mascot

Project Number:	IST-026905
Project Title:	Multiple-Access Space-Time Coding Testbed
Project Coordinator:	C.F. Mecklenbräuker
Deliverable Number:	D3.1.1

Title of Deliverable:	Performance tradeoffs in MU-MIMO systems
Workpackage:	WP-3
Nature:	R
Dissemination level:	PU
Editor:	Jialai Weng, Giorgio Taricco
Authors:	see list inside
Contractual Date of Deliverable:	Jun. 30, 2007
Actual Date of Delivery:	Jun. 30, 2007

Abstract:

This deliverable discusses several performance tradeoffs in MU-MIMO systems. We start by surveying the key diversity-multiplexing tradeoff for narrowband MIMO channels developed by Zheng and Tse. Then, the tradeoff is extended to the time, frequency and time-frequency selective case by applying the concept of "Jensen channel." Next, we elaborate the effect of multi-access interference by resorting to interference functions, which are used to optimize specific quality-of-service goals.

Contents

1 The Diversity-Multiplexing Tradeoff **7**
- 1.1 Diversity and Multiplexing for the Coherent Channel 9
  - 1.1.1 System Model and Known Results 9
  - 1.1.2 Optimal Tradeoff: $T \geq \mathrm{M_T} + \mathrm{M_R} - 1$ case 17

2 Diversity-Multiplexing Tradeoff in Selective-Fading **19**
- 2.1 Introduction 19
  - 2.1.1 Contributions 19
  - 2.1.2 Notation 20
- 2.2 Channel and signal model 20
- 2.3 Diversity-multiplexing tradeoff 21
  - 2.3.1 Preliminaries 21
  - 2.3.2 Jensen channel and Jensen outage event 23
- 2.4 Jensen-optimal code design criterion 24
  - 2.4.1 Code design criterion 25
  - 2.4.2 Application to the frequency-selective case 29
- 2.5 Conclusions 29

3 Introduction to Interference Function **31**
- 3.1 QoS-based power and resource allocation 32
  - 3.1.1 Interference functions 32

4 Advanced Interference Calculus **36**
- 4.1 Introduction 36
- 4.2 Interference Functions 37
  - 4.2.1 Adaptive Receive Strategies 38
  - 4.2.2 Robust Designs 39
  - 4.2.3 SIR Balancing and Feasible Sets 40
- 4.3 Elementary Representations and Bounds 41
  - 4.3.1 General Interference Functions 42
  - 4.3.2 Convex Interference Functions 42

4.3.3 Log-Convex Interference Functions 43
4.4 Proportional Fairness and Bargaining 44

Authors

Pedro Coronel and Helmut Bölcskei
Eidgenössische Technische Hochschule
8092 Zürich, Switzerland

Holger Boche and Martin Schubert
Fraunhofer Institute for Telecommunications HHI,
Einsteinufer 37, 10587 Berlin, Germany

Jialai Weng
Politecnico di Torino – Dipartimento di Elettronica,
corso Duca degli Abruzzi 24, 10129 Torino, Italy

Executive Summary

This deliverable investigates several performance tradeoffs for MU-MIMO systems deriving from the key diversity-multiplexing tradeoff. The diversity-multiplexing tradeoff characterizes the best compromise between error performance (related to the diversity order) and communication rate (related to the multiplexing gain) which can be attained simultaneously on a given system.

The deliverable is organized into four chapters. Chapter 1 is introductory and describes the diversity-multiplexing tradeoff for narrowband MIMO communication systems. After introducing some basic assumptions on coded MIMO systems, we review its fundamental performance indicators represented by channel capacity and pairwise error probability. Multiplexing capabilities are related with channel capacity and the system diversity is embedded in the pairwise error probability so that their joint study allows one to establish the required tradeoff. Both parameters (multiplexing and diversity order) are formally defined in the setting of asymptotically large SNR. The chapter concludes with its main result consisting in the characterization of the optimal tradeoff curve $d^*(r)$ (Theorem 1), which is given by the piecewise linear interpolation of a set of integer-valued points in the multiplexing-diversity plane [57].

In Chapter 2, we establish the optimal diversity-multiplexing (DM) tradeoff of coherent time, frequency and time-frequency selective-fading MIMO channels and provide a code design criterion for DM-tradeoff optimality. Our results are based on the analysis of the "Jensen channel" associated to a given selective-fading MIMO channel. While the original problem seems analytically intractable due to the mutual information being a sum of correlated random variables, the Jensen channel (equivalent to the original channel in the sense of the DM-tradeoff) lends itself nicely to analytical treatment. From these results we find that the classical rank criterion for space–time code design (in selective-fading MIMO channels) ensures optimality also in the sense of the DM-tradeoff.

The effect of multi-access interference is addressed in Chapter 3. It is observed that the quality-of-service (QoS) of a communication link is affected by the existence of other links, which are experienced as interference. A proper allocation of the system resources allows to limit the average effect of interference on users. The problem is addressed by introducing a specific QoS-based power and resource allocation model based on interfering functions depending on the power allocation itself.

The framework presented in Chapter 3 is used as a background for Chapter 4, where we identify the key properties leading to the observed convergence behavior. It is known that the optimization of interference-coupled multiuser networks is complicated by adaptive transmit or receive strategies, which often result in non-convex problem formulations. However, among the studied properties, we show that convexity is not the most basic one and sometimes, globally convergent algorithms can be derived from certain monotonicity and scalability properties alone.

Chapter 1

The Diversity-Multiplexing Tradeoff

Multiple-input multiple-output (MIMO) technology represents an important means of increasing the performance of wireless communication systems. It is widely known that in a system with multiple transmit and receive antennas (MIMO channel) the spectral efficiency can be greatly increased from that of the conventional single antenna channels. Recent research shows that by using multiple antennas at both the transmitter and receiver, the performance of the system can be greatly improved. Multiple antennas can improve the system performance in two aspects, which are reliability and supporting higher data rate.

Different design criteria of MIMO communication systems are based on different ways to understand fading channel. Multiple antennas can be used to increase diversity. In this situation each pair of transmit and receive antennas provides a signal path from the transmitter to the receiver. By sending signals that carry the same information through a number of different paths, multiple independently replicated data symbols can be obtained at the receiver; by combining these replicated signals, we can have more reliability in transmission. For example, in a slow Rayleigh fading environment with 1 transmit and $\mathrm{M_R}$ receive antennas, the transmitted signal is passed through $\mathrm{M_R}$ different paths. It is well known that if the fading is independent, a maximal diversity gain of $\mathrm{M_R}$ can be achieved: an error occurs only when all of the $\mathrm{M_R}$ paths are in simultaneous deep fading, compare to 1 channel is in deep fading; then lower error probability is achievable. In a system with $\mathrm{M_T}$ transmit and $\mathrm{M_R}$ receive antennas, the maximal diversity gain is $\mathrm{M_T M_R}$.

A different direction of utilizing MIMO suggests that in a MIMO channel, fading can in fact be beneficial through increasing the degrees of freedom available for communication. Essentially, if the path gains between individual transmit-receive antenna pairs fade independently, the channel matrix can be well-conditioned, in which case the channel is essentially multiple parallel spatial channels. By transmitting independent information in parallel through the spatial channels, the data rate can be increased. This effect is also called spatial multiplexing, and is particularly important in the high signal-to-noise ratio (SNR) regime where the system is degree-of-freedom-limited. It is well known that in the high SNR regime, the capacity of a channel with $\mathrm{M_T}$ transmit, $\mathrm{M_R}$ receive antennas and i.i.d. Rayleigh faded gains between each antenna pair is scaled by $\min\{\mathrm{M_T}, \mathrm{M_R}\}$. We say the spatial multiplexing gain of the system is $\min\{\mathrm{M_T}, \mathrm{M_R}\}$.

A MIMO system can provide two types of gains: diversity gain and spatial multiplexing gain. In the design of a MIMO system, maximizing one type of gain may not necessarily maximize the other. For example, the space time coding structure from the orthogonal designs, while achieving the full diversity gain, reduces the achievable spatial multiplexing gain. In fact, each of the two design goals addresses only one aspect of the problem. This makes it difficult to compare the performance between diversity-based and multiplexing-based schemes.

The work of Zheng and Tse [57] on diversity and multiplexing tradeoff put forth a new viewpoint: given a MIMO channel, both gains can in fact be simultaneously obtained, but there is a fundamental tradeoff between the two types of gain: higher spatial multiplexing gain yields lower diversity. Their seminal work provides an approach to characterize the optimal tradeoff curve achievable by any scheme. In the high SNR regime, a scheme is said to have a spatial multiplexing gain r and a diversity gain d if the rate of the scheme scales like $r \log SNR$ and the average error probability falls on $1/SNR^d$ rate. The spatial multiplexing gain indicates the fraction of the capacity that is achieved; and the diversity gain indicates the reliability. The optimal tradeoff curve shows for each multiplexing gain r the optimal diversity advantage $d^*(r)$ achievable by any scheme; hence gives the performance limit in the data rate and the reliability.

For the i.i.d. Rayleigh flat fading channel, the optimal tradeoff is in very simple form for most system parameters. Consider a slow block fading channel in which the channel is random but remains unchanged for a duration of T symbols, and assume a coherent reception, where the receiver has the channel gain information. The work of Zheng and Tse shows that as long as $T \geq \mathrm{M_T} + \mathrm{M_R} - 1$, the optimal diversity gain $d^*(r)$ achievable by any coding scheme of block length T and multiplexing gain r is $(\mathrm{M_T} - r)(\mathrm{M_R} - r)$. This suggests an appealing interpretation: out of the total resource of $\mathrm{M_T}$ transmit and $\mathrm{M_R}$ receive antennas, it is as though r transmit and r receive antennas were used for multiplexing and the remaining $\mathrm{M_T} - r$ transmit and $\mathrm{M_R} - r$

receive antennas provided the diversity. Thus, by adding one transmit and one receive antenna, the spatial multiplexing gain can be increased by one while maintaining the same diversity level. It should also be observed that this optimal tradeoff does not depend on T as long as $T \geq \mathrm{M_T} + \mathrm{M_R} - 1$; hence, no more diversity gain can be extracted by coding over block lengths greater than $\mathrm{M_T} + \mathrm{M_R} - 1$ than using a block length equal to $\mathrm{M_T} + \mathrm{M_R} - 1$.

