation between the different CSI schemes and different levels of transmit correlation. We omit the super-index ^{T} for convenience, because in MISO systems there is no receive correlation. Assume that the correlation vector $\boldsymbol{\lambda}^2$ majorizes $\boldsymbol{\lambda}^1$, i.e. $\boldsymbol{\lambda}^1 \prec \boldsymbol{\lambda}^2$. We define the fully correlated vector $\boldsymbol{\lambda}_{cc} = [n_T, 0, ..., 0]^T$ and the completely uncorrelated vector as $\boldsymbol{\lambda}_{nc} = [1, 1, ..., 1]^T$. Note, that the vector $\boldsymbol{\lambda}_{cc}$ majorizes all other vectors and that the vector $\boldsymbol{\lambda}_{nc}$ is majorizes by all other vectors.

Corollary 12: For the average mutual information and the ergodic capacities in MISO systems with different levels of correlation and different CSI at the transmitter, we have the following inequalities:

$$\begin{aligned} C^{noCSI}_{MISO}(\boldsymbol{\lambda}_{cc}) \leq C^{noCSI}_{MISO}(\boldsymbol{\lambda}^2) \leq C^{noCSI}_{MISO}(\boldsymbol{\lambda}^1) \leq C^{noCSI}_{MISO}(\boldsymbol{\lambda}_{nc}) = \\ C^{cfCSI}_{MISO}(\boldsymbol{\lambda}_{nc}) \leq C^{cfCSI}_{MISO}(\boldsymbol{\lambda}^1) \leq C^{cfCSI}_{MISO}(\boldsymbol{\lambda}^2) \leq C^{cfCSI}_{MISO}(\boldsymbol{\lambda}_{cc}) = \\ C^{pCSI}_{MISO}(\boldsymbol{\lambda}_{cc}) \leq C^{pCSI}_{MISO}(\boldsymbol{\lambda}^2) \leq C^{pCSI}_{MISO}(\boldsymbol{\lambda}^1) \leq C^{pCSI}_{MISO}(\boldsymbol{\lambda}_{nc}). \end{aligned} \tag{2.80}$$

This corollary follows from the Schur-concavity of the average mutual information for no CSI and perfect CSI, as well as from the Schur-convexity of the ergodic channel capacity for covariance knowledge.

Loss and gain due to correlation

The worst case scenario is the uninformed transmitter with fully correlated channels $C^{noCSI}_{MISO}(\boldsymbol{\lambda}_{cc})$. The best case scenario is the perfectly informed transmitter with completely uncorrelated channels $C^{pCSI}_{MISO}(\boldsymbol{\lambda}_{nc})$. In the following, we characterise the quantitative capacity gain. We explicitly list the capacities from (2.80):

$$\begin{aligned} C^{noCSI}_{MISO}(\boldsymbol{\lambda}_{cc}) &= \mathbb{E}\log(1+\rho w_1) \\ C^{noCSI}_{MISO}(\boldsymbol{\lambda}_{nc}) &= \mathbb{E}\log\left(1+\frac{\rho}{n_T}\sum_{k=1}^{n_T} w_k\right) = C^{cfCSI}_{MISO}(\boldsymbol{\lambda}_{nc}) \\ C^{cfCSI}_{MISO}(\boldsymbol{\lambda}_{cc}) &= \mathbb{E}\log(1+n_T\rho w_1) = C^{pCSI}_{MISO}(\boldsymbol{\lambda}_{cc}) \\ C^{pCSI}_{MISO}(\boldsymbol{\lambda}_{nc}) &= \mathbb{E}\log\left(1+\rho\sum_{k=1}^{n_T} w_k\right) \end{aligned}$$

We define the difference between the best and the worst case scenario for no CSI as

$$\Delta^{cr}_{nc}(\rho, n_T) = C^{opt}_{nc}(\rho, n_T) - C^{opt}_{cr}(\rho) \tag{2.81}$$

For no CSI at the transmitter, the capacity difference in the single-user MISO system for uncorrelated transmit antenna versus fully correlated transmit antennas for SNR approaching infinity is given by

$$\Delta^{cr}_{nc}(\infty, n_T) = \frac{1}{\log_e(2)}\left(\gamma - \log(n_T) + \Psi(n_T)\right). \tag{2.82}$$

This follows from Theorem 4. As a result, we obtain for the limit as $n_T \to \infty$

$$\lim_{n_T\to\infty} \Delta^{cr}_{nc,noCSI}(n_T) = \frac{\gamma}{\log_e(2)}. \tag{2.83}$$

For the case in which the transmitter knows the covariance matrix, the capacity gain due to correlation increases with the number of transmit antennas n_T and with the SNR ρ.

The capacity is bounded for increasing SNR, but grows unbounded with the number of transmit antennas. Analogue to (2.82), we obtain the difference

$$\Delta^{cr}_{nc,cov}(n_T) = \frac{1}{\log(2)}(\gamma - 2\cdot\log(n_T) + \Psi(n_T)) \tag{2.84}$$

For n_T approaching infinity, the expression in (2.84) behaves like $-\frac{\log(n_T)}{\log(2)}$. In the covariance feedback scenario, the maximum capacity gain due to correlation grows unbounded with the number of transmit antennas.

If the transmitter has perfect CSI the capacity loss due to correlation increases with the SNR ρ and the number of transmit antennas n_T. It is bounded with increasing SNR. The capacity loss is given by

$$\begin{aligned} \Delta^{cr}_{nc,pCSI}(\rho, n_T) &= C^{pCSI}_{MISO}(\boldsymbol{\lambda}_{nc}) - C^{pCSI}_{MISO}(\boldsymbol{\lambda}_{cc}) \\ &= \mathbb{E}\log(1+\rho n_T w_1) - \mathbb{E}\log(1+\rho\omega) \end{aligned} \tag{2.85}$$

with χ^2 distributed ω with n_T degrees of freedom. Notice that the difference between the scenario without CSI and (2.85) consists only in the denominator n_T. As a result, we obtain the same capacity difference as in the case without CSI

$$\Delta^{cr}_{nc,pCSI}(n_T) = \frac{1}{\log(2)}(\gamma - \log(n_T) + \Psi(n_T))$$

and the same upper bound

$$\Delta^{cr}_{nc,pCSI}(n_T) \leq \frac{\gamma}{\log_e(2)}.$$

Illustration of relationship between different CSI schemes

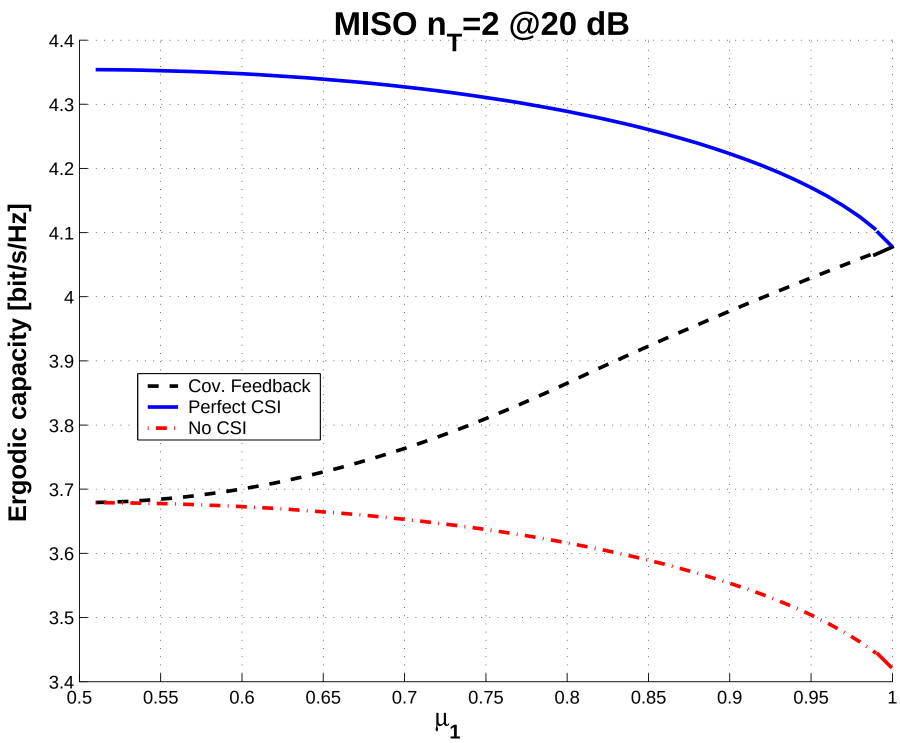

Figure 2.3: Capacity as a function of correlation for MISO 2×1 system with different levels of CSI

Figure (2.3) illustrates the results from the last sections. For a MISO system with two transmit antennas, we show the average mutual information over the largest eigenvalue of the channel covariance matrix μ_1. The second channel covariance matrix eigenvalues is given by $1 - \mu_1$. The left side $\mu_1 = 0.5$ corresponds to the uncorrelated scenario while the right side $\mu_1 = 1$ corresponds to the completely correlated scenario. In figure (2.3), we consider three types of CSI at the transmitter, no CSI, covariance feedback, and perfect CSI and show the open-loop average mutual information, the closed-loop, and covariance feedback ergodic channel capacity.

At first, we observe in figure (2.3) that the ergodic capacity is lowest for no CSI and highest for perfect CSI regardless of the correlation μ_1. In addition to this, the ergodic capacity for no CSI and covariance feedback for completely uncorrelated transmit antennas is equal because the optimal power allocation with uncorrelated transmit antennas and covariance feedback is equal power allocation. In the case in which the transmit antennas are completely correlated, the ergodic capacity for covariance feedback and perfect CSI is equal because in the completely correlated case only one signal dimension is available for power allocation with perfect CSI or with covariance knowledge.

Finally, we illustrate the inequality chain from (2.80). We observe that the capacity for this scenario can be increased by more CSI and by more or less correlation. Starting with no CSI from 3,44 bit/s/Hz (100 %) the capacity increases with less correlation to 3.68 bit/s/Hz (107 %). With covariance feedback we start uncorrelated at 3.68 bit/s/Hz and increase the capacity up to 4.08 bit/s/Hz (119 %) with more correlation. In the scenario with perfect CSI, we start at this value (4.08 bit/s/Hz) completely correlated and gain up to 4.30 bit/s/Hz (125 %) with less correlation. For SNR approaching infinity, the upper bounds for the capacity difference between completely correlated and completely uncorrelated, read $\Delta_{noCSI} = \Delta_{pCSI} \leq 0.4427$ bit/s/Hz and $\Delta_{cfCSI} \leq 0.5573$ bit/s/Hz.

2.3.5 Comparison of average performance results between MIMO and MISO systems

Comparison of optimum transmit strategies between MIMO and MISO

In this section, we conclude the results for optimal transmission from the last sections and compare them to the optimal strategies in MISO systems. In table 2.1, these results are summarised. In curly brackets is the reference for the general result, the result with respect to average mutual information, and the result with respect to average sum MSE. A cross stands for no reference.

In the case with uninformed transmitters, the optimum transmission strategies for MIMO and MISO systems are equal. Intuitively, one would say that if we have no idea where our conversational partner is, we call out in any direction. The authors in [PCL03b] handle the case in which the channel is unknown to the transmitter but belongs to a class of channels which is known at the transmitter. It is shown that the instantaneous mutual information is maximised by uniform power allocation with respect to the worst case channel from the class. In the literature, this kind of channel model is characterised by the compound channel capacity [Wol78]. The capacity in the case where the family of channels consists of memoryless channels is studied in [BBT59] and where the family consists of finite-state channel in [LT98].

In the case with covariance knowledge at the transmitter, the transmission strategies of MIMO and MISO systems are quite similar but differ in their power allocation programming problem.

In general we conclude: The more antennas are available and the more uncorrelated

CSI	MISO
no CSI	equal power allocation in all directions { Theorem 3, [BJ04a], x }
covariance	direction of eigenvectors of covariance matrix { Lemma 2, [JVG01, JB02b] , x }, power allocation { Algorithm 1, [JB04c], x }
perfect CSI	beamforming in direction of instantaneous channel realization { Lemma 3, [JB04c], x }
CSI	**MIMO**
no CSI	equal power allocation { Theorem 3, [JB03e], [PCL03a] }
covariance	direction of eigenvectors of covariance matrix { Lemma 2, [JB03b], [PCL03a] }, power allocation { algorithm 1, [JVG01] and [JB03b], x }
perfect CSI	waterfilling in direction of eigenvectors of instantaneous channel realization { Lemma 2, [CT91], [SSB$^+$02] }

Table 2.1: Optimal transmission schemes in MIMO and MISO systems with respect to the average performance with different types of CSI at the transmitter and perfect CSI at the receiver.

the antennas on the receiver side are, the higher the number of parallel transmit data streams. This conclusion is valid for partial as well as for perfect CSI at the transmitter. This conclusion has an important impact: Let us imagine the case in which the transmit antennas are not allowed to cooperate, i.e. we have a multiuser SIMO system. One important question is, how many users are supported at the same time, i.e. how many data streams are transmitted in parallel in order to maximise the sum capacity of this corresponding multiuser system. The conclusions from the single-user MIMO system indicate, that the number of users which are allowed to simultaneously transmit grows with the number of (uncorrelated) receive antennas. Or we can ask the other way round, when can the sum capacity in multiuser SIMO systems be achieved by TDMA schemes [3]. Answers to this and related questions can be found in chapter 3 of this thesis.

Comparison of impact on correlation in MISO and MIMO systems

In this section, we conclude the results on the impact of correlation on the average MISO and MIMO channel performance for different types an CSI. The reference for the general result is in curly brackets, the result with respect to average mutual information, and the result with respect to average sum MSE. A cross stands for no reference.

In table (2.2), the impact of transmit correlation in MISO systems on the average performance with different types of CSI at the transmitter and perfect CSI at the receiver is summarised.

In table 2.3, the impact of transmit correlation in MIMO systems on the average performance with different types of CSI at the transmitter and perfect CSI at the receiver is summarised.

In table 2.4, the impact of receive correlation in MIMO systems on the average performance with different types of CSI at the transmitter and perfect CSI at the

[3] In this type of analysis, TDMA in multiuser SIMO systems corresponds to beamforming in single user MIMO systems.

CSI	MISO
no CSI	Schur-concave { Theorem 3 , [JB04c] , x }
covariance	Schur-convex (only for average mutual information){ x, Theorem 6, x }
perfect CSI	Schur-concave { Theorem 3 , [JB04c], x }

Table 2.2: Impact of transmit correlation in MISO systems on the average performance with different types of CSI at the transmitter and perfect CSI at the receiver.

CSI	MIMO: Tx correlation
no CSI	Schur-concave { Theorem 3 , [JB04e], x }
covariance	for small SNR Schur-convex, for high SNR Schur-concave { Lemma 6, x , x }
perfect CSI	for small SNR Schur-convex, for high SNR Schur-concave { section 2.3.3, x, x }

Table 2.3: Impact of transmit correlation in MIMO systems on the average mutual information and on the ergodic capacity with different types of CSI at the transmitter and perfect CSI at the receiver.

receiver is summarised.

CSI	MIMO: Rx correlation
no CSI	Schur-concave { Corollary 7, [JB04e], x }
covariance	Schur-concave { Theorem 5, x, x }
perfect CSI	for small SNR Schur-convex, for high SNR Schur-concave { section 2.3.3, x, x}

Table 2.4: Impact of receive correlation in MIMO systems on the average performance with different types of CSI at the transmitter and perfect CSI at the receiver.

In addition to the results which can be found in table 2.2, 2.3, and 2.4 it is worth mention that the loss due to correlation in MISO systems is bounded by some small constant for the open-loop and closed-loop channel capacity. Whereas the loss in MIMO system due to correlation in open-loop systems grows unbounded with the number of transmit antennas and the SNR. However, in some cases it is possible to exploit the correlation if partial or perfect CSI is available.

2.3.6 Numerical results and discussion

Beamforming range in MIMO systems with covariance knowledge

Inequality (2.49) can be solved for SNR ρ. The necessary SNR over the largest eigenvalue $\lambda_1^T = 1 - \lambda_2^T$ can be computed. In figure (2.4) the largest SNR at which beamforming achieves capacity given the channel eigenvalues for different numbers of receive antennas is plotted. The eigenvalues of the channel covariance matrix are normalised, i.e. $\sum_{i=1}^{n_T} \lambda_i^T = 1$. If we look at $\lambda_1^T = 0.6$, we can calculate the SNR value ($n_R = 2$) at which the inequality is tight. For $\lambda_1^T \to 1$ the largest SNR at which beamforming achieves capacity approaches infinity and for $\lambda_1^T \to 0.5$ the SNR at which beamforming achieves capacity approaches minus infinity. The more (uncorrelated) receive antennas are used the smaller is the largest

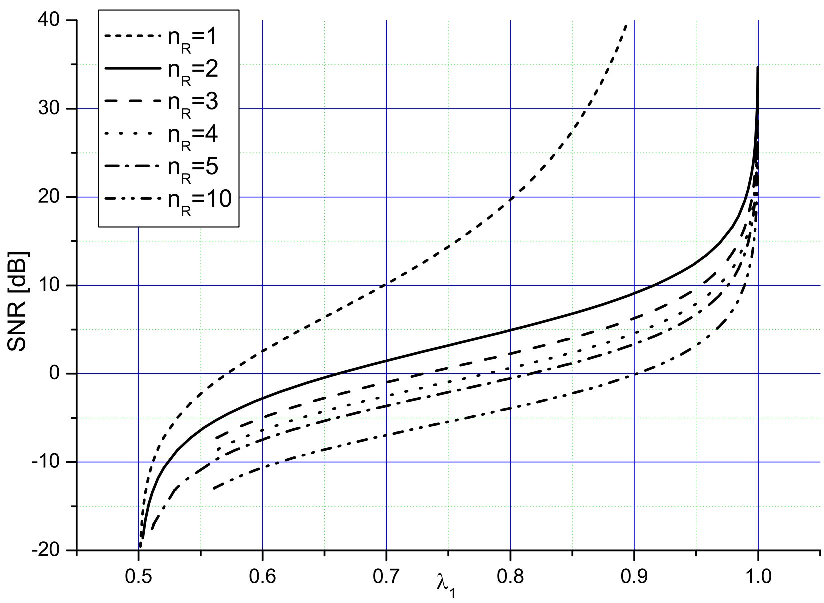

Figure 2.4: SNR range of beamforming over correlation λ_1^T.

SNR at which beamforming can achieve capacity. This behaviour follows from the expectation value of $||\boldsymbol{w}_1||^2$ that is a function of the number of receive antennas $\mathbb{E}(||\boldsymbol{w}_1||^2) = \mathbb{E}(||\boldsymbol{w}_2||^2) = n_R$. From the observations in figure (2.4) it follows that in the MIMO scenario additional directions are supported at lower SNR. Especially, the beamforming range in MISO systems is larger than in MIMO systems. In [JB03b], this result has been extended to the case in which the receive antennas are potentially correlated, too. The more correlated the receive antennas are, the larger is the beamforming range [JB03b, Lemma 2].

MIMO average mutual information and ergodic channel capacities for different system architectures under transmit correlation

In figure (2.5) we show the average mutual information (no CSI) and the ergodic channel capacity (covariance knowledge and perfect CSI) for a two times two MIMO system with different types of CSI at the transmitter. The two receive antennas are completely uncorrelated. On the x-axis the largest eigenvalues of the transmit correlation matrix is varied from $\lambda_1^T = 0.5...1$. The second eigenvalue of the transmit correlation matrix is given by $\lambda_2^T = 1 - \lambda_1^T$.

In figure (2.5), we have two SNR scenarios, low SNR at 0 dB and medium high SNR at 15 dB. The behaviour which was described in table 2.3 and 2.4 can be observed in figure (2.5): For small SNR values, we observe that the open-loop MIMO average mutual information decreases with transmit correlation whereas the covariance feedback and closed-loop MIMO ergodic channel capacities increase with transmit correlation. If knowledge about the transmit correlation is available, correlation can be utilised at low SNR values.

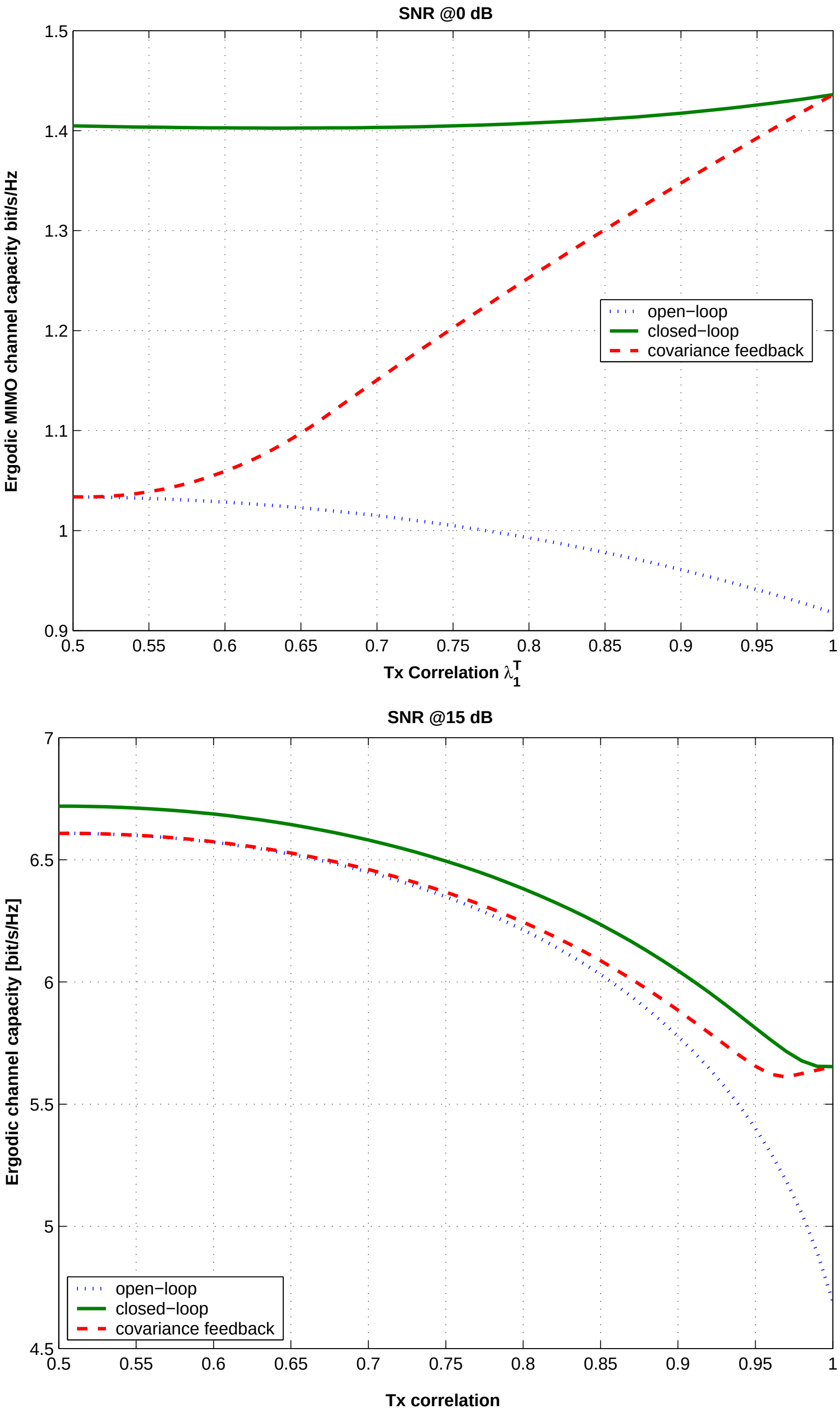

Figure 2.5: Open-loop average mutual information, closed-loop and covariance feedback ergodic MIMO channel capacity over transmit correlation with uncorrelated receive antennas.

For higher SNR values in (2.5), we observe that the open-loop MIMO average mutual information decreases with transmit correlation. The covariance feedback and closed-loop MIMO ergodic capacities are best for uncorrelated transmit antennas.

The curve for the covariance feedback is not monotonic with λ_1^T. For higher SNR the minimum moves to the right edge and the covariance feedback MIMO ergodic channel capacity becomes Schur-concave with respect to the transmit correlation.

Furthermore, note that for completely uncorrelated transmit antennas the ergodic channel capacity with covariance feedback and the average mutual information for no CSI are equal, because the transmit strategy is equal power allocation in both cases. For completely correlated transmit antennas, the ergodic channel capacity with perfect CSI and covariance feedback are equal because only one eigenvalue is supported and the transmit strategy is the same in both cases (beamforming).

MIMO average mutual information and ergodic channel capacities for different system architectures under receive correlation

In figure (2.6) we show the average mutual information (for no CSI) and the ergodic channel capacity (for covariance knowledge and perfect CSI) for a two times two MIMO system. The two transmit antennas are slightly correlated with $\lambda_1^T = 0.6$ and $\lambda_2^T = 0.4$. On the x-axis the largest eigenvalues of the receive correlation matrix is varied from $\lambda_1^R = 0.5...1$. The second eigenvalue of the transmit correlation matrix is given by $\lambda_2^R = 1 - \lambda_1^R$.

First, we note in figure (2.6), that at the boundary points with no correlation $\lambda_1^R = 0.5$ and with complete correlation $\lambda_1^R = 1$ the average mutual information and the ergodic channel capacities for the different types of CSI does not equal, because we have some transmit correlation which is utilised the more CSI is available at the transmitter.

For low SNR, we observe in figure (2.6), that the open-loop average mutual information and the covariance feedback MIMO ergodic channel capacity decrease with correlation, whereas the closed-loop capacity increases with correlation. For high SNR, all mutual informations decrease with correlation.

Concluding remarks

In this subsection, we studied a MIMO system with different types of CSI at the transmitter and correlation at the transmitter and the receiver side. We assumed perfect CSI at the receiver and analysed the average mutual information and the ergodic channel capacity, respectively.

Without CSI, we derived the open-loop MIMO average mutual information and the optimal transmission strategy. Equal power allocation turned out to be most robust against worst case channel correlation. Next, we showed that the average open-loop MIMO mutual information is Schur-concave with respect to correlation at the transmit array and Schur-concave with respect to correlation at the receive antenna array. In addition to this, we computed the loss due to correlation in open-loop MIMO systems.

With covariance knowledge at the transmitter, the transmit correlation can be utilised in order to increase the average mutual information. We derived the optimal transmission strategy and showed that the ergodic covariance feedback MIMO channel capacity can be Schur-convex (for small SNR values) or can be Schur-concave (for high SNR values) in contrast to the ergodic covariance feedback MISO channel capacity which is strict Schur-convex with respect to transmit correlation. The ergodic covariance feedback MIMO channel capacity is Schur-concave with respect to receive correlation.

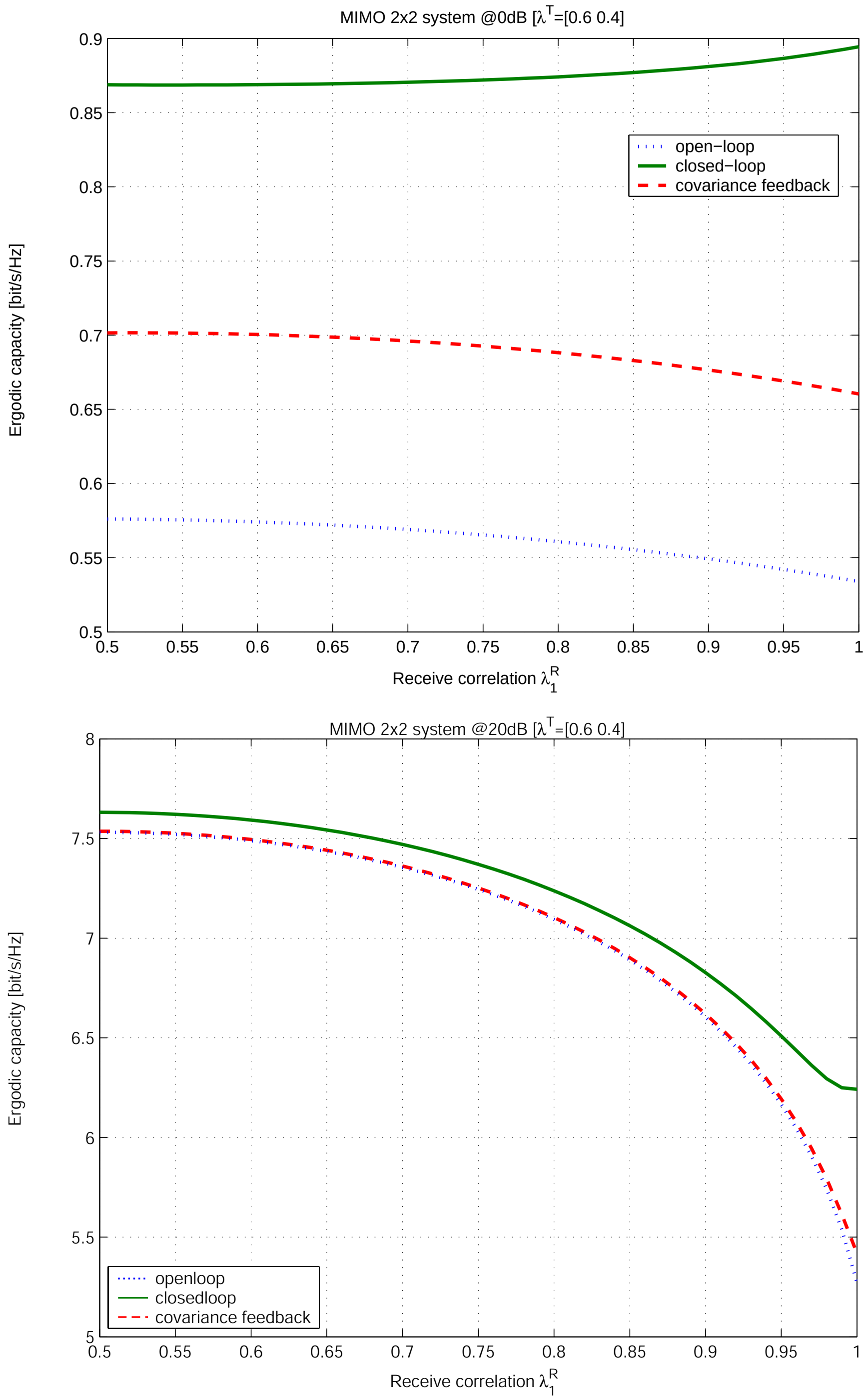

Figure 2.6: Open-loop average mutual information, closed-loop and covariance feedback ergodic MIMO channel capacities over receive correlation with correlated transmit antennas $\boldsymbol{\lambda}^T = [0.6; 0.4]$.