1.1 Diversity and Multiplexing for the Coherent Channel

1.1.1 System Model and Known Results

We consider the wireless link with $\mathrm{M_T}$ transmit and $\mathrm{M_R}$ receive antennas. We assume a flat fading environment. The fading coefficient h_{ij} is the complex path gain from transmit antenna j to receive antenna i. We assume that the coefficients are independently Rayleigh distributed with unit variance, and write $\mathbf{H} = [h_{ij}] \in \mathcal{C}^{\mathrm{M_R} \times \mathrm{M_T}}$. We assume that the channel matrix $\mathbf{H}$ remains constant within a block of T symbols. Under these assumptions, the channel, within one block, can be written as:

$$Y = \sqrt{\frac{SNR}{\mathrm{M_T}}}\mathbf{HX} + \mathbf{W} \tag{1.1}$$

where $\mathbf{X} \in \mathcal{C}^{\mathrm{M_T} \times T}$ has entries $x_{mt}; m = 1, \ldots, \mathrm{M_T}, t = 1, \ldots, T$ being the signals transmitted from antenna m at time t; $\mathbf{Y} \in \mathcal{C}^{\mathrm{M_R} \times T}$ has entries $\mathbf{y}_{nt}, n = 1, \ldots, \mathrm{M_R}, t = 1, \ldots, T$ being the signals received from antenna n at time t; the additive noise $\mathbf{W}$ has i.i.d. entries $\mathbf{w}_{nt} \sim \mathcal{C}N(0, 1)$; SNR is the average signal to noise ratio at each receive antenna.

The transmitted signal $\mathbf{X}$ is normalized to have the average transmitted power at each antenna in each symbol period to be 1. The power constraint on a codebook $\mathbf{C}$ can be written as

$$\frac{1}{|\mathbf{C}|}\sum_i \sum_{m=1}^{\mathrm{M_T}} \sum_{t=1}^{T} ||\mathbf{x}_{mt}(i)||^2 \leq \mathrm{M_T}T$$

where the first sum is over all the codewords in the codebook.

Based on the applications in different fading environment, this multiple antenna channel is studied under both the coherent and the non-coherent assumptions. In this chapter, we will focus on the coherent channel.

The coherent channel model is based on the following physical scenario. In a fixed wireless environment, the fading coefficients vary slowly. The transmitter can periodically send pilot (training) signals to allow the receiver to estimate the fading coefficients accurately. The time that one spends in estimating the channel is usually negligible comparing to the time that one can communicate before these estimates become inaccurate. In this case, we can usually ignore the acquiring of the channel knowledge, and simply assume that the channel fading coefficients in $\mathbf{H}$ are perfectly known at the receiver end. This assumption is referred as the "coherent assumption".

In the following of this section, we review some of the known results for the coherent channel.

Strictly speaking, the Shannon capacity of channel (1.1) is not defined for finite values of block length T. In the case that $T = \infty$, the Shannon capacity is 0, since for any given data rate, there is a strictly positive probability that the channel matrix $\mathbf{H}$ is too bad (close to 0) to support it; in other words, a channel outage occurs. In this case, one can write the mutual information of the channel as a function of the channel matrix $\mathbf{H}$. The notion of outage capacity is studied [50] to answer the questions such as "what data rate one can communicate with no more than 1 percent probability of channel outage".

In this report, we will mainly focus on the ergodic capacity of multiple antenna channels. To do that, we look on (1.1) as one channel use, with input and output super- symbols of dimension $\mathrm{M_T} \times T$ and $\mathrm{M_R} \times T$ respectively. We assume that after each block of T symbol periods, the channel matrix $\mathbf{H}$ changes into an independent realization, and remains constant for another block. The ergodic channel capacity is thus the highest data rate that can be reliably transferred by coding over infinitely many such super-symbols. In the rest of this dissertation, we will without causing confusion refer to "capacity" of channel (1.1) as the ergodic capacity. The coherent capacity of the channel (1.1) is computed by Telatar [50] which is summarized in the following lemma.

Lemma 1. *Coherent Capacity*

Assume the fading coefficient matrix $\mathbf{H}$ *is known to the receiver, the channel capacity (bps/Hz) of a system with* $\mathrm{M_T}$ *transmit and* $\mathrm{M_R}$ *receive antennas is given by*

$$C_{coherent}(SNR) = E\left[\log_2 \det(I_{\mathrm{M_R}} + \frac{SNR}{\mathrm{M_T}}\mathbf{H}\mathbf{H}^\dagger)\right]$$

Defining $K = \min\{\mathrm{M_T}, \mathrm{M_R}\}$, $K' = \max\{\mathrm{M_T}, \mathrm{M_R}\}$, then a lower bound can be derived:

$$C_{coherent}(SNR) \leq K \log_2 \frac{SNR}{\mathrm{M_T}} + \sum_{i=K'-K+1}^{K'} E[\log_2 \chi^2_{2i}];$$

where χ^2_{2i} is chi-square random variable with dimension $2i$. Moreover, this lower bound is asymptotically tight at high SNR. We observe that this is equivalent to the capacity of $K = \min\{\mathrm{M_T}, \mathrm{M_R}\}$ sub-channels. In other words, the multiple antenna channel has K degrees of freedom to communicate.

For the case $\mathrm{M_T} = \mathrm{M_R}$, at high SNR,

$$C_{coherent}(SNR) = \mathrm{M_T} \log_2 \frac{SNR}{\mathrm{M_T}} + \sum_{i=1}^{\mathrm{M_T}} E[\log_2 \chi^2_{2i}] + o(1)$$

If we let the number of transmit antennas $\mathrm{M_T}$ increase to infinity, the high SNR capacity increases linearly with $\mathrm{M_T}$, and

$$\lim_{\mathrm{M_T}\to\infty} \lim_{SNR\to\infty} \left[\frac{C_{coherent}(SNR)}{\mathrm{M_T}} - \log_2(\frac{SNR}{e}) \right] = 0$$

This capacity can be achieved by using a "layered space-time architecture" which is discussed in detail in [18].

It is of practical interests to compute the probability of detection error when communicating over the multiple antenna channel with a given finite codeword length. Most of the current effort in the error probability analysis is focused on the case when one can code over a single block as given in (1.1).

Note that when we compute the error probability for codes over one single block, the assumption of independently faded blocks given in previously is irrelevant. This assumption is introduced only to develop the notion of the ergodic capacity. In the error probability analysis, we do not need to assume anything for the blocks that follows. Nevertheless, in the rest of this dissertation, we will frequently talk about the channel capacity and error probability, without repeating the difference in the assumptions.

Due to the difficulty to compute the average error probability, the current results are mainly on the pairwise error probability. The following lemma summarizes the result from [48].

Lemma 2. *Pairwise Error Probability, Coherent Case*

Assume that $X_0, X_1 \in \mathcal{C}^{\mathrm{M_T}\times T}$ are two codewords, and X_0 is transmitted over channel (1.1). The probability that a maximum likelihood detector would make error in favor of X_1 is bounded by

$$P(X_0 \to X_1) \le \prod_{i=1}^{\mathrm{M_T}} \left(1 + \frac{SNR}{4}\lambda_i\right)^{-\mathrm{M_R}}$$

where $\lambda_i, i = 1, \ldots, \mathrm{M_T}$ *are the eigenvalues of the matrix* $A \triangleq (X_0 - X_1)(X_0 - X_1)^\dagger$.

Let r denote the rank of matrix A, and $\lambda_1, \ldots, \lambda_r$ be the non-zero eigenvalues of A, then it follows that

$$P(X_0 \to X_1) \le \left(\prod_{i=1}^{r} \lambda_i\right)^{-\mathrm{M_R}} \left(\frac{SNR}{4}\right)^{-rN}$$

Following this argument, one reaches to the rank criterion of designing space-time codes: in order to minimize the error probability at high SNR, the difference between each pair of possible codeword matrices has to be full rank.

As shown in the previous sections, by using multiple antenna channel, the performance of a wireless link can be greatly improved, both in terms of having a better reliability and supporting a higher data rate. In other words, the channel provides two types of performance gains. Unfortunately, in the multiple antenna research community, these two types of gains are studied separately in most cases, and the relation between them is not studied. In this section, we will start to find a simple way to characterize these two types of gains, and study the relation between them.

The fact that multiple antennas can be used to improve the reliability. It can be understood as providing spatial diversity. The basic idea is to supply to the receiver with multiple independently faded replicas of the same information symbol, so that the probability that all the signal components fade simultaneously is reduced.

As an example, consider uncoded binary PSK signals over a single antenna fading channel ($\mathrm{M_T} = \mathrm{M_R} = T = 1$ in the above model). It is well known that the probability of error at high SNR (averaged over the fading gain $\mathbf{H}$ as well as the additive noise) is

$$P_e(SNR) \approx \frac{1}{4} SNR^{-1}$$

In contrast, transmitting the same signal to a receiver equipped with 2 antennas, the error probability is

$$P_e(SNR) \approx \frac{3}{16} SNR^{-2}$$

Here we observe that by having the extra receive antenna, the error probability decreases with SNR at a faster speed of SNR^{-2}. This phenomenon implies that at high SNR, the error probability is much smaller. Similar results can be obtained if we change the binary PSK signals to other constellations. The reason of this improvement is that the data symbol passes through two different fading channels, one to each receive antenna, to the receiver end; hence diversity is obtained.

Since the performance gain at high SNR is dictated by the SNR exponent of the error probability, this exponent is called the diversity gain.

Besides providing diversity to improve reliability, multiple antenna channels can also support a higher data rate than single antenna channels. As an evidence of this, consider an ergodic block fading channel in which each block is as in (1.1) and the channel matrix is independent and identically distributed across blocks. The ergodic capacity (bps/Hz) of this channel is well-known [18, 50]:

$$C(SNR) = \mathcal{E}\left[\log\det\left(I + \frac{SNR}{\mathrm{M_T}}\mathbf{H}\mathbf{H}^\dagger\right)\right]$$

At high SNR

$$C(SNR) = K\log SNR + \sum_{i=K'-K+1}^{K'} \mathcal{E}[\log\chi^2_{2i}] + o(1)$$

where $K = \min\{\mathrm{M_T}, \mathrm{M_R}\}$, $K' = \max\{\mathrm{M_T}, \mathrm{M_R}\}$. We observe that at high SNR, the channel capacity increases with SNR as $K\log SNR$ (bps/Hz), in contrast to $\log SNR$ for single antenna channels. This result suggests that the multiple antenna channel can be viewed as K parallel spatial channels; hence the number $K = \min\{\mathrm{M_T}, \mathrm{M_R}\}$ is the total number of degrees of freedom to communicate. Now one can transmit independent information symbols in parallel through the spatial channels. This idea is also called spatial multiplexing.

Reliable communication at rates arbitrarily close to the ergodic capacity requires averaging across many independent realizations of the channel gains over time. Since we are considering coding over only a single block, we must lower the data rate and step back from the ergodic capacity to cater for the randomness of the channel $\mathbf{H}$. Since the channel capacity increases linearly with $\log SNR$, in order to achieve a certain fraction of the capacity at high SNR, we should consider schemes that support a data rate which also increases with SNR. Here, we think of a scheme as a family of codes $\{\mathbf{C}(SNR)\}$ of block length T, one at each SNR level. Let $R(SNR)$ (bits/symbol) be the rate of the code $\mathbf{C}(SNR)$. We say that a scheme achieves a spatial multiplexing gain of r if the supported data rate

$$R(SNR) \approx r \log SNR (\text{bps/Hz})$$

One can think of spatial multiplexing as achieving a non-vanishing fraction of the degrees of freedom in the channel. According to this definition, any fixed-rate scheme has a zero multiplexing gain, since eventually at high SNR, any fixed data rate is only a vanishing fraction of the capacity.

Now to formalize, we have the following definition.

Definition 1. A communication channel is said to achieve spatial multiplexing gain r and diversity gain d if the data rate satisfies

$$\lim_{SNR \to \infty} \frac{R(SNR)}{\log SNR} = r$$

and the average error probability satisfies

$$\lim_{SNR \to \infty} \frac{\log P_e(SNR)}{\log SNR} = -d \tag{1.2}$$

For each r, define $d^*(r)$ as the supremum of the diversity gain achieved over all channels. We also define

$$d^*_{max} \triangleq d^*(0)$$

$$r^*_{max} \triangleq \sup\{r : d^*(r) > 0\}$$

which are respectively the maximal diversity gain and the maximal spatial multiplexing gain in the channel.

Throughout the rest of the chapter, we will use the special symbol $\doteq$ to denote exponential equality, i.e., we write $f(SNR) \doteq SNR^b$ to denote

$$\lim_{SNR \to \infty} \frac{\log f(SNR)}{\log SNR} = b$$

(1.2) can thus be written as

$$P_e(SNR) \doteq SNR^{-d}$$

The error probability $P_e(SNR)$ is averaged over the additive noise $\mathbf{W}$, the channel matrix $\mathbf{H}$ and the transmitted codewords (assumed equally likely).

The definition of diversity gain here differs from the standard definition in the space-time code literature (see for example [48]) in two important ways.

This is the actual error probability of a code, and not the pairwise error probability between two codewords as is commonly used as a diversity criterion in space-time code design.

In the standard formulation, diversity is a performance metric for one fixed code. It indicates how fast the error probability, when using this code, decays with SNR. The conventional notion of diversity corresponds to the number of independently faded paths that a symbol passes through; in other words, the number of independent fading coefficients that can be averaged over to detect the symbol. In a general system with $\mathrm{M_T}$ transmit and $\mathrm{M_R}$ receive antennas, there are in total $\mathrm{M_T} \times \mathrm{M_R}$ random fading coefficients to be averaged over; hence the maximal (full) diversity gain provided by the channel is $\mathrm{M_T M_R}$. In our formulation, diversity gain is a performance metric of a scheme, the data rate of which also increases with SNR. It still indicates how fast the error probability decays with SNR, but the effect of an increasing data rate is included. This is a natural generalization of the conventional concept, since in terms of the improving the reliability, what matters is not only how many times one repeats the data symbol, but certainly also how far the system is operating below the capacity. Under our formulation, a fixed code can be viewed as a scheme with multiplexing gain $r = 0$; therefore the conventional notion of diversity gain corresponds to $d^*(0)$ in our definition.

The spatial multiplexing gain can also be thought as the data rate normalized with respect to the SNR level. A common way to characterize the performance of a communication scheme is to compute the error probability as a function of SNR for a fixed data rate. However, different schemes may support different data rate. In order to compare these schemes fairly, Forney proposed to plot error probability against the normalized SNR:

$$SNR_{norm} \triangleq \frac{SNR}{C^{-1}(R)}$$

where $C(SNR)$ is the capacity of the channel as a function of SNR. That is, SNR_{norm} measures how far the SNR is above the minimal required to support the target data rate. A dual way to characterize the performance is to plot the error probability as a function of the data rate, for a fixed SNR level. Analogous to Forneys formulation, to take into consideration of the effect of SNR, one should use the normalized data rate R_{norm} instead of R:

$$R_{norm} \triangleq \frac{R}{C(SNR)}$$

which indicates how far a system is operating from the Shannon limit. Notice that at high SNR, the capacity of the multiple antenna channel is $C(SNR) \approx$

$K \log SNR$; hence the spatial multiplexing gain

$$r = \frac{R}{\log SNR} \approx KR_{norm}$$

is just a constant multiple of R_{norm}.

Definition 1 contains the key idea of the diversity multiplexing tradeoff. Essentially the research on multiple antenna channels is split into two branches. Information theoretical study views multiple antennas as the source of multiple degrees of freedom, which provides a significant gain in terms of the Shannon capacity at high SNR. On the other hand, the space-time code designers characterize the performance gain by using multiple antennas by the spatial diversity gain, which leads to a lower probability of detection error at high SNR. By giving definition 1, we try to put these two types of gains in one picture, and study the relation between them. The conventional methods, the capacity study and error probability analysis, can be thought as some extreme ways of using the channel resource. For example, we say that for each 3dB increase of the SNR the capacity gain of a Gaussian channel is 1 (bps/Hz), without considering the error probability. In order to put forward a unified picture, we introduced the spatial multiplexing gain and the generalized version of the diversity gain, which indicate the fractions of the total number of degrees of freedom and the full diversity gain that are actually achieved by a given scheme. This formulation allows us to move continuously from one extreme to the other by including a large family of intermediate schemes: with each 3dB of SNR increase, we can choose to increase the data rate by r (bps/Hz) and decrease the error probability by 2^{-d} simultaneously, for certain values of r and d.

The diversity-multiplexing tradeoff essentially reflects how one uses the channel resource; that is, how to translate each dB of SNR increase into improvement of the performance, in terms of both the data rate and the reliability. The optimal tradeoff curve $d^*(r)$, gives the fundamental limit of this utilization.

The conventional code design approach usually fixes the data rate, and compute the speed that the error probability decays with SNR (diversity order). This approach fails to provide a sound base to discuss the tradeoff between the data rate and the diversity. In fact, at any fixed rate, one can always design codes with full diversity, hence another performance metric, the "determinant criterion" [48]. must be used. In contrast, in our formulation, the channel resource, the data rate, and the reliability are all characterized in an incremental fashion; thus a uniform context to discuss the tradeoff is provided.

1.1.2 Optimal Tradeoff: $T \geq \mathrm{M_T} + \mathrm{M_R} - 1$ case

In this section, we give the optimal tradeoff between the diversity gain and the spatial multiplexing gain that any scheme can achieve in a given multiple antenna channel. We will focus on the case that the block length $T \geq \mathrm{M_T} + \mathrm{M_R} - 1$.

The main result of Zheng and Tse's work [57] is stated by the following theorem.

Theorem 1. *Assume $T \geq \mathrm{M_T} + \mathrm{M_R} - 1$. Let $K = \min\{\mathrm{M_T}, \mathrm{M_R}\}$. The optimal tradeoff curve $d^*(r)$ is given by the piecewise linear function connecting the points $(k, d^*(k))$, $k = 0, \ldots, K$, where*

$$d^*(k) = (\mathrm{M_T} - k)(\mathrm{M_R} - k)$$

As a consequence $d^*_{max} = MN$, and $r^*_{max} = \min\{\mathrm{M_T}, \mathrm{M_R}\}$.

The optimal tradeoff curve intersects the r axis at $K = \min\{\mathrm{M_T}, \mathrm{M_R}\}$. This means that the maximum achievable spatial multiplexing gain $r^*_{max} = K$, which is the total number of degrees of freedom provided by the channel, as suggested by the ergodic capacity result in (1.2). Theorem 4 says that at this point, however, no positive diversity gain can be achieved. Intuitively, as $r \to K$, the data rate approaches the ergodic capacity and there is no protection against the randomness in the fading channel.

On the other hand, the curve intersects the d axis at the maximal diversity gain $d^*_{max} = \mathrm{M_T}\mathrm{M_R}$, corresponding to the total number of random fading coefficients that a scheme can average over. There are known designs that achieve the maximal diversity gain at a fixed data rate [1]. Theorem 1 says that in order to achieve the maximal diversity gain, no positive spatial multiplexing gain can be obtained at the same time.

The optimal tradeoff curve $d^*(r)$ bridges the gap between the above two design criteria, by connecting the two extreme points: $(0, d^*_{max})$ and $(r^*_{max}, 0)$. This result says that positive diversity gain and spatial multiplexing gain can be achieved simultaneously. However, increasing the diversity advantage comes at a price of decreasing the spatial multiplexing gain, and vice versa. The tradeoff curve is thus a more complete concept than the two extreme points corresponding to the maximum diversity gain and multiplexing gain. For example, the ergodic capacity result suggests that by increasing the minimum of the number of transmit and receive antennas, $K = \min\{\mathrm{M_T}, \mathrm{M_R}\}$, by one, the channel gains one more degree of freedom, corresponds to r^*_{max} is increased by 1; Theorem 1 makes a more informative statement: if we increase both $\mathrm{M_T}$ and $\mathrm{M_R}$ by 1, the entire tradeoff curve is shifted to the right by 1, i.e., for a given diversity gain requirement d, the supported spatial multiplexing gain is increased by 1.

This chapter we give an introduction to the fundamental diversity and multiplexing tradeoff in MIMO fading channel. In the following chapter, we study the diversity multiplexing tradeoff in selective fading MIMO channels.

Chapter 2

Diversity-Multiplexing Tradeoff in Selective-Fading MIMO Channels

2.1 Introduction

The diversity-multiplexing (DM) tradeoff framework introduced by Zheng and Tse [57] allows to efficiently characterize the information-theoretic performance limits of communication over multiple-input multiple-output (MIMO) fading channels. In addition, the results in [57] have triggered significant activity on the design of DM-tradeoff optimal space-time codes. In particular, the *non-vanishing* determinant criterion [2, 54] on codeword difference matrices has been shown to represent a sufficient condition for DM-tradeoff optimality in flat-fading MIMO channels with two transmit and two or more receive antennas [54]; this criterion has led to the construction of space-time codes based on constellation rotation [54, 17] and cyclic division algebras [3]. In [19] lattice-based space-time codes have been shown to be DM-tradeoff optimal. The DM-tradeoff optimality of *approximately universal* space-time codes was established in [49].

2.1.1 Contributions

While the results mentioned above focus on frequency-flat block-fading channels, extensions to frequency-selective channels can be found in [21,31]. However, a general characterization of the optimal DM-tradeoff in time, frequency or time-frequency selective-fading MIMO channels, in the following simply referred to as selective-fading MIMO channels, remains an open problem. The present chapter solves this problem for the coherent case (i.e., for perfect channel state information (CSI) at the receiver) and provides a code design criterion guaranteeing DM-tradeoff optimality. Our results are based on exponentially tight (in the sense of exhibiting the same DM-tradeoff behavior) upper and lower bounds on the mutual information of (coherent) selective-fading MIMO channels. In particular, we show that the DM-tradeoff of this class of channels can be obtained by solving the analytically tractable problem of computing the DM-tradeoff curve corresponding to the associated "Jensen channel".