With perfect CSI at the transmitter, the optimal transmit strategy is waterfilling. Because the waterfilling power allocation is computed at each instantaneous channel realization, it is difficult to characterise the impact of transmit or receive correlation on the ergodic capacity. For transmit correlation we can argue the same lines as for the covariance feedback case, i.e. for small SNR values closed-loop ergodic channel capacity is Schur-convex while for high SNR values it is Schur-concave.

In all three scenarios, the transmitter consists of a Gaussian codebook from which the codewords are drawn, and of a signal processing unit with power allocation and beamforming entity. Depending on the type of CSI the power allocation and beamforming directions are adapted to the instantaneous channel realizations in each time slot with perfect CSI, or they are adapted for a long time period to the average transmit and receive correlation with covariance knowledge at the transmitter, or the transmit strategy is kept fixed for all time without CSI.

2.4 Proofs

2.4.1 Proof of Lemma 2

The optimisation problem $\max \mathbb{E}\ \mathrm{tr}\ \phi(\rho\tilde{\boldsymbol{W}}\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}\tilde{\boldsymbol{W}}^H)$ can be rewritten using the representation of matrix monotone functions as

$$\max_{\substack{\boldsymbol{Q}\succeq 0\\ \mathrm{tr}\ \boldsymbol{Q}\leq P}} \int_0^\infty \mathbb{E}\ \mathrm{tr}\ \left(s\rho\tilde{\boldsymbol{W}}\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}\tilde{\boldsymbol{W}}^H\left[\rho\tilde{\boldsymbol{W}}\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}\tilde{\boldsymbol{W}}^H+s\boldsymbol{I}\right]^{-1}\right)d\mu(s).$$

Next, we maximise the integrand for all $s > 0$ with respect to $\boldsymbol{Q}$. This maximises the integral itself. The resulting maximisation problem is given by

$$\begin{aligned} V &= \max_{\substack{\boldsymbol{Q}\succeq 0\\ \mathrm{tr}\ \boldsymbol{Q}\leq P}} \mathbb{E}\ \mathrm{tr}\ \left(s\rho\tilde{\boldsymbol{W}}\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}\tilde{\boldsymbol{W}}^H\left[\rho\tilde{\boldsymbol{W}}\boldsymbol{R}_T^{1/2}\boldsymbol{Q}\boldsymbol{R}_T^{1/2}\tilde{\boldsymbol{W}}^H+s\boldsymbol{I}\right]^{-1}\right)\\ &= \max_{\substack{\hat{\boldsymbol{Q}}\succeq 0\\ \mathrm{tr}\ \boldsymbol{R}_T^{-1/2}\hat{\boldsymbol{Q}}\boldsymbol{R}_T^{-1/2}\leq P}} \mathbb{E}\ \mathrm{tr}\ \left(s\rho\tilde{\boldsymbol{W}}\hat{\boldsymbol{Q}}\tilde{\boldsymbol{W}}^H\left[\rho\tilde{\boldsymbol{W}}\hat{\boldsymbol{Q}}\tilde{\boldsymbol{W}}^H+s\boldsymbol{I}\right]^{-1}\right) \end{aligned} \tag{2.86}$$

for all real numbers $s \geq 0$. If $\boldsymbol{R}_T$ has full rank, there exists a $\tilde{\boldsymbol{Q}}$ with

$$\mathrm{tr}\ \boldsymbol{R}_T^{-1/2}\tilde{\boldsymbol{Q}}\boldsymbol{R}_T^{-1/2} \leq P$$

which solves (2.86)

$$V = \mathbb{E}\ \mathrm{tr}\ \left(s\rho\tilde{\boldsymbol{W}}\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{W}}^H\left[\rho\tilde{\boldsymbol{W}}\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{W}}^H+s\boldsymbol{I}\right]^{-1}\right). \tag{2.87}$$

With the eigenvalue decomposition of $\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{U}}^H$ and using the fact that the random matrix $\tilde{\boldsymbol{W}}$ is invariant against multiplication with unitary matrix from the right, we obtain

$$V = \mathbb{E}\ \mathrm{tr}\ \left(s\rho\tilde{\boldsymbol{W}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{W}}^H\left[\rho\tilde{\boldsymbol{W}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{W}}^H+s\boldsymbol{I}\right]^{-1}\right). \tag{2.88}$$

Finally, we have to show that the trace condition is satisfied by

$$\mathrm{tr}\ \boldsymbol{D}_T^{-1/2}\tilde{\boldsymbol{\Lambda}}\boldsymbol{D}_T^{-1/2} \leq P.$$

This follows from the fact that the trace of the sum of two Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with eigenvalues $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq ... \geq \beta_n$, respectively, is bounded from below and above by

$$\sum_{k=1}^{n} \alpha_k \beta_{n-k+1} \leq \text{ tr } \boldsymbol{AB} \leq \sum_{k=1}^{n} \alpha_k \beta_k. \tag{2.89}$$

The upper bound in (2.89) is derived in [MO79, 9.H.1.g] and the lower bound is proven in [MO79, 9.H.1.h]. This completes the proof.

□

2.4.2 Proof of Theorem 1

We name the optimal power allocation with $\hat{\boldsymbol{p}}$, i.e. from (2.39)

$$\hat{\boldsymbol{p}} = \arg \max_{||\boldsymbol{p}|| \leq P, p_i \geq 0} \Phi(\boldsymbol{p}, \boldsymbol{\lambda}^R, \boldsymbol{\lambda}^T).$$

Let the correlation matrix eigenvalues $\lambda_1^T, ..., \lambda_{n_T}^T$ and $\lambda_1^R, ..., \lambda_{n_R}^R$ be fixed. We define the parametrised power allocation

$$\boldsymbol{p}(t) = (1-t)\hat{\boldsymbol{p}} + t\boldsymbol{p} \tag{2.90}$$

with arbitrary $\boldsymbol{p} : ||\boldsymbol{p}|| \leq P, p_i \geq 0$. The parametrised average performance metric is given by

$$\Phi(t) \quad = \quad \mathbb{E} \text{ tr } \phi \left(\rho \sum_{k=1}^{n_T} \hat{p}_k \lambda_k \boldsymbol{w}_k \boldsymbol{w}_k^H + \rho t \sum_{k=1}^{n_T} (p_k - \hat{p}_k) \lambda_k \boldsymbol{w}_k \boldsymbol{w}_k^H \right). \tag{2.91}$$

The first derivative of (2.91) at the point $t = 0$ is given by

$$\left. \frac{\delta \Phi(t)}{\delta t} \right|_{t=0} = \sum_{k=1}^{n_T} (p_k - \hat{p}_k) \alpha_k(\hat{p}) \tag{2.92}$$

with $\alpha_k(\hat{p})$ defined in (2.41). This follows from the Lemma 9 in appendix 2.4.3. It is easily shown that the second derivative of $\Phi(t)$ is always smaller to zero for all $0 \leq t \leq 1$, because the inner performance function $\phi(\boldsymbol{X})$ is matrix-concave by assumption. Hence, it suffices to show that the first derivative of $\Phi(t)$ at the point $t = 0$ is less or equal to zero, i.e.

$$\sum_{k=1}^{n_T} (p_k - \hat{p}_k) \alpha_k(\hat{p}) \leq 0. \tag{2.93}$$

We split the proof into two parts. In the first part, we will show that the condition in (2.44) is sufficient. We assume that (2.44) is fulfilled. We can rewrite the first derivative of $\Phi(t)$ at the point $t = 0$ as

$$\begin{aligned} Q \quad &= \quad \sum_{k=1}^{n_T} (\hat{p}_k - p_k) \alpha_k(\hat{p}_k) = \sum_{k=1}^{n_T} \hat{p}_k \alpha_k(\hat{p}) \\ &= \quad \max_{k \in \{1,...,n_T\}} \alpha_k(\hat{p}) \sum_{l \in \mathcal{I}(\hat{p})} \hat{p}_l - \sum_{l=1}^{n_T} p_l \alpha_l(\hat{p}). \end{aligned} \tag{2.94}$$

But we have

$$\sum_{l=1}^{n_T} p_l \alpha_l(\hat{p}) \leq \sum_{l=1}^{n_T} p_l \max_{k\in\{1,\ldots,n_T\}} \alpha_l(\hat{p}).$$

Therefore, it follows for Q in (2.95)

$$Q \geq 0 \tag{2.95}$$

In order to show that condition (2.44) is a necessary condition for the optimality of power allocation $\hat{\boldsymbol{p}}$, we study two cases and prove them by contradiction.

1. Assume (2.44) is not true. Then we have a $k \in \mathcal{I}(\hat{p})$ and $k_0 \in \mathcal{I}(\hat{p})$ with the following properties:

$$\max_{1\leq k\leq n_T} \alpha_k(\hat{p}) = \alpha_{k_0(\hat{p})}$$

and $\alpha_k(\hat{p}) < \alpha_{k_0}(\hat{p})$. We set $\tilde{p}_{k_0} = 1$ and $\tilde{p}_{i\in\{1,\ldots,n_T\}k_0} = 0$. It follows that

$$\sum_{l=1}^{n_T} (\hat{p}_k - \tilde{p}_k)\alpha_k(\hat{p}) < 0$$

which is a contradiction.

2. Assume $\exists k_0 : \alpha_{k_0} > \alpha_k$ with $k_0 \notin \mathcal{I}(\hat{p})$ and $k \in \mathcal{I}(\hat{p})$, then set $\tilde{p}_{k_0} = 1$ and $\tilde{o}_{l\in\{1,\ldots,n_T\}k_0} = 0$. Then we have the contradiction

$$\sum_{k=1}^{n_T} (\hat{p}_k - \tilde{p}_k)\alpha_k < 0.$$

This completes the proof.

□

2.4.3 Lemma 9 and its proof

Lemma 9: The first derivative of the function $\operatorname{tr} F(\boldsymbol{C} + \epsilon\boldsymbol{D})$ at the point $\epsilon = 0$ is given by

$$\frac{\partial}{\partial\epsilon} \operatorname{tr} \left[F(\boldsymbol{C} + \epsilon\boldsymbol{D})\right]\Big|_{\epsilon=0} = \operatorname{tr} \left(F^{[1]}(\boldsymbol{C}) \cdot \boldsymbol{D}\right) \tag{2.96}$$

Proof: The function F is analytic, because F is matrix-monotone and the representation in (2.28) can be used. Therefore, the function can be approximated arbitrary well with a polynomial. Both sides of (2.96) in Lemma 1 are linear in F. Therefore, it suffices to prove equation (2.96) for the powers $F(t) = t^p$ with $p \in \mathbb{N}^+$. It holds

$$\begin{aligned}
\operatorname{tr}\left(\frac{\partial}{\partial\epsilon} \operatorname{tr} \left[F(\boldsymbol{C} + \epsilon\boldsymbol{D})\right]\Big|_{\epsilon=0}\right) &= \operatorname{tr}\left(\sum_{k=1}^{p} \boldsymbol{C}^{k-1}\boldsymbol{D}\boldsymbol{C}^{p-k}\right) \\
&= \operatorname{tr}\left(\boldsymbol{D}\boldsymbol{C}^{p-1} + \boldsymbol{C}\boldsymbol{D}\boldsymbol{C}^{p-2} + \boldsymbol{C}^2\boldsymbol{D}\boldsymbol{C}^{p-3} + \ldots\right) \\
&= p \cdot \operatorname{tr}\left(\boldsymbol{D}\boldsymbol{C}^{p-1}\right) \\
&= \operatorname{tr}\left(F^{[1]}(\boldsymbol{C}) \cdot \boldsymbol{D}\right).
\end{aligned}$$

□

2.4.4 Proof of Theorem 2

First, we have to prove the following: If the optimum is achieved by beamforming, the inequality (2.49) is fulfilled. In the case of beamforming, the full transmit power is used to transmit in direction of the largest channel covariance matrix eigenvalue. We express the SNR as $\rho = \frac{P}{\sigma_n^2}$. We allocate power $1-p$ to the first eigenvalue of the transmit covariance matrix, i.e. $p_1 = (1-p)$ and choose a power allocation $\tilde{p}_2, ..., \tilde{p}_{n_T}$ which sums up to p, i.e. $\sum_{k=2}^{n_T} \tilde{p}_k = p$. This normalisation can be achieved by $p_2, ..., p_{n_T}$ which sum up to one and are multiplied by p. Therefore, we have the following parameterised average performance metric

$$\Phi(p) = \mathbb{E}\ \mathrm{tr}\ \phi\left(\rho(1-p)\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H + \rho p \sum_{k=2}^{n_T} p_k \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H\right). \tag{2.97}$$

By assumption, the optimum can be achieved with beamforming. This gives

$$\Phi(p) \leq \Phi(0) \tag{2.98}$$

for all $0 \leq p \leq 1$. Note, that $\Phi(p)$ is concave with respect to p. It follows that the first derivate of $\Phi(p)$ at the point $p = 0$ is less or equal to zero, because (2.98) holds. The first derivate of $\Phi(p)$ at the point $p = 0$ is given by

$$\begin{aligned} \left.\frac{\partial \Phi(p)}{\partial p}\right|_{p=0} &= \mathbb{E}\ \mathrm{tr}\ \left(\left[-\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H + \rho \sum_{k=2}^{n_T} p_k \lambda_k^T \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H\right] \cdot \phi^{[1]}\left(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H\right)\right) \\ &\leq \mathbb{E}\ \mathrm{tr}\ \left(\left[-\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H + \rho\lambda_2^T \tilde{\boldsymbol{w}}_2 \tilde{\boldsymbol{w}}_2^H\right] \cdot \phi^{[1]}\left(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H\right)\right). \end{aligned} \tag{2.99}$$

The first inequality in (2.99) follows from the choice $p_2 = p$ and $p_3 = p_4 = ... = p_{n_T} = 0$ which provides an upper bound. With the upper bound of the first derivative (2.99) at the point $p = 0$ follows the condition

$$\mathbb{E}\ \mathrm{tr}\ \left(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H \phi^{[1]}(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H)\right) \geq \mathbb{E}\ \mathrm{tr}\ \left(\rho \boldsymbol{R}_R \phi^{[1]}(\rho\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H)\right) \tag{2.100}$$

which exactly corresponds to the condition in (2.48).

For the reverse direction, we have to prove that if the inequality is fulfilled, then the optimum can be achieved by beamforming. From equation (2.100), we know that if λ_1^T and λ_2^T fulfill the inequality (2.48) then the first derivative of the parameterised performance (2.97) at the point $p = 0$ is less or equal to zero. Now assume that beamforming can not achieve the optimum. Then there exists an optimal power allocation (P_1, P_2) with $P_2 > 0$. From Lemma 10 in the section 2.4.5 follows that for all $0 \leq p \leq P_1$

$$\Phi(p) > \Phi(0). \tag{2.101}$$

This yields $\Phi'(0) > 0$ which is a contradiction to (2.48). It follows that (2.48) is also sufficient and the proof is completed.

□

2.4.5 Lemma 10 and its proof

Lemma 10: We assume that the optimal transmission covariance matrix $\boldsymbol{Q}$ has rank 2, i.e. beamforming can not achieve the optimum. It follows that the optimum achieving power allocation has the form (P_1, P_2) with $P_2 > 0$. We can write

$$\Phi_{opt} = \mathbb{E} \text{ tr } \phi(\rho P_1 \lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H + \rho P_2 \lambda_2^T \tilde{\boldsymbol{w}}_2 \tilde{\boldsymbol{w}}_2^H).$$

Then for all $0 < p \leq P_2$ and

$$\Phi(p) = \mathbb{E} \text{ tr } \phi(\rho((1-p)\lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H + p\lambda_2^T \tilde{\boldsymbol{w}}_2 \tilde{\boldsymbol{w}}_2^H))$$

it follows

$$\Phi(p) > \Phi(0). \tag{2.102}$$

Proof: With $0 < p < P_2$ arbitrary define $\nu = 1 - \frac{p}{P_2}$. It follows $0 < \nu < 1$. Further on, define the matrix

$$\boldsymbol{A}_1 = \rho \lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H \tag{2.103}$$

and

$$\boldsymbol{A}_2 = \rho(P_1 \lambda_1^T \tilde{\boldsymbol{w}}_1 \tilde{\boldsymbol{w}}_1^H + P_2 \lambda_2^T \tilde{\boldsymbol{w}}_2 \tilde{\boldsymbol{w}}_2^H). \tag{2.104}$$

Now consider the linear combination of (2.103) and (2.104)

$$\boldsymbol{A}(\nu) = \nu \boldsymbol{A}_1 + (1-\nu)\boldsymbol{A}_2. \tag{2.105}$$

With (2.105) and $\nu = 1 - \frac{p}{P_2}$ the performance function can be parameterised

$$\begin{aligned}
\Phi(p) &= \mathbb{E} \text{ tr } \phi(\boldsymbol{A}(\nu)) \\
&= \mathbb{E} \text{ tr } (\nu \boldsymbol{A}_1 + (1-\nu)\boldsymbol{A}_2) \\
&\geq \nu \mathbb{E} \text{ tr } \phi(\boldsymbol{A}_1) + (1-\nu)\mathbb{E} \text{ tr } \phi(\boldsymbol{A}_2).
\end{aligned}$$

The inequality follows by observing that tr $\Phi(\boldsymbol{A}(\nu))$ is strict concave on the set of positive definite matrices (with the usual ordering). The expectation operator is linear and order preserving and therefore follows the inequality in (2.102). Hence, if the rank of the optimal $\boldsymbol{Q}$ is equal to two than the performance metric $\Phi(p)$ is greater than $\Phi(0)$ for all $0 < p < P_2$. The case $p = P_2$ is trivial.

□

2.4.6 Proof of Lemma 3

The Lagrangian for the optimisation problem in (2.53) is given by

$$L(p_1, .., p_{n_T}, \mu, \psi_1, ..., \psi_{n_T}) = \sum_{k=1}^{n_T} \phi(\lambda_k p_k) + \sum_{k=1}^{n_T} \psi_k p_k - \mu \left(P - \sum_{k=1}^{n_T} p_k \right) \tag{2.106}$$

with Lagrangian multipliers $\mu > 0$ and $\psi_k = 0$ if $p_k > 0$, otherwise $\psi_k > 0$ if $p_k = 0$. The KKT optimality condition is then

$$\lambda_k \phi^{[1]}(\lambda_k p_k) = \mu - \psi_k \qquad \text{for all } 1 \leq k \leq n_T.$$

From this follows directly the optimal power allocation in (2.54).

□

2.4.7 Proof of Lemma 4

Assume that two different power allocations $\boldsymbol{p}^1$ and $\boldsymbol{p}^2$ are optimal with $\sum_{k=1}^{n_T} p_k^1 = \sum_{k=1}^{n_T} p_k^2 = P$. Note, that if $p_1^1 = p_1^2$ then the Lagrangian multiplier μ for user one and user two is equal and therefore the complete power allocation $p_k^1 = p_k^2$ would be equal. As a result, we assume without loss of generality, that $p_1^1 > p_2^2$. Then it follows that the Lagrangian multiple $\nu^1 < \nu^2$ because

$$\nu^1 = \rho\lambda_1\phi^{[1]}(p_1^1\lambda_1) < \rho\lambda_1\phi^{[1]}(p_1^2\lambda_1) = \nu^2$$

and because $\phi^{[1]}$ is monotonically decreasing. Set L^1 and L^2 the number of 'active' powers, i.e. for all $1 \leq j \leq L^1$ it holds $p_1^1, ..., p_j^1 > 0$. From the representation in (2.55) follows for all $l > L^1$ that $\lambda_l \leq \mu^1$. For $\lambda_l > \mu^1$ holds $p_l > 0$. This set contains the set of all k with $\lambda_k \geq \mu^2$. As a result, we have $L^1 \geq L^2$.

For all $1 \leq l \leq L^2$ holds $p_l^1 > p_l^2$ because $\mu^1 < \mu^2$. Therefore, we have

$$\sum_{l=1}^{L^2} p_l^1 > \sum_{l=1}^{L^2} p_l^2 = P. \tag{2.107}$$

The inequality in (2.107) is a contradiction to the assumption, that $\sum p_k^1 = \sum p_k^2 = P$. This completes the proof.

□

2.4.8 Proof of Theorem 3

For fixed receive correlation, we show that Schur's condition in (2.11) is fulfilled with respect to transmit correlation vector $\boldsymbol{\lambda}^T$. In order to verify Schur's condition, the first derivative of (2.63) with respect to λ_1^T and λ_2^T is important. These partial derivatives are given by

$$\frac{\partial\Phi(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R)}{\partial\lambda_1^T} = \mathbb{E}\,\mathrm{tr}\left[\rho\tilde{\boldsymbol{w}}_1\tilde{\boldsymbol{w}}_1^H \cdot \phi^{[1]}\left(\rho\sum_{k=1}^{n_T}\lambda_k^T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right)\right] \tag{2.108}$$

$$\frac{\partial\Phi(\boldsymbol{\lambda}^T, \boldsymbol{\lambda}^R)}{\partial\lambda_2^T} = \mathbb{E}\,\mathrm{tr}\left[\rho\tilde{\boldsymbol{w}}_2\tilde{\boldsymbol{w}}_2^H \cdot \phi^{[1]}\left(\rho\sum_{k=1}^{n_T}\lambda_k^T\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right)\right] \tag{2.109}$$

The first two largest eigenvalues λ_1^T and λ_2^T are parameterised by

$$\lambda_1^T = \lambda + t \qquad \text{and} \qquad \lambda_2^T = \lambda - t.$$

Then, the difference between the first derivatives in (2.108) and (2.109) is a function of t and is given by

$$\Gamma(t) = \mathbb{E}\,\mathrm{tr}\left[\boldsymbol{\Delta} \cdot \phi^{[1]}\left(\boldsymbol{R} + \rho t\boldsymbol{\Delta}\right)\right] \tag{2.110}$$

with

$$\begin{aligned}\boldsymbol{W}_k &= \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^H \\ \boldsymbol{R} &= \rho\lambda\left(\boldsymbol{W}_1 - \boldsymbol{W}_2\right) + \rho \sum_{k=3}^{n_T} \lambda_k^T \boldsymbol{W}_k \\ \boldsymbol{\Delta} &= \boldsymbol{W}_1 - \boldsymbol{W}_2.\end{aligned}$$

Next, the matrix monotone function $\phi(\boldsymbol{A})$ can be written as

$$\phi(\boldsymbol{A}) = \int_0^\infty s\boldsymbol{A}\left[s\boldsymbol{I} + \boldsymbol{A}\right]^{-1} d\mu(s)$$

with probability distribution $d\mu(s)$. The 'first derivative' as defined in section 2.3.1 is then given as

$$\phi^{[1]}(\boldsymbol{A}) = \int_0^\infty s^2\left[s\boldsymbol{I} + \boldsymbol{A}\right]^{-2} d\mu(s). \tag{2.111}$$

The result from (2.111) is set into $\Gamma(t)$ in (2.110) and integration and summation is exchanged to obtain

$$\Gamma(t) = \int_0^\infty s^2 \mathbb{E} \text{ tr } \left(\boldsymbol{\Delta}\cdot\left[\boldsymbol{R}(s) + \rho t\boldsymbol{\Delta}\right]^{-2}\right) d\mu(s) \tag{2.112}$$

with $\boldsymbol{R}(s) = s\boldsymbol{I} + \rho\lambda\left(\boldsymbol{W}_1 + \boldsymbol{W}_2\right) + \rho\sum_{k=3}^{n_T}\lambda_k^T\boldsymbol{W}_k$. Finally, we study the trace expression in (2.112) as a function with respect to t. Therefore, we define for all $s \geq 0$

$$M(t) = \text{ tr } \left(\boldsymbol{\Delta}\cdot\left[\boldsymbol{R}(s) + \rho t\boldsymbol{\Delta}\right]^{-2}\right). \tag{2.113}$$

Note, that $\Gamma(0) = 0$. The first derivative of the function $M(t)$ with respect to t is for all $s \geq 0$ and all $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ smaller than or equal to zero. Therefore, the integral in (2.112) is smaller than or equal to zero, too, because the outer integral is over a probability distribution function, which is positive for all s by definition and has only positive steps. With $\boldsymbol{T} = \left[\boldsymbol{R}(s) + \rho t\boldsymbol{\Delta}\right]^{-1}$ it holds

$$\frac{\partial M(t)}{\partial t} = -\rho \text{ tr } \left(\boldsymbol{\Delta}\boldsymbol{T}\boldsymbol{T}\boldsymbol{\Delta}\boldsymbol{T}\right) - \rho \text{ tr } \left(\boldsymbol{\Delta}\boldsymbol{T}\boldsymbol{\Delta}\boldsymbol{T}\boldsymbol{T}\right) = -2\rho \text{ tr } \left(\boldsymbol{T}\underbrace{\boldsymbol{\Delta}\boldsymbol{T}\boldsymbol{\Delta}}_{\boldsymbol{Q}}\boldsymbol{T}\right). \tag{2.114}$$

Note, that the matrix $\boldsymbol{T}$ is positive definite. Finally, the matrix $\boldsymbol{Q}$ can be written as

$$\begin{aligned}\boldsymbol{Q} &= \boldsymbol{W}_1\boldsymbol{T}\boldsymbol{W}_1 - \boldsymbol{W}_2\boldsymbol{T}\boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{T}\boldsymbol{W}_2 + \boldsymbol{W}_2\boldsymbol{T}\boldsymbol{W}_2 \\ &= \underbrace{\left[\boldsymbol{W}_1\boldsymbol{T}^{1/2} - \boldsymbol{W}_2\boldsymbol{T}^{1/2}\right]}_{\boldsymbol{C}}\left[\boldsymbol{T}^{1/2}\boldsymbol{W}_1 - \boldsymbol{T}^{1/2}\boldsymbol{W}_2\right] = \boldsymbol{C}\boldsymbol{C}^H \succeq \boldsymbol{0}. \end{aligned} \tag{2.115}$$

Inequality (2.115) shows that the matrix $\boldsymbol{Q}$ is positive definite and therefore the first derivative of $\boldsymbol{M}(t)$ with respect to t in (2.114) is smaller than or equal to zero. Therefore, the function $\Phi(t)$ in (2.112) is smaller than or equal to zero and this verifies Schur's condition and completes the proof.

□

2.4.9 Proof of Lemma 5

The difference in (2.64) can be rewritten with $\boldsymbol{X} = \boldsymbol{W}\boldsymbol{W}^H$ and $Y = n_T n_R w_1$ as

$$\begin{aligned}
\text{tr } \Phi(\rho \boldsymbol{X}) - \text{ tr } \Phi(\rho Y) &= \mathbb{E} \int_0^\infty s \text{ tr } \left(\boldsymbol{X} \left[\frac{s}{\rho} \boldsymbol{I} + \boldsymbol{X} \right]^{-1} - Y \left(\frac{s}{\rho} 1 + Y \right)^{-1} \right) d\mu(s) \\
&= \mathbb{E} \int_0^\infty \tilde{s}\rho \text{ tr } \left(\boldsymbol{X} \left[\tilde{s} \boldsymbol{I} + \boldsymbol{X} \right]^{-1} - Y (\tilde{s} + Y)^{-1} \right) d\mu(\tilde{s}) \\
&= \rho \left(\text{ tr } \tilde{\Phi}(\boldsymbol{X}) - \text{ tr } \tilde{\Phi}(Y) \right). \qquad (2.116)
\end{aligned}$$

The second step follows from the substitution $\tilde{s} = \frac{s}{\rho}$. The difference tr $\tilde{\Phi}(\boldsymbol{X})$ − tr $\tilde{\Phi}(Y)$ in (2.116) is positive, since $\tilde{\Phi}(\boldsymbol{X})$ is matrix-monotone, too. As a result, the difference is monotonically increasing with SNR ρ.

□

2.4.10 Proof of Theorem 4

We can lower bound the difference in (2.67) by

$$\begin{aligned}
\Delta_{nc}^{cc}(\rho, n_T, n_R) &= \mathbb{E} \log \det \left(\frac{1}{\rho} \boldsymbol{I} + \boldsymbol{W}\boldsymbol{W}^H \right) - \mathbb{E} \log \left(\frac{1}{\rho} + n_R n_T w_1 \right) + \log(\rho^{n_T - 1}) \\
&\geq \mathbb{E} \log \det \left(\boldsymbol{W}\boldsymbol{W}^H \right) - \mathbb{E} \log(n_T n_R w_1) + \log(\rho^{n_T - 1}). \qquad (2.117)
\end{aligned}$$

Note that for SNR approaching infinity this bound gets tight. The first two terms can be calculated as follows. The expectation of $\log(n_T n_R w_1)$ is given by

$$\mathbb{E} \log(n_T n_R w_1) = \log(n_T n_R) - \gamma$$

with Euler's constant γ [GR80]. The pdf of the determinant of the Wishart matrix is the same as the pdf of the sum of independent chi squared distributed random numbers with different degrees of freedom [Goo63, GN00]

$$\det \left(\sum_{k=1}^{n_T} \boldsymbol{w}_k \boldsymbol{w}_k^H \right) \sim \sum_{k=1}^{n_T} \chi^2_{2(n_R - k + 1)} - n_T \cdot \log(2). \qquad (2.118)$$

Using (2.118) and the following identity for a chi squared random variable with ζ degrees of freedom

$$\mathbb{E} \log \chi^2_\zeta = \log(2) + \Psi(\frac{\zeta}{2}),$$

we obtain

$$\begin{aligned}\mathbb{E}\log\det\big(\sum_{k=1}^{n_T}\boldsymbol{w}_k\boldsymbol{w}_k^H\big) &= \sum_{k=1}^{n_T}\mathbb{E}\log\chi^2_{2(n_R-k+1)} - n_T\log(2)\\ &= \sum_{k=1}^{n_T}(\log(2)+\Psi(n_R-k+1)) - n_T\log(2)\\ &= \sum_{k=1}^{n_T}\Psi(n_R-k+1) \qquad (2.119)\end{aligned}$$

with the Psi-function which can be recursively defined for integer arguments (see section 8.365 in [GR80])

$$\begin{aligned}\Psi(1) &= -\gamma\\ \Psi(k+1) &= \Psi(k)+\frac{1}{k}. \qquad (2.120)\end{aligned}$$

For the lower bound on the capacity loss due to correlation of the single user MIMO system with an uninformed transmitter and perfect CSI (2.119) yield in (2.117) the desired equation.