2.1.2 Notation

$\mathrm{M_T}$ and $\mathrm{M_R}$ denote the number of transmit and receive antennas, respectively. For $x \in \mathbb{R}$, we let $[x]^+ := \max(0, x)$. The superscripts ^{T}, ^{H} and * stand for transposition, conjugate transposition and complex conjugation, respectively. $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{A} \odot \mathbf{B}$ denote, respectively, the Kronecker and Hadamard products of the matrices $\mathbf{A}$ and $\mathbf{B}$, and $\mathbf{A} \succeq \mathbf{B}$ stands for the positive semidefinite ordering. If $\mathbf{A}$ has columns $\mathbf{a}_k$ $(k=1, 2, \ldots, m)$, $\mathrm{vec}(\mathbf{A}) = [\mathbf{a}_1^T \; \mathbf{a}_2^T \; \ldots \; \mathbf{a}_m^T]^T$. For the $n \times m$ matrices $\mathbf{A}_k$ $(k=0, 1, \ldots, K-1)$, $\mathrm{diag}\{\mathbf{A}_k\}_{k=0}^{K-1}$ denotes the $nK \times mK$ block-diagonal matrix with the kth diagonal entry given by $\mathbf{A}_k$. If $\mathcal{S}$ is a set, $|\mathcal{S}|$ denotes its cardinality. For index sets $\mathcal{S}_1 \subseteq \{1, 2, \ldots, n\}$ and $\mathcal{S}_2 \subseteq \{1, 2, \ldots, m\}$, $\mathbf{A}(\mathcal{S}_1, \mathcal{S}_2)$ stands for the (sub)matrix consisting of the rows of $\mathbf{A}$ indexed by $\mathcal{S}_1$ and the columns of $\mathbf{A}$ indexed by $\mathcal{S}_2$. The eigenvalues of the $n \times n$ Hermitian matrix $\mathbf{A}$, sorted in ascending order, are denoted by $\lambda_k(\mathbf{A})$, $k = 1, 2, \ldots, n$. The Kronecker delta function is defined as $\delta(m) = 1$ for $m = 0$ and zero otherwise. If X and Y are random variables (RVs), $X \sim Y$ denotes equality in distribution and $\mathbb{E}_X$ is the expectation operator with respect to (w.r.t.) the RV X. The random vector $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$ is multivariate circularly symmetric zero-mean complex Gaussian with $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{C}$. $f(x)$ and $g(x)$ are said to be exponentially equal, denoted by $f(x) \doteq g(x)$, if $\lim_{x \to \infty} \frac{\log f(x)}{\log x} = \lim_{x \to \infty} \frac{\log g(x)}{\log x}$. Exponential inequality, denoted by $\dot{\geq}$ and $\dot{\leq}$, is defined analogously.

2.2 Channel and signal model

The input-output relation for the class of MIMO channels considered in this chapter is given by

$$\mathbf{y}_n = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}} \mathbf{H}_n \mathbf{x}_n + \mathbf{z}_n, \quad n = 0, 1, \ldots, N-1 \tag{2.1}$$

where the index n corresponds to a time, frequency or time-frequency slot and SNR denotes the signal-to-noise ratio at each receive antenna. The vectors $\mathbf{y}_n$, $\mathbf{x}_n$ and $\mathbf{z}_n$ denote, respectively, the corresponding $\mathrm{M_R} \times 1$ receive signal vector, $\mathrm{M_T} \times 1$ transmit signal vector, and $\mathrm{M_R} \times 1$ zero-mean circularly symmetric complex Gaussian noise vector satisfying $\mathbb{E}\{\mathbf{z}_n \mathbf{z}_{n'}^H\} = \delta(n-n')\,\mathbf{I}_{\mathrm{M_R}}$. We restrict our analysis to spatially uncorrelated Rayleigh fading channels so that, for a given n, $\mathbf{H}_n$ has i.i.d. $\mathcal{CN}(0,1)$ entries. We do allow, however, for correlation across n, assuming, for simplicity, that each scalar subchannel has the same correlation function, i.e., $\mathbb{E}\{\mathbf{H}_n(i,j)(\mathbf{H}_{n-m}(i,j))^*\} = r_{\mathbb{H}}(m)$, $(i=1,2,\ldots,\mathrm{M_R}, j=1,2,\ldots,\mathrm{M_T})$. Defining $\mathbf{H} = [\mathbf{H}_0\ \mathbf{H}_1\ \ldots\ \mathbf{H}_{N-1}]$, we therefore have

$$\mathbb{E}\{\mathrm{vec}(\mathbf{H})\,(\mathrm{vec}(\mathbf{H}))^H\} = \mathbf{R}_{\mathbb{H}} \otimes \mathbf{I}_{\mathrm{M_T M_R}} \tag{2.2}$$

where the covariance matrix $\mathbf{R}_{\mathbb{H}}(i,j) = r_{\mathbb{H}}(i-j)$ $(i,j=0,1,\ldots,N-1)$ follows from the channel's scattering function [4]. In the purely frequency-selective case, e.g., assuming an orthogonal frequency-division multiplexing (OFDM) system [33] with N tones and hence $\mathbf{H}_n = \sum_{l=0}^{L-1} \mathbf{H}(l)\, e^{-j\frac{2\pi}{N}ln}$, where the uncorrelated (across l) matrix-valued taps $\mathbf{H}(l)$ have i.i.d. $\mathcal{CN}(0,\sigma_l^2)$ entries, we obtain $r_{\mathbb{H}}(m) = \sum_{l=0}^{L-1} \sigma_l^2 e^{-j\frac{2\pi}{N}lm}$ $(m=0,1,\ldots,N-1)$. In the remainder of this chapter, we use the definition $\rho := \mathrm{rank}(\mathbf{R}_{\mathbb{H}})$.

2.3 Diversity-multiplexing tradeoff

2.3.1 Preliminaries

Assuming perfect CSI in the receiver, the mutual information of the channel in (2.1) is given by

$$\mathrm{I}(\mathsf{SNR}) = \frac{1}{N}\sum_{n=0}^{N-1} \log\det\left(\mathbf{I}_{\mathrm{M_R}} + \frac{\mathsf{SNR}}{\mathrm{M_T}} \mathbf{H}_n \mathbf{C}_n \mathbf{H}_n^H\right) \tag{2.3}$$

where the transmit signal vectors are uncorrelated across n and satisfy $\mathbf{x}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_n)$ with power constraint $\mathrm{Tr}\,(\mathbf{C}_n) \leq \mathrm{M_T}$, $n=0,1,\ldots,N-1$. The DM-tradeoff realized by a family (w.r.t. SNR) of codes $\mathcal{C}_r$ with rate $R(\mathsf{SNR}) = r\log\mathsf{SNR}$, where $r \in [0, \min\{\mathrm{M_T}, \mathrm{M_R}\}]$, is given by the function

$$d_{\mathcal{C}}(r) = -\lim_{\mathsf{SNR}\to\infty} \frac{\log P_e(r,\mathsf{SNR})}{\log \mathsf{SNR}}$$

where $P_e(r,\mathsf{SNR})$ is the error probability obtained through ML detection. At a given SNR, the corresponding codebook $\mathcal{C}_r(\mathsf{SNR})$ contains SNR^{Nr} codewords $\mathbf{X} = [\mathbf{x}_0\ \mathbf{x}_1\ \ldots\ \mathbf{x}_{N-1}]$. We say that such a family of codes $\mathcal{C}_r$ operates at multiplexing rate r. The optimal tradeoff curve $d^\star(r) = \sup_{\mathcal{C}_r} d_{\mathcal{C}}(r)$, where the supremum is taken over all families of codes satisfying $R(\mathsf{SNR}) = r \log \mathsf{SNR}$, quantifies the maximum achievable diversity gain as a function of r. Since the outage probability $P_{\mathcal{O}}(r, \mathsf{SNR})$ is a lower bound to the error probability [57], we have

$$d^{\star}(r) \leq d_{\mathcal{O}}(r) = -\lim_{\mathsf{SNR}\to\infty} \frac{\log P_{\mathcal{O}}(r, \mathsf{SNR})}{\log \mathsf{SNR}}.$$

Extending the arguments that lead to [57, Eq. (9)] to the case $N > 1$, we can conclude that setting $\mathbf{C}_n = \mathbf{I}_{\mathrm{M_T}}$ $(n = 0, 1, \ldots, N-1)$ in (2.3) does not alter the exponential behavior of mutual information. Hence

$$P_{\mathcal{O}}(r, \mathsf{SNR}) \doteq \mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1} \log\det\left(\mathbf{I}_{\mathrm{M_R}} + \mathsf{SNR}\,\mathbf{H}_n\mathbf{H}_n^H\right) < r \log \mathsf{SNR}\right) \tag{2.4}$$

where we used the fact that the factor $1/\mathrm{M_T}$ in (2.3) can be neglected in the scale of interest. Let $\boldsymbol{\mu}(n) := [\mu_1(n)\ \mu_2(n)\ \ldots\ \mu_{\min\{\mathrm{M_T},\mathrm{M_R}\}}(n)]$ $(n = 0, 1, \ldots, N-1)$, with the singularity levels defined as

$$\mu_k(n) = -\frac{\log \lambda_k(\mathbf{H}_n\mathbf{H}_n^H)}{\log \mathsf{SNR}}, \quad k = 1, 2, \ldots, \min\{\mathrm{M_T}, \mathrm{M_R}\}$$

and note that [57]

$$P_{\mathcal{O}}(r, \mathsf{SNR}) \doteq \mathbb{P}\left(\mathcal{O}(r)\right) \tag{2.5}$$

where

$$\mathcal{O}(r) = \left\{\boldsymbol{\mu}(n) \in \mathbb{R}_+^{\min\{\mathrm{M_T},\mathrm{M_R}\}}, n = 0, 1, \ldots, N-1 : \frac{1}{N}\sum_{n=0}^{N-1}\sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}} [1 - \mu_k(n)]^+ < r\right\} \tag{2.6}$$

and $\mathbb{R}_+^{\min\{\mathrm{M_T},\mathrm{M_R}\}}$ denotes the nonnegative orthant. Unlike the frequency-flat fading case treated in [57], characterizing $d_{\mathcal{O}}(r)$ for the selective-fading case seems analytically intractable with the main difficulty stemming from the fact that one has to deal with the sum of correlated (recall that the $\mathbf{H}_n$ are correlated across n) terms in (2.4). It turns out, however, that one can find lower and upper bounds on $\mathrm{I}(\mathsf{SNR})$ which are exponentially tight (and, hence, preserve the DM-tradeoff behavior) and analytically tractable. The next section formalizes this idea.