□

2.4.11 Proof of Lemma 6

In this proof we cannot directly use Schur's condition, because the optimal power allocation strategy depends on the transmit correlation. Therefore, we have to prove first that the additional term which occurs for the derivative of the optimal power allocation with respect to the correlation eigenvalues, vanishes.

The proof is constructed in the following way: At first, we consider two arbitrary transmit correlation vectors $\boldsymbol{\lambda}^1$ and $\boldsymbol{\lambda}^2$ which satisfy $\boldsymbol{\lambda}^1 \succ \boldsymbol{\lambda}_2$. The we construct all possible linear combinations of $\boldsymbol{\lambda}^1$ and $\boldsymbol{\lambda}^2$. Notice that all linear combinations lie in the space of comparable correlation vectors. Next, we study the parametrised performance as a function of the linear combination parameter t. We show that the first derivative of the parametrised performance with respect to t is less than or equal to zero for all t if the condition (2.71) in Lemma 6 is fulfilled. This result can be generalised and can be applied for every (comparable) correlation vector.

With arbitrary $\boldsymbol{\lambda}^1$ and $\boldsymbol{\lambda}^2$ which satisfy $\boldsymbol{\lambda}^1 \succ \boldsymbol{\lambda}^2$, define the eigenvalue vector

$$\boldsymbol{\lambda}(t) = t\boldsymbol{\lambda}^2 + (1-t)\boldsymbol{\lambda}^1.$$

Assume that the optimum power allocation is given by $p_1(t), ..., p_{n_T}(t)$. The parametrised average performance is then given by

$$\begin{aligned}f(t) &= \mathbb{E}\,\mathrm{tr}\,\phi\left(\rho\sum_{k=1}^{n_T}\lambda_k(t)p_k(t)\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right) \qquad (2.121)\\ &= \mathbb{E}\,\mathrm{tr}\,\phi\left(\rho\sum_{k=1}^{n_T}(\lambda_k^1+t(\lambda_k^2-\lambda_k^1))p_k(t)\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\right).\end{aligned}$$

The first derivative of (2.121) with respect to t is given by

$$\begin{aligned}\frac{df(t)}{dt} &= \mathbb{E}\,\mathrm{tr}\,\Big(\phi^{[1]}(\boldsymbol{A})\big(\rho\sum_{k=1}^{n_T} p_k(t)(\lambda_k^2-\lambda_k^1)\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H \\ &\qquad +\rho\sum_{k=1}^{n_T}\frac{dp_k(t)}{dt}\left(\lambda_k^1+t(\lambda_k^2-\lambda_k^1)\right)\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H\big)\Big) \end{aligned} \tag{2.122}$$

with

$$\boldsymbol{A}=\rho\sum_{k=1}^{n_T} p_k(t)\cdot\left(\lambda_k^1+t(\lambda_k^2-\lambda_k^1)\right)\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^H.$$

Let us consider the second term in (2.122) first. Define

$$\psi_k(t)=(\lambda_k^2+t(\lambda_k^1-\lambda_k^2))\ \ \forall k=1...n_T.$$

Then we have

$$\sum_{k=1}^{n_T}\frac{dp_k(t)}{dt}\mathbb{E}\left(\psi_k(t)\tilde{\boldsymbol{w}}_k^H\phi^{[1]}(\boldsymbol{A})\tilde{\boldsymbol{w}}_k\right)=\sum_{k=1}^{n_T}\frac{dp_k(t)}{dt}\alpha_k(t). \tag{2.123}$$

In order to show that (2.123) is equal to zero, we define the index m for which holds

$$\frac{dp_k(t)}{dt}\neq 0\quad \forall 1\leq k\leq m. \tag{2.124}$$

We split the sum in (2.123) in two parts, i.e.

$$\sum_{k=1}^{m}\frac{dp_k(t)}{dt}\alpha_k(t)+\sum_{k=m+1}^{n_T}\frac{dp_k(t)}{dt}\alpha_k(t). \tag{2.125}$$

For all $1\leq k\leq m$ we have from (2.124) that either

$$p_k(t+\epsilon) > 0\quad\forall\epsilon>0 \tag{2.126}$$
$$\alpha_k(t+\epsilon) = \alpha_{max}(t+\epsilon)=\alpha_1(t+\epsilon)\quad\forall\epsilon>0 \tag{2.127}$$

or

$$p_k(t-\epsilon) > 0\quad\forall\epsilon>0 \tag{2.128}$$
$$\alpha_k(t-\epsilon) = \alpha_{max}(t-\epsilon)=\alpha_1(t-\epsilon)\quad\forall\epsilon>0 \tag{2.129}$$

Equation (2.127) and (2.129) follows from (2.126) and (2.128) using the result in (2.44) for all active k. In (2.127), direction k is supported for all $\tau\geq t$, i.e. direction k occurs at t. In (2.129), direction k is supported for all $\tau\leq t$, i.e. direction k vanishes at t.

Assume the case in (2.127). The set of points t for which $\alpha_k(t)=\alpha_1(t)$ is closed because the preimages of closed sets are closed. Using continuity, it holds

$$\alpha_k(t)=\lim_{\epsilon\to 0}\alpha_k(t+\epsilon)=\lim_{\epsilon\to 0}\alpha_1(t+\epsilon)=\alpha_1(t). \tag{2.130}$$

For the case in (2.129), it holds

$$\alpha_k(t)=\lim_{\epsilon\to 0}\alpha_k(t-\epsilon)=\lim_{\epsilon\to 0}\alpha_1(t-\epsilon)=\alpha_1(t). \tag{2.131}$$

The consequence from (2.130) and (2.131) is that all active directions at point t and all directions which occur or vanish at point t fulfill $\alpha_1(t) = \alpha_2(t) = ... = \alpha_m(t)$. Therefore, the first addend in (2.125) is

$$\sum_{k=1}^{m} \frac{dp_k(t)}{dt} = \alpha_1(t) \sum_{k=1}^{m} \frac{dp_k(t)}{dt} = 0.$$

The second addend in (2.125) is obvious equal to zero. We obtain for (2.122)

$$\frac{df(t)}{dt} = \sum_{k=1}^{n_T} (\lambda_k^2 - \lambda_k^1) p_k(t) \mathbb{E}\left(\tilde{\boldsymbol{w}}_k^H \phi^{[1]}(\boldsymbol{A}) \tilde{\boldsymbol{w}}_k\right).$$

We are going to show that

$$\sum_{k=1}^{n_T} (\lambda_k^2 - \lambda_k^1) \mathbb{E}\left(p_k(t) \tilde{\boldsymbol{w}}_k^H \phi^{[1]}(\boldsymbol{A}) \tilde{\boldsymbol{w}}_k\right) \leq 0. \tag{2.132}$$

We define

$$\begin{aligned} a_k &= \lambda_k^1 - \lambda_k^2 && (2.133) \\ s_k &= \sum_{l=1}^{k} a_l && (2.134) \\ s_{n_T} &= \sum_{k=1}^{n_T} a_k = 0 = s_0. && (2.135) \end{aligned}$$

Therefore, it holds $s_k \geq 0$ for all $1 \leq k \leq n_T$. We can reformulate (2.132) and obtain

$$\sum_{l=1}^{n_T - 1} s_l (b_l(t) - b_{l+1}(t)) \geq 0 \tag{2.136}$$

with

$$b_l(t) = \mathbb{E}\left(p_l(t) \tilde{\boldsymbol{w}}_l^H \phi^{[1]}(\boldsymbol{A}) \tilde{\boldsymbol{w}}_l\right). \tag{2.137}$$

The inequality in (2.136) is fulfilled if

$$b_l(t) \geq b_{l+1}(t). \tag{2.138}$$

The term b_l in (2.137) is related to α_l from (2.44) by

$$b_l(t) = \frac{p_l(t)}{\lambda_l(t)} \alpha_l(t).$$

As a result, we obtain the sufficient condition for the monotony of the parametrised average performance $f(t)$

$$\frac{p_l(t)}{\lambda_l(t)} \geq \frac{p_{l+1}(t)}{\lambda_{l+1}(t)}. \tag{2.139}$$

In order to show the necessity of the condition, we argue that if the optimal power allocation does not fulfill the condition then the first derivative of the parameterised average performance in (2.122) is not smaller than or equal to zero and therefore, the average performance is not Schur-convex. This completes the proof.

□

2.4.12 Sketch of Proof of Theorem 5

We give here only the sketch of the proof, because it does not provide new insights but is rather a repetition of technicalities already used. The proof follows from the steps in the proof of Lemma 6 and the proof of Theorem 3. Here, the average performance is parameterised with respect to two arbitrary receive correlation vectors. The first derivative of the parameterised average performance with respect to the linear combination parameter t consists of two terms. The second term arises from the power allocation and depends on the transmit *and* receive correlation as well. Following the derivation in the proof of Lemma 6 it can be shown that the second term vanishes. The first term directly corresponds to $b_1(t)$ and $b_2(t)$ in the proof of Theorem 3 and it can be shown that the first derivative is always smaller than or equal to zero.

2.4.13 Proof of Lemma 7

For the proof of optimality of uncorrelated receivers, we write the channel capacity with covariance knowledge as

$$\begin{aligned} C &= \max_{\mathrm{tr}(\boldsymbol{\Lambda})=P} \mathbb{E} \log\det\left(\boldsymbol{I} + \rho \boldsymbol{D}_R^{\frac{1}{2}} \boldsymbol{W} \boldsymbol{D}_T^{\frac{1}{2}} \boldsymbol{\Lambda} \boldsymbol{D}_T^{\frac{1}{2}} \boldsymbol{W}^H \boldsymbol{D}_R^{\frac{1}{2}}\right) \\ &= \max_{\mathrm{tr}(\boldsymbol{\Lambda})=P} \mathbb{E} \log\det\left(\boldsymbol{I} + \rho \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{D}_T^{\frac{1}{2}} \boldsymbol{W}^H \boldsymbol{D}_R \boldsymbol{W} \boldsymbol{D}_T^{\frac{1}{2}} \boldsymbol{\Lambda}^{\frac{1}{2}}\right), \end{aligned}$$

by reciprocity. For fixed $\boldsymbol{D}_T$ and $\boldsymbol{\Lambda}$, the optimisation problem for $\boldsymbol{D}_R$ is the same as for transmit covariance matrix without channel state information. The result in [Tel99] can be applied for this scenario, too.

$$\boldsymbol{D}_R^{opt} = \arg \max_{\mathrm{tr}(\boldsymbol{D}_R)=n_R} \mathbb{E} \log\det\left(\boldsymbol{I} + \rho \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{D}_T^{\frac{1}{2}} \boldsymbol{W}^H \boldsymbol{D}_R \boldsymbol{W} \boldsymbol{D}_T^{\frac{1}{2}} \boldsymbol{\Lambda}^{\frac{1}{2}}\right).$$

For this case, we know already that the optimal strategy is to transmit in all direction with equal power. Hence, the largest capacity is achieved with uncorrelated receive antennas $\boldsymbol{D}_R^{opt} = \boldsymbol{I}$.

□

2.4.14 Proof of Theorem 6

The proof follows the same line as the proof of Lemma 6 which is proven in appendix 2.4.11. We start with the necessary and sufficient condition for Schur-convexity in (2.139).

$$\frac{p_l}{\lambda_l} \geq \frac{p_{l+1}}{\lambda_{l+1}}. \tag{2.140}$$

We show that the condition in (2.140) is always fulfilled by the optimum $\boldsymbol{p}$ for all correlations $\boldsymbol{\lambda}$. The necessary and sufficient condition for the optimal $\boldsymbol{p}$ is that for active $p_l > 0$ and $p_{l+1} > 0$ it holds

$$\alpha_l - \alpha_{l+1} = 0,$$

i.e.

$$\int_0^\infty e^{-t} f(t) \frac{\lambda_l}{1+\rho t \lambda_l p_l} dt - \int_0^\infty e^{-t} f(t) \frac{\lambda_{l+1}}{1+\rho t \lambda_{l+1} p_{l+1}} dt \quad = \quad 0 \qquad (2.141)$$

with

$$f(t) = \prod_{k=1}^{n} \frac{1}{1+\rho t \lambda_k p_k}.$$

From (2.141) it follows that

$$\int_0^\infty e^{-t} f(t) g_l(t) \left(\lambda_l - \lambda_{l+1} - (\rho t \lambda_{l+1} \lambda_l)(p_l - p_{l+1})\right) dt = 0$$

with

$$g_l(t) = (1+\rho t \lambda_l p_l)^{-1} (1+\rho t \lambda_{l+1} p_{l+1})^{-1}.$$

This gives

$$\int_0^\infty e^{-t} f(t) g_l(t) \left(\frac{\lambda_l - \lambda_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \lambda_l \lambda_{l+1}} - t \right) dt = 0$$

and

$$\frac{\lambda_l - \lambda_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \lambda_l \lambda_{l+1}} \int_0^\infty e^{-t} f(t) g_l(t) dt - \int_0^\infty e^{-t} f(t) g_l(t) t dt = 0. \qquad (2.142)$$

Note the following facts about the functions $f(t)$ and $g_l(t)$

$$\begin{aligned} g_l(t) &\geq 0 \quad \forall\, 0 \leq t \leq \infty \quad\quad f(t) \geq 0 \quad \forall\, 0 \leq t \leq \infty \\ \frac{dg_l(t)}{dt} &\leq 0 \quad \forall\, 0 \leq t \leq \infty \quad\quad \frac{df(t)}{dt} \leq 0 \quad \forall\, 0 \leq t \leq \infty. \end{aligned} \qquad (2.143)$$

By partial integration we obtain the following inequality

$$\begin{aligned} \int_0^\infty f(t) g_l(t) (1-t) e^{-t} dt &= \left(f(t) g_l(t) t e^{-t} \right)_{t=0}^{\infty} \\ &\quad - \int_0^\infty \frac{d(f(t) g_l(t))}{dt} t e^{-t} dt \geq 0. \end{aligned} \qquad (2.144)$$

From (2.144) and the properties of $f(t)$ and $g_l(t)$ in (2.143) follows that

$$\int_0^\infty e^{-t} f(t) g_l(t) dt \geq \int_0^\infty t e^{-t} f(t) g_l(t) dt.$$

Now we can lower bound the equality in (2.142) by

$$\begin{aligned} 0 &= \frac{\lambda_l - \lambda_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \lambda_l \lambda_{l+1}} \int_0^\infty e^{-t} f(t) g_l(t) dt - \int_0^\infty e^{-t} f(t) g_l(t) t dt \\ &\geq \frac{\lambda_l - \lambda_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \lambda_l \lambda_{l+1}} - 1. \end{aligned} \qquad (2.145)$$

From (2.145) follows

$$1 \geq \frac{\lambda_l - \lambda_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \lambda_l \lambda_{l+1}}$$

and further on

$$\lambda_l - \lambda_{l+1} \leq (p_l - p_{l+1}) \rho \lambda_l \lambda_{l+1}. \tag{2.146}$$

From (2.146) follows

$$\lambda_l (1 - \rho \lambda_{l+1} p_l) \leq \lambda_{l+1} (1 - \rho \lambda_l p_{l+1})$$

and finally

$$\rho \lambda_{l+1} p_l \geq \rho \lambda_l p_{l+1}. \tag{2.147}$$

From (2.147) follows the inequality in (2.140). This completes the proof.

□

3 Optimal transmission strategies for multiple antenna multiple access and broadcast channels

3.1 Introduction

In wireless point-to-point links, one applies multiple antennas to increase the spectral efficiency and the performance of wireless systems. In multiuser scenarios, multiple antennas at the base or even at the mobiles require the development of new transmission strategies in order to achieve the benefits of using the spatial domain. In multiple input multiple output (MIMO) multiple access channels (MAC), the optimum transmission strategy depends on the objective function, the power constraints, the channel realisation, and the SNR range.

The optimisation problems are divided into two classes: In one class, the objective function measures a global performance criteria of the system. In order to increase the average throughput of the MIMO MAC, the ergodic sum capacity can be maximised or the average normalised mean-square error can be minimised. The solution of this class' optimisation problems leads to transmission strategies which can be quite unfair for some users. If they experience poor channel conditions for long periods of time, they are not allowed to transmit. Therefore, the other class of optimisation problems deals with the fulfilment of rate, SINR, or MSE requirements with minimal power. In order to solve problems of this class it is necessary to understand the geometry of the achievable rate, SINR, or MSE region. In both classes of optimisation problems, the constraints can be either individual power constraints of each user or a sum power constraint. The second class of programming problems are non-convex non-linear programming problems which are notoriously complicate to analyse. The big number of degrees of freedom in the temporal as well as the spatial domain increases the number of parameters which can be controlled. In order to simplify the analysis, it is of advantage to divide the programming problem into parts which can be solved in an iterative fashion.

The analysis of multiuser MIMO systems is very important because usually more than one user is involved in cellular as well as ad-hoc systems. Up to now, only little has been found out about MIMO multiuser systems. The achievable rates and the transmission strategy depend on the following:

- *Structure of the wireless MIMO system:* In the common cellular approach, many mobiles share one base station which controls the scheduling and transmission strategies, e.g. power control in a centralised manner. In cellular systems the inter- and intracell interference can be controlled by spectrum and time allocation. In MIMO systems an additional dimension, namely the space, is available for allocation purposes.
- *Transmit strategies:* Obviously, the transmit strategies of the participating mobiles influence the achievable rate and the properties of the complete system. In turn, the transmit strategies depend on the type of channel state information (CSI) at the transmitter, i.e. the more CSI is known about the own channel as well as about the other users and the interference, the more adaptive and smart transmission strategies can be applied. If no CSI is available at the transmitter, it is best to use multiuser space-time (-spreading)

codes.

- *Receiver strategies:* Different decoding and detection strategies can be used at the receiver. The range leads from single-user detection algorithms which treat the other users a noise up to linear and even non-linear multiuser detection algorithms. Of course, the receiver architecture depends on the type of CSI, too.

- *System parameters:* In general, an important factor is the scenario in which the wireless system works. In home or office scenarios the system parameter heavily differ from parameters in public access, hot-spots, or high velocity scenarios. User parameters, resource parameters, and especially channel parameters have to be taken into account. The achievable performance and throughput depend on those system parameters.

The impact of interference in single-cell multiuser MIMO systems was studied in [BJ02g, BSJ03, BSJ02]. Joint processing lead to a set of optimal transmit covariance matrices which maximise the sum capacity. These results are only valid for perfect CSI at both sides of the link, and for successive interference cancellation (SIC) for the uplink or Costa precoding for the downlink. For independent decoding or precoding, the difficult optimisation problem was studied in [SB03]. Under the assumption of a multiuser MMSE receiver, the minimisation of the average sum MSE and the structure of the individual MSE region were analysed in [JB03g].

Our first approach in understanding a point-to-point link in a multiuser scenario is to treat the impact of interference as additional noise. The question here is, how much performance is lost by worst case interference? Obviously, the maximum performance loss depends on the constraints which are imposed on the noise plus interference. We model the impact of the mentioned effects on the system by a special noise covariance matrix and analyse the structure and the performance of the resulting MIMO channels. We do not assume an a priori structure of the interference. It could be the uplink or downlink transmission, the interference could be an intercell or an intracell interference. The receiver noise, the intercell, and the intracell interference restrict the achievable capacity. We will open this section by this worst case noise analysis because it provides a nice transition from single-user systems to multiuser systems. The performance function is the same as in the single-user scenario only the noise structure changes. Furthermore, this approach provides insights into the structure of multiuser multiple-antenna communication which is utilised in the succeeding sections.

In the second part of this chapter, we analyse sum performance optimisation problems. The development from the single-antenna MAC to the MIMO MAC is shown and the differences and the common ground between the single-antenna and the multiple-antenna cases are stressed. Furthermore, we focus on the connections between the different objective functions and their corresponding programming problems. We give a generalisation of the iterative waterfilling algorithm from [YRBC04, YRBC01] for arbitrary matrix-monotone performance functions. An iterative algorithm, which performs power allocation and iterative generalised waterfilling, maximises the sum performance of the multiuser MIMO channel. Finally, all theoretical results and algorithms are illustrated by numerical simulations.

In the SISO multiuser MAC it has been shown that the maximum sum capacity is achieved when only the best user is allowed to transmit [KH95] under a sum power constraint[1]. In [YRBC04], the authors maximise the ergodic sum capacity of the

[1] The sum power constraint can be either a long-term sum power constraint or a short-term power constraint. In this work, we are considering only short-term power constraints, because the

MIMO MAC for fixed individual power constraints for the transmit covariance matrices. It was shown that the optimal transmit covariance matrices are characterised by an iterative water-filling solution which treats the other users like noise under individual power constraints. We study the sum performance of a Gaussian MIMO MAC and BC under individual and sum power constraints. The MIMO MAC model appears in the uplink transmission from multiple users to the multi-antenna base station. Each user is equipped with multiple antennas. Furthermore, we assume that the base station and all mobiles have perfect channel state information (CSI). Recently, the perfect CSI assumption was made for the MIMO MAC and the MIMO BC in [RC01, RC03, YRBC04, BJ02g] and [VJG02a, VT03]. In [RC01], the ergodic sum capacity of a multi-antenna Gaussian multiple-access channel is defined, and the impact of the number of transmit and receive antennas, as well as the number of users are analysed. The special case in which covariance information is only available at the mobiles is considered in [JG01a]. In [RC03], the ergodic capacity, the ergodic sum rate region, the outage capacity, and algorithms for the vector MAC are analysed.

The system model, which is dual to the MIMO MAC, is the multiuser MIMO downlink transmission. It leads to the BC [Cov72, Cov98], the multi-antenna Gaussian non-degraded broadcast channel [CS01], or vector broadcast channel [YC01]. Recently, the sum capacity and achievable region of the multiuser MIMO BC were studied in [VJG02a, VJG02b] and [VT03]. An upper bound on the capacity region of the BC was derived in [Sat78]. The bound is found by computing the capacity of the cooperative system under worst case noise. The structure of the worst case noise for the MIMO BC is analysed in [Yu03]. In [VJG02b], [BS02a], and [TV02], the duality between the multiuser uplink and downlink channel was studied. It was shown that the achievable capacity region of the downlink transmission collapses with the capacity region of the uplink. In addition to this, the maximum sum rate point on the envelope of the capacity region can be characterised by the capacity of the equivalent cooperative MIMO system with the worst case noise [Sat78]. It has to be stressed that the capacity region of the non-degraded BC is still an open research problem. In [VKS$^+$03], the authors show that all points below their proposed upper bound are achievable under the assumption that Gaussian code books are optimal. This assumption has been derived in [WSS04] and it has been shown that the capacity region of the nondegraded MIMO BC equals that of the MIMO MAC.

We proceed as follows: First, the worst case noise performance of MIMO systems is analysed. And next, the sum performance and individual performance measures are optimised with respect to individual or sum power constraints.

3.2 Worst Case Noise Analysis in Multiuser MIMO Systems

3.2.1 Motivation and related results

Let us assume, that we have a two-player game in which the transmit player wants to increase the performance of the point-to-point link, while the noise plus interference player wants to reduce the performance. The transmit player is first. He perfectly knows the channel realisation as well as the noise covariance matrix. The noise plus interference player is second. This game leads to the a minimax problem of the

extension to a long-term constraint is straight forward and lead to an additional water-filling across the temporal dimension. All our results can be further improved by this technique. The analysis in [KH95] is under a long-term power constraint.

following type

$$\min_{\boldsymbol{Z}\in\mathcal{Z}} \max_{\boldsymbol{Q}\in\mathcal{Q}} \text{ tr } F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2}) \tag{3.1}$$

in which the noise $\boldsymbol{Z}$ is in some set of admissible noise or interference $\mathcal{Z}$ and the transmit strategy $\boldsymbol{Q}$ belongs to some set of admissible transmit strategies $\mathcal{Q}$ and F is a matrix monotone function. The matrix $\boldsymbol{H}$ is fixed and depends on the transmission medium. If the AWGN channel is studied, $\boldsymbol{H} = \boldsymbol{I}$. Otherwise, the matrix $\boldsymbol{H}$ can be a flat-fading MIMO channel matrix, or a frequency-selective channel matrix, etc. In general, we assume that the transmit strategy is power limited, i.e. $\mathcal{Q} = \{\boldsymbol{Q} : \text{ tr } (\boldsymbol{Q}) \leq P\}$. However, the noise $\boldsymbol{Z}$ can be created by a variety of effects, e.g. thermal noise, intercell-, intracell-, inter-symbol-, or inter-carrier-interference. The concrete structure of $\boldsymbol{Z}$ depends on the application, transmit- and receive strategies. We do not determine a concrete scenario or strategy but consider three noise scenarios which are of general interest. Furthermore, we only restrict the objective function to be matrix-monotone. In section 3.2.3, concrete examples from wireless transmission illustrate the application of the theoretical results.

The vector channel capacity [Tel99] or the MSE can be used direclty as an objective function because both are traces of matrix-monotone functions. In [VBG03], expressions like (3.1), in which F was the channel capacity, were studied under different admissible sets $\mathcal{Z}$ and $\mathcal{Q}$. In order to characterise the minimax points, the authors of [VBG03] used the dual Lagrangian approach. Based on the characterisation of the broadcast rate region in [Sat78], the authors in [VT03] established a duality and reciprocity theory between the SIMO multiple access sum capacity point, MIMO uplink capacity, MIMO downlink capacity and MISO broadcast sum capacity point for systems which apply SIC (uplink) and Costa precoding (downlink). The duality between the SIMO MAC sum capacity point and MIMO uplink capacity corresponds to the noise constraints which will be given in Scenario III. For independent coding and decoding, the uplink - downlink duality for SIMO was shown in [BS02b]. In [Yu03], the MIMO broadcast channel was studied, and the structure of the worst case noise of the corresponding cooperative MIMO system which minimises the Sato upper bound was analysed.

In [Blu03], the author analyses the MIMO channel capacity with unknown interfering users and no CSI at the transmitter as well as perfect CSI at the receiver. The optimum signalling for achieving the channel capacity is characterised by analysing the second derivative of the mutual information and by showing that in some cases it is negative and in some cases it is positive. Therefore, in these cases in which the interference is sufficiently weak or sufficiently strong, the optimum signalling is either equal power allocation across all antennas or single-antenna allocation. In [YB03], the mutual information of a MIMO system with multiple users and perfect CSI at both sides is considered and different signalling approaches are considered.

A minimax approach in [PCL03b] studies the maximum of the mutual information with respect to the transmit covariance matrix and the minimum with respect to the channel realization of the instantaneous capacity in a flat-fading MIMO channel. In addition to this, the worst case capacity of a MIMO system is studied in [GHIM01].

3.2.2 Noise scenarios

We are interested in the general limits of the MIMO channel performance under different types of noise plus interference. Therefore, we consider three scenarios in which the noise is subject to different constraints. In the first scenario the noise covariance matrix is trace constrained. This is the less restrictive constraint, because only the sum noise power is kept fixed and their eigenvalues and diagonal entries are free to choose. This leads to the worst case noise under a trace constraint. In the second scenario the eigenvalues of the noise covariance matrix are fixed. This leads to the notion of worst case directions. These are the eigenvectors of the noise covariance matrix $\boldsymbol{Z}$. Finally, in the third scenario the diagonal elements of the noise covariance matrix are fixed. This yields the worst case coloured noise. We show for all three scenarios that the achievable minimax performance fulfils the saddle point property. Furthermore, the worst case noise in Scenario I and Scenario II leads to two different types of worst case orthogonal channels. In Scenario I the complete CSI at the transmitter is lost and therefore the cooperation, too, because CSI is necessary for successful cooperation at the transmit side. In Scenario III, the performance of the MIMO channel with worst case coloured noise equals the capacity of a multiuser SIMO channel with white noise, i.e. the transmitter cooperation gets lost.

1. *Trace Constraint:* The trace of the noise covariance matrix is constraint to $n_R \cdot \sigma_N^2$, i.e.

$$\mathrm{tr}\,(\boldsymbol{Z}) \leq \sigma_N^2 n_R. \tag{3.2}$$

   In this scenario, the sum noise power which arrives at the base station is kept fixed. The noise has no additional constraints. This model corresponds to a scenario in which the inter- and intracell interference dominates. The noise has fewest constraints in comparison to the other scenarios. We denote the noise in (3.2) as *worst case noise under trace constraints*.

2. *Fixed Eigenvalues:* The eigenvalues of the noise covariance matrix are fixed. The diagonal matrix

$$\boldsymbol{\Lambda}_Z = \mathrm{diag}(\lambda_1(\boldsymbol{Z}), ..., \lambda_m(\boldsymbol{Z}))$$

   is fixed.

   Here, the average power (eigenvalues of noise covariance matrix $\boldsymbol{\Lambda}_Z$) is fixed, while the dominant directions of the noise (eigenvectors of the noise covariance matrix $\mathbf{U}$) vary. This constraint leads to the *worst case noise directions*.

3. *Diagonal constraint:* The diagonal of the noise covariance matrix is constraint to be less or equal to some constant σ_N^2, i.e.

$$\mathrm{diag}(\boldsymbol{Z}) = [\sigma_N^2, ..., \sigma_N^2].$$

   In this scenario, we fix the noise power at each receive antenna at the base station. The receiver noise at each receive antenna has equal power but the intra- and intercell interference creates the correlation or the colour of the noise. The free parameter is the correlation of the noise. This scenario provides the *worst case coloured noise*.

3.2.3 Applications

The analysis in this work can be applied to the following MIMO MAC scenario. In addition to this, it can be applied to MIMO BC by duality.

MIMO MAC

We consider the typical flat-fading MIMO MAC model with n_T transmit and n_R receive antennas in cellular multiuser uplink transmission. In figure (3.1), we show an uplink transmission from user (U) to the base (B). On the one hand, intercell interference comes from neighbour cells (K) and on the other hand user in the same cell create intracell interference (S).

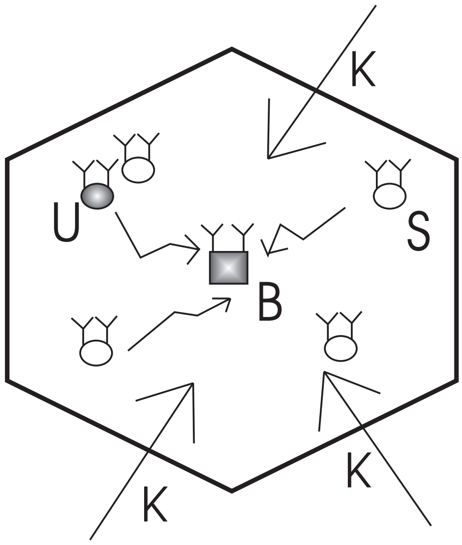

Figure 3.1: Cellular MIMO multiuser uplink.

Following the flat block-fading cellular MIMO multiuser uplink model considered above, the received signal is given by

$$\boldsymbol{y} = \underbrace{\boldsymbol{H}\boldsymbol{x}}_{\text{user signal}} + \underbrace{\sum_{k \in \mathcal{S}} \boldsymbol{H}_k \boldsymbol{x}_k}_{\text{intracell interference}} + \underbrace{\sum_{k \in \mathcal{K}} \boldsymbol{H}_k \boldsymbol{x}_k}_{\text{intercell interference}} + \boldsymbol{n} \tag{3.3}$$

with white Gaussian noise $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma_N^2 \boldsymbol{I})$. We collect all noise and interference terms in one vector $\boldsymbol{z}$ and obtain

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{z} \tag{3.4}$$

where $\boldsymbol{x}$ is the transmitted signal, $\boldsymbol{H}$ is the channel matrix, $\boldsymbol{z}$ is the interference plus noise. The mobile has n_T transmit antennas and the base n_R receive antennas. The number of transmit antennas is assumed to be equal to or smaller than the number of receive antennas, i.e. $n_T \leq n_R$. The channel matrices and signals of the intercell and intracell interfering users can have arbitrary number of transmit dimensions.

We assume that the interference plus noise is complex Gaussian distributed with covariance matrix $\boldsymbol{Z}$. This assumption is motivated by the law of large numbers for a large number of interfering users. Furthermore, the zero-mean complex Gaussian distribution is the worst case noise distribution under a variance constraint. Therefore, we model the interference plus noise as zero-mean complex Gaussian distributed with covariance matrix $\boldsymbol{Z}$, i.e. $\boldsymbol{z} \sim \mathcal{CN}(0, \boldsymbol{Z})$. Note that the noise covariance matrix has always full rank, otherwise the noise-free dimension could be used for transmission and we would have arbitrary high rate.

Then the optimum input distribution which maximises the capacity of the channel in (3.4) is the zero-mean complex Gaussian distribution, too, i.e. $\boldsymbol{x} \sim \mathcal{CN}(0, \boldsymbol{Q})$. The transmit covariance matrix is given by $\boldsymbol{Q} = \mathcal{E}(\boldsymbol{x}\boldsymbol{x}^H)$. The mutual information of the channel in (3.4) is [Tel99]

$$C(\boldsymbol{Q}, \boldsymbol{Z}) = \log \frac{\det(\boldsymbol{Z} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H)}{\det(\boldsymbol{Z})}. \tag{3.5}$$

We assume that the sum transmit power is constrained to P, i.e. $tr(\boldsymbol{Q}) \leq P$.

MIMO BC

We consider the typical flat-fading MIMO BC model with n_T transmit and n_R receive antennas in cellular multiuser downlink transmission. In figure (3.2), we show an downlink transmission from the base (B) to user (U). The base might simultaneously transmit to the other user in the cell (S). On the one hand, intercell interference comes from neighbour cells (K) and on the other hand user in the same cell create intracell interference (S).

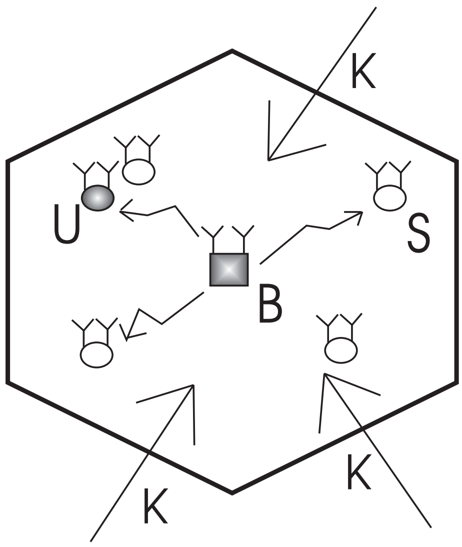

Figure 3.2: Cellular MIMO multiuser downlink

Following the flat block-fading cellular MIMO multiuser downlink model considered above, the received signal is given by

$$\boldsymbol{y} = \underbrace{\boldsymbol{H}\boldsymbol{x}}_{\text{user signal}} + \underbrace{\sum_{k \in \mathcal{S}} \boldsymbol{H}\boldsymbol{x}_k}_{\text{intracell interference}} + \underbrace{\sum_{k \in \mathcal{K}} \boldsymbol{H}\boldsymbol{x}_k}_{\text{intercell interference}} + \boldsymbol{n} \tag{3.6}$$

with white Gaussian noise $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma_N^2 \boldsymbol{I})$. We collect all noise and interference terms in one vector $\boldsymbol{z}$ and obtain

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{z} \tag{3.7}$$

where $\boldsymbol{x}$ is the transmitted signal, $\boldsymbol{H}$ is the channel matrix, $\boldsymbol{z}$ is the interference plus noise. The mobile has n_T transmit antennas and the base n_R receive antennas.

The channel matrix and signals of the intercell and intracell interfering users can have arbitrary number of transmit dimensions.

Under the same assumptions as in the MIMO MAC case above, we arrive at the mutual information of the channel in (3.7)

$$C(\boldsymbol{Q}, \boldsymbol{Z}) = \log \frac{\det(\boldsymbol{Z} + \boldsymbol{HQH}^H)}{\det(\boldsymbol{Z})}.$$

We assume that the sum transmit power is constrained to P, i.e. $tr(\boldsymbol{Q}) \leq P$, too.

Examples for performance metrics

Examples for performance metrics of the point-to-point link which belong to the large class of trace of matrix-monotone functions are:

- Channel Capacity: The capacity of vector system

  $$C(\boldsymbol{Q}, \boldsymbol{Z}) = \log \left(\frac{\det\left(\boldsymbol{Z} + \boldsymbol{HQH}^H\right)}{\det \boldsymbol{Z}} \right) \tag{3.8}$$

  can be transformed into

  $$\text{tr } F(\boldsymbol{Z}^{-1/2}\boldsymbol{HQH}^H\boldsymbol{Z}^{-1/2}) = \text{ tr } \log\left(\boldsymbol{I} + \boldsymbol{Z}^{-1/2}\boldsymbol{HQH}^H\boldsymbol{Z}^{-1/2}\right).$$

  The capacity $C(\boldsymbol{Q}, \boldsymbol{Z})$ in (3.8) is convex with respect to $\boldsymbol{Z}$ [DC01, Lemma II.3] and concave with respect to $\boldsymbol{Q}$ [CT88, Theorem 1]. The capacity of MIMO MAC, BC with worst case noise is analysed in [BJ03b].

- Normalised MSE: The normalised MSE

  $$G(\boldsymbol{Z}, \boldsymbol{Q}) = n_T - tr\left(\left[\boldsymbol{Z} + \boldsymbol{HQH}^H\right]^{-1} \boldsymbol{HQH}^H\right)$$

  can be transformed into

  $$\text{tr } F(\boldsymbol{Z}^{-1/2}\boldsymbol{HQH}^H\boldsymbol{Z}^{-1/2}) = n_T + \text{ tr } \left(\left[\boldsymbol{I} + \boldsymbol{Z}^{-1/2}\boldsymbol{HQH}^H\boldsymbol{Z}^{-1/2}\right]^{-1}\right).$$

  The MSE is to be minimized or $-MSE$ is to be maximized. $-G$ is a concave function with respect to $\boldsymbol{Q}$ and convex function with respect to $\boldsymbol{Z}$, too. M is a constant which depends on the system parameter, e.g. the number of antennas.

3.2.4 Preliminaries

In order to characterise the impact of $\boldsymbol{Z}$ and $\boldsymbol{Q}$ on the performance metric in (3.1), note that tr $F(\boldsymbol{Z}^{-1/2}\boldsymbol{HQH}^H\boldsymbol{Z}^{-1/2})$ is concave with respect to $\boldsymbol{Q}$. In addition to this, we show in the following that tr $F(\boldsymbol{Z}^{-1/2}\boldsymbol{HQH}^H\boldsymbol{Z}^{-1/2})$ is convex with respect to $\boldsymbol{Z}$. We define $\boldsymbol{M} = \boldsymbol{HQH}^H$ for convenience and obtain

$$F(\boldsymbol{Z}^{-1/2}\boldsymbol{MZ}^{-1/2}) = \int_0^\infty s\boldsymbol{Z}^{-1/2}\boldsymbol{MZ}^{-1/2}\left(s\boldsymbol{I} + \boldsymbol{Z}^{-1/2}\boldsymbol{MZ}^{-1/2}\right)^{-1} d\mu(s). \tag{3.9}$$

The representation in (3.9) follows from Löwners Theorem. The same assumptions as in section 2.3.1 are valid. Applying the trace operator on (3.9) and swap integration and finite summation, we obtain

$$\mathrm{tr}\ \left(F(\boldsymbol{Z}^{-1/2}\boldsymbol{M}\boldsymbol{Z}^{-1/2})\right) = \int_0^\infty s\ \mathrm{tr}\ \left(\boldsymbol{M}\left(s\boldsymbol{Z}+\boldsymbol{M}\right)^{-1}\right) d\mu(s). \tag{3.10}$$

Define the parameterised noise matrix $\boldsymbol{Z}(\lambda) = (1-\lambda)\boldsymbol{Z}_1 + \lambda\boldsymbol{Z}_2$. From [MO79, Proposition 16.E.7.C] follows that $\boldsymbol{A}^{-1}$ is matrix-convex and therefore we have for (3.9) and (3.10) that

$$\begin{aligned} &-\ \mathrm{tr}\ \boldsymbol{M}\left(s\boldsymbol{Z}(\lambda) +\ \mathrm{tr}\ \boldsymbol{M}\right)^{-1} +\ \mathrm{tr}\ (1-\lambda)\boldsymbol{M}\left(s\boldsymbol{Z}_1+\boldsymbol{M}\right)^{-1} \\ &+\ \mathrm{tr}\ \lambda\boldsymbol{M}\left(s\boldsymbol{Z}_2+\boldsymbol{M}\right)^{-1} \geq 0. \end{aligned} \tag{3.11}$$

And finally with inequality (3.11), it holds

$$\begin{aligned} &\mathrm{tr}\ \left(F\left(\boldsymbol{Z}^{-1/2}(\lambda)\boldsymbol{M}\boldsymbol{Z}^{-1/2}(\lambda)\right)\right) \\ \leq (1-\lambda)\ \mathrm{tr}\ &\left(F\left(\boldsymbol{Z}_1^{-1/2}\boldsymbol{M}\boldsymbol{Z}_1^{-1/2}\right)\right) + \lambda\ \mathrm{tr}\ \left(F\left(\boldsymbol{Z}_2^{-1/2}\boldsymbol{M}\boldsymbol{Z}_2^{-1/2}\right)\right). \end{aligned}$$

This shows that the trace of every matrix-monotone function

$$\mathrm{tr}\ F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2})$$

is convex with respect to $\boldsymbol{Z}$ and concave with respect to $\boldsymbol{Q}$. The set of all admissible matrices $\boldsymbol{Q}$ with trace constraint is convex. The set of noise matrices $\boldsymbol{Z}$ in Scenario I and III is convex, too. For the minimax problem in (3.1) follows, that the minimax problem with noise constraint I and III fulfils the Saddle-point property. In order to decide whether the min-max expressions satisfy the saddle-point property

$$\min_{x\in X}\max_{y\in Y} f(x,y) = \max_{y\in Y}\min_{x\in X} f(x,y) \tag{3.12}$$

we use Theorem 1 in [Fan53]. One result in [Fan53, Theorem 1] states, that (3.12) is fulfilled, if f is convex on x and concave on y and if the sets X and Y are convex, too.

In order to characterise the solution of minimax problems of type (3.1), we need the following theorem which characterises the worst case noise under trace constraint.
Theorem 7: For positive semidefinite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with eigenvalues $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq ... \geq \beta_n$ and arbitrary matrix-monotone function F it holds

$$\begin{aligned} \min_\pi\ \mathrm{tr}\ F(\mathrm{diag}(\alpha_1, ..., \alpha_n)\mathrm{diag}(\beta_{\pi_1}, ..., \beta_{\pi_n})) \leq\ &\mathrm{tr}\ F(\boldsymbol{B}^{1/2}\boldsymbol{A}\boldsymbol{B}^{1/2}) \leq \\ &\max_\pi\ \mathrm{tr}\ F(\mathrm{diag}(\alpha_1, ..., \alpha_n)\mathrm{diag}(\beta_{\pi_1}, ..., \beta_{\pi_n})) \end{aligned}$$

with permutation π.

The proof of Theorem 7 can be found in appendix 3.4.1. Theorem 7 is a generalisation of [Fie71].

3.2.5 Worst case noise with trace constraint

In this section, the worst case noise with the trace constraint from scenario 1 is characterised. The optimisation problem for the scenario 1 is given by

$$\Phi_I = \min_{\operatorname{tr}(\boldsymbol{Z})\leq\sigma_N^2 n_R} \max_{\operatorname{tr}(\boldsymbol{Q})\leq P} \operatorname{tr} F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2}). \tag{3.13}$$

The problem in (3.13) fulfils the saddle-point properties [Fan53]. Therefore we can switch the min-max problem into max-min, i.e.

$$\begin{aligned}&\min_{\operatorname{tr}(\boldsymbol{Z})\leq\sigma_N^2 n_R} \max_{\operatorname{tr}(\boldsymbol{Q})\leq P} \operatorname{tr} F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2}) = \\ &\quad \max_{\operatorname{tr}(\boldsymbol{Q})\leq P} \min_{\operatorname{tr}(\boldsymbol{Z})\leq\sigma_N^2 n_R} \operatorname{tr} F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2}).\end{aligned}$$

Let us describe the first result (Theorem 8): In Scenario I, the vector MIMO channel transforms into orthogonal channels in figure 3.3. The input streams are weighted by the eigenvalues of the transmit covariance matrix $\lambda_i(\boldsymbol{Q})^{1/2}$ then weighted by the channel matrix eigenvalues $\lambda_i(\boldsymbol{H})^{1/2}$. The additive noise $\mathcal{CN}(0,1)$ is weighted by minimal noise covariance matrix eigenvalues $\lambda_i(\boldsymbol{Z})^{1/2}$ which are computed according to (3.67). The vector MIMO channel with perfect CSI at the transmitter transforms into a MIMO channel without CSI and SNR $\rho = \frac{P}{n_R\sigma_N^2}$.

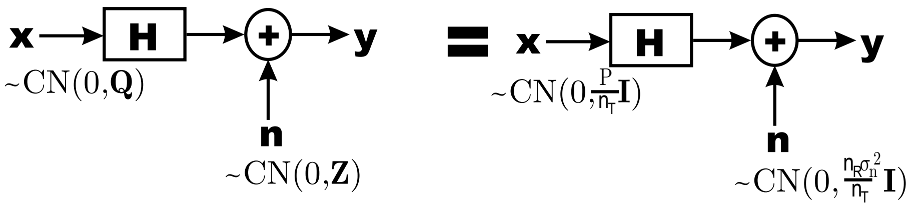

Figure 3.3: Worst Case Noise with Trace Constraint: Vector MIMO channel Φ_I with perfect CSI and worst case noise and corresponding vector MIMO channel without CSI and white noise (Theorem 8).

In figure (3.3), the correspondence is shown between the closed-loop MIMO system (transmit covariance matrix $\boldsymbol{Q}$) with worst case noise (noise covariance matrix $\boldsymbol{Z}$) with trace constraint and the open-loop MIMO system (transmit covariance matrix $\boldsymbol{I}$) with white noise with SNR ρ.

The following Theorem 8 solves the minimax problem in (3.13).
Theorem 8: The value of the minimax problem in (3.13) is given by

$$\Phi_I = \operatorname{tr} F(\rho\boldsymbol{H}\boldsymbol{H}^H) \tag{3.14}$$

with SNR $\rho = \frac{P}{n_R\sigma_N^2}$.

The proof can be found in appendix 3.4.2.

Remark 1: The RHS of equation (3.14) in Theorem 8 is the value of the objective function for uncorrelated white noise and equal power allocation divided by the dimension of the noise matrix. This surprising result of Theorem 8 is interpreted in the following way: The value of the trace of an arbitrary matrix-monotone objective function of an transmission system with perfect information and cooperation at the transmitter under worst case noise with trace constraint is equal to the value of the system with an uninformed transmitter and with uncorrelated white noise and SNR ρ.

Remark 2: The optimal transmit strategy can be further characterised: It is given by

$$\lambda_k^*(\boldsymbol{Q}) = \frac{\lambda_k(\boldsymbol{H})}{\mu}\rho^2 F^{[1]}(\rho\lambda_k(\boldsymbol{H})). \tag{3.15}$$

The Lagrangian μ is chosen such that $\sum_{k=1}^{n_T} \lambda_k^*(\boldsymbol{Q}) = 1$. As a result we have

$$\mu = \rho^2 \sum_{k=1}^{n_T} \lambda_k(\boldsymbol{H}) F^{[1]}(\rho\lambda_k(\boldsymbol{H})). \tag{3.16}$$

Equation (3.15) and μ in (3.16) yields

$$\lambda_k^*(\boldsymbol{Q}) = \frac{\lambda_k(\boldsymbol{H}) F^{[1]}(\rho\lambda_k(\boldsymbol{H}))}{\sum_{l=1}^{n_T} \lambda_l(\boldsymbol{H}) F^{[1]}(\rho\lambda_l(\boldsymbol{H}))}. \tag{3.17}$$

For small SNR values,

$$\lim_{\sigma_N^2 \to \infty} \lambda_i^*(\boldsymbol{Q}) = \frac{\lambda_i(\boldsymbol{H})}{\sum_{k=1}^{\nu} \lambda_k(\boldsymbol{H})} \tag{3.18}$$

because $F^{[1]}(0) = c$. This follows from Lemma 13 in appendix 3.4.5.

3.2.6 Worst case noise directions

We assume that the noise eigenvalues are fixed. We study the impact of the directions or eigenvectors of $\boldsymbol{Z}$, i.e. of the unitary matrix $\boldsymbol{U}_Z$. We write the set of unitary $n_R \times n_R$ matrices as $\mathcal{U}(n_R)$. Let us define the optimisation problem as

$$\Phi_{II} = \min_{\boldsymbol{W} \in \mathcal{U}(n_R)} \max_{\text{tr }(\boldsymbol{Q}) \leq P} \text{ tr } F(\boldsymbol{\Lambda}_Z^{-1/2} \boldsymbol{W}^H \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H \boldsymbol{W} \boldsymbol{\Lambda}_Z^{-1/2}). \tag{3.19}$$

Furthermore, we define

$$\Phi_{II}^D = \max_{\sum_{i=1}^n \lambda_i(\boldsymbol{Q}) \leq P} \sum_{i=1}^n F\left(\frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}{\lambda_i(\boldsymbol{Z})}\right). \tag{3.20}$$

Obviously, the solution in (3.20) is the waterfilling solution for $\boldsymbol{Q}$.

The result in this section (Theorem 9) is that Φ_{II} and Φ_{II}^D are equal. In figure (3.4), the correspondence between the closed-loop MIMO system with worst case noise directions with noise covariance matrix $\boldsymbol{Z}$ and the system with parallel SISO channels $\lambda_1(\boldsymbol{H}), ..., \lambda_{n_R}(\boldsymbol{H})$ and noise variances $\lambda_1(\boldsymbol{Z}), ..., \lambda_{n_R}(\boldsymbol{Z})$ is shown.

In Scenario II, the worst case directions de-construct the MIMO channel into n_R orthogonal channels. The capacity Φ_{II}^D in (3.20) can be easily computed.

We collect this result in the following theorem that states that the capacities in (3.19) and (3.20) are equal. Using Theorem 7, one can easily prove the following Theorem 9.
Theorem 9: The worst case noise directions $\boldsymbol{W}$ in (3.19), are given by a permutation matrix $\boldsymbol{\Pi}$. As a result, $\Phi_{II} = \Phi_{II}^D$.

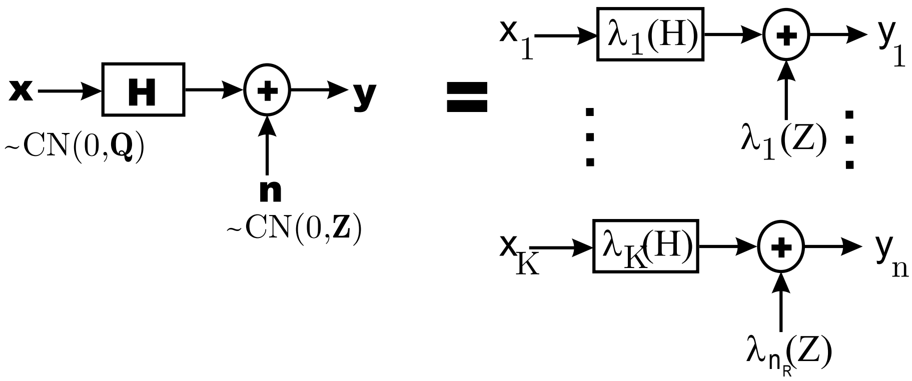

Figure 3.4: Worst Case Noise Directions: Vector MIMO channel Φ_{II} and corresponding diagonalised orthogonal channels Φ_{II}^D

At this point, the special case in which the performance metric is the instantaneous channel capacity has to be discussed. If $\Phi(\boldsymbol{X}) = \log(1+\boldsymbol{X})$ then a stronger version of Theorem 9 can be proven, because the worst case and best case permutation of the eigenvalues is fixed and does not vary with their values. We assume that the noise eigenvalues are fixed and ordered, i.e. $\lambda_1(\boldsymbol{Z}) \geq \lambda_2(\boldsymbol{Z}) \geq ... \geq \lambda_n(\boldsymbol{Z})$. Let us define the optimisation problem as

$$C_{II} = \min_{\boldsymbol{W}\in\mathcal{U}(n)} \max_{\text{tr }(\boldsymbol{Q})\leq P} \log \frac{\det(\boldsymbol{W}\boldsymbol{\Lambda}_Z\boldsymbol{W}^H + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H)}{\det(\boldsymbol{W}\boldsymbol{\Lambda}_Z\boldsymbol{W}^H)}. \tag{3.21}$$

Furthermore, we define

$$C_{II}^D = \max_{\sum_{i=1}^n \lambda_i(\boldsymbol{Q})\leq P} \sum_{i=1}^n \log\left(1 + \frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}{\lambda_i(\boldsymbol{Z})}\right). \tag{3.22}$$

Obviously, the solution in (3.22) is the waterfilling solution.
Theorem 10: The capacity C_{II} in (3.21) and the capacity C_{II}^D in (3.22) are equal.

The proof can be found in appendix 3.4.6.

3.2.7 Worst case coloured noise

In this scenario, the diagonal entries of the noise covariance matrix are equal to σ_N^2. We define the set of all noise covariance matrices with constant σ_N^2 entries on the diagonal as

$$\mathcal{Z} = \{\boldsymbol{Z} : \boldsymbol{Z} \succeq 0, \text{diag}(\boldsymbol{Z}) = [\sigma_N^2 ... \sigma_N^2]\}. \tag{3.23}$$

We define the value of the minimax problem as

$$\Phi_{III} = \min_{\boldsymbol{Z}\in\mathcal{Z}} \max_{\text{tr }(\boldsymbol{Q})\leq P} \text{tr } F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2}). \tag{3.24}$$

Furthermore, we define the following programming problem

$$\Phi_{III}^D = \max_{\sum_{k=1}^{n_T} p_k = P} \text{tr } F(\boldsymbol{H}\text{Diag}([p_1, ..., p_{n_T}])\boldsymbol{H}^H). \tag{3.25}$$

The next Theorem shows that the values of the optimisation problems in (3.24) and (3.25) are equal.
Theorem 11: The value of (3.24) equals the value of (3.25), i.e. $\Phi_{III} = \Phi^D_{III}$.

The proof of Theorem 11 can be found in appendix 3.4.7. The transmitter looses its cooperation, i.e. the matrix $\boldsymbol{Q}$ cannot be chosen arbitrarily. However, the diagonal entries can be controlled. The worst case coloured noise system has equal value as a system without cooperation and only power control and white uncorrelated noise.

The result of this section (Theorem 11) is that Φ_{III} and Φ^D_{III} are equal. Let us first describe the result: In figure (3.5), the correspondence between the closed-loop MIMO system with worst case coloured noise and the SIMO MAC with white noise is shown. In Scenario III, there is cooperation at the transmit side only in terms of power control.

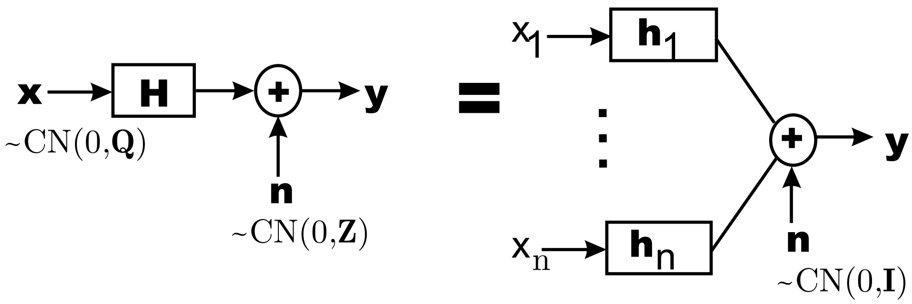

Figure 3.5: Worst Case Coloured Noise: Vector MIMO channel Φ_{III} and corresponding SIMO MAC Φ^D_{III}.

The worst case colour of the noise reduces the achievable performance of a MIMO system with n_T cooperating transmit antennas to n_T users who perform only power control. The achievable performance Φ_{III} for the MIMO channel with worst case coloured noise equals the sum performance of the multiuser SIMO MAC. This fact has been used in [Sat78] in the context of capacities to derive an upper bound on the capacity region of the broadcast channel.

Example for worst case coloured noise

In the following, we give two examples for the computation of the worst case coloured noise saddle point by solving the simple SIMO MAC problem. The performance metric is the capacity, i.e. $F(\boldsymbol{X}) = \log(\boldsymbol{I} + \boldsymbol{X})$. In the first example both users are supported. In the second example, only one mobile user.

Example A

Let the channel matrix given by

$$\boldsymbol{H} = \begin{pmatrix} 0.1 & 0.5 \\ 0.8 & 0.2 \end{pmatrix}$$

with transmit power constraint $P = 10$ and noise power $\sigma_N^2 = 1$. We apply the following steps:

1. We compute the optimal power allocation for the SIMO MAC by MAXDET [WVB96]

$$\boldsymbol{S}^* = \text{diag}([3.5457, 6.4543]).$$

The Lagrangian multiplier is $\lambda_1 = \lambda_2 = 0.124$.

2. We compute the worst case coloured noise in (3.82). The corresponding KKT condition yields

$$\boldsymbol{Z}^* = \begin{pmatrix} 1 & 0.151 \\ 0.151 & 1 \end{pmatrix}.$$

3. Waterfilling with respect to the effective channel provides the optimal MIMO transmission strategy:

$$\boldsymbol{Q}^* = \begin{pmatrix} 6.2675 & 1.0151 \\ 1.0151 & 3.7325 \end{pmatrix}$$

Next, we verify this result by computing the sum capacity of the SIMO MAC with $\boldsymbol{S}^*$ and by computing the capacity of the MIMO channel with worst case coloured noise $\boldsymbol{Z}^*$ and transmit strategy $\boldsymbol{Q}^*$:

$$\begin{aligned} C_{III}^D &= \log\det\left(\boldsymbol{I} + \boldsymbol{H}^H \boldsymbol{S}^* \boldsymbol{H}\right) = 3.2653. \\ C_{III} &= \log\det\left(\boldsymbol{Z}^* + \boldsymbol{H}\boldsymbol{Q}^*\boldsymbol{H}^H\right) - \log\det \boldsymbol{Z} = 3.2653. \end{aligned}$$

And we have

$$C_{III}^D = C_{III}.$$

Example B

We consider the same channel as in example A and the same noise variance. We choose the transmit power $P = 1$.

1. The optimal power allocation for the SIMO MAC is given by

$$\boldsymbol{S}^* = \mathrm{diag}([0, 1])$$

and the Lagrangian multiplier is $\lambda_1 = 0.2407 < 0.4048 = \lambda_2$.