2.3.2 Jensen channel and Jensen outage event

We start by noting that applying Jensen's inequality yields

$$\mathrm{I}(\mathsf{SNR}) = \frac{1}{N}\sum_{n=0}^{N-1} \log\det\left(\mathbf{I}_{\mathrm{M_R}} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{H}_n\mathbf{H}_n^H\right) \leq$$
$$\log\det\left(\mathbf{I}_{\min\{\mathrm{M_T},\mathrm{M_R}\}} + \frac{\mathsf{SNR}}{\mathrm{M_T}N}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^H\right) := \mathrm{J}(\mathsf{SNR}) \quad (2.7)$$

where the "Jensen channel" is defined as

$$\boldsymbol{\mathcal{H}} = \begin{cases} [\mathbf{H}_0\ \mathbf{H}_1\ \ldots\ \mathbf{H}_{N-1}], & \text{if } \mathrm{M_R} \leq \mathrm{M_T}, \\ [\mathbf{H}_0^H\ \mathbf{H}_1^H\ \ldots\ \mathbf{H}_{N-1}^H], & \text{if } \mathrm{M_R} > \mathrm{M_T}. \end{cases}$$

In the following, we say that a Jensen outage event occurs if the Jensen channel $\boldsymbol{\mathcal{H}}$ is in outage w.r.t. the rate $R(\mathsf{SNR}) = r\log\mathsf{SNR}$, i.e., if $\mathrm{J}(\mathsf{SNR}) < R(\mathsf{SNR})$. The corresponding outage probability will be denoted as $P_{\mathcal{J}}(r,\mathsf{SNR})$ and clearly satisfies $P_{\mathcal{J}}(r,\mathsf{SNR}) \leq P_{\mathcal{O}}(r,\mathsf{SNR})$. The operational significance of a Jensen outage will be established at the end of this section. We shall first focus on characterizing the Jensen outage event analytically. Using (2.2), it is readily seen that $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_w(\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{\max\{\mathrm{M_T},\mathrm{M_R}\}})$, where $\boldsymbol{\mathcal{H}}_w$ is an i.i.d. $\mathcal{CN}(0,1)$ matrix with the same dimensions as $\boldsymbol{\mathcal{H}}$. Noting that $\boldsymbol{\mathcal{H}}_w\mathbf{U} \sim \boldsymbol{\mathcal{H}}_w$ for $\mathbf{U}$ unitary and using the eigendecomposition $\mathbf{R}_{\mathbb{H}} \otimes \mathbf{I}_{\max\{\mathrm{M_T},\mathrm{M_R}\}} = \mathbf{U}(\boldsymbol{\Lambda} \otimes \mathbf{I}_{\max\{\mathrm{M_T},\mathrm{M_R}\}})\mathbf{U}^H$, where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1(\mathbf{R}_{\mathbb{H}}), \lambda_2(\mathbf{R}_{\mathbb{H}}), \ldots, \lambda_\rho(\mathbf{R}_{\mathbb{H}}), 0, \ldots, 0\}$, it follows that

$$\begin{aligned}\mathrm{J}(\mathsf{SNR}) &= \log\det\left(\mathbf{I}_{\min\{\mathrm{M_T},\mathrm{M_R}\}} + \frac{\mathsf{SNR}}{\mathrm{M_T}N}\boldsymbol{\mathcal{H}}_w(\mathbf{R}_{\mathbb{H}} \otimes \mathbf{I}_{\max\{\mathrm{M_T},\mathrm{M_R}\}})\boldsymbol{\mathcal{H}}_w^H\right) \\ &\sim \log\det\left(\mathbf{I}_{\min\{\mathrm{M_T},\mathrm{M_R}\}} + \frac{\mathsf{SNR}}{\mathrm{M_T}N}\boldsymbol{\mathcal{H}}_w\left(\boldsymbol{\Lambda} \otimes \mathbf{I}_{\max\{\mathrm{M_T},\mathrm{M_R}\}}\right)\boldsymbol{\mathcal{H}}_w^H\right).\end{aligned}$$

Next, observe that the following positive semidefinite ordering holds

$$\begin{aligned}\lambda_1(\mathbf{R}_{\mathbb{H}})\,\mathrm{diag}\left\{\mathbf{I}_{\rho\max\{\mathrm{M_T},\mathrm{M_R}\}}, \mathbf{0}\right\} &\preceq \boldsymbol{\Lambda} \otimes \mathbf{I}_{\max\{\mathrm{M_T},\mathrm{M_R}\}} \\ &\preceq \lambda_\rho(\mathbf{R}_{\mathbb{H}})\,\mathrm{diag}\left\{\mathbf{I}_{\rho\max\{\mathrm{M_T},\mathrm{M_R}\}}, \mathbf{0}\right\}. \quad (2.8)\end{aligned}$$

Since $f(\mathbf{A}) = \log\det(\mathbf{I}+\mathbf{A})$ is increasing over the cone of positive semidefinite matrices [15], we get the following bounds on the Jensen outage probability

$$\begin{aligned}&\mathbb{P}\left(\log\det\left(\mathbf{I}_{\min\{\mathrm{M_T},\mathrm{M_R}\}} + \lambda_\rho(\mathbf{R}_{\mathbb{H}})\frac{\mathsf{SNR}}{\mathrm{M_T}N}\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\right) < r\log\mathsf{SNR}\right) \\ &\leq P_{\mathcal{J}}(r,\mathsf{SNR}) \quad (2.9) \\ &\leq \mathbb{P}\left(\log\det\left(\mathbf{I}_{\min\{\mathrm{M_T},\mathrm{M_R}\}} + \lambda_1(\mathbf{R}_{\mathbb{H}})\frac{\mathsf{SNR}}{\mathrm{M_T}N}\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\right) < r\log\mathsf{SNR}\right)\end{aligned}$$

where $\overline{\boldsymbol{\mathcal{H}}}_w = \boldsymbol{\mathcal{H}}_w([1\!:\!\min\{\mathrm{M_T},\mathrm{M_R}\}],[1\!:\!\rho\max\{\mathrm{M_T},\mathrm{M_R}\}])$. Taking the exponential limit (in SNR) in (2.9), it follows readily that

$$P_{\mathcal{J}}(r,\mathsf{SNR}) \doteq \mathbb{P}\left(\log\det\left(\mathbf{I}_{\min\{\mathrm{M_T},\mathrm{M_R}\}} + \mathsf{SNR}\,\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\right) < r\log\mathsf{SNR}\right). \quad (2.10)$$

For later use, we define $\boldsymbol{\alpha} := [\alpha_1\ \alpha_2\ \ldots\ \alpha_{\min\{\mathrm{M_T},\mathrm{M_R}\}}]$ with the singularity levels

$$\alpha_k = -\frac{\log\lambda_k(\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H)}{\log\mathsf{SNR}}, \quad k=1,2,\ldots,\min\{\mathrm{M_T},\mathrm{M_R}\} \quad (2.11)$$

and note that $P_{\mathcal{J}}(r,\mathsf{SNR}) \doteq \mathbb{P}\left(\mathcal{J}(r)\right)$, where

$$\mathcal{J}(r) = \left\{\boldsymbol{\alpha}\in\mathbb{R}_+^{\min\{\mathrm{M_T},\mathrm{M_R}\}}\colon \alpha_1\geq\alpha_2\geq\ldots\geq\alpha_{\min\{\mathrm{M_T},\mathrm{M_R}\}},\ \sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}}[1-\alpha_k]^+ < r\right\}.$$

It is now natural to define the Jensen outage curve as

$$d_{\mathcal{J}}(r) = -\lim_{\mathsf{SNR}\to\infty}\frac{\log P_{\mathcal{J}}(r,\mathsf{SNR})}{\log\mathsf{SNR}}.$$

Based on (2.10), we can conclude that $d_{\mathcal{J}}(r)$ is nothing but the DM-tradeoff curve of an effective MIMO channel with $\rho\max\{\mathrm{M_T},\mathrm{M_R}\}$ transmit and $\min\{\mathrm{M_T},\mathrm{M_R}\}$ receive antennas. We can therefore directly apply the results in [57] to infer that the Jensen outage curve is the piecewise linear function connecting the points $(r, d_{\mathcal{J}}(r))$ for $r=0,1,\ldots,\min\{\mathrm{M_T},\mathrm{M_R}\}$, with

$$d_{\mathcal{J}}(r) = (\rho\max\{\mathrm{M_T},\mathrm{M_R}\}-r)(\min\{\mathrm{M_T},\mathrm{M_R}\}-r). \quad (2.12)$$

Since, as already noted, $P_{\mathcal{J}}(r,\mathsf{SNR}) \leq P_{\mathcal{O}}(r,\mathsf{SNR})$, we obtain

$$d_{\mathcal{C}}(r) \leq d^\star(r) \leq d_{\mathcal{O}}(r) \leq d_{\mathcal{J}}(r), \quad r\in[0,\min\{\mathrm{M_T},\mathrm{M_R}\}], \quad (2.13)$$

for any family of codes $\mathcal{C}_r$. The optimal DM-tradeoff curve $d^\star(r)$ will be established in the next section by showing that codes satisfying $d_{\mathcal{C}}(r) = d_{\mathcal{J}}(r)$ do exist and hence $d^\star(r) = d_{\mathcal{J}}(r)$.

2.4 Jensen-optimal code design criterion

The goal of this section is to derive a sufficient condition for a family of codes to achieve $d_{\mathcal{J}}(r)$, and hence, by virtue of (2.13), to be DM-tradeoff optimal.

2.4.1 Code design criterion

Theorem 2. *Consider a family of codes $\mathcal{C}_r$ with block length $N \geq \rho \mathrm{M_T}$ that operates over the channel (2.1). If, for any codebook $\mathcal{C}_r(\mathsf{SNR}) \in \mathcal{C}_r$ and any two codewords $\mathbf{X}$, $\mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})$, the codeword difference matrix $\mathbf{E} = \mathbf{X} - \mathbf{X}'$ is such that*

$$\mathrm{rank}\big(\mathbf{R}_{\mathbb{H}} \odot \mathbf{E}^H \mathbf{E}\big) = \rho \mathrm{M_T} \tag{2.14}$$

then the error probability (for ML decoding) satisfies

$$P_e(r, \mathsf{SNR}) \doteq \mathsf{SNR}^{-d_{\mathcal{J}}(r)}.$$

Proof. We start by deriving an upper bound on the average (w.r.t. the random channel) pairwise error probability (PEP). Assuming that $\mathbf{X}$ was transmitted, the probability of the ML decoder mistakenly deciding in favor of codeword $\mathbf{X}'$ can be upper-bounded in terms of the codeword difference vectors $\mathbf{e}_n = \mathbf{x}_n - \mathbf{x}'_n$ $(n = 0, 1, \ldots, N-1)$ as

$$\begin{aligned}\mathbb{P}\left(\mathbf{X} \to \mathbf{X}'\right) &\leq \mathbb{E}_{\mathbf{H}}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}} \sum_{n=0}^{N-1} \|\mathbf{H}_n \mathbf{e}_n\|^2\right)\right\} \\ &= \mathbb{E}_{\mathbf{H}}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}} \mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Upsilon} \mathbf{H}_w^H\right)\right)\right\}\end{aligned}$$

where

$$\boldsymbol{\Upsilon} = (\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{\mathrm{M_T}})\, \mathrm{diag}\big\{\mathbf{e}_n \mathbf{e}_n^H\big\}_{n=0}^{N-1} (\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{\mathrm{M_T}})$$

and $\mathbf{H}_w$ denotes an $\mathrm{M_R} \times \mathrm{M_T} N$ i.i.d. $\mathcal{CN}(0,1)$ matrix. Straightforward manipulations reveal that $\mathrm{rank}(\boldsymbol{\Upsilon}) = \mathrm{rank}\big(\mathbf{R}_{\mathbb{H}} \odot \mathbf{E}^H \mathbf{E}\big)$ so that the assumption (2.14) implies $\mathrm{rank}(\boldsymbol{\Upsilon}) = \rho \mathrm{M_T}$. With the eigendecomposition $\boldsymbol{\Upsilon} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^H$, we have $\mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Upsilon} \mathbf{H}_w^H\right) \sim \mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Lambda} \mathbf{H}_w^H\right)$, and hence

$$\mathbb{P}\left(\mathbf{X} \to \mathbf{X}'\right) \leq \mathbb{E}_{\mathbf{H}}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}} \mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Lambda} \mathbf{H}_w^H\right)\right)\right\}.$$