2. The corresponding worst case coloured noise for the single-user MIMO system is

$$\boldsymbol{Z}^* = \begin{pmatrix} 1(0.5946) & 0.2647 \\ 0.2647 & 1 \end{pmatrix}.$$

Here, the entry $(1, 1)$ in $\boldsymbol{Z}^*$ was filled up from 0.5946 to 1. Note, that the noise covariance matrix which has not been filled up can be used to compute the optimal transmit covariance matrix $\boldsymbol{Q}^*$ and the capacity in the next steps [Yu03].

3. The waterfilling solution yields

$$\boldsymbol{Q}^* = \begin{pmatrix} 0.9412 & 0.2353 \\ 0.2353 & 0.0588 \end{pmatrix}.$$

An as in the previous example, we have

$$C_{III}^D = 0.7485 = C_{III}.$$

3.2.8 Interpretation and discussion of worst case noise analysis

In all three scenarios it is shown that the achievable performance with optimal transmit covariance matrix under worst case noise (with different constraints) equals the achievable performance without transmit cooperation and iid white noise. The differences between CSI and cooperation are further discussed in the following.

Discussion of results

We have considered a scenario in which the transmitter has perfect CSI $\boldsymbol{H}$ and furthermore knowledge of the interference covariance matrix $\boldsymbol{Z}$. Obviously, both the performance Φ as well as the optimal transmit covariance matrix $\boldsymbol{Q}$ are a function of the noise covariance matrix $\boldsymbol{Z}$. In all three noise scenarios, we have searched for the global minimum of Φ with respect to $\boldsymbol{Z}$. From a multiuser point of view, this corresponds to the question 'what is the worst case interference, that limit the performance of the point to point link?'. Closely related are the questions 'how much performance can be guaranteed even in worst case noise?' and 'what is lost due to worst case noise?'. In general MIMO systems, the transmit antennas can cooperate, i.e. beamforming in addition to spatial multiplexing. A necessary condition for cooperation is some kind of CSI at the transmit side. This means that there are different stages of transmit operation. Without CSI, there is obviously no cooperation.

In order to summarise the results and answer the three questions from the last paragraph, the following summary is proposed.

- *Worst case noise with trace constraint:* Regarding the performance, a MIMO system with perfect CSI about $\boldsymbol{H}$ and about interference $\boldsymbol{Z}$ at both sides of the link equals a MIMO system without CSI at the transmitter and with white slightly amplified noise. Due to worst case noise, the transmitter loses its CSI and hence its cooperation.
- *Worst case noise directions:* In this scenario, the MIMO system with perfect CSI about $\boldsymbol{H}$ and $\boldsymbol{Z}$ transforms under worst case noise directions with fixed noise covariance matrix eigenvalues to a system with parallel fading channels with effective channel gains $\frac{\lambda_i(\boldsymbol{H})}{\lambda_i(\boldsymbol{Z})}$. At the transmitter, CSI and cooperation was available and necessary to diagonalise the channel. The last optimisation step is power allocation according to waterfilling against the effective channel.
- *Worst case coloured noise:* Regarding the channel performance, a MIMO system with perfect CSI about $\boldsymbol{H}$ and $\boldsymbol{Z}$ under worst case coloured noise equals a MISO MAC with CSI at the transmitters and white noise. In this scenario, the CSI is still available at the transmit antennas. In exchange, the cooperation at the transmit side is lost.

3.2.9 Comparison of worst case noise capacities

In Scenario I, the worst case noise has the same directions as in Scenario II. Additionally, the eigenvalues of the noise covariance matrix are chosen to minimise the performance. The optimal noise covariance matrix eigenvalues are explicitly given in (3.66). The minimax problem in (3.13) fulfils the saddle-point property. Therefore, we have for all $\boldsymbol{Q}$ and $\boldsymbol{Z}$ and with optimal pair $(\boldsymbol{Q}^*, \boldsymbol{Z}^*)$

$$F(\boldsymbol{Q}, \boldsymbol{Z}^*) \leq F(\boldsymbol{Q}^*, \boldsymbol{Z}^*) \leq F(\boldsymbol{Q}^*, \boldsymbol{Z}).$$

For fixed $\boldsymbol{Z}$, the optimal $\boldsymbol{Q}$ is the waterfilling solution and for fixed $\boldsymbol{Q}$ the noise covariance matrix which minimises the performance is given by (3.66). In general, we have for the eigenvectors of the optimal $\boldsymbol{Q}^*$ and $\boldsymbol{Z}^*$

$$\boldsymbol{Q}^* = \boldsymbol{V}_H \boldsymbol{\Lambda}_{\boldsymbol{Q}^*} \boldsymbol{V}_H^H$$
$$\boldsymbol{Z}^* = \boldsymbol{U}_H \boldsymbol{\Lambda}_{\boldsymbol{Z}^*} \boldsymbol{U}_H^H$$

In Scenario I the set of admissible noise covariance matrices is larger than in Scenario II. The additional choice of eigenvalue distribution reveals this. Therefore, Φ_I is smaller than or equal to Φ_{II}.

Obviously, Φ_I is smaller than or equal to Φ_{III} and Φ_I is smaller than or equal to Φ_{III}. This follows from the fact, that the set of feasible noise covariance matrices in Scenario I and Scenario II is larger than the set of feasible noise covariance matrices in Scenario III. We obtain

$$\Phi_I \leq \Phi_{II}$$

and

$$\Phi_I \leq \Phi_{III}.$$

Unfortunately, we cannot compare the capacities Φ_{II} and Φ_{III}, because neither the set of noise covariance matrices in Scenario II is a subset of the set in Scenario III and vice versa.

At this point, the characterisation of the worst case noise performance of point to point MIMO link is complete. In the next section, we go one step beyond and incooperate the transmit strategies of all other users in the cell or system in order to optimise the sum of the performances of all users. The results from this section will help, because single-user optimisation under coloured noise will be one key ingredient in multi-user optimisation algorithms.

3.3 Sum Performance Analysis of Multiuser MIMO Systems

In this section, we take all users and their transmit strategies into account in order to optimise the overall performance of the multiuser MIMO system. At first, we present the signal model and define the relevant performance measures in section 3.3.1. The performance measures split into two classes. The first class is the sum performance of the overall system and individual performances of each single user. The focus in this section is on the sum performance of the overall system. The second class of individual performance requirements and their fulfilment is left for the future research section in chapter 4.

In section 3.3.2, we propose the problem statements associated with the sum performance measures and derive the generalised representation of the sum performance as the trace of an arbitrary matrix-monotone function. The optimisation of this function is performed in section 3.3.3. The properties of the optimal transmit strategies of the users are analysed in 3.3.4.

3.3.1 Signal model and sum performance measures

In this section, we present the signal model of the MIMO MAC and BC. The performance is either measured in terms of sum capacity for the MAC with SIC and the BC with Costa Precoding, or if the receiver applies the linear MMSE receiver at the base station, in terms of the average sum MSE.

MIMO MAC

Consider the multiple access channel in figure (3.6). The communication channel between each user and the base station is modelled by a block-flat fading MIMO channel as in the single-user scenario in figure (1.1).

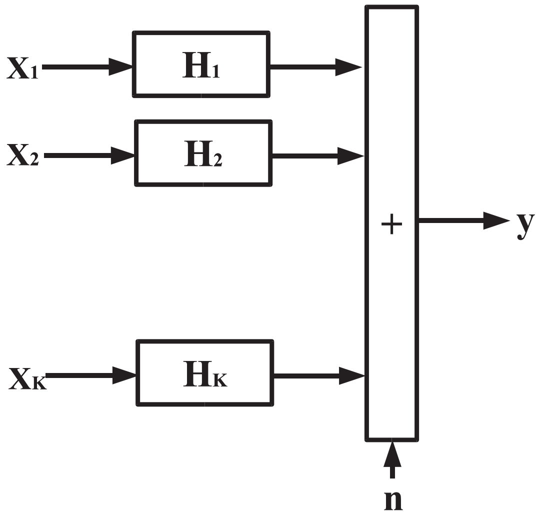

Figure 3.6: MIMO MAC system.

We have K mobiles with n_T antennas each. We can easily extend the results to the case in which every mobile has a different number of transmit antennas. The base station owns n_R receive antennas. In the discrete time model, the received vector $\boldsymbol{y}$ at any one time at the base station can be described by

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{n} \tag{3.26}$$

with the receiver noise $\boldsymbol{n} \in \mathbb{C}^{n_R \times 1}$ which is additive white Gaussian (AWG) noise, flat fading channel matrices $\boldsymbol{H}_k \in \mathbb{C}^{n_R \times n_T}$, and transmit signals $\boldsymbol{x}_k \in \mathbb{C}^{n_T \times 1}$. We assume uncorrelated noise with covariance $\sigma_n^2 \boldsymbol{I}_{n_R}$. The inverse noise power is denoted by $\rho = \frac{1}{\sigma_n^2}$.

Equation (3.26) can be rewritten in compact form as

$$\boldsymbol{y} = \hat{\boldsymbol{H}} \hat{\boldsymbol{x}} + \boldsymbol{n} \tag{3.27}$$

with $\hat{\boldsymbol{H}} = [\boldsymbol{H}_1, \boldsymbol{H}_2, ..., \boldsymbol{H}_K]$ and $\hat{\boldsymbol{x}} = [\boldsymbol{x}_1^T, ..., \boldsymbol{x}_K^T]^T$. We collect the transmit covariance matrices in

$$\hat{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{Q}_1 & 0 & 0 & ... & 0 \\ 0 & \boldsymbol{Q}_2 & 0 & ... & 0 \\ & & \ddots & & \\ 0 & 0 & & & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{Q}_K \end{pmatrix}. \tag{3.28}$$

The performance measures of the MIMO MAC were introduced in section 1.3.2. The sum capacity of the MIMO MAC with SIC applied at the base station is given by

$$C(\mathcal{Q}, \mathcal{H}, \rho) = \log\det\left(\boldsymbol{I} + \rho \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^H\right)$$

with the set of covariance matrices $\mathcal{Q}$ and the set of channel realisations $\mathcal{H}$

$$\mathcal{Q} = \{\boldsymbol{Q}_1, \boldsymbol{Q}_2, ..., \boldsymbol{Q}_K\} \quad \text{and} \quad \mathcal{H} = \{\boldsymbol{H}_1, \boldsymbol{H}_2, ..., \boldsymbol{H}_K\}.$$

In order to derive the normalised MSE of the linear multiuser MMSE receiver, we follow the definition and derivation of the normalised MSE in [VAT99] for the synchronous CDMA system. The linear MMSE multiuser receiver computes the data estimate

$$\tilde{\boldsymbol{x}} = \hat{\boldsymbol{Q}}\hat{\boldsymbol{H}}^H \left(\sigma_n^2 \boldsymbol{I} + \hat{\boldsymbol{H}}\hat{\boldsymbol{Q}}\hat{\boldsymbol{H}}^H\right)^{-1} \boldsymbol{y}. \tag{3.29}$$

The covariance matrix of the estimation error ϵ is given as

$$\boldsymbol{K}_\epsilon = \hat{\boldsymbol{Q}} - \hat{\boldsymbol{Q}}\hat{\boldsymbol{H}}^H \left(\hat{\boldsymbol{H}}\hat{\boldsymbol{Q}}\hat{\boldsymbol{H}}^H + \sigma_n^2 \boldsymbol{I}\right)^{-1} \hat{\boldsymbol{H}}\hat{\boldsymbol{Q}}. \tag{3.30}$$

For convenience, we define the matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \sigma_n^2 \boldsymbol{I} + \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^H. \tag{3.31}$$

From (3.30), we define the normalised MSE as

$$\begin{aligned} MSE(\mathcal{Q}, \mathcal{H}, \rho) &= tr(\hat{\boldsymbol{Q}}^{-1/2}\boldsymbol{K}_\epsilon \hat{\boldsymbol{Q}}^{-1/2}) = Kn_T - \sum_{i=1}^{n_R} \frac{\mu_i}{\sigma_n^2 + \mu_i} \\ &= Kn_T - n_R + \sigma_n^2 \sum_{i=1}^{n_R} \frac{1}{\sigma_n^2 + \mu_i} \\ &= Kn_T - n_R + \sigma_n^2 \operatorname{tr}\left(\boldsymbol{A}^{-1}\right) \end{aligned} \tag{3.32}$$

with μ_i as the eigenvalues of $\hat{\boldsymbol{H}}\hat{\boldsymbol{Q}}\hat{\boldsymbol{H}}^H$. The MSE is reduced by minimising the sum in the RHS of (3.32). It is worth mentioning that the term $\sum_{i=1}^{n_R} \frac{1}{\sigma_n^2+\mu_i}$ is a Schur-convex function with respect to the μ_i [MO79].

MIMO BC

In figure (3.7), the MIMO BC is depicted. Here, we study the downlink transmission from the base station to the mobiles. The base station is equipped with n_T transmit antennas and each mobile has n_R antennas. The channel matrices in the downlink transmission correspond to the Hermitian channel matrices from the uplink, i.e. $\boldsymbol{H}_i^{dl} = \boldsymbol{H}_i^H$ (reciprocity).

The received vector $\boldsymbol{y}_k$ at each mobile k can be written as

$$\boldsymbol{y}_k = \boldsymbol{H}_k^H \boldsymbol{x}_k + \sum_{l=1, k\neq l}^{K} \boldsymbol{H}_k^H \boldsymbol{x}_l + \boldsymbol{n}_k \tag{3.33}$$

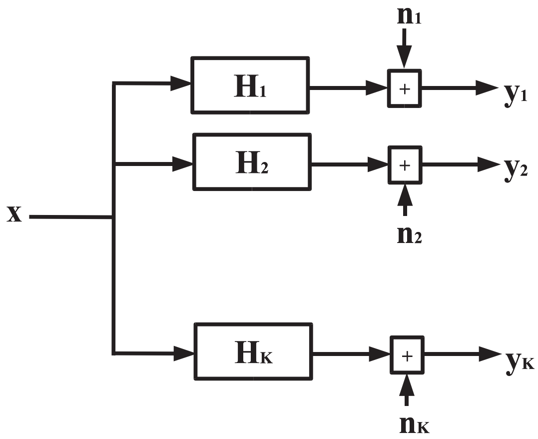

Figure 3.7: MIMO BC system.

with the flat fading channel matrices $\boldsymbol{H}_k$, the AWG receiver noises $\boldsymbol{n}_k$, and the transmit signal $\boldsymbol{x}_k$ which is intended for mobile k. The noise at the mobiles is assumed uncorrelated and independent identically distributed. In equation (3.33) the first term is the signal for user k, the second term is the interference from the signals for the other users, and the last term is the noise.

In order to end up with the same expression for the sum capacity as in the MIMO MAC case, we use the following results:

- The capacity regions of the MIMO MAC and MIMO BC are equal [VBG03, WSS04].
- The set of transmit covariance matrices in the MIMO MAC (with fixed decoding order) can be transformed into a set of transmit covariance matrices for the MIMO BC (with flipped channels and reverse precoding order) which achieves the same capacity point [Vis03, Section 3.3].

It suffices to consider the sum capacity optimisation of the MIMO MAC because the following steps lead to the optimal transmit covariance matrices for the MIMO BC:

- Solve the sum capacity optimisation problem for the MIMO MAC with flipped channels $\boldsymbol{H}_k^H$.
- Transform the transmit covariance matrix $\boldsymbol{Q}_j^M$ for each user j according to the rule in [Vis03]

$$\boldsymbol{Q}_j^B = \boldsymbol{B}_j^{-1/2}\boldsymbol{F}_j\boldsymbol{G}_j^H\boldsymbol{A}_j^{1/2}\boldsymbol{Q}_j^M\boldsymbol{A}_j^{1/2}\boldsymbol{G}_j\boldsymbol{F}_j^H\boldsymbol{B}_j^{-1/2}$$

  with $\boldsymbol{F}_j$ as the left eigenvectors of the channel matrix $\boldsymbol{H}_j$ and $\boldsymbol{G}_j$ as the right eigenvectors of the channel matrix $\boldsymbol{H}_j$ and

$$\boldsymbol{B}_j = \boldsymbol{I} + \rho \sum_{l=j+1}^{K} \boldsymbol{H}_l\boldsymbol{Q}_l^M\boldsymbol{H}_l^H$$

  and

$$\boldsymbol{A}_j = \boldsymbol{I} + \rho\boldsymbol{H}_j^H \sum_{l=1}^{j-1} \boldsymbol{Q}_l^B\boldsymbol{H}_j.$$

These transmit covariance matrices achieve the sum capacity of the MIMO BC.

General structure of performance function

The sum capacity of the MIMO MAC with SIC and MIMO BC with Costa Precoding as well as the average normalised sum MSE for the MIMO MAC can be written in the following generalised form using an arbitrary matrix-monotone inner performance function $\phi(\boldsymbol{X})$

$$\Phi(\rho, \mathcal{Q}, \mathcal{H}) = \operatorname{tr} \phi \left(\rho \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^H \right). \tag{3.34}$$

In order to obtain the sum capacity from (3.34), the inner performance function $\phi(\boldsymbol{X})$ has to be chosen as

$$\phi_1(\boldsymbol{X}) = \log\det\left(\boldsymbol{I} + \boldsymbol{X}\right)$$

This corresponds to the choice of the inner objective function ϕ_1 in the single-user scenario in equation (2.17). The inner performance function for the normalised sum MSE is given by

$$\tilde{\phi}_2(\boldsymbol{X}) = \left(\frac{Kn_T}{n_R} - 1\right)\boldsymbol{I} + \left[\boldsymbol{I} + \boldsymbol{X}\right]^{-1} = \frac{Kn_T}{n_R}\boldsymbol{I} - \boldsymbol{X}\left[\boldsymbol{I} + \boldsymbol{X}\right]^{-1}.$$

In the case in which the[2] normalised sum MSE is considered, the inner objective function is chosen as

$$\phi_2(\boldsymbol{X}) = \boldsymbol{X}\left[\boldsymbol{I} + \boldsymbol{X}\right]^{-1}. \tag{3.35}$$

Note the similarity to the single-user MSE performance function in (2.18). Again, we restrict the class of inner performance functions to be matrix-monotone as in the single-user case. Therefore, the same assumptions as in section 2.3.1 hold.

In the following three sections, we are going to maximise the sum performance in (3.34) by choosing the optimal transmit covariance matrices $\boldsymbol{Q}_1, ..., \boldsymbol{Q}_K$. At first, we consider the case in which the transmit powers of all users are constrained, i.e. $\operatorname{tr}\left(\boldsymbol{Q}_k\right) \leq p_k$. In order to get a better performance we apply adaptive power allocation under a sum power constraint, i.e. $\sum_{k=1}^{K} p_k \leq P$. These two cases directly lead to an iterative algorithm which solves the general problem of sum performance optimisation under a sum power constraint.

3.3.2 Problem statements: Sum performance optimisation under different power constraints

We optimise the transmit covariance matrices in order to maximise the sum performance in (3.34). First, each transmit covariance matrix is constrained in its power to P_k, i.e. $0 \leq \operatorname{tr}\left(\boldsymbol{Q}_k\right) \leq p_k$. This is motivated in the MAC by an admissible transmit power at the mobiles.
Problem 1: In order to maximise the sum performance in (3.34) for fixed and known channel realizations $\boldsymbol{H}_k$, find the optimal transmit covariance matrices $\boldsymbol{Q}_1^*, ..., \boldsymbol{Q}_K^*$,

[2] We take the last term of the difference in order to obtain an equivalent maximisation problem.

i.e. solve

$$\max \ \mathrm{tr}\ \phi\left(\rho\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right)$$
$$\text{subject to} \quad \mathrm{tr}\ \boldsymbol{Q}_k \leq p_k \ \text{ and } \ \boldsymbol{Q}_k \succeq 0, \quad 1 \leq k \leq K. \tag{3.36}$$

In the next step, an additional power allocation under sum power constraint for fixed transmit covariance matrices is performed. This leads to the next problem statement.
Problem 2: The channel realizations $\boldsymbol{H}_k$ of all users k are assumed to be known. Keep the transmit covariance matrices fixed $\boldsymbol{Q}'_1, \boldsymbol{Q}'_2, ..., \boldsymbol{Q}'_K$. Distribute an fixed amount of transmit power P across the mobiles, i.e. solve

$$\max \ \mathrm{tr}\ \phi\left(\rho\sum_{k=1}^{K}p_k\boldsymbol{H}_k\boldsymbol{Q}'_k\boldsymbol{H}_k^H\right)$$
$$\text{subject to} \quad \sum_{k=1}^{K} p_k \leq P \ \text{ and } \ p_k > 0, \quad 1 \leq k \leq K. \tag{3.37}$$

Combining these two step, power allocation in Problem 2 and transmit covariance matrix optimisation in Problem 1, then one arrives at the general problem of sum performance optimisation under a sum power constraint. This problem arises for example in the downlink transmission in which the base station can allocate an amount of power P for the transmit signals of the users. The corresponding problem statement is given in Problem 3.
Problem 3: Assume that the channel realizations $\boldsymbol{H}_k$ are known and fixed. Solve the sum performance optimisation under the sum power constraint, i.e.

$$\max \ \mathrm{tr}\ \phi\left(\rho\sum_{k=1}^{K}\boldsymbol{H}_k\tilde{\boldsymbol{Q}}_k\boldsymbol{H}_k^H\right)$$
$$\text{subject to} \quad \sum_{k=1}^{K} \mathrm{tr}\ \tilde{\boldsymbol{Q}}_k \leq P \ \text{ and } \ \tilde{\boldsymbol{Q}}_k \succeq 0, \quad 1 \leq k \leq K. \tag{3.38}$$

3.3.3 Optimisation of sum performance

In the following, we analyse the structure of the optimisation problem in Problem 1, 2, and 3 using the Karush-Kuhn-Tucker (KKT) optimality conditions. All problems are convex optimisations. Besides, the set of covariance matrices $\boldsymbol{Q}_1, ..., \boldsymbol{Q}_K$ with the trace constraints is convex, the set of power allocations $p_1, ...p_K$ is convex, and the set of covariance matrices $\tilde{\boldsymbol{Q}}_1, ..., \tilde{\boldsymbol{Q}}_K$ with sum trace constraint is convex. Therefore, Slater's condition is satisfied. And as a result, the KKT conditions are sufficient and necessary for the optimum solution [BV03].

Covariance optimisation with individual power constraints

In the following, we assume that the powers $p_1, ..., p_K$ are fixed and study the optimisation in (3.36). We show that the optimal covariance matrices can be found by iterative single-user performance optimisation with coloured noise. This approach corresponds with the iterative waterfilling approach in [YRBC04] for sum capacity optimisation in which single-user waterfilling is iteratively performed treating the other users as noise in order to maximise the sum capacity. On the one hand this approach provides insight into the structure of the optimum transmit covariance matrices and on the other hand under specific conditions this approach is computational more efficient than the joint optimisation of the transmit covariance matrices. If the number of users is large in comparison to the number of transmit antennas of the users, the joint optimisation is computational more complex than the iterative optimisation of each user separately.

The Lagrange function for the optimisation problem in (3.36) is given by

$$\begin{aligned}\mathcal{L}_1(\mathcal{Q},\mathcal{P},\nu_k) &= \operatorname{tr}\phi\left(\rho\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right)\\ &\quad+\sum_{k=1}^{K}\operatorname{tr}\left(\boldsymbol{\Psi}_k\boldsymbol{Q}_k\right)+\sum_{k=1}^{K}\nu_k(p_k-tr(\boldsymbol{Q}_k)) \qquad (3.39)\end{aligned}$$

with the set of covariance matrices and the set of Lagrangian multipliers

$$\mathcal{Q}=(\boldsymbol{Q}_1,...,\boldsymbol{Q}_K) \quad \text{and} \quad \mathcal{P}=(\boldsymbol{\Psi}_1,...,\boldsymbol{\Psi}_K).$$

The first derivative of (3.39) with respect to $\boldsymbol{Q}_i$ is given by

$$\frac{d\mathcal{L}_1(\mathcal{Q},\mathcal{P},\nu_k)}{d\boldsymbol{Q}_i}=\rho\boldsymbol{H}_i^H\phi^{[1]}\left(\rho\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_i^H\right)\boldsymbol{H}_i+\boldsymbol{\Psi}_i-\nu_i\boldsymbol{I}. \qquad (3.40)$$

As a result, the KKT conditions for the optimal covariance matrices which solve optimisation problem in (3.36) are

$$\begin{aligned}\rho\boldsymbol{H}_i^H\phi^{[1]}\left(\rho\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_i^H\right)\boldsymbol{H}_i &= \nu_i\boldsymbol{I}-\boldsymbol{\Psi}_i \quad 1\le i\le K\\ \operatorname{tr}\left(\boldsymbol{\Psi}_i\boldsymbol{Q}_i\right) &= 0 \quad 1\le i\le K\\ \boldsymbol{\Psi}_k &\succeq 0 \quad 1\le i\le K\\ \boldsymbol{Q}_i &\succeq 0 \quad 1\le i\le K\\ \nu_i &\ge 0 \quad 1\le i\le K\\ p_i-tr(\boldsymbol{Q}_i) &\ge 0 \quad 1\le i\le K\\ \nu_i\left(p_i-tr(\boldsymbol{Q}_i)\right) &= 0 \quad 1\le i\le K \qquad (3.41)\end{aligned}$$

Based on these optimality conditions, we will show next that a kind of iterative single-user optimisation solves (3.36). For the k-th user, we write the noise plus interference as

$$\boldsymbol{Z}_k=\boldsymbol{I}+\rho\sum_{\substack{l=1\\ l\neq k}}^{K}\boldsymbol{H}_l\boldsymbol{Q}_l\boldsymbol{H}_l^H. \qquad (3.42)$$

At this point, the structure of the performance metric has an impact on the choice of the single-user optimisation problem.

Sum-Capacity

In the case in which the performance metric is the sum capacity, the single-user problem which is iteratively solved is the waterfilling with respect to the effective channel $\boldsymbol{Z}_k^{-1/2}\boldsymbol{H}_k$. The performance metric is $\phi(\boldsymbol{X}) = \log(\boldsymbol{I}+\boldsymbol{X})$, i.e. the sum performance metric is the given by

$$\begin{aligned} &\max \ \operatorname{tr}\ \log\left(\boldsymbol{I}+\rho\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right)\\ &\text{subject to} \quad \operatorname{tr}\boldsymbol{Q}_k \le p_k \ \text{ and } \ \boldsymbol{Q}_k \succeq 0, \quad 1 \le k \le K. \end{aligned} \tag{3.43}$$

For each user $k \in [1...K]$ we solve the optimisation problem

$$\begin{aligned} &\max \ \operatorname{tr}\ \log\left(\boldsymbol{I}+\rho\overbrace{\boldsymbol{Z}_k^{-1/2}\boldsymbol{H}_k}^{\tilde{\boldsymbol{H}}_k}\boldsymbol{Q}_k\boldsymbol{H}_k^H\boldsymbol{Z}_k^{-1/2}\right)\\ &\text{subject to} \quad \operatorname{tr}\ (\boldsymbol{Q}_k) \le p_k \ \text{ and } \ \boldsymbol{Q}_k \succeq 0. \end{aligned} \tag{3.44}$$

The next theorem shows that the single-user covariance optimisations for all users $1 \le k \le K$ in (3.44) mutually solve the optimisation problem (3.43). This theorem corresponds to Theorem 3 in [YRBC04].
Theorem 12: If all covariance matrices $\boldsymbol{Q}_k^*$ mutually solve the optimisation problem in (3.44) for $\boldsymbol{Z}_k$ in (3.42), then they solve optimisation problem in (3.43) for the sum capacity, too.

The result follows from the fact, that the optimisation problem in (3.44) has the same optimality conditions as the original problem in (3.43).

Remark: In order to prove convergence of the iterative single-user waterfilling, note that the objective in (3.44) differs from the objective in (3.43) only by a constant which is independent of $\boldsymbol{Q}_k$. Therefore, in each single-user waterfilling step the sum capacity is increased. The channel matrix $\tilde{\boldsymbol{H}}_k = \boldsymbol{Z}_k^{-1/2}\boldsymbol{H}_k$ in (3.44) is the effective channel which is weighted by the inverse noise. The iterative single-user performance algorithm in (3.44) solves the original optimisation problem in (3.36). However, in contrast to the iterative waterfilling algorithm proposed for capacity maximisation, we cannot derive a simple algorithm which solves the generalised single-user performance problem because of the dependence on the noise covariance matrix $\boldsymbol{Z}_k$ in (3.44).

Sum MSE

In the case in which the performance metric is the sum MSE, the single-user problem which is iteratively solved is the original sum MSE problem for fixed transmit strategies of the other users. The performance metric is $\phi(\boldsymbol{X}) = \boldsymbol{X}(\boldsymbol{I}+\boldsymbol{X})$, i.e. the sum performance metric is the given by

$$\begin{aligned} &\max \ \operatorname{tr}\ \left(\rho\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\left[\boldsymbol{I}+\rho\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right]^{-1}\right)\\ &\text{subject to} \quad \operatorname{tr}\boldsymbol{Q}_k \le p_k \ \text{ and } \ \boldsymbol{Q}_k \succeq 0, \quad 1 \le k \le K. \end{aligned} \tag{3.45}$$

For each user $k \in [1...K]$ we solve the optimisation problem

$$\begin{aligned} &\max \ \operatorname{tr}\ -\left(\left[\boldsymbol{I}+\boldsymbol{Z}_k+\rho\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right]^{-1}\right)\\ &\text{subject to} \quad \operatorname{tr}\ (\boldsymbol{Q}_k) \le p_k \ \text{ and } \ \boldsymbol{Q}_k \succeq 0. \end{aligned} \tag{3.46}$$

The next theorem shows that the single-user covariance optimizations for all users $1 \leq k \leq K$ in (3.46) solve the optimization problem (3.45).
Theorem 13: If all covariance matrices $\mathbf{Q}_k^*$ mutually solve the optimization problem in (3.46) for

$$\mathbf{Z}_k = \sigma_n^2 \mathbf{I} + \sum_{l=1, l \neq k}^{K} \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^H, \tag{3.47}$$

then they solve optimization problem in (3.45), too.

The proof can be found in section 3.4.8.

Remark: Note that the single-user optimization problem in (3.44) has an interesting interpretation: Assume the single-user MSE optimization with colored noise $\mathbf{Z}_k = \mathbf{U}_Z \mathbf{\Lambda}_Z \mathbf{U}_Z^H$. We can write

$$\begin{aligned} tr\left(\left[\mathbf{Z}_k + \rho \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^H\right]^{-1}\right) &= tr\left(\mathbf{\Lambda_Z}\left[\mathbf{I} + \rho \tilde{\mathbf{H}}_\mathbf{k} \mathbf{Q}_k \tilde{\mathbf{H}}_\mathbf{k}^H\right]^{-1}\right) \\ &= \sum_{l=1}^{n_R} \lambda_Z^{-1}(l) \left(\left[\mathbf{I} + \rho \tilde{\mathbf{H}}_\mathbf{k} \mathbf{Q}_k \tilde{\mathbf{H}}_\mathbf{k}^H\right]^{-1}\right)_{l,l}. \end{aligned} \tag{3.48}$$

The channel matrix $\tilde{\mathbf{H}}_\mathbf{k} = \mathbf{Z}_k^{-1/2} \mathbf{H}_k$ in (3.48) is the weighted effective channel. The iterative single-user MSE algorithm solves the original optimization problem in (3.45). However, in contrast to the iterative waterfilling algorithm we cannot derive a simple algorithm which solves the single-user MSE problem because of the dependence on the noise eigenvalues in (3.48).

In the last section, we have assumed that the transmit power of each user k is constrained by some individual power constraint p_k. Next, we assume that the available transmit power can be distributed between all mobiles with a sum power constraint, i.e. $\sum_{k=1}^{K} p_k \leq P$.

Power allocation

The sum performance can be further improved if adaptive power control is applied. During the power allocation we keep the transmit covariance matrices fixed. For fixed transmit covariance matrices $\boldsymbol{Q}'_1, ..., \boldsymbol{Q}'_K$ solve programming problem (3.37).

The transmit covariance matrices $\boldsymbol{Q}'_k$ are fixed and power normalised, i.e. $tr(\boldsymbol{Q}'_k)) = 1$. The Lagrangian for the optimisation problem in (3.37) is given by

$$\begin{aligned} \mathcal{L}_2(\boldsymbol{p}, \boldsymbol{\lambda}, \mu) &= \operatorname{tr} \phi\left(\rho \sum_{k=1}^{K} p_k \boldsymbol{H}_k \boldsymbol{Q}'_k \boldsymbol{H}_k^H\right) \\ &\quad + \sum_{k=1}^{K} (\lambda_k p_k) + \mu\left(P - \sum_{k=1}^{K} p_k\right). \end{aligned} \tag{3.49}$$

with the power vector and Lagrangian multiplier vector

$$\boldsymbol{p} = (p_1, ..., p_K) \quad \text{and} \quad \boldsymbol{\lambda} = (\lambda_1, ..., \lambda_K).$$

The first derivative of (3.49) with respect to p_i is given by

$$\frac{d\mathcal{L}_2(\boldsymbol{p}, \boldsymbol{\lambda}, \mu)}{dp_i} = \text{tr}\left(\rho \boldsymbol{H}_i \boldsymbol{Q}'_i \boldsymbol{H}_i^H \phi^{[1]} \left[\rho \sum_{k=1}^{K} p_k \boldsymbol{H}_k \boldsymbol{Q}'_k \boldsymbol{H}_k^H\right]\right) + \lambda_i - \mu. \tag{3.50}$$

As a result, the KKT conditions of the power allocation optimisation problem in (3.37) are

$$\begin{aligned}
\text{tr}\left(\rho \boldsymbol{H}_i \boldsymbol{Q}'_i \boldsymbol{H}_i^H \phi^{[1]} \left[\rho \sum_{k=1}^{K} p_k \boldsymbol{H}_k \boldsymbol{Q}'_k \boldsymbol{H}_k^H\right]\right) &= \mu - \lambda_i \quad 1 \le i \le K \\
\lambda_k p_k &= 0 \quad 1 \le k \le K \\
\lambda_k &\ge 0 \quad 1 \le k \le K \\
\mu &\ge 0 \\
p_k &\ge 0 \quad 1 \le k \le K \\
P - \sum_{k=1}^{K} p_k &\ge 0 \\
\mu\left(P - \sum_{k=1}^{K} p_k\right) &= 0
\end{aligned} \tag{3.51}$$

These conditions in (3.51) are fulfilled by the optimum power allocation vector for fixed covariance matrices. Observe, that for all active users l the Lagrangian multiplier λ_l is equal to zero. Therefore, the condition

$$\text{tr}\left(\rho \boldsymbol{H}_i \boldsymbol{Q}'_i \boldsymbol{H}_i^H \phi^{[1]} \left[\rho \sum_{k=1}^{K} p_k \boldsymbol{H}_k \boldsymbol{Q}'_k \boldsymbol{H}_k^H\right]\right) = \mu$$

is fulfilled. We use these conditions twofold. First, it can be used to derive the properties of the optimal transmit strategies. In section (3.3.4), the SNR range in which only the best user gets the complete power is studied. Second, the condition can be used to verify that the solution by the interior point algorithm satisfies the KKT. If it did not satisfy the KKT, we would restart the search algorithm with another starting point in the interior.

Iterative power and covariance optimisation

The final step in optimising the performance of the multiuser MIMO system is to connect the two optimisation problems from the last sections. Consequently, this will lead us to an iterative algorithm which solves the performance maximisation problem with a sum power constraint. This problem is given in (3.38).

In the following, we study the structure of the optimisation problem in (3.38). The Lagrangian of (3.38) is given by

$$\begin{aligned}
\mathcal{L}(\tilde{\mathbf{Q}}, \mathcal{P}, \mu) &= \text{tr}\, \phi\left(\rho \sum_{k=1}^{K} \boldsymbol{H}_k \tilde{\boldsymbol{Q}}_k \boldsymbol{H}_k^H\right) \\
&\quad + \sum_{k=1}^{K} \text{tr}\,(\tilde{\boldsymbol{Q}}_k \boldsymbol{\Psi}_k) + \mu\left(P - \sum_{k=1}^{K} \text{tr}\,(\tilde{\boldsymbol{Q}}_k)\right).
\end{aligned} \tag{3.52}$$

with the set of covariance matrices and Lagrangian multipliers

$$\tilde{\mathfrak{Q}} = \{\tilde{\boldsymbol{Q}}_1, ..., \tilde{\boldsymbol{Q}}_K\} \quad \text{and} \quad \boldsymbol{\mathcal{P}} = \{\boldsymbol{\Psi}_1, ..., \boldsymbol{\Psi}_K\}.$$

The first derivative of (3.52) with respect to $\boldsymbol{Q}_i$ is given by

$$\frac{d\mathcal{L}(\tilde{\mathfrak{Q}}, \boldsymbol{\mathcal{P}}, \mu)}{d\tilde{\boldsymbol{Q}}_i} = \left(\rho \boldsymbol{H}_i^H \phi^{[1]} \left[\rho \sum_{k=1}^{K} \boldsymbol{H}_k \tilde{\boldsymbol{Q}}_k \boldsymbol{H}_k^H \right] \boldsymbol{H}_i \right) + \Psi_i - \mu \boldsymbol{I}. \tag{3.53}$$

As a result, the KKT conditions for the optimisation problem in (3.38) are

$$\begin{aligned}
\rho \boldsymbol{H}_i^H \phi^{[1]} \left[\rho \sum_{k=1}^{K} \boldsymbol{H}_k \tilde{\boldsymbol{Q}}_k \boldsymbol{H}_k^H \right] \boldsymbol{H}_i &= \mu \boldsymbol{I} - \boldsymbol{\Psi}_i \quad 1 \leq i \leq K \\
\operatorname{tr}\left(\tilde{\boldsymbol{Q}}_i \boldsymbol{\Psi}_i\right) &= 0 \quad 1 \leq i \leq K \\
\boldsymbol{\Psi}_i &\succeq 0 \quad 1 \leq i \leq K \\
\tilde{\boldsymbol{Q}}_i &\succeq 0 \quad 1 \leq i \leq K \\
\mu &\geq 0 \\
P - \sum_{k=1}^{K} \operatorname{tr}\left(\tilde{\boldsymbol{Q}}_k\right) &\geq 0 \qquad (3.54) \\
\mu \left(P - \sum_{k=1}^{K} \operatorname{tr}\left(\tilde{\boldsymbol{Q}}_k\right) \right) &= 0. \qquad (3.55)
\end{aligned}$$

These conditions in (3.55) are fulfilled by the optimum transmit covariance matrices $\tilde{\boldsymbol{Q}}_i$ and by the corresponding Lagrangian multipliers $\boldsymbol{\Psi}_i$ and μ. We use these KKT conditions in (3.55) to analyse the behaviour at small SNR values. We characterise the 'best' user and its transmission strategy and provide a necessary and sufficient condition for the optimality of single-user transmission in section (3.3.4).

In the following, we describe the optimisation strategy which leads to the optimum power allocation and transmit covariance matrices. The structure we derive is a generalisation of the solution of the sum capacity optimisation of the MIMO MAC [YRBC04, BJ02g].

The optimisation consists of two steps: We start with fixed covariance matrices $\boldsymbol{Q}_k^0 = \boldsymbol{I}$. Then we perform iteratively power allocation (3.37) and covariance matrix optimisation (3.36). This yields Algorithm (2).

Algorithm 2 Iterative power allocation and covariance matrix optimisation

initialise $\boldsymbol{Q}_1^0 = ... = \boldsymbol{Q}_K^0 = \boldsymbol{I}$.
while required accuracy not reached **do**
  From last step $n-1$ given covariance matrices $\boldsymbol{Q}_1^n, ...\boldsymbol{Q}_K^n$.
  Solve the power allocation optimisation (3.37):
  $p_1^n, ..., p_K^n = \arg\max_{p_1,...,p_K} \ \operatorname{tr} \phi \left(\rho \sum_{k=1}^K p_k \boldsymbol{H}_k \boldsymbol{Q}_k^n \boldsymbol{H}_k^H \right)$
  subject to $p_k \geq 0$ and $\sum_{k=1}^K p_k = P$.
  With fixed power allocation $p_1^n, ..., p_K^n$ solve the covariance matrix optimisation (3.36)
  $\boldsymbol{Q}_1^n, ..., \boldsymbol{Q}_K^n = \arg\min_{\bar{\boldsymbol{Q}}_1,...,\bar{\boldsymbol{Q}}_K} \ \operatorname{tr} \phi \left(\rho \sum_{k=1}^K p_k^n \boldsymbol{H}_k \bar{\boldsymbol{Q}}_k \boldsymbol{H}_k^H \right)$
  subject to $\boldsymbol{Q}_k \succeq 0$ and $tr(\boldsymbol{Q}_k) = 1$ using iterative single-user covariance optimisation (see Theorem 12 for capacity and Theorem 13 for MSE).
end while

Description of the algorithm: At first, we initialise the transmit covariance matrices as identity matrices. Next, we iteratively perform the loop while the difference in the sum performance between two steps is larger than some small constant. Inside the loop, we perform the power allocation step at first. With the new fixed powers we optimise the covariance matrices with the algorithm from the last section. Note, that the performance is increasing with each step.

In the case of sum capacity optimisation, the inner covariance matrix optimisation specialises to the iterative waterfilling [YRBC04] and the complete algorithm 1 specialises to the power allocation plus iterative waterfilling [BJ02g].

Optimality of iterative approach

In order to check whether the iterative algorithm 2 solves the problem (3.37), we propose the following theorem. We define the set of active users, i.e. the users with $p_k > 0$, for convenience

$$\mathcal{I} = \{k \in \{1, ..., K\} : p_k > 0\}$$

Theorem 14: Suppose that the set of covariance matrices $\{\boldsymbol{Q}_k^*\}$ solves (3.36) for given $\{p_k^*\}$ and $\{p_k^*\}$ solves (3.37) for given covariance matrices $\{\boldsymbol{Q}_k^*\}$. The covariance matrices $\tilde{\boldsymbol{Q}}_k = p_k^* \boldsymbol{Q}_k^*$ solve problem (3.38) if and only if there exists a $\bar{\mu} \geq 0$ such that

$$\frac{\nu_k}{p_k^*} = \bar{\mu}, \quad k \in \mathcal{I} \tag{3.56}$$

$$\bar{\mu}\boldsymbol{I} - \rho \boldsymbol{H}_i^H \phi^{[1]} \left[\rho \sum_{k=1}^{K} p_k^* \boldsymbol{H}_k \boldsymbol{Q}_k^* \boldsymbol{H}_k^H \right] \boldsymbol{H}_i \succeq 0 \quad k \in \{1, ..., k\} \backslash \mathcal{I}. \tag{3.57}$$

The proof of Theorem 14 can be found in section 3.4.9.

The solution of problem (3.38) can imply that either one or a small number of users is allowed to transmit in order to maximise sum performance. This depends on the SNR and on the channel matrices. Of course, this solution is unfair for users which want to transmit but are not allowed to for a larger period of time. Especially, if the users have a certain service requirement for voice of image transmission, some MSE requirements occur.

3.3.4 Properties of optimal transmit strategy

In this section, we show that for small SNR values the complete transmission power is allocated to the best user. We characterise the best user and show that the user with the maximum channel eigenvalue is the best user. Next, we characterise the single-user SNR range in which the sum performance is maximised by one single user. Finally, we provide the complete characterisation of the single-user range.

In the following, we consider the two best users with channel $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$. However, the derivation can be easily extended to the arbitrary multiuser case. In the following theorem, we show that for arbitrary small SNR values both users are supported if their maximum channel eigenvalues are equal. As a corollary it follows that the user with the maximum eigenvalue gets the complete transmission power for low

SNR values. We use this result in the next section for the analysis of the single-user range, i.e. the SNR range in which only one user is supported.

Theorem 15: Let the users be ordered according to their maximum channel eigenvalues in decreasing order. Let $\boldsymbol{Q}_1^s$ be the optimal transmit covariance matrix for the single-user channel $\boldsymbol{H}_1$ according to generalized single-user water filling from Lemma 3. For fixed $\hat{\rho}$, it is optimal to allocate the complete sum power to user one, i.e. the user with the largest maximum channel eigenvalue, if and only if the following condition is satisfied

$$\lambda_{max}\left(\boldsymbol{H}_2^H \phi^{[1]}\left[\hat{\rho}\boldsymbol{H}_1\boldsymbol{Q}_1^s(\hat{\rho})\boldsymbol{H}_1^H\right]\boldsymbol{H}_2\right) \leq \lambda_{max}\left(\boldsymbol{H}_1^H \phi^{[1]}\left[\hat{\rho}\boldsymbol{H}_1\boldsymbol{Q}_1^s(\hat{\rho})\boldsymbol{H}_1^H\right]\boldsymbol{H}_1\right) \quad (3.58)$$

then the optimal set of covariance matrices is $\boldsymbol{Q}_1(\hat{\rho}) = \boldsymbol{Q}_1^s(\hat{\rho})$ and $\boldsymbol{Q}_2(\hat{\rho}) = \boldsymbol{Q}_3(\hat{\rho}) = \ldots = \boldsymbol{Q}_K(\hat{\rho})$. This transmit strategy is also optimal for all $\rho \leq \hat{\rho}$.

The proof can be found in section 3.4.10.

We know that the first user is the user with the largest maximum channel eigenvalue and that the next user is the user with the second maximal channel eigenvalue. This is somewhat surprising, since the average transmitted power is given by the trace of the channel covariance matrix, i.e. the Frobenius norm. The order which decides who is the best user is the induced l_2-norm.

3.3.5 Discussion of sum performance optimisation algorithm

Let us briefly review the construction of the solution to the sum performance optimisation of multiuser MIMO systems. The original problem of transmit strategy optimisation was decomposed into two subproblems, namely power allocation and covariance matrix optimisation under individual power constraints. This scheme is illustrated in figure (3.8).

The outer loop is between power allocation and covariance matrix optimisation under individual power constraints. The covariance matrix optimisation can be decomposed into an inner loop in which single-user covariance matrix optimisation with respect to the effective channel is performed.

In the case in which the sum performance is measured by the sum capacity, the inner single-user waterfilling algorithm can be derived in closed form [YRBC04]. Then, the covariance matrix optimisation corresponds to iterative waterfilling.

3.4 Proofs

3.4.1 Proof of Theorem 7

We prove Theorem 7 by contradiction. Assume that the optimal eigenvectors are chosen such that $\boldsymbol{A}$ and $\boldsymbol{B}$ do not commute. Then the directional derivative is larger (smaller) than zero and therefore other eigenvectors achieve a better higher (smaller) value.

The objective function $F(\boldsymbol{B}^{1/2}\boldsymbol{A}\boldsymbol{B}^{1/2})$ can be rewritten using the eigenvalues decomposition $\boldsymbol{A} = \boldsymbol{U}_A\boldsymbol{\Lambda}_A\boldsymbol{U}_A^H$, $\boldsymbol{B} = \boldsymbol{U}_B\boldsymbol{\Lambda}_B\boldsymbol{U}_B^H$, and $\boldsymbol{U} = \boldsymbol{U}_B^H\boldsymbol{U}_A$. Without loss of generality, we assume that both $\boldsymbol{A}$ and $\boldsymbol{B}$ are full rank. The objective function is rewritten as

$$\mathrm{tr}\, F(\boldsymbol{B}^{1/2}\boldsymbol{A}\boldsymbol{B}^{1/2}) = \mathrm{tr}\, F(\boldsymbol{\Lambda}_B^{1/2}\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H\boldsymbol{\Lambda}_B^{1/2}).$$

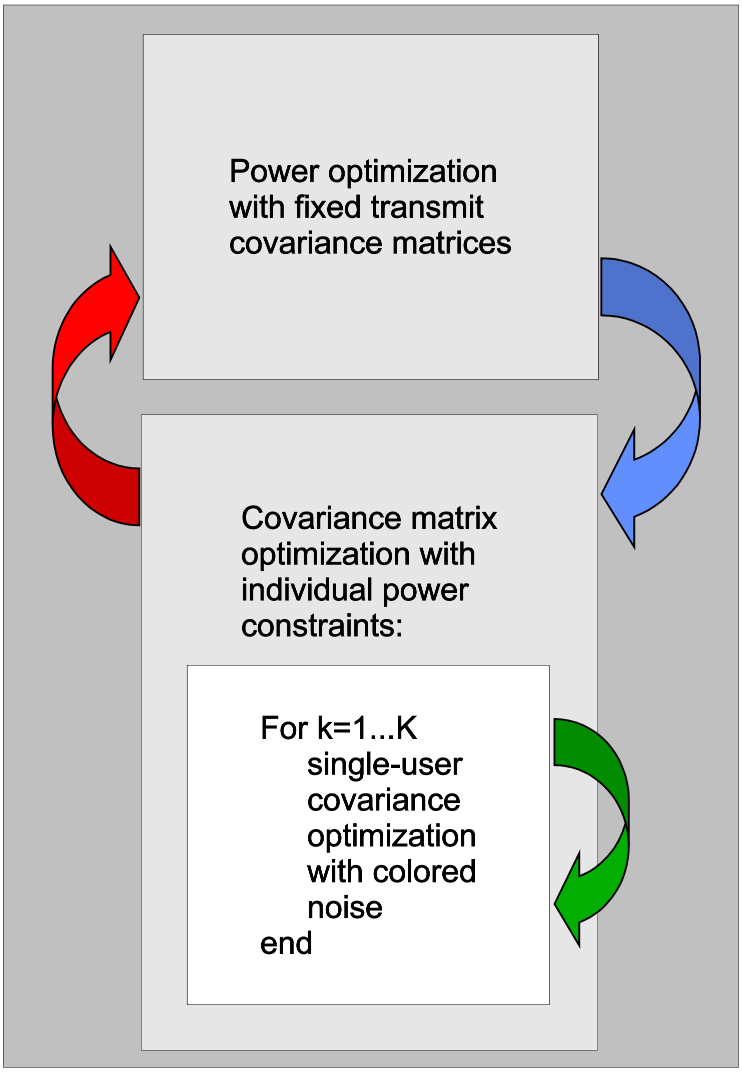

Figure 3.8: Sum Performance Optimisation Algorithm

Let $\boldsymbol{B}_0 = \boldsymbol{\Lambda}_B$ and $\boldsymbol{A}_1 = \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H$. We go similar lines as in the proof of [Fie71]. Note that $\boldsymbol{B}_0$ and $\boldsymbol{A}_1$ do not commutate, i.e. $\boldsymbol{B}_0\boldsymbol{A}_1 \neq \boldsymbol{A}_1\boldsymbol{B}_0$. Next, we parameterise an unitary matrix $\boldsymbol{U} = e^{\epsilon \boldsymbol{S}}$ with specific choice of $\boldsymbol{S}$ with $\boldsymbol{S} = -\boldsymbol{S}^H$. We show that the directional derivative of the parameterised objective function

$$F(\epsilon) = F(\boldsymbol{B}_0^{1/2} e^{\epsilon \boldsymbol{S}} \boldsymbol{A}_1 e^{-\epsilon \boldsymbol{S}} \boldsymbol{B}_0^{1/2}) \tag{3.59}$$

at the point $\epsilon = 0$ is either larger than zero by a specific choice of $\boldsymbol{S}$ or smaller than zero for another choice of $\boldsymbol{S}$. Therefore, it can neither be the global maximum nor the global minimum of the objective function. This is a contradiction and it follows that the maximum and minimum is attained for commutating matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. For the maximum, we chose

$$\begin{aligned} \boldsymbol{S} &= \boldsymbol{B}_0^{1/2} F^{[1]}(\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2})\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1 \\ &\quad -\boldsymbol{A}_1\boldsymbol{B}^{1/2}F^{[1]}(\boldsymbol{B}^{1/2}\boldsymbol{A}_1\boldsymbol{B}^{1/2})\boldsymbol{B}^{1/2} \neq 0. \end{aligned} \tag{3.60}$$

The inequality in (3.60) is fulfilled for all full rank matrices $\boldsymbol{A}_1$ and $\boldsymbol{B}_0$ which do not commute and for all matrix-monotone function which can be represented as in section 2.3.1. Note that $\boldsymbol{S}^H = -\boldsymbol{S}$ in (3.60). The Taylor series expansion of $F(\epsilon)$ is given by

$$F(\epsilon) = F\left(\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2} + \epsilon\left(\boldsymbol{B}_0^{1/2}\boldsymbol{S}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2} - \boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{S}\boldsymbol{B}_0^{1/2}\right) + O(\epsilon^2)\right). \quad (3.61)$$

The trace of the first derivative of $F(\epsilon)$ in (3.61) with respect to ϵ at the point $\epsilon = 0$ is given by

$$\begin{aligned}
& \operatorname{tr} \frac{\partial}{\partial \epsilon} F(\epsilon)\Big|_{\epsilon=0} \\
= \; & \operatorname{tr}\left(F^{[1]}(\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2})\left(\boldsymbol{B}_0^{1/2}\boldsymbol{S}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2} - \boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{S}\boldsymbol{B}_0^{1/2}\right)\right) \\
= \; & \operatorname{tr}\left(\boldsymbol{A}_1\boldsymbol{B}_0^{1/2}F^{[1]}(\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2})\boldsymbol{B}^{1/2}\boldsymbol{S} - \boldsymbol{B}_0^{1/2}F^{[1]}(\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2})\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{S}\right) \\
= \; & \operatorname{tr}\left(\left[\boldsymbol{A}_1\boldsymbol{B}_0^{1/2}F^{[1]}(\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2})\boldsymbol{B}_0^{1/2} - \boldsymbol{B}_0^{1/2}F^{[1]}(\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2})\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\right]\boldsymbol{S}\right) \\
= \; & \operatorname{tr}\left(\boldsymbol{S}\boldsymbol{S}^H\right) > 0. \qquad (3.62)
\end{aligned}$$

The same approach proves that the derivative at point $\boldsymbol{B}_0^{1/2}\boldsymbol{A}_1\boldsymbol{B}_0^{1/2}$ is less than zero for a corresponding choice of the matrix $\boldsymbol{S}$. This is a contradiction and completes the proof.

□

3.4.2 Proof of Theorem 8

We give an overview over the proof: First, we prove that the minimax performance equals the minimax performance of the expression

$$\Phi_I^D = \min_{\substack{\boldsymbol{\Lambda}_Z \succ 0 \\ \operatorname{tr} \boldsymbol{\Lambda}_Z \le n_R \sigma_N^2}} \max_{\substack{\boldsymbol{\Lambda}_Q \succeq 0 \\ \operatorname{tr} \boldsymbol{\Lambda}_Q \le P}} \operatorname{tr} F\left(\boldsymbol{\Lambda}_Z^{-1/2}\boldsymbol{\Lambda}_H^{1/2}\boldsymbol{\Lambda}_Q\boldsymbol{\Lambda}_H^{1/2}\boldsymbol{\Lambda}_Z^{-1/2}\right) \quad (3.63)$$

with channel matrix eigenvalues $\boldsymbol{\Lambda}_H = \boldsymbol{U}_H^H\boldsymbol{H}\boldsymbol{H}^H\boldsymbol{U}_H$. This is done using the Karush-Kuhn-Tucker (KKT) conditions for optimality of $\boldsymbol{Z}$ and $\boldsymbol{Q}$ in the following way: For fixed noise covariance matrix $\boldsymbol{Z}$, the optimal transmit covariance matrix $\boldsymbol{Q}^*$ is characterised by the corresponding KKT conditions. For fixed transmit covariance matrix $\boldsymbol{Q}$, the worst case noise covariance matrix $\boldsymbol{Z}^*$ is characterised by the corresponding KKT conditions. The pair of covariance matrices $(\boldsymbol{Q}^*, \boldsymbol{Z}^*)$ is on the saddle point if and only if for fixed noise covariance matrix $\boldsymbol{Z}^*$, the KKT conditions are fulfilled by transmit covariance matrix $\boldsymbol{Q}^*$ and the other way round, if for fixed transmit covariance matrix $\boldsymbol{Q}^*$, the KKT conditions are fulfilled by the noise covariance matrix $\boldsymbol{Z}^*$. First, we need the following Lemma 11.
Lemma 11: The minimax expression Φ_I in (3.13) and the minimax expression in Φ_I^D in (3.63) are equal for fixed channel matrix $\boldsymbol{H}$, i.e. $\Phi_I = \Phi_I^D$.

The proof of Lemma 11 is given in appendix 3.4.3.

Remark: For fixed noise covariance matrix eigenvalues and channel eigenvalues, the optimum transmit covariance matrix eigenvalues are given by the generalised waterfilling solution (see section 2.3.2). For fixed channel eigenvalues and transmit covariance matrix eigenvalues, the noise eigenvalues which minimise Φ_I^D can be easily found. Let ν denote the rank of the transmit covariance matrix $\mathbf{Q}$, i.e.

$\lambda_1(\boldsymbol{Q}) \geq ... \geq \lambda_\nu(\boldsymbol{Q}) > \lambda_{\nu+1}(\boldsymbol{Q}) = ... = \lambda_n(\boldsymbol{Q}) = 0$. We start with the Lagrangian of the minimisation problem

$$\begin{aligned} \mathcal{L}(\hat{\boldsymbol{\Lambda}}_Z, \xi_k, \mu) &= \sum_{i=1}^{\nu} F\left(\frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}{\lambda_i(\boldsymbol{Z})}\right) + \mu\left(\sum_{l=1}^{\nu}\lambda_l(\boldsymbol{Z}) - n_R\sigma_N^2\right) \\ &\quad + \sum_{k=1}^{\nu}\xi_k\lambda_k(\boldsymbol{Z}). \end{aligned} \tag{3.64}$$

The Lagrangian multiplier ξ_k which ensure that the eigenvalues of the noise covariance matrix $\mathbf{Z}$ are greater than or equal to zero, are all equal to zero, because $\lambda_k(\boldsymbol{Z}) > 0$ for all $1 \leq k \leq \nu$. Otherwise the performance would be infinity. Since the optimisation problem is convex with respect to the noise eigenvalues, we have the necessary and sufficient Karush-Kuhn-Tucker (KKT) condition from (3.64)

$$\frac{\partial\mathcal{L}(\hat{\boldsymbol{\Lambda}}_Z, \mu)}{\partial\lambda_i(\boldsymbol{Z})} = -\frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})F^{[1]}\left(\frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}{\lambda_i(\boldsymbol{Z})}\right)}{\lambda_i(\boldsymbol{Z})^2} + \mu = 0. \tag{3.65}$$

We express (3.65) as

$$F^{[1]}\left(\frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}{\lambda_i^*(\boldsymbol{Z})}\right) = \mu\frac{\lambda_i^*(\boldsymbol{Z})^2}{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}. \tag{3.66}$$

The Lagrangian multiplier μ has to be chosen that $\sum_{i=1}^{n}\lambda_i^*(\boldsymbol{Z}) = n_R\sigma_N^2$.

In the second part of the proof, we further characterise the worst case noise covariance eigenvalues in the following Lemma 12. The SNR is $\rho = \frac{P}{n_R\sigma_N^2}$.
Lemma 12: The worst case noise eigenvalues in (3.66) correspond to the weighted optimal transmit covariance matrix eigenvalues which are given by the water-filling solution

$$\lambda_i^*(\boldsymbol{Z}) = \frac{1}{\rho}\lambda_i^*(\boldsymbol{Q}). \tag{3.67}$$

The proof of Lemma 12 is given in appendix 3.4.4. Finally, we set $\lambda_k^*(\boldsymbol{Q})$ into the performance function and obtain (3.14) completing the proof.