Setting $\overline{\mathbf{H}}_w = \mathbf{H}_w([1{:}\mathrm{M_R}], [1{:}\rho \mathrm{M_T}])$ and denoting the smallest nonzero eigenvalue of $\boldsymbol{\Upsilon}$ as λ, we note that

$$\mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Lambda} \mathbf{H}_w^H\right) \geq \lambda\, \mathrm{Tr}\left(\overline{\mathbf{H}}_w \overline{\mathbf{H}}_w^H\right) \tag{2.15}$$

and thus

$$\mathbb{P}\left(\mathbf{X} \to \mathbf{X}'\right) \leq \mathbb{E}_{\overline{\mathbf{H}}_w}\left\{\exp\left(-\frac{\lambda\, \mathsf{SNR}}{4\mathrm{M_T}} \mathrm{Tr}\left(\overline{\mathbf{H}}_w \overline{\mathbf{H}}_w^H\right)\right)\right\}. \tag{2.16}$$

Next, note that

$$\begin{aligned}\mathrm{Tr}\left(\overline{\mathbf{H}}_w\overline{\mathbf{H}}_w^H\right) &= \mathrm{Tr}\left(\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\right)\\ &= \sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}} \lambda_k(\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H)\\ &= \sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}} \mathsf{SNR}^{-\alpha_k} \qquad (2.17)\end{aligned}$$

where (2.17) follows from (2.11). We can now write the PEP upper-bound in (2.16) in terms of the singularity levels α_k ($k=1,2,\ldots,\min\{\mathrm{M_T},\mathrm{M_R}\}$) characterizing the Jensen outage event:

$$\mathbb{P}\left(\mathbf{X}\to\mathbf{X}'\right) \leq \mathbb{E}_{\boldsymbol{\alpha}}\left\{\exp\left(-\frac{\lambda}{4\mathrm{M_T}}\sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}}\mathsf{SNR}^{1-\alpha_k}\right)\right\}. \qquad (2.18)$$

Next, consider a realization of the random vector $\boldsymbol{\alpha}$ and let $\mathcal{S}=\{k:\alpha_k\leq 1\}$. We have

$$\begin{aligned}\sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}}\mathsf{SNR}^{1-\alpha_k} &\geq \sum_{k\in\mathcal{S}}\mathsf{SNR}^{1-\alpha_k}\\ &\overset{\text{(i)}}{\geq} |\mathcal{S}|\ \mathsf{SNR}^{\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}}(1-\alpha_k)}\\ &\overset{\text{(ii)}}{=} |\mathcal{S}|\ \mathsf{SNR}^{\frac{1}{|\mathcal{S}|}\sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}}[1-\alpha_k]^+} \qquad (2.19)\end{aligned}$$

where (i) follows from the arithmetic-geometric mean inequality and (ii) follows from the definition of $\mathcal{S}$. Using (2.19) in (2.18), we get

$$\mathbb{P}\left(\mathbf{X}\to\mathbf{X}'\right) \leq \mathbb{E}_{\boldsymbol{\alpha}}\left\{\exp\left(-\frac{\lambda\,|\mathcal{S}|}{4\mathrm{M_T}}\,\mathsf{SNR}^{\frac{1}{|\mathcal{S}|}\sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}}[1-\alpha_k]^+}\right)\right\}. \qquad (2.20)$$

The dependence of the PEP upper bound (2.20) on the singularity levels characterizing the Jensen outage event suggests to split up the overall error probability according to

$$\begin{aligned}P_e(r,\mathsf{SNR}) &= \mathbb{P}\left(\mathrm{error},\boldsymbol{\alpha}\in\mathcal{J}(r)\right)+\mathbb{P}\left(\mathrm{error},\boldsymbol{\alpha}\notin\mathcal{J}(r)\right)\\ &= \mathbb{P}\left(\boldsymbol{\alpha}\in\mathcal{J}(r)\right)\mathbb{P}\left(\mathrm{error}|\boldsymbol{\alpha}\in\mathcal{J}(r)\right)\\ &\quad +\mathbb{P}\left(\boldsymbol{\alpha}\notin\mathcal{J}(r)\right)\mathbb{P}\left(\mathrm{error}|\boldsymbol{\alpha}\notin\mathcal{J}(r)\right)\\ &\leq \mathbb{P}\left(\boldsymbol{\alpha}\in\mathcal{J}(r)\right)\\ &\quad +\mathbb{P}\left(\boldsymbol{\alpha}\notin\mathcal{J}(r)\right)\mathbb{P}\left(\mathrm{error}|\boldsymbol{\alpha}\notin\mathcal{J}(r)\right). \qquad (2.21)\end{aligned}$$

For any $\boldsymbol{\alpha} \notin \mathcal{J}(r)$, we have $\sum_{k=1}^{\min\{\mathrm{M_T},\mathrm{M_R}\}}[1-\alpha_k]^+ \geq r$ and $|\mathcal{S}| \geq 1$, which upon noting that $|\mathcal{C}_r(\mathsf{SNR})| = \mathsf{SNR}^{Nr}$, yields the following union bound based on the PEP in (2.20)

$$\mathbb{P}\left(\text{error}|\boldsymbol{\alpha} \notin \mathcal{J}(r)\right) \leq \mathsf{SNR}^{Nr} \exp\left(-\frac{\lambda}{4\mathrm{M_T}}\,\mathsf{SNR}^{r/\min\{\mathrm{M_T},\mathrm{M_R}\}}\right)$$

where we used $|\mathcal{S}| \leq \min\{\mathrm{M_T},\mathrm{M_R}\}$. Hence, for any $r > 0$, $\mathbb{P}\left(\text{error}|\boldsymbol{\alpha} \notin \mathcal{J}(r)\right)$ decays exponentially in SNR and we have

$$\begin{aligned}\mathbb{P}\left(\text{error}, \boldsymbol{\alpha} \notin \mathcal{J}(r)\right) &= \underbrace{\mathbb{P}\left(\boldsymbol{\alpha} \notin \mathcal{J}(r)\right)}_{\leq 1} \mathbb{P}\left(\text{error}|\boldsymbol{\alpha} \notin \mathcal{J}(r)\right)\\ &\leq \mathsf{SNR}^{Nr} \exp\left(-\frac{\lambda}{4\mathrm{M_T}}\,\mathsf{SNR}^{r/\min\{\mathrm{M_T},\mathrm{M_R}\}}\right). \qquad (2.22)\end{aligned}$$

Consequently, noting that $\mathbb{P}\left(\boldsymbol{\alpha} \in \mathcal{J}(r)\right) \doteq P_{\mathcal{J}}(r, \mathsf{SNR})$ and using (2.22) in (2.21), we obtain

$$P_e(r, \mathsf{SNR}) \,\dot{\leq}\, P_{\mathcal{J}}(r, \mathsf{SNR}).$$

Since $P_{\mathcal{J}}(r, \mathsf{SNR}) \leq P_{\mathcal{O}}(r, \mathsf{SNR})$, it follows trivially that $P_{\mathcal{J}}(r, \mathsf{SNR}) \,\dot{\leq}\, P_{\mathcal{O}}(r, \mathsf{SNR})$. In addition, for a specific family of codes $\mathcal{C}_r$, we have $P_{\mathcal{O}}(r, \mathsf{SNR}) \leq P_e(r, \mathsf{SNR})$ and hence $P_{\mathcal{O}}(r, \mathsf{SNR}) \,\dot{\leq}\, P_e(r, \mathsf{SNR})$. Putting the pieces together, we finally obtain

$$P_{\mathcal{O}}(r, \mathsf{SNR}) \,\dot{\leq}\, P_e(r, \mathsf{SNR}) \,\dot{\leq}\, P_{\mathcal{J}}(r, \mathsf{SNR}) \,\dot{\leq}\, P_{\mathcal{O}}(r, \mathsf{SNR})$$

which implies

$$P_e(r, \mathsf{SNR}) \doteq P_{\mathcal{J}}(r, \mathsf{SNR})$$

and hence (by definition of $d_{\mathcal{J}}(r)$)

$$P_e(r, \mathsf{SNR}) \doteq \mathsf{SNR}^{-d_{\mathcal{J}}(r)}.$$

■

As a direct consequence of Theorem 2, a family of codes that satisfies (2.14) for all codeword difference matrices in any codebook $\mathcal{C}_r(\mathsf{SNR}) \in \mathcal{C}_r$ realizes a DM-tradeoff curve $d_{\mathcal{C}}(r) = d_{\mathcal{J}}(r)$ and hence, by (2.13)

$$d_{\mathcal{J}}(r) \leq d^{\star}(r) \leq d_{\mathcal{J}}(r)$$

which implies

$$d^{\star}(r) = d_{\mathcal{J}}(r). \qquad (2.23)$$

The optimal DM-tradeoff curve for selective-fading MIMO channels is therefore given by the DM-tradeoff curve of the associated Jensen channel. Put differently, Theorem 2 shows that, even though $\mathcal{J}(r) \subseteq \mathcal{O}(r)$ by definition, we still have

$$\mathbb{P}(\mathcal{J}(r)) \doteq \mathbb{P}(\mathcal{O}(r))$$

which essentially says that the "original" channel has the same high-SNR outage behavior as its associated Jensen channel.

The code design criterion in Theorem 2 provides a sufficient condition for achieving the DM-tradeoff curve. Interestingly, the classical rank criterion [48, 47, 14, 13, 12, 30], aimed at maximizing the diversity gain for $r = 0$, can be shown [16] to be equivalent to the criterion in Theorem 2. We emphasize, however, that optimality w.r.t. the DM-tradeoff at multiplexing rate r requires that (2.14) is satisfied for all codeword difference matrices in any codebook $\mathcal{C}_r(\mathsf{SNR}) \in \mathcal{C}_r$, in particular also for $\mathsf{SNR} \to \infty$. We next state a sufficient condition for DM-tradeoff optimality which makes this aspect explicit and establishes a connection to the approximately universal code design criterion in [49].