□

3.4.3 Proof of Lemma 11

The singular value decomposition of $\boldsymbol{H}$ is given by $\boldsymbol{H} = \boldsymbol{U}_H\boldsymbol{\Lambda}_H^{1/2}\boldsymbol{V}_H^H$. At first, we show that $\Phi_I \leq \Phi_I^D$. We have

$$\max_{\mathrm{tr}\,(\boldsymbol{Q})\leq P} \mathrm{tr}\, F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2}) = \max_{\mathrm{tr}\,(\boldsymbol{Q})\leq P} \mathrm{tr}\, F(\boldsymbol{Z}^{-1/2}\boldsymbol{U}_H\boldsymbol{\Lambda}_H^{1/2}\boldsymbol{Q}\boldsymbol{\Lambda}_H^{1/2}\boldsymbol{U}_H^H\boldsymbol{Z}^{-1/2}).$$

Now, we choose $\hat{\boldsymbol{Z}} = \boldsymbol{U}_H\boldsymbol{\Lambda}_Z\boldsymbol{U}_H^H$ fixed, then it directly follows

$$\begin{aligned} \Phi_I &= \min_{\mathrm{tr}\,(\boldsymbol{Z})\leq\sigma_N^2 n_R}\ \max_{\mathrm{tr}\,(\boldsymbol{Q})\leq P} \mathrm{tr}\, F(\boldsymbol{Z}^{-1/2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{Z}^{-1/2}) \\ &\leq \min_{\mathrm{tr}\,(\boldsymbol{Z})\leq\sigma_N^2 n_R}\ \max_{\mathrm{tr}\,(\boldsymbol{Q})\leq P} \mathrm{tr}\, F(\boldsymbol{\Lambda}_Z^{-1/2}\boldsymbol{\Lambda}_H^{1/2}\boldsymbol{Q}\boldsymbol{\Lambda}_H^{1/2}\boldsymbol{\Lambda}_Z^{-1/2}) = \Phi_I^D \end{aligned} \tag{3.68}$$

Next, we use Theorem 7 to show that $\Phi_I \geq \Phi_I^D$. With Theorem 7 we have

$$\text{tr } F(\mathbf{\Lambda}_Z^{-1/2}\mathbf{\Lambda}_H^{1/2}\boldsymbol{Q}\mathbf{\Lambda}_H^{1/2}\mathbf{\Lambda}_Z^{-1/2}) \geq \min_{\pi} \sum_{i=1}^{m} \text{tr } F(\lambda_{\pi_i}(\boldsymbol{Z})\lambda_i(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H)). \tag{3.69}$$

The maximum over $\boldsymbol{Q}$ of the term in (3.69) is greater or equal to the term with the choice of $\boldsymbol{U}_Q = \boldsymbol{U}_H^H$, i.e.

$$\max_{\text{tr }(\boldsymbol{Q})\leq P} \text{tr } F(\mathbf{\Lambda}_Z^{-1/2}\mathbf{\Lambda}_H^{1/2}\boldsymbol{Q}\mathbf{\Lambda}_H^{1/2}\mathbf{\Lambda}_Z^{-1/2}) \geq \min_{\pi} \sum_{i=1}^{m} \text{tr } F(\lambda_{\pi_i}(\boldsymbol{Z})\lambda_i(\boldsymbol{H})\lambda_i(\hat{\boldsymbol{Q}})) \tag{3.70}$$

Inequality (3.70) is valid for all $\boldsymbol{Z}$. Therefore, we have

$$\begin{aligned}\min_{\text{tr }(\boldsymbol{Z})\leq n_R\sigma_N^2} \max_{\text{tr }(\boldsymbol{Q})\leq P} \text{tr } F(\mathbf{\Lambda}_Z^{-1/2}\mathbf{\Lambda}_H^{1/2}\boldsymbol{Q}\mathbf{\Lambda}_H^{1/2}\mathbf{\Lambda}_Z^{-1/2}) \geq \\ \min_{\text{tr }(\boldsymbol{Z})\leq n_R\sigma_N^2} \max_{\text{tr }(\boldsymbol{Q})\leq P} \sum_{i=1}^{m} \text{tr } F(\lambda_i(\boldsymbol{Z})\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})).\end{aligned} \tag{3.71}$$

From (3.71) it follows

$$\Phi_I \geq \Phi_I^D. \tag{3.72}$$

From (3.68) and (3.72) follows $\Phi_I = \Phi_I^D$. This completes the proof.

□

3.4.4 Proof of Lemma 12

We denote the optimal transmit covariance matrix eigenvalues with $\lambda_i^*(\boldsymbol{Q})$ and the worst case noise covariance matrix eigenvalues with $\lambda_i^*(\boldsymbol{Z})$. The water-filling solution of the transmit covariance matrix eigenvalues is given for all $\lambda_i^*(\boldsymbol{Q}) > 0$ as

$$F^{[1]}\left(\frac{\lambda_k(\boldsymbol{H})\lambda_k^*(\boldsymbol{Q})}{\lambda_k(\boldsymbol{Z})}\right) = \tilde{\mu}\frac{\lambda_k(\boldsymbol{Z})}{\lambda_k(\boldsymbol{H})} \tag{3.73}$$

with $\tilde{\mu} \geq 0$. We show that the choice $\lambda_i^*(\boldsymbol{Z}) = \frac{1}{\rho}\lambda_i^*(\boldsymbol{Q})$ fulfils both optimality conditions (3.66) and (3.73) simultaneously. This result is derived by computing the Lagrangian multiplier for (3.73) and (3.66) by noting that

$$\tilde{\mu}\lambda_k^*(\boldsymbol{Q}) = \mu\lambda_k^*(\boldsymbol{Z}). \tag{3.74}$$

Finally, the two Lagrangian multiplier are related from (3.74) by

$$\tilde{\mu} = \frac{n_R\sigma_N^2}{P}\mu.$$

□

3.4.5 Lemma 13 and the sketch of its proof

Lemma 13: Consider the operator-monotone function $f(t) = \int_0^\infty st(s+t)^{-1} d\mu(s)$. Then $f(0+) = 0$ and the expression

$$\lim_{t \to 0} \frac{f(t) - f(0)}{t} = f'(0)$$

exists and $0 < f'(0) < +\infty$.
Proof: We give here only the sketch of the proof, because the proof contains many mathematical technicalities which does not provide new information or new insights.

The properties of the function $f(t)$ clearly depend on the measure $d\mu(s)$. We split the positive measure $d\mu(s)$ into two parts

$$d\mu(s) = d\mu_0(s) + d\hat{\mu}(s). \tag{3.75}$$

The first part $d\mu_0(s)$ is a discrete measure with a step at $s = 0$ and the second part $d\hat{\mu}(s)$ is a continuous measure around the zero.

For the first part of Lemma 13, we consider

$$f(t) = \underbrace{\int_0^\infty \frac{ts}{s+t} d\mu_0(s)}_{=0 \text{ for all } t>0} + \int_0^\infty \frac{ts}{s+t} d\hat{\mu}(s). \tag{3.76}$$

From now on, we have to study only the continuous part, i.e. $\mu = \hat{\mu}$. For all $t > 0$ holds $ts(s+t)^{-1} \leq t$. Therefore, we have with finite positive number $c = \hat{\mu}((0, \infty))$

$$0 \leq \lim_{t \to 0} f(t) \leq \lim_{t \to 0} t \cdot c = 0.$$

For the second part of Lemma 13, we have the following identities

$$\begin{aligned} f'(t) &= \lim_{\tau \to 0} \frac{f(t+\tau) - f(t)}{\tau} \\ f'(0) &= \lim_{\tau \to 0} \frac{f(\tau) - f(0)}{\tau} \\ &= \lim_{\tau \to 0} \frac{f(\tau)}{\tau} \\ &= \lim_{\tau \to 0} \int_0^\infty \frac{s}{s+\tau} d\hat{\mu}(s). \end{aligned}$$

Using the continuity of the measure around zero we can show that interchanging the limit and the integral is allowed, and it follows

$$f'(0) = \int_0^\infty d\hat{\mu}(s) = c < \infty.$$

This completes the sketch of the proof.

3.4.6 Proof of Theorem 10

The singular value decomposition of $\boldsymbol{H}$ is given by $\boldsymbol{H} = \boldsymbol{U}_H \boldsymbol{\Lambda}_H^{1/2} \boldsymbol{V}_H^H$. First, we show that $C_{II} \leq C_{II}^D$. We choose $\hat{\boldsymbol{W}} = \boldsymbol{V}_H$. Then it follows that

$$\begin{aligned} C_{II} &\leq \max_{\operatorname{tr}\,(\boldsymbol{Q}) \leq P} \log \frac{\det(\hat{\boldsymbol{W}} \boldsymbol{\Lambda}_Z \hat{\boldsymbol{W}}^H + \boldsymbol{V}_H \boldsymbol{\Lambda}_H^{1/2} \boldsymbol{Q} \boldsymbol{\Lambda}_H^{1/2} \boldsymbol{V}_H^H)}{\det(\hat{\boldsymbol{W}} \boldsymbol{\Lambda}_Z \hat{\boldsymbol{W}}^H)} \\ &= \max_{\operatorname{tr}\,(\boldsymbol{\Lambda}_Q) \leq P} \sum_{i=1}^m \log\left(1 + \frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}{\lambda_i(\boldsymbol{Z})}\right) = C_{II}^D. \end{aligned} \tag{3.77}$$

Next, we use the following Theorem [Fie71, Theorem 1] For positive semidefinite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with eigenvalues $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq ... \geq \beta_n$ it holds

$$\prod_{i=1}^n (\alpha_i + \beta_i) \leq \det(\boldsymbol{A} + \boldsymbol{B}) \leq \prod_{i=1}^n (\alpha_i + \beta_{n+1-i})$$

This Theorem is stronger as Theorem 7, because it characterises the best and worst case permutation. However, it is only applicable to the special case of $F(\boldsymbol{X}) = \log(\boldsymbol{I} + \boldsymbol{X})$.

Hence, applying the Theorem, we show that $C_{II} \geq C_{II}^D$. We have

$$\begin{aligned} C_{II} &\geq \max_{\operatorname{tr}\, Q \leq P} \sum_{i=1}^m \log\left(1 + \frac{\lambda_i(\boldsymbol{\Lambda}_H^{1/2} \boldsymbol{Q} \boldsymbol{\Lambda}_H^{1/2})}{\lambda_i(\boldsymbol{Z})}\right) = \\ &= \max_{\operatorname{tr}\,(\boldsymbol{\Lambda}_Q) \leq P} \sum_{i=1}^m \log\left(1 + \frac{\lambda_i(\boldsymbol{H})\lambda_i(\boldsymbol{Q})}{\lambda_i(\boldsymbol{Z})}\right) = C_{II}^D. \end{aligned} \tag{3.78}$$

From (3.77) and (3.78) follows $C_{II} = C_{II}^D$. This completes the proof.

□

3.4.7 Proof of Theorem 11

At first, we define the maximisation problem

$$\Phi_{IIIa}(\boldsymbol{Z}) = \max_{\operatorname{tr}\,(\boldsymbol{Q}\boldsymbol{Z}) \leq P} \operatorname{tr} F(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H). \tag{3.79}$$

Obviously, we have the following inequality chain

$$\Phi_{III}^D \leq \Phi_{III} \leq \Phi_{IIIa}(\boldsymbol{Z}). \tag{3.80}$$

Next, we find a $\boldsymbol{Z}^*$ which is in $\mathcal{Z}$ and for which $\Phi_{IIIa}(\boldsymbol{Z}^*) = \Phi_{III}^D$, i.e. the optimal input $\boldsymbol{Q}^* = \operatorname{diag}(p_1, ..., p_n)$ for Φ_{III}^D is also optimal for Φ_{III}. The sufficient condition for optimality of $\boldsymbol{Q}^*$ is given by the KKT conditions [BV03]. The Lagrangian of the optimisation problem in (3.79) is given by

$$\begin{aligned} L(\boldsymbol{Q}, \lambda, \boldsymbol{\Psi}) &= \int_0^\infty s \operatorname{tr}\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H \left[s\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right]^{-1}\right) d\mu(s) \\ &\quad -\lambda \operatorname{tr}\,(\boldsymbol{Q}\boldsymbol{Z} - P) + \operatorname{tr} \boldsymbol{\Psi}\boldsymbol{Q}. \end{aligned}$$

As a result, the KKT conditions are given by

$$\begin{aligned} \operatorname{tr} \boldsymbol{Q}\boldsymbol{Z} &\le P \\ \operatorname{tr} \boldsymbol{\Psi}\boldsymbol{Q} &= 0 \\ \frac{\partial}{\partial \boldsymbol{Q}} L(\boldsymbol{Q},\lambda)\Big|_{\boldsymbol{Q}=\boldsymbol{Q}^*} = \int_0^\infty s^2 \left[s\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}^*\boldsymbol{H}^H\right]^{-2} d\mu(s) - \lambda \boldsymbol{Z} + \boldsymbol{\Psi} &= 0 \end{aligned} \tag{3.81}$$

Next, we find a noise covariance matrix such that the corresponding optimal transmit covariance matrix $\boldsymbol{Q}$ is diagonal and corresponds to the optimal power alloction from the corresponding MAC problem. The condition in (3.81) is fulfilled by

$$\boldsymbol{Z}^* = \frac{1}{\lambda}\int_0^\infty s^2 \left[s\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}^*\boldsymbol{H}^H\right]^{-2} d\mu(s) + \frac{1}{\lambda}\boldsymbol{\Psi}. \tag{3.82}$$

Note that the worst case noise covariance matrix and the Lagrangian multiplier $\boldsymbol{\Psi}$ for positive semidefiniteness of $\boldsymbol{Q}$ fulfill

$$\begin{aligned} \operatorname{tr} \boldsymbol{Q}^*\boldsymbol{\Psi} &= 0 \\ \operatorname{tr} \boldsymbol{Z}^*\boldsymbol{Q}^* &= P. \end{aligned}$$

The Lagrangian multiplier λ is derived from (3.25) as

$$\lambda = \int_0^\infty \frac{s^2}{\left(s + \lambda_k(\boldsymbol{H})p_k\right)^2} d\mu(s) \tag{3.83}$$

for all k for which $p_k > 0$. The solution in (3.83) is a generalised waterfilling solution. From (3.82), we have the worst case noise for which $\Phi_{III}^D = \Phi_{III} = \Phi_{IIIa}(\boldsymbol{Z}^*)$ and for which the optimal $\boldsymbol{Q}^* = \operatorname{diag}(p_1, ..., p_n)$ fulfils the trace constraint. This completes the proof.

□

3.4.8 Proof of Theorem 13

The necessary and sufficient condition for the optimality of $\mathbf{Q}_k^*$ for $\mathbf{Z}_k^*$ from (3.47) is with respect to the single-user optimization problem in (3.46)

$$\begin{aligned} \mathbf{H}_i^H \left[\mathbf{Z}_k + \mathbf{H}_k\mathbf{Q}_k^*\mathbf{H}_k^H\right]^{-1} \mathbf{H}_i &= \mu_i \mathbf{I} - \boldsymbol{\Psi}_i \\ tr(\mathbf{Q}_i\boldsymbol{\Psi}_i) &= 0. \end{aligned} \tag{3.84}$$

We conclude that (3.84) holds for all $\mathbf{Q}_i^*$ for all $i \in \{1, ..., K\}$. Since (3.84) exactly corresponds with (3.41), the KKT from (3.45) are fulfilled, too. The KKT which are derived from (3.45) are necessary and sufficient for the global optimum. This completes the proof.

□

3.4.9 Proof of Theorem 14

Suppose that (3.56) and (3.57) are fulfilled. With the KKT condition for the covariance matrix optimisation in (3.41) this implies

$$\rho \boldsymbol{H}_i^H \phi^{[1]} \left(\rho \sum_{k=1}^K p_k^* \boldsymbol{H}_k \boldsymbol{Q}_k^* \boldsymbol{H}_k^H\right) \boldsymbol{H}_i = \bar{\mu}\boldsymbol{I} - \bar{\boldsymbol{\Psi}}_i \quad i \in \mathfrak{I}, . \tag{3.85}$$

where $\bar{\boldsymbol{\Psi}}_i = \boldsymbol{\Psi}_i / p_i^*$ is positive semidefinite. With the condition $tr(\boldsymbol{Q}_i^* \bar{\boldsymbol{\Psi}}_i) = 0$ from (3.41) we can choose $\boldsymbol{Q}_i^* = 0$ for $i \in \{1, ..., K\} \backslash \mathcal{I}$. Therefore the KKT condition $tr(\boldsymbol{Q}_i^* \bar{\boldsymbol{\Psi}}_i) = 0$ is fulfilled for all positive semidefinite $\bar{\boldsymbol{\Psi}}_i$. Since $\bar{\boldsymbol{\Psi}}_i$ is positive semidefinite, (3.85) will also be fulfilled for $i \in \{1, ..., K\} \backslash \mathcal{I}$. This is an immediate consequence of (3.57). Hence, the KKT conditions (3.55) are fulfilled for all i, which implies optimality with respect to the original problem (3.38).

To prove the reverse direction, we assume that $\tilde{\boldsymbol{Q}}_i$ is optimal with respect to the original problem in (3.38). Then, the KKT in (3.55) are fulfilled. Further on, there is a decomposition $\tilde{\boldsymbol{Q}}_i = p_i^* \boldsymbol{Q}_i^*$ which solves the partial problem (3.37) and (3.36), respectively. Otherwise, it would be possible to achieve a performance higher than the optimum of the original problem. This would lead to a contradiction. From the KKT condition (3.55) immediately follows the KKT in (3.51) and (3.41). This completes the proof.

□

3.4.10 Proof of Theorem 15

Using the KKT conditions from (3.51), the necessary and sufficient condition in Theorem 15 can be easily derived. Next, we will show that the users are sorted by their maximum channel matrix eigenvalues and not by the Frobenius norm. Hence, we start with the KKT conditions for the power optimisation with fixed optimal transmit covariance matrices $\boldsymbol{Q}_1, \boldsymbol{Q}_2$ with tr $\boldsymbol{Q}_1 =$ tr $\boldsymbol{Q}_2 = 1$. The first KKT condition is given in (3.55)

$$\begin{aligned} \rho \boldsymbol{H}_1^H \phi^{[1]} \left[\rho p_1 \boldsymbol{H}_1 \boldsymbol{Q}_1 \boldsymbol{H}_1^H + \rho p_2 \boldsymbol{H}_2 \boldsymbol{Q}_2 \boldsymbol{H}_2^H \right] \boldsymbol{H}_1 &= \mu \boldsymbol{I} - \boldsymbol{\Psi}_1 \\ \rho \boldsymbol{H}_2^H \phi^{[1]} \left[\rho p_1 \boldsymbol{H}_1 \boldsymbol{Q}_1 \boldsymbol{H}_1^H + \rho p_2 \boldsymbol{H}_2 \boldsymbol{Q}_2 \boldsymbol{H}_2^H \right] \boldsymbol{H}_2 &= \mu \boldsymbol{I} - \boldsymbol{\Psi}_2 \end{aligned} \tag{3.86}$$

The normalised covariance matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are the optimum transmit covariance matrices. The Lagrangian multipliers $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ are positive definite. As a result, we have for all $\boldsymbol{w} : ||\boldsymbol{w}||^2 = 1$

$$\rho \boldsymbol{w}^H \left(\boldsymbol{H}_1^H \phi^{[1]} \left[\rho p_1 \boldsymbol{H}_1 \boldsymbol{Q}_1 \boldsymbol{H}_1^H + \rho p_2 \boldsymbol{H}_2 \boldsymbol{Q}_2 \boldsymbol{H}_2^H \right] \boldsymbol{H}_1 \right) \boldsymbol{w} \leq \mu ||\boldsymbol{w}||^2 = \mu. \tag{3.87}$$

Taking the supremum of the LHS in (3.87) with the eigenvector of $\boldsymbol{H}_1$ which equals the maximum eigenvalue of $\boldsymbol{H}_1^H \phi^{[1]} \left[\rho p_1 \boldsymbol{H}_1 \boldsymbol{Q}_1 \boldsymbol{H}_1^H + \rho p_2 \boldsymbol{H}_2 \boldsymbol{Q}_2 \boldsymbol{H}_2^H \right] \boldsymbol{H}_1$, we obtain

$$\rho \lambda_{max} \left(\boldsymbol{H}_1^H \phi^{[1]} \left[\rho p_1 \boldsymbol{H}_1 \boldsymbol{Q}_1 \boldsymbol{H}_1^H + \rho p_2 \boldsymbol{H}_2 \boldsymbol{Q}_2 \boldsymbol{H}_2^H \right] \boldsymbol{H}_1 \right) \leq \mu. \tag{3.88}$$

Let the transmit covariance matrix of user one given as

$$\boldsymbol{Q}_1 = \sum_{l=1}^{L_1} \lambda_{\boldsymbol{Q}_1}(l) \boldsymbol{u}_l^{(1)} \boldsymbol{u}_l^{(1)H}$$

with $rank(\boldsymbol{Q}_1) = L_1$. Then from the optimality condition $tr(\boldsymbol{Q}_1 \boldsymbol{\Psi}_1) = 0$ follows $\boldsymbol{u}_l^{(1)H} \boldsymbol{\Psi}_1 \boldsymbol{u}_l^{(1)} = 0$ for all $1 \leq l \leq L$. Using (3.86) this shows

$$\rho \boldsymbol{u}_l^{(1)H} \boldsymbol{H}_1^H \phi^{[1]} \left[\rho p_1 \boldsymbol{H}_1 \boldsymbol{Q}_1 \boldsymbol{H}_1^H + \rho p_2 \boldsymbol{H}_2 \boldsymbol{Q}_2 \boldsymbol{H}_2^H \right] \boldsymbol{H}_1 \boldsymbol{u}_l^{(2)} = \mu. \tag{3.89}$$

With (3.88) and (3.89) we have

$$\rho\lambda_{max}\left(\boldsymbol{H}_1^H\phi^{[1]}\left[\rho p_1\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H+\rho p_2\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right]\boldsymbol{H}_1\right)=\mu. \tag{3.90}$$

By analogy arguments we obtain for user 2

$$\rho\lambda_{max}\left(\boldsymbol{H}_2^H\phi^{[1]}\left[\rho p_1\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H+\rho p_2\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right]\boldsymbol{H}_2\right)=\mu.$$

As a result with (3.86), we obtain the optimality condition which is fulfilled by optimal p_1 and p_2

$$\begin{aligned}&\lambda_{max}\left(\boldsymbol{H}_1^H\phi^{[1]}\left[\rho p_1\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H+\rho p_2\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right]\boldsymbol{H}_1\right)\\&=\lambda_{max}\left(\boldsymbol{H}_2^H\phi^{[1]}\left[\rho p_1\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H+\rho p_2\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right]^{-2}\boldsymbol{H}_2\right).\end{aligned} \tag{3.91}$$

Next, we study the convergence behaviour of the maximum eigenvalues in (3.91) for $P\to 0$. We can show by Lemma 13 that the maximum eigenvalues in (3.91) converge to

$$\begin{aligned}\lambda_{max}\left(\boldsymbol{H}_1^H\phi^{[1]}\left[\rho p_1\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H+\rho p_2\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right]\boldsymbol{H}_1\right)&\longrightarrow_{\rho\to 0}\lambda_{max}(\boldsymbol{H}_1\boldsymbol{H}_1^H)\\\lambda_{max}\left(\boldsymbol{H}_2^H\phi^{[1]}\left[\rho p_1\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H+\rho p_2\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right]\boldsymbol{H}_2\right)&\longrightarrow_{\rho\to 0}\lambda_{max}(\boldsymbol{H}_2\boldsymbol{H}_2^H)\end{aligned}$$

Therefore, from (3.91) follows for $\rho\to 0$ that

$$\lambda_{max}(\boldsymbol{H}_1\boldsymbol{H}_1^H)=\lambda_{max}(\boldsymbol{H}_2\boldsymbol{H}_2^H).$$

This proves the first part of the Theorem 15. The second part of the Theorem 15 follows from the KKT condition already used in (3.91) for $p_2=0$. In order to prove the last part of the theorem, we start with the RHS of (3.91) for $p_2=0$. With the singular value decomposition of the channel matrix $\boldsymbol{H}_1=\boldsymbol{V}\boldsymbol{\Lambda}_H\boldsymbol{U}^H$, we can write

$$\begin{aligned}&\rho\lambda_{max}\left(\boldsymbol{H}_2^H\phi^{[1]}\left[\hat{\rho}\boldsymbol{H}_1\boldsymbol{Q}_1^s(\hat{\rho})\boldsymbol{H}_1^H\right]\boldsymbol{H}_2\right)\\&=\rho\max_{||\boldsymbol{u}||^2=1}\boldsymbol{u}^H\boldsymbol{H}_2^H\boldsymbol{V}_1\phi^{[1]}\left[\hat{\rho}\mathrm{diag}(p_k\lambda_k(\boldsymbol{H}))\right]\boldsymbol{V}_1^H\boldsymbol{H}_2\boldsymbol{u}.\end{aligned} \tag{3.92}$$

The set $\mathcal{M}$ is defined as $\mathcal{M}=\{\boldsymbol{m}:\boldsymbol{m}^H=\boldsymbol{u}^H\boldsymbol{H}_2^H\boldsymbol{V}_1\Big|||\boldsymbol{u}||^2=1\}$. The optimal single-user transmit strategy $\boldsymbol{Q}_1^s(\hat{\rho})$ was derived in Lemma 3 as $\rho\lambda_k\phi^{[1]}(\rho p_k\lambda_k)=\mu$ for all active $k\le R(\rho)$. This yields for (3.92)

$$\begin{aligned}\rho\max_{\boldsymbol{w}\in\mathcal{M}}\boldsymbol{w}^H\phi^{[1]}\left[\hat{\rho}\mathrm{diag}(p_k\lambda_k(\boldsymbol{H}))\right]\boldsymbol{w}&=\max_{\boldsymbol{w}\in\mathcal{M}}\mu\sum_{l=1}^{R(\hat{\rho})}\frac{||\boldsymbol{w}_l||^2}{\lambda_l(\boldsymbol{H})}\\&=\mu\max_{\boldsymbol{w}\in\mathcal{M}}\sum_{l=1}^{R(\hat{\rho})}\frac{||\boldsymbol{w}_l||^2}{\lambda_l(\boldsymbol{H})}.\end{aligned} \tag{3.93}$$

By assumption $\hat{\rho}$ was chosen such that the LHS of (3.92) is smaller than or equal to μ. It follows

$$\max_{\boldsymbol{w}\in\mathcal{M}}\sum_{l=1}^{R(\hat{\rho})}\frac{||\boldsymbol{w}_l||^2}{\lambda_l(\boldsymbol{H})}\le 1. \tag{3.94}$$

The rank function $R(\rho)$ is monotonic increasing with ρ. Therefore, (3.94) holds for all $\rho \leq \hat{\rho}$. This completes the proof.

$\square$

4 Conclusions and future research

4.1 Conclusions

In this thesis, the performance of single-user and multiple-user multiple-antenna wireless systems was studied. We considered the average and instantaneous sum capacity as well as the average normalised MSE and the normalised sum MSE as performance metrics. The following topics were covered and the following results were derived:

- Based on the well established MIMO block flat fading channel model with correlated transmit and receive antennas, we provided a partial order of correlation scenarios based on majorization. This allowed us to compare the achievable performance of correlated MIMO systems.

- We identified the general structure of the performance functions in MIMO systems. It was possible to write the considered performance measures as the trace of a matrix-monotone function. Using Löwner's representation of those functions, we developed a general theory for solving performance optimisation problems in MIMO systems.

- Under three different types of CSI at the transmitter, namely perfect CSI, knowledge of the correlation matrices, and no CSI, we derived the optimal transmit strategies with respect to the class of performance functions for single user MIMO systems. We showed that the optimal signal processing structure consists of a Gaussian codebook, a power allocation entity, and a set of beamformers.

- The impact of the correlation on the achievable average performance of single-user MIMO systems depends on the type of CSI available at the transmitter. We characterised the behaviour of the average performance and quantified the loss or gain due to correlation.

- The step in direction of multi-user MIMO performance analysis was set up by considering a point-to-point link with inter-cell and intra-cell interference modelled by coloured noise. We modelled three representative scenarios by different constraints on the noise covariance matrix. The transmitter was assumed to have perfect CSI and knowledge of the noise variance. We proved that worst case noise robs the CSI and deconstructs the cooperation at the transmit side.

- Incooperating the transmit strategies of all participating users, we optimised the sum performance of the multi-user MIMO system and derived an iterative algorithm which efficiently computes the optimal transmit strategy. In addition to this, we characterised the optimal transmit strategy at low SNR values and derived the single-user optimality range.

Note, that the result in chapter 3 can be directly applied to wired communications systems, e.g. DSM for DSL [YGC01].

4.2 Future research

There are countless possibilities for future research in the important and interesting area of (network) information theory and wireless communications. In the following, we will mention two representative and challenging directions:

4.2.1 Outage capacity and delay limited capacity

In contrast to the average quantities in (1.8), (1.9), and (1.13) a more detailed measure of information provides the *outage probability* which is described next. Let us assume that the system works at a given transmission rate R. Obviously, an *outage* occurs if the information cannot be reliably transmitted with this rate, i.e. if the instantaneous mutual information $f(\boldsymbol{Q}, \boldsymbol{H}, \rho)$ is below that value R. The probability of this event is called the *outage probability* of the MIMO system

$$\Omega(\rho, \boldsymbol{Q}) = Pr\left[f(\boldsymbol{Q}, \boldsymbol{H}, \rho) < R\right] \tag{4.1}$$

for transmit strategy $\boldsymbol{Q}$ or with respect to the instantaneous channel capacity

$$\Omega(\rho) = Pr\left[c(\boldsymbol{H}, \rho) < R\right].$$

In [Tel99], the outage probability was derived and analysed by numerical simulations. By using the outage probability as a measure for performance evaluation, one can design a system which works for a given percentage of time. On the other hand, for a given percentage of outage ϵ, the rate R with the constraint that the outage probability is smaller than or equal to ϵ, i.e.

$$\mathbb{R}(\epsilon) = \max R \;\text{ s.t. }\; \Omega(\rho) \leq \epsilon.$$

The quantity $\mathbb{R}(\epsilon)$ is called the *ϵ-capacity* of the channel. The 0-capacity is called the *delay-limited capacity*, i.e.

$$\mathbb{D} = \max_{\boldsymbol{Q}:\mathbb{E}\,\mathrm{tr}\,\boldsymbol{Q} \leq P} R \;\text{ s.t. }\; Pr\left[f(\boldsymbol{Q}, \boldsymbol{H}, \rho) < R\right] = 0. \tag{4.2}$$

The delay-limited capacity was derived and analysed in [CTB99] for SISO and single-sided diversity channels. The work was extended in [BCT01] for block-fading channels with multiple antennas. The notions of capacity vs. outage and delay-limited capacity were discussed in [Ber00, Chapter 3].

These measures of instantaneous, outage, and delay limited performance have to be analysed for single-user MIMO systems. Based on these performance measures it is then possible to incooperate higher layer aspects into the analysis and design of MIMO systems [BJH03]. The impact of correlation and the optimal transmission strategies for MISO systems were studied in [JB03a, BJ04d]. A generalization to MIMO systems is an open research problem.

4.2.2 Individual QoS requirements and network optimisation

Both individual rates without and with SIC R_k and R_k^{SIC} respectively, of one user k depend on the rates of the other users. Consider a two user MAC. The achievable rates with SIC if user one is detected first and subtracted before user two is detected are given by

$$R_1^{SIC} = \log\det\left(\boldsymbol{I} + \rho\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H + \rho\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right) - \log\det\left(\boldsymbol{I} + \rho\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right),$$

$$R_2^{SIC} = \log\det\left(\boldsymbol{I} + \rho\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right).$$

In general, the rate of user two can be expressed as a function of user one and vice versa. It is possible to plot the rate of user two over rate of user one. This results in the capacity region of the two user MAC. For all rate tupes (r_1, r_2) inside the capacity region, there exist codes of length n for which the maximal probability of error tends to zero with increasing code length $n \to 0$ [CT91, section 14.3]. In this case, it is assumed that the receiver applies the optimal detector.

Each user drives its own service, e.g. speech connection-oriented, or a multi-media stream-oriented service, which requires a minimum quality of service (QoS). On the physical layer, these QoS requirements transform into SINR or rate requirements. On the DLL, there are delay and reliability constraints. The problem statement then is: Allocate minimum sum transmit power in order to fulfill the QoS requirements of all users. This problem has been solved for multiuser SIMO systems in [Sch02]. Actually, the extension to MIMO MAC and BC remains an open research problem [BSJ02, JB02c]. Furthermore, higher layer aspects, like scheduling based on this spatial characterisation of the physical layer lead to the newly coined cross-layer design and network optimisation [GW02, Yeh03, BJH03]. This will be a major research topic in the future.

4.2.3 Extension to multi-carrier communications

If we extend the class of channels to frequency-selective channels, one would think that additional efforts are necessary in order to deal with intersymbol-interference. Orthogonal Frequency Division Multiplexing (OFDM) is an effective engineering approach to deal with frequency-selective channels. From an information theoretic point of view, we have to exploit the additional spectral dimension in order to improve the performance, throughput, and reliability of the transmission.

The difference of the spectral dimension in contrast to the spatial dimension is that the carriers are orthogonal and does not interfere with each other (ideally). The spatial signals do interfere. In order to better understand the properties of the MIMO OFDM channel [BGP02] it will be necessary to study the impact of the power delay profile on the average and instantaneous performance. In addition to this, the correlations between the carriers depend on the number of taps and on the power delay profile. The analysis of the performance metrics from section 4.2.1 with respect to power delay profile as well as spatial correlation will provide further insights into the maximum achievable performance and throughput. Furthermore, it will lead to the development of optimum transmit strategies which exploit the spectral dimension as well as the spatial dimension.

Publication List

[BJ02a] H Boche and E. A. Jorswieck. Analysis of diversity and multiplexing trade-off for multi-antenna systems with covariance feedback. *Vehicular Technology Conference*, 2:864–868, Sept. 2002.

[BJ02c] H. Boche and E. A. Jorswieck. On the optimality range of beamforming for MIMO systems with covariance feedback. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E85-A(11):2521–2528, November 2002.

[BJ02d] H. Boche and E. A. Jorswieck. On the schur-concavity of the ergodic and outage capacity with respect to correlation in multi-antenna systems with no csi at the transmitter. *Proceedings of Allerton Conference*, 2002.

[BJ02e] H. Boche and E. A. Jorswieck. Optimal transmit strategies for MIMO systems with partial channel state information. *Proceedings of SPIE International Symposium ITCOM*, 2002.

[BJ02f] H. Boche and E. A. Jorswieck. Optimal transmit strategies for mimo systems with partial channel state information. *Proc. of SPIE International Symposium ITCOM, Boston, Massachusetts, USA*, 2002.

[BJ02g] H. Boche and E. A. Jorswieck. Sum capacity optimization of the MIMO gaussian MAC. *5th International Symposium on Wireless Personal Multimedia Communications, invited paper*, 1:130–134, Oct. 2002.

[BJ03a] H. Boche and E. A. Jorswieck. Channel aware scheduling for multiple antenna multiple access channels. *Proc. of Asilomar Conference*, 2003.

[BJ03b] H. Boche and E. A. Jorswieck. Multiuser MIMO systems: Worst case noise and transmitter cooperation. *Proc. of ISSSPIT, invited paper*, 2003.

[BJ03c] H. Boche and E. A. Jorswieck. Optimal power allocation for MISO systems and complete characterization of the impact of correlation on the capacity. *Proc. of IEEE ICASSP*, 4:IV – 373–6, April 2003.

[BJ03d] H. Boche and E. A. Jorswieck. Optimum power allocation and complete characterization of the impact of correlation on the capacity of miso systems with different csi at the transmitter. *Proc. of IEEE International Symposium on Information Theory*, pages 353–353, July 2003.

[BJ03e] H. Boche and E. A. Jorswieck. Uplink sumrate maximization with different types of channel state information at the transmitters. *Proc. of IEEE ISSPIT*, 2003.

[BJ04a] H. Boche and E. A. Jorswieck. On the ergodic capacity as a function of the correlation properties in systems with multiple transmit antennas without csi at the transmitter. *will appear in IEEE Trans. on Communications, April*, 2004.

[BJ04b] H. Boche and E. A. Jorswieck. Multiple antenna multiple user channels: Optimisation in low SNR. *Proceedings of IEEE Wireless Communications and Networking Conference*, 2004.

[BJ04d] H. Boche and E. A. Jorswieck. Outage probability of multiple antenna systems: Optimal transmission and impact of correlation. *Proc. IEEE International Zürich Seminar*, 2004.

[BJ04e] H. Boche and E. A. Jorswieck. Optimization of Matrix Monotone Functions: Saddle-Point, Worst Case Noise Analysis, and Applications. *Will appear in Proc. of IEEE International Symposion on Information Theory*, 2004.

[BJ04f] H. Boche and E. A. Jorswieck. Universal Approach to Performance Optimization of Multiuser MIMO Systems *Will appear in Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications*, 2004.

[BJ04g] H. Boche and E. A. Jorswieck. Analysis of Multiuser MIMO Systems with MMSE Receiver Based on Worst Case Noise *Proc. of ITG Workshop on Smart Antennas*, 2004

[BJ04h] H. Boche and E. A. Jorswieck. Multiple Antenna Multiple User Channels: Optimisation in low SNR *Proc. of IEEE Wireless Communications and Networking Conference*, 2004.

[BJH03] H. Boche, E. A. Jorswieck, and T. Haustein. Channel aware scheduling for multiple antenna multiple access channels. *Proc. Asilomar Conference*, 2003.

[BSJ02] H. Boche, M. Schubert, and E. A. Jorswieck. Trace balancing for multiuser MIMO downlink transmission. *Proc. Asilomar Conference*, 2002.

[BSJ03] H. Boche, M. Schubert, and E. A. Jorswieck. Throughput maximization for the multiuser MIMO broadcast channel. *Proc. of IEEE ICASSP*, 4:808–8011, April 2003. invited paper.

[HvHJ+01] T. Haustein, C. von Helmolt, E. A. Jorswieck, V. Jungnickel, U. Krueger, and V. Pohl. Bit error rates for a MIMO system in rayleigh and rician channels. *IEEE 54th Vehicular Technology Conference*, 4:1984–1987, Oct. 2001.

[HvHJ+02] T. Haustein, C. von Helmolt, E. A. Jorswieck, V. Jungnickel, and V. Pohl. Performance of MIMO systems with channel inversion. *IEEE 55th Vehicular Technology Conference*, 1:35–39, May 2002.

[JB02a] E. A. Jorswieck and H. Boche. Characterization of the optimality-range of beamforming for multiple transmit and receive antennas with covariance-feedback. *In proceedings of 36th Annual Conference on Information Sciences and Systems*, 2002.

[JB02b] E. A. Jorswieck and H. Boche. On transmit diversity with imperfect channel state information. *Proceedings of IEEE ICASSP*, 3:III–2181–III–2184, May 2002.

[JB02c] E. A. Jorswieck and H. Boche. Rate balancing for the multi-antenna gaussian broadcast channel. *Proceedings of the 7th IEEE International Symposium on Spread Spectrum Techniques and Applications*, 2:545–549, Sept. 2002.

[JB03a] E. A. Jorswieck and H. Boche. Behavior of outage probability in MISO systems with no channel state information at the transmitter. *Proceedings of IEEE Information Theory Workshop*, pages 353–356, 2003.

[JB03b] E. A. Jorswieck and H. Boche. Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *will appear in IEEE Trans. on Wireless Communication*, 2003.

[JB03c] E. A. Jorswieck and H. Boche. On the impact of correlation on the capacity in MIMO systems without CSI at the transmitter. *Proc. of Conference on Information Sciences and Systems*, 2003.

[JB03d] E. A. Jorswieck and H. Boche. On the optimal transmission strategy for the MIMO MAC with MMSE receiver. *Proc. of IEEE ICASSP*, 5:109–112, April 2003.

[JB03e] E. A. Jorswieck and H. Boche. Optimal transmission strategy for multiple antenna systems with uninformed transmitter and correlation. *Proc. of Asilomar Conference*, 2003.

[JB03f] E. A. Jorswieck and H. Boche. Optimal transmission with imperfect channel state information at the transmit antenna array. *Wireless Personal Communications*, 27(1):33–56, 2003.

[JB03g] E. A. Jorswieck and H. Boche. Transmission strategies for the MIMO MAC with MMSE receiver: Average MSE optimization and achievable individual MSE region. *IEEE Trans. on Signal Processing*, 51(11):2872–2881, November 2003.

[JB04c] E. A. Jorswieck and H. Boche. Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information. *will appear in IEEE Trans. on Signal Processing*, 2004.

[JB04d] E. A. Jorswieck and H. Boche. Performance analysis of capacity of multiuser mimo systems based on worst case noise behaviour. *submitted to Eurasip WCN Special Issue MIMO Multiuser systems*, 2004.

[JB04e] E. A. Jorswieck and H. Boche. Average Mutual Information in Spatial Correlated MIMO Systems with Uninformed Transmitter *Proc. of Conference on Information Sciences and Systems*, 2004.

[JBS04] E. A. Jorswieck, H. Boche, and M. Schubert. Analysis of multiuser mimo systems with mmse receiver based on worst case noise. *will appear in ITG Workshop on Smart Antennas*, 2004.

[JHJ+01] V. Jungnickel, T. Haustein, E. A. Jorswieck, V. Pohl, and C. von Helmolt. Performance of a mimo system with overlay pilots. *Global Telecommunications Conference*, 1:25–29, Nov. 2001.

[JHJ+02] E. A. Jorswieck, T. Haustein, V. Jungnickel, V. Pohl, and C. von Helmolt. On the performance of signal detection algorithms in flat-fading and frequency-selective MIMO channels. *IEEE 55th Vehicular Technology Conference*, 1:40–44, May 2002.

[JHJvH01] V. Jungnickel, T. Haustein, E. A. Jorswieck, and C. von Helmolt. A MIMO WLAN based on linear channel inversion. *IEE Seminar on MIMO: Communications Systems from Concept to Implementations*, pages 20/1–20/6, Dec. 2001.

[JHJvH02] V. Jungnickel, T. Haustein, E. A. Jorswieck, and C. von Helmolt. On linear pre-processing in multi-antenna systems. *IEEE Global Telecommunications Conference*, 1:1012–1016, Nov. 2002.

[JS04] E. A. Jorswieck and A. Sezgin. Impact of spatial correlation on the performance of orthogonal space-time block codes. *IEEE Communications Letters*, 8(1):21–23, Jan. 2004.

[JWJH01] E. A. Jorswieck, G. Wunder, V. Jungnickel, and T. Haustein. Inverse eigenvalue statistics for rayleigh and ricean MIMO channels. *IEE Seminar* *on MIMO: Communications Systems from Concept to Implementations*, pages 3/1–3/6, Dec. 2001.

[PJJvH02] V. Pohl, V. Jungnickel, E. A. Jorswieck, and C. von Helmolt. Zero forcing equalizing filter for mimo channels with intersymbol interference. *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 3:1037–1041, Sept. 2002.

[SJ02] A. Sezgin and E. A. Jorswieck. Joint decoding and channel estimation for low-complexity stc. *Proc. of Asilomar*, 2002.

[SJ03a] A. Sezgin and E. A. Jorswieck. Maximum diversity detection for layered space-time codes. *Vehicular Technology Conference*, 2:833–837, April 2003.

[SJ03b] A. Sezgin and E. A. Jorswieck. On optimal constellations for quasi-orthogonal space-time codes. *IEEE ICASSP*, 4:IV – 345–8, April 2003.

[SJB03] A. Sezgin, E. A. Jorswieck, and H. Boche. Performance criteria analysis and further performance results for quasi-orthogonal space-time block codes. *Proc. of IEEE ISSPIT*, 2003.

[SJB04] A. Sezgin, E. A. Jorswieck, and H. Boche. On exit-chart analysis of coherent and non-coherent space-time codes. *will appear in ITG Workshop on Smart Antennas*, 2004.

[SJJ03] A. Sezgin, E. A. Jorswieck, and P. Jung. Analysis of turbo-coded mimo systems for noncoherent communications using exit-charts. *Proc. of Conference on Information Sciences and Systems*, 2003.

References

[3gp01] 3GPP TR 25.848 V4.0.0. 2001.

[3gp02] 3GPP TR 24.858 V1.0.4. Jan. 2002.

[AFTS01] I. A. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz). The capacity of dicrete-time memoryless Rayleigh fading channels. *IEEE Transactions on Information Theory, vol. 47, no. 4, May*, 2001.

[BMO04] S. Bergman, C. Martin, and B. Ottersten. Bit and Power Loading for Spatial Multiplexing using Partial Channel State Information. *ITG Workshop on Smart Antennas*, 2004.

[BBT59] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Ann. Math. Stat.*, 30:1229–1241, Dec. 1959.

[BCT01] E. Biglieri, G. Caire, and G. Taricco. Limiting performance of block-fading channels with multiple antennas. *IEEE Trans. on Information Theory*, 47(4):1273–1289, May 2001.

[Ber00] R. A. Berry. *Power and delay trade-offs in fading channels*. PhD thesis, MIT, Cambridge, MA, June 2000.

[BGP02] H. Bölcskei, D. Gesbert, and A. J. Paulraj. On the capacity of ofdm-based spatial multiplexing systems. *IEEE Trans. on Communications*, 50(2):224–234, Feb. 2002.

[Bha97] R. Bhatia. *Matrix Analysis*. Springer - Verlag, 1997.

[Blu03] R. S. Blum. MIMO capacity with interference. *IEEE Journal on Sel. Areas in Communications*, 21(5):793–801, June 2003.

[BS01] H. Boche and M. Schubert. A new approach to power adjustment for spatial covariance based downlink beamforming. *Proc. of IEEE ICASSP*, 5:2957–2960, May 2001.

[BS02a] H. Boche and M. Schubert. A general theory for uplink and downlink beamforming. *Proceedings of IEEE 56th Vehicular Technology Conference*, 1:87–91, Sept. 2002.

[BS02b] H. Boche and M. Schubert. Solution of the SINR downlink beamforming problem. *Proc. Conf. on Information Sciences and Systems (CISS)*, March 2002.

[BV03] S. Boyd and L. Vanderberghe. Convex optimization. *Course reader for EE364 (Stanford) and EE236 (UCLA)*, 2003.

[CFGV02] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela. Keyholes, correlations, and capacities of multielement transmit and receive antennas. *IEEE Trans. on Wireless Communications*, 1(2):361–368, April 2002.

[CFV00] D. Chizhik, G. J. Foschini, and R. A. Valenzuela. Capacities of multielement transmit and receive antennas. *IEE Electronics Letters*, 36:1099–1100, June 2000.

[Cos83] M. Costa. Writing on dirty paper. *IEEE Trans. on Information Theory*, 29:439–441, May 1983.

[Cov72] T. M. Cover. Broadcast channels. *IEEE Trans. Inform. Theory*, 18(1):2–14, Jan. 1972.

[Cov98] T. M. Cover. Comments on broadcast channels. *IEEE Trans. Inform. Theory*, 44(6):2524–2530, Oct. 1998.

[CS01] G. Caire and S. Shamai. On the multiple-antenna broadcast channel. *36th Asilomar Conference on Signals, Systems and Computers*, 2001.

[CT88] Thomas M. Cover and Joy A. Thomas. Determinant inequalities via information theory. *SIAM J. Matrix Analysis*, 9(3):384–392, July 1988.

[CT91] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley & Sons, 1991.

[CTB99] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Trans. on Information Theory*, 45(5):1468–1489, July 1999.

[CTK02] C.-N. Chuah, D. N. C. Tse, and J. M. Kahn. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. on Information Theory*, 48(3):637–650, March 2002.

[DC01] S. N. Diggavi and T. M. Cover. The worst additive noise under a covariance constraint. *IEEE Trans. on Information Theory*, 47(7):3072–3081, November 2001.

[Fan53] Ky Fan. Minimax theorems. *Proc. N. A. S.*, 39:42–47, 1953.

[FG98] G.J. Foschini and M.J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications 6: 311-335*, 1998.

[Fie71] M. Fiedler. Bounds for the determinant of the sum of hermitian matrices. *Proceedings of the American Mathematical Society*, 30(1):27–31, September 1971.

[GBGP02] D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj. Outdoor mimo wireless channels: Models and performance prediction. *IEEE Trans. on Information Theory*, 50(12):1926–1934, Dec. 2002.

[GHIM01] M. Godavarti, A. O. Hero-III, and T. Marzetta. Min-capacity of a multiple-antenna wireless channel in a static rician fading environment. *Proceedings of IEEE Symposion on Information Theory*, page 57, 2001.

[GN00] A. K. Gupta and D. K. Nagar. *Matrix variate distributions*. Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics, 2000.

[Goo63] N. R. Goodman. The distribution of the determinant of a complex wishart distributed matrix. *Annals of Mathematical Statistics*, 34:178–180, 1963.

[GP94] D. Gerlach and A. Paulraj. Adaptive transmitting antenna methods for multipath environments. *Global Telecommunications Conference*, 1:425–429, Dec. 1994.

[GR80] I. S. Gradshteyn and I. M. Ryzhik. *Tables of Integrals, Series and Products*. Academic Press, 1980.

[GV97] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. on Information Theory*, 43(6):1986–1992, November 1997.

[GW02] A. J. Goldsmith and S. B. Wicker. Design challenges for energy-constrained ad hoc wireless networks. *IEEE Wireless Communications Journal*, 9(4):8–27, Aug. 2002.

[HB03] T. Haustein and H. Boche. On optimal power allocation and bit-loading strategies for the MIMO transmitter with channel knowledge. *Proceedings of IEEE ICASSP 2003, vol. IV, pp. 405-409*, 2003.

[HP82] F. Hansen and Gert K. Pedersen. Jensen's inequality for operators and Löwner's theorem. *Math. Ann.*, 258:229–241, 1982.

[HJ85] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[HJ91] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[Jak74] W.C. Jakes. *Microwave Mobile Communication*. New York: IEEE Press, 1974.

[JG01a] S. A. Jafar and A. Goldsmith. Vector MAC capacity region with covariance feedback. *International Symposium on Information Theory*, page 54, 2001.

[JG01b] S.A. Jafar and A. Goldsmith. On optimality of beamforming for multiple antenna systems with imperfect feedback. *International Symposium on Information Theory*, page 321, 2001.

[JG02] S. Jafar and A. Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems. *submitted to Trans. on Wireless Communication*, 2002.

[JG03] S. Jafar and A. Goldsmith. Multiple antenna capacity in correlated rayleigh fading with channel covariance information. *submitted to IEEE Trans. on Wireless Communication*, 2003.

[JVG01] S.A. Jafar, S. Vishwanath, and A. Goldsmith. Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback. *International Conference onf Communications*, 7:2266–2270, June 2001.

[KF70] A. Kolmogorov and S. Fomin. *Introductory Real Analysis*. Dover Publications, Inc New York, 1970.

[KH95] R. Knopp and P.A. Humblet. Information capacity and power control in single-cell multiuser communications. *International Conference on Communications*, 1:331–335, June 1995.

[Lö34] K. T. Löwner. Über monotone Matrixfunktionen. *Mathematische Zeitschrift*, 38:177–216, 1934.

[LT98] A. Lapidoth and I. E. Telatar. The compound channel capacity of a class of finite-state channels. *IEEE Trans. on Information Theory*, 44(3):973–983, May 1998.

[MH99] T.L. Marzetta and B.M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Trans. on Information Theory*, 45(1):139–157, January 1999.

[MO79] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Application*. Mathematics in Science and Engineering Vol. 143, Academic Press, Inc. (London) Ltd., 1979.

[MO02] C. Martin and B. Ottersten. Analytic approximations of eigenvalue moments and mean channel capacity for MIMO channels. *IEEE ICASSP*, 3:2389–2392, May 2002.

[MBO04] C. Martin, S. Bergman, and B. Ottersten. Simple Spatial Multiplexing Based on Imperfect Channel Estimates. *IEEE ICASSP*, May 2004.

[Mol04] A. F. Molisch. A generic model for MIMO wireless propagation channels in macro- and microcells. *IEEE Trans. on Signal Processing*, 52(1):61–71, January 2004.

[MSS01] A. L. Moustakas, S. H. Simon, and A. M. Sengupta. MIMO capacity through correlated channels in the presence of correlated interfers and noise: A (not so) large n analysis. *IEEE Trans. on Information Theory*, 49(10):2545–2561, October 2001.

[NBP02] R. U. Nabar, H. Bölcskei, and A. J. Paulraj. Outage performance of space-time block codes for generalized MIMO channels. *submitted to IEEE Trans. on Inf. Theory*, March 2002.

[NLTW98] A. Narula, M.J. Lopez, M.D. Trott, and G.W. Wornell. Efficient use of side information in multiple-antenna data transmission over fading channels. *IEEE Journal on selected areas in communications*, 16(8):1423–1436, October 1998.

[NTW99] A. Narula, M.J. Trott, and G.W. Wornell. Performance limits of coded diversity methods for transmitter antenna arrays. *IEEE Trans. on Information Theory*, 45(7):2418–2433, November 1999.

[ONBP02] Ö. Oyman, R. U. Nabaer, H. Bölcskei, and A. J. Paulraj. Tight lower bounds on the ergodic capacity of rayleigh fading MIMO channels. *Global Telecommunications Conference*, 2:1172–1176, Nov. 2002.

[PCL03a] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas. Joint TX-RX beamforming design for multicarrier MIMO channels: A unified framework for convex optimization. *IEEE Trans. on Signal Processing*, 51(9):2381–2401, September 2003.

[PCL03b] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas. Uniform power allocation in MIMO channels: a game-theoretic approach. *IEEE Trans. on Information Theory*, 49(7):1707–1727, July 2003.

[PF04] D. P. Palomar and J. R. Fonollosa. Practical algorithms for a family of waterfilling solutions. *will appear in IEEE Trans. on Signal Processing*, 2004.

[RC01] W. Rhee and J. M. Cioffi. Ergodic capacity of multi-antenna Gaussian multiple-access channels. *Proc. of the 35th Asilomar Conference on Signals, Systems and Computers*, November 2001.

[RC03] W. Rhee and J. M. Cioffi. On the capacity of multiuser wireless systems with multiple antennas. *IEEE Trans. on Information Theory*, 49(10):2580–2595, October 2003.

[RFLT98] F. Rashid-Farrokhi, K.J.R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE Journal on Selected Areas in Communications*, 16(8):1437–1450, October 1998.

[Sat78] H. Sato. An outer bound to the capacity region of broadcast channels. *IEEE Transaction on Information Theory*, 24(3):374–376, May 1978.

[Say02] A. M. Sayeed. Deconstructing multiantenna fading channels. *IEEE Trans. on Signal Processing*, 50(10):2563–2579, October 2002.

[SB95] J.W. Silverstein and Z.D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis 54(2): 175-192*, 1995.

[SB01] M. Schubert and H. Boche. Solvability of coupled downlink beamforming problems. *Global Telecommunications Conference*, 1:614–618, Nov. 2001.

[SB03] M. Schubert and H. Boche. Throughput maximization for uplink and downlink beamforming with independent coding. In *Proceedings Conf. on Information Sciences and Systems (CISS)*, March 2003.

[Sch02] M. Schubert. *Power-Aware Spatial Multiplexing with Unilateral Antenna Cooperation*. PhD thesis, Technical University of Berlin, 2002.

[SFGK00] D. Shiu, G.J. Foschini, M.J. Gans, and J.M. Kahn. Fading correlation and its effect on the capacity of multi-element antenna systems. *IEEE Trans. on Communications*, 48(3):502–513, March 2000.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.

[SM00] A. M. Sengupta and P. P. Mitra. Capacity of multivariate channels with multiplicative noise: I. random matrix techniques and large-n expansions for full transfer matrices. *Tech. Rep., ATT Bell Labs*, 2000.

[SM02] S. Simon and A. Moustakas. Optimality of beamforming in multiple transmitter multiple receiver communication systems with partial channel knowledge. In *Proc. of DIMACS Conference on Signal Processing for Wireless Transmission*, 2002.

[SSB$^+$02] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath. Optimal designs for space-time linear precoders and decoders. *IEEE Trans. on Signal Processing*, 50(5):1051–1064, May 2002.

[Tel99] E. Telatar. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, Nov/Dec 1999.

[TV02] D. N. Tse and P. Viswanath. Downlink-uplink duality and effective bandwidths. *International Symposium on Information Theory*, page 52, 2002.

[VAT99] P. Viswanath, V. Anantharam, and D. N. C. Tse. Optimal sequences, power control and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers. *IEEE Transactions on Information Theory*, 45(6):1968–1983, Sept. 1999.

[VBG03] S. Vishwanath, S. Boyd, and A. Goldsmith. Worst-case capacity of Gaussian vector channels. In *Proocedings of IEEE Canadian Workshop on Information Theory*, 2003.

[VG97] M. K. Varanasi and T. Guess. Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the gaussian multiple-access channel. In *Proc. Asilomar Conf. on Signals, Systems and Computers*, pages 1405–1409, 1997.

[VJG02a] S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Trans. on Information Theory*, 49(10):2658–2668, Oct. 2002.

[VJG02b] S. Vishwanath, N. Jindal, and A. Goldsmith. On the capacity of multiple input multiple output broadcast channels. *Proceedings of International Conference on Communications*, 3:1444–1450, April 2002.

[VKS+03] S. Vishwanath, G. Kramer, S. Shamai (Shitz), S. Jafar, and A. Goldsmith. Capacity bounds for Gaussian vector broadcast channels. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 2003.

[Vis03] S. Vishwanath. *Multi-User Multiple Antenna Systems: Theoretical Limits and Practical Strategies*. PhD thesis, Stanford University, 2003.

[VM01] E. Visotsky and U. Madhow. Space-time transmit precoding with imperfect feedback. *IEEE Trans. on Information Theory*, 47(6):2632–2639, September 2001.

[VT03] P. Viswanath and D. N. C. Tse. Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Trans. on Information Theory*, 49(8):1912–1921, August 2003.

[WSS04] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). The capacity region of the gaussian mimo broadcast channel. *Proc. of Conference on Computer Systems and Sciences*, pages 7–12, 2004.

[WFGV98] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela. V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel. *Proceedings URSI International Symposium on Signals, Systems, and Electronics., (IEEE, New York, NY, USA), pp. 295–300*, 1998.

[Win98] J. H. Winters. The diversity gain of transmit diversity in wireless systems with Rayleigh fading. *IEEE Trans. on Vehicular Technology*, 47(1):119–123, February 1998.

[Wol78] J. Wolfowitz. *Coding Theorems of Information Theory*, volume 3rd ed. Berlin, Germany: Springer-Verlag, 1978.

[WVB96] S. P. Wu, L. Vandenberghe, and S. Boyd. *MAXDET: software for determinant maximization problems*, May 1996.

[YB03] S. Ye and R. S. Blum. Optimized signaling for MIMO interference systems with feedback. *IEEE Trans. on Signal Processing*, 51(11):2839–2848, November 2003.

[YC01] W. Yu and J. M. Cioffi. Sum capacity of gaussian vector broadcast channels. *sub, to IEEE Trans. on Inf. Theory*, November 2001.

[Yeh03] E. M. Yeh. An inter-layer view of multiaccess communications. *Proc. of International Symposium on Information Theory*, page 112, 2003.

[YGC01] W. Yu, G. Ginis, and J. M. Cioffi. An adaptive multiuser power control algorithm for vdsl. *Proc. of IEEE Global Telecommunications Conference*, 2001.

[YRBC01] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi. Iterative water-filling for vector multiple access channels. In *International Symposium on Information Theory*, page 322, June 2001.

[YRBC04] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi. Iterative water-filling for Gaussian vector multiple-access channels. *will appear in IEEE Transactions on Inf. Theory*, 2004.

[Yu03] W. Yu. The structure of the worst-noise in Gaussian vector broadcast channels. *American Mathematical Society - Discrete Mathematics and Theoretical Computer Science (DIMACS) series on Network Information Theory*, 2003.