Corollary 1. *A family of codes $\mathcal{C}_r$ of block length $N \geq \rho \mathrm{M_T}$ is DM-tradeoff optimal if there exists an $\epsilon > 0$ such that*

$$\lambda^{\min\{\mathrm{M_T},\mathrm{M_R}\}}(\mathsf{SNR}) \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)} \tag{2.24}$$

where

$$\lambda(\mathsf{SNR}) = \min_{\substack{k=1,2,\ldots,\rho\mathrm{M_T} \\ \mathbf{E}=\mathbf{X}-\mathbf{X}',\, \mathbf{X},\mathbf{X}'\in\mathcal{C}_r(\mathsf{SNR})}} \left\{ \lambda_k(\mathbf{R}_{\mathbb{H}} \odot \mathbf{E}^H\mathbf{E}) > 0 \right\}.$$

Proof. Using (2.24) in (2.22), we obtain

$$\mathbb{P}(\text{error}, \boldsymbol{\alpha} \notin \mathcal{J}(r)) \leq \mathsf{SNR}^{Nr} \exp\left(- \frac{\mathsf{SNR}^{\epsilon/\min\{\mathrm{M_T},\mathrm{M_R}\}}}{4\mathrm{M_T}} \right)$$

which, following the same logic as in the proof of Theorem 2, implies that $P_e(r, \mathsf{SNR}) \doteq \mathsf{SNR}^{-d_{\mathcal{J}}(r)}$. ∎

Note that the quantity $\lambda^{\min\{\mathrm{M_T},\mathrm{M_R}\}}(\mathsf{SNR})$ is trivially a lower bound on the product of the $\min\{\mathrm{M_T}, \mathrm{M_R}\}$ smallest nonzero eigenvalues of any codeword difference matrix in the codebook $\mathcal{C}_r(\mathsf{SNR})$. Consequently, in the case of non-selective fading, where $\mathbf{R}_{\mathbb{H}} \odot \mathbf{E}^H\mathbf{E} = \mathbf{E}^H\mathbf{E}$, any family of codes $\mathcal{C}_r$ satisfying (2.24) will also be approximately universal in the sense of [49, Th. 3.1]. Moreover, if $\lambda(\mathsf{SNR})$ remains strictly positive as $\mathsf{SNR} \to \infty$, $\mathcal{C}_r$ fulfills the non-vanishing determinant criterion [2, 54] and will, by (2.22) and the same arguments as in the proof of Theorem 2, be DM-tradeoff optimal.

2.4.2 Application to the frequency-selective case

As an example, we shall next specialize our results to frequency-selective fading MIMO channels, recovering the results reported previously in [21,31]. For the sake of simplicity of exposition, we shall employ a cyclic signal model, as obtained in an OFDM system for example. The channel's transfer function is given by

$$\mathbf{H}(e^{j2\pi\theta}) = \sum_{l=0}^{L-1} \mathbf{H}(l)\, e^{-j2\pi l\theta}, \quad 0 \leq \theta < 1$$

where the $\mathbf{H}(l)$ have i.i.d. $\mathcal{CN}(0, \sigma_l^2)$ entries and satisfy

$$\mathbb{E}\Big\{\mathrm{vec}(\mathbf{H}(l))\, \mathrm{vec}(\mathbf{H}(l'))^H\Big\} = \sigma_l^2\, \delta(l-l')\, \mathbf{I}_{\mathrm{M_T M_R}}.$$

With $\mathbf{H}_n = \mathbf{H}(e^{j2\pi\frac{n}{N}})$, $n=0,1\ldots,N-1$, the channel's covariance matrix follows as

$$\mathbf{R}_{\mathbb{H}} = \mathbf{F}\, \mathrm{diag}\big\{\sigma_0^2, \sigma_1^2, \ldots, \sigma_{L-1}^2, 0, \ldots, 0\big\}\, \mathbf{F}^H$$

where $\mathbf{F}$ is the $N \times N$ FFT matrix. Since $\mathrm{rank}(\mathbf{R}_{\mathbb{H}}) = L$, inserting $\rho = L$ into (2.12) and using (2.23) yields the optimal DM-tradeoff curve as the piecewise linear function connecting the points $(r, d^{\star}(r))$ for $r = 0, 1, \ldots, \min\{\mathrm{M_T}, \mathrm{M_R}\}$, with

$$d^{\star}(r) = (L\max\{\mathrm{M_T}, \mathrm{M_R}\} - r)(\min\{\mathrm{M_T}, \mathrm{M_R}\} - r). \tag{2.25}$$

This is the optimal DM-tradeoff curve for frequency-selective fading MIMO channels reported previously in [31]. Specializing (2.25) to the case $\mathrm{M_T} = \mathrm{M_R} = 1$ and noting that $d^{\star}(r) = (L-r)(1-r) = L(1-r)$ for $r = \{0, 1\}$, yields the results reported in [21]. We note that the proof techniques employed in [21, 31] are different from the approach taken in this chapter and seem to be tailored to the frequency-selective case. In addition, our approach is not limited to large code lengths as (2.14) can be guaranteed for any $N \geq L\mathrm{M_T}$.

2.5 Conclusions

Analyzing the high-SNR outage behavior of the Jensen channel instead of the original channel was found to be an effective tool to establish the DM-tradeoff in selective-fading MIMO channels. We showed that satisfying extensions (to the selective-fading MIMO case) of the approximately universal code design criterion [49] and the non-vanishing determinant criterion [2, 54] results in DM-tradeoff optimal codes. Finally, we note that the concepts introduced in this chapter can be extended to multiple-access selective-fading MIMO channels and to the analysis of the DM-tradeoff properties of specific (suboptimal) receivers.

Chapter 3

Introduction to Interference Function

The wireless channel is a broadcast medium, so each communication link is possibly interfered by other users transmitting at the same resource. The traditional way of handling interference is to assign all links orthogonal resources, in time (TDMA), frequency (FDMA), or code space (CDMA). This considerably simplifies the system design since the links are no longer coupled by interference. However, reserving each link a fixed resource often comes at the cost of sacrificing spectral efficiency. The available bandwidth is generally best exploited by letting transmitted signals interfere with each other in a controlled way. Also, orthogonality may be lost because of system imperfections and the effects of the time-varying multipath channel. It can be said that interference and power constraints are the main hurdle in achieving a high per-user throughput in heavily loaded multiuser networks, as will be required in the future.

Since interference plays an important role in the optimal exploitation of the given bandwidth, it is generally not sufficient to regard the system as a collection of point-to-point communication links. The quality-of-service (QoS) of each link depends on its own transmission power, but also on the power levels of the other links, which are experienced as interference. This results in a competitive situation, where all users try to compensate interference by increasing its own transmission power, which in turn increases the overall interference in the system. A transmission strategy which neglects these interdependencies is likely to cause uncontrollable and exceeding interference, which means a waste of the overall system efficiency. Thus, it is desirable to find a suitable equilibrium that optimally exploits the available resources. This requires a joint optimization of all communication links.

Optimization can be performed with respect to various design goals, like the overall efficiency, max-min fairness, proportional fairness, network utility maximization, etc. There is no such thing as “the” optimal communication strategy. There exists a great deal of literature on resource allocation from various points of view. For example, there are network-centric strategies, which aim at finding a stable performance trade-off by bidding strategies, accounting for traffic, channel quality, and revenues. User-centric strategies, which are closely related to power control, aim at fulfilling user-specific QoS requirements. Both strategies have in common that they are determined and limited by the QoS feasible region (the set of jointly achievable QoS).

We provide a fundamental theoretical framework which helps to understand the underlying effects of interference coupling, and to characterize the QoS feasible region. A fundamental question in this context is: what is the region of jointly achievable QoS, and how can certain points be achieved in a spectrally efficient way? This question is closely related to the classical power allocation problem, but in this text we will go one step further in assuming that interference not only depends on the power allocation, but also on adaptive receive strategies, like interference filtering or channel assignment. The additional optimization of the receive strategy adds new degrees of freedom to the problem of resource allocation. Thus, new concepts and algorithms are required.

Since power allocation and receive strategies are intricately intertwined, our approach is to use abstract models, which provide a better understanding of the underlying structure of the problem at hand. In this respect, the work can be seen as a theoretical basis, which can be applied to solve existing problems in wireless communications.

3.1 QoS-based power and resource allocation

In this section we give an overview on some aspects of QoS-Based power allocation. We start by introducing the basic model used throughout this text, which will be refined later on.

3.1.1 Interference functions

Consider a network with K communication links, whose transmission powers are collected in a power allocation vector

$$p = [p_1, \dots, p_K]^T > 0$$

The interference power experienced by the kth user can be modeled by a function $I_k(p)$. The functions $I_1, \dots, I_K$ describe how the links are affected by mutual cross-talk. Different definitions of $I_k(p)$ and the resulting QoS region will be analyzed in this text. It should be noted that the mapping $\mathcal{I}_k : \mathbb{R}_+^K \to \mathbb{R}_+$ can be linear or non-linear, and it can also model the impact of adaptive receiver designs, like MMSE or interference cancellation, as well as other system aspects. A few examples are listed in the following.

- $I_k(p) = [\Psi p]_k$, where the positive coupling matrix $\Psi > 0$ contains interference coefficients, which determine in which way the users are affected by cross-talk (interference). This is a common model in power control theory.

- $I_k(p) = \min_{z \in Z}[\Psi(z)p]_k$, where the adjustable receive strategy z (from a compact set of possible strategies Z) has impact on the interference structure. This specific model, which holds e.g. for multiantenna beamforming or CDMA designs, and many more.

- $I_k(p) = \max_c f_k(p,c)$, where $f_k(p,c)$ is the interference for a given power allocation p under some interference uncertainty c. This definition can be used, e.g. to model worst-case interference under imperfect channel knowledge.

But instead of focusing on a particular model, this text aims at characterizing basic properties, which are a common for a wide range of interference functions. To this end, we introduce an axiomatic characterization of interference functions. This generic model contains the above examples as special cases. The axiomatic framework will be gradually refined in the following sections. By introducing additional properties, more results can be shown.

The signal-to-interference ratio (SIR) of the kth user is

$$SIR_k(p) = \frac{p_k}{I_k(p)}, \qquad k \le k \le K \tag{3.1}$$

where p_k is the desired transmission power of the kth user. Note, that the function $I_k(p)$ can include receiver noise. If noise is part of the assumed model, then we will emphasize this by using "SINR" instead of "SIR". If we use SIR, then we discuss the general case where noise can be included or not. In this case, we need $I_k(p) > 0$ to ensure that (3.1) is well defined.

The term "QoS" is commonly used to describe the performance and reliability of a communication link. In order to keep the results as general as possible, we do not make any specific assumption on QoS, except that it is related to the SIR by a monotonic and bijective function ϕ:

$$QoS_k(p) = \phi(SIR_k(p)), \qquad 1 \leq k \leq K \tag{3.2}$$

Some examples are BER: $\phi(x) = Q(\sqrt{x})$, MMSE: $\phi(x) = 1/(1+x)$, BER-slope for α-fold diversity: $\phi(x) = x^{-\alpha}$, or capacity: $\phi(x) = \log(1+x)$. Let γ be the inverse function of ϕ, then

$$\gamma_k = \gamma(Q_k), \qquad 1 \leq k \leq K$$

is the minimum SIR level needed by the kth user to satisfy the QoS target Q_k. Thus, the problem of achieving certain QoS requirements, carries over to the problem of achieving SIR targets $\gamma > 0, \forall k$. In the following we will also summarize the targets in a diagonal matrix

$$\mathbf{\Gamma} = \mathbf{diag}\{\gamma_\mathbf{1}, \ldots, \gamma_\mathbf{K}\}$$

It is desirable to find a power allocation $p > 0$ such that $SIR_k(p) \leq \gamma_k$, for all $k = 1, \ldots, K$. This can be rewritten as $\min_k SIR_k(p)/\gamma_k \geq 1$ or equivalently as $\max_k \gamma_k I_k(p)/p_k \leq 1$. We say that the target Γ is feasible if and only if $C(\mathbf{\Gamma}) \leq \mathbf{1}$, where

$$C(\mathbf{\Gamma}) = \inf_{\mathbf{p}>\mathbf{0}} \left(\max_{\mathbf{1} \leq \mathbf{k} \leq \mathbf{K}} \frac{\gamma_\mathbf{k} \mathbf{I_k}(\mathbf{p})}{\mathbf{p_k}} \right) \tag{3.3}$$

In the following we will refer to (3.3) as the "min-max balancing problem".

The optimum $C(\mathbf{\Gamma})$ provides a single measure for the joint feasibility of the targets $\mathbf{\Gamma}$. Note that the optimization is over $p > 0$ to ensure that $I_k(p)/p_k$ is always defined. However, this does not restrict the generality of the results since p can be made arbitrarily small.

The min-max optimum $C(\mathbf{\Gamma})$ can be used to characterize the QoS feasible region:

$$Q = \{[\phi(\gamma_1), \ldots, \phi(\gamma_K)] : C(\mathbf{\Gamma}) \leq \mathbf{1}\}$$

A fundamental problem in resource allocation theory is to find a feasible point $[Q_1, \ldots, Q_K] \in \mathcal{Q}$ according to certain design criteria, like network efficiency, stability, or fairness. The optimization strategy can depend on many parameters, like operator revenue, user requests, queuing lengths, individual link priorities. But there exists no joint optimization framework. They actual problem structure strongly depends on the geometry of Q and on the definition of the underlying interference function.

So the purpose of this text is not to give a comprehensive overview on allocation strategies, but rather to provide a theoretical framework which helps to understand underlying principles. Most of the optimization problems illustrated are directly connected with the min-max balancing problem and the associated QoS feasible region $\mathcal{Q}$.

In the following we will study the QoS (resp. SIR) feasible region for different interference functions $I_k(p)$, including adaptive receive strategies and worst-case designs. But before we start with the most basic (axiomatic) interference model we provide some additional motivation by discussing the relationship of the generic interference model with problems in wireless communications.

Chapter 4

Advanced Interference Calculus: A General Framework for Interference Coordination

4.1 Introduction

Adaptivity and robustness are important design principles for interference-coupled communication systems. But adaptivity also complicates the development of interference management strategies. In order to fully exploit the available capacity of the system, a joint optimization of transmission powers and adaptive strategies is typically required. Such optimization problems are often difficult to solve because of the many degrees of freedom involved.

Convexity is commonly considered as the dividing line between "easy" and "difficult" problems [29]. In the context of signal processing for communications, there are many examples for a successful application of convex optimization theory. Single-user transceiver optimization for MIMO [32], multiuser beamforming [5, 53, 39], and robust designs [52, 28, 39]. So when searching for efficient algorithmic solutions, a typical approach is to search for some hidden convexity.

For the beamforming problem, there is an alternative approach [36, 51], which is not based on convexity. This can be regarded as a special case of Yates' framework of *interference functions* [56]. Instead of relying on convexity, this approach is based on certain monotonicity and scaling properties, which lead to contractive fixed-point iterations.

This observation points to another fundamental principle which allows a characterization of "easy" problems in the context of coupled multiuser systems. Even when convexity cannot be exploited, sometimes solutions can be derived within the framework of interference functions [56, 38]. This *interference calculus* is based on a few elementary properties, which will be outlined in the next section.

4.2 Interference Functions

Consider an interference-coupled system, where K users transmit with powers $\boldsymbol{p} = [p_1, \ldots, p_K]^T$. We say that $\mathcal{I} : \mathbb{R}_+^K \mapsto \mathbb{R}_+$ is an *interference function* if it fulfills the axioms

A1 (non-negativity) $\mathcal{I}(\boldsymbol{p}) \geq 0$,

A2 (scale-invariance) $\mathcal{I}(\alpha\boldsymbol{p}) = \alpha\mathcal{I}(\boldsymbol{p})$ for $\alpha \geq 0$,

A3 (monotonicity) $\mathcal{I}(\boldsymbol{p}') \geq \mathcal{I}(\boldsymbol{p})$ if $\boldsymbol{p}' \geq \boldsymbol{p}$.

Property A1 is clear since $\mathcal{I}$ stands for a power level. Property A3 means that if one component of $\boldsymbol{p}$ is increased then the resulting interference cannot decrease. The same properties were used in Yates' framework [56].

But A2 differs slightly from the definition in [56], where $\mathcal{I}(\alpha\boldsymbol{p}) < \alpha\mathcal{I}(\boldsymbol{p})$ for all $\alpha > 1$ was assumed. The difference lies in how receiver noise is treated. In [56], the assumption of noise is part of the model. In order to include noise in our model A1-A3, we first need to introduce an extended power allocation $\overline{\boldsymbol{p}} = [p_1, \ldots, p_K, \sigma_n^2]^T$, where σ_n^2 stands for noise power. If the following condition (4.1) is fulfilled in addition to A1–A3, then we obtain an interference model which is equivalent to Yates' framework [56].

$$\mathcal{I}(\overline{\boldsymbol{p}}') > \mathcal{I}(\overline{\boldsymbol{p}}) \quad \text{if } \overline{p}'_{K+1} > \overline{p}_{K+1} \text{ and } \overline{\boldsymbol{p}}' \geq \overline{\boldsymbol{p}} \,. \tag{4.1}$$

The reason for choosing the model A1-A3 is its generality. The axioms A1–A3 can be used as a basis for different interference functions, of which [56] is a special case. In the following we will combine A1–A3 with various additional properties, each leading to efficient and globally optimal algorithmic solutions.

As an example, consider K interference functions $\mathcal{I}_1(\overline{\boldsymbol{p}}), \ldots, \mathcal{I}_K(\overline{\boldsymbol{p}})$ fulfilling A1–A3 plus (4.1). The last component $\overline{p}_{K+1} = \sigma_n^2 > 0$ is a constant noise power. Then, the sum-power minimization problem

$$\min_{\boldsymbol{p}>0} \sum_{l=1}^{K} p_l \quad \text{s.t.} \quad \mathrm{SINR}_k(\overline{\boldsymbol{p}}) = \frac{p_k}{\mathcal{I}_k(\overline{\boldsymbol{p}})} \geq \gamma_k, \;\; \forall k \,, \tag{4.2}$$

with feasible SINR targets $\gamma_1, \dots, \gamma_K$, can be solved by the fixed-point iteration [56]

$$p_k^{(n+1)} = \gamma_k \mathcal{I}_k(\overline{\boldsymbol{p}}^{(n)}), \quad k = 1, 2, \dots, K \,. \tag{4.3}$$

The property (4.1) plays an important role for proving global convergence of this iteration (see e.g. [38]). A1–A3 alone are insufficient for this purpose.

The goal of this chapter is a general discussion of properties of interference function that allow for optimal algorithm designs. Property (4.1) is not the only such property. Even though (4.1) is typically fulfilled in practice, it is not always essential (some examples will be discussed in the following). By adding further properties to A1–A3, more specific results will be shown. This approach brings our clearly the underlying properties being responsible for certain algorithmic behaviors observed in the literature.

4.2.1 Adaptive Receive Strategies

The aforementioned multiuser beamforming problem can be solved by using the framework of interference functions. Consider an uplink scenario with optimum receive beamforming, then the interference of the k-th user is

$$\mathcal{I}_k(\overline{\boldsymbol{p}}) = \frac{1}{\boldsymbol{h}_k^H \left(\sigma_n^2 \mathbf{I} + \sum_{l \neq k} p_l \boldsymbol{h}_l \boldsymbol{h}_l^H\right)^{-1} \boldsymbol{h}_k} \,. \tag{4.4}$$

The interference function (4.4) fulfills A1–A3 plus (4.1), so we know from [56] that problem (4.2) can be solved by the fixed-point iteration (4.3). This approach was used in [36, 51].

However, the fixed-point iteration only achieves linear convergence [25, 23, 6]. A better convergence behavior is achieved by exploiting additional properties of $\mathcal{I}_k$. Namely, (4.4) is locally Lipschitz continuous and concave. For such 'semi-smooth' functions, it was shown in [6, 35] that problem (4.2) can be solved by a Newton-type iteration with super-linear convergence. If $\mathcal{I}_k$ are semi-continuous of degree 2, then the convergence is even quadratic [35]. This explains the rapid convergence observed in [40].

The beamforming problem can be understood in a more general context. As shown in [6, 37], the problem (4.2) can be solved with super-linear convergence for all interference functions of the form

$$\mathcal{I}_k(\overline{\boldsymbol{p}}) = \min_{z_k \in \mathcal{Z}_k} \left[\boldsymbol{p}^T \boldsymbol{v}_k(z_k) + \sigma_n^2 n_k(z_k)\right] \,. \tag{4.5}$$

Here, $\boldsymbol{v}_k(z_k) \geq 0$ is a coefficient vector, which models how the users are coupled by interference. The coefficient $n_k(z_k) > 0$ stands for a possible noise amplification, e.g., caused by the noise enhancement effect. These coefficients depend on a 'receive strategy' z_k, which is chosen from a compact set $\mathcal{Z}_k$. For the above beamforming example, z_k would be a vector of filter coefficients (the "beamformer"), chosen from a set $\{\boldsymbol{w} \in \mathbb{C}^K : \|\boldsymbol{w}\| = 1\}$. This set can be further restricted, e.g. by adding 'shaping constraints' as in [22]. Also, $\mathcal{Z}_k$ can be a discrete set, as in [55, 24, 36], where the interference is minimized over a set of possible base station assignments. These examples can all be interpreted as special cases of the interference model (4.5), for which the power minimum (4.2) can be found by the iteration

$$\boldsymbol{p}^{(n+1)} = \sigma_n^2 \big(\mathbf{I} - \mathbf{\Gamma V}(z^{(n)})\big)^{-1} \mathbf{\Gamma} \boldsymbol{n}(z^{(n)}) \tag{4.6}$$

$$\text{where } z_k^{(n)} = \arg\min_{z_k \in \mathcal{Z}_k} \big[(\boldsymbol{p}^{(n)})^T \boldsymbol{v}_k(z_k) + \sigma_n^2 n_k(z_k)\big] \tag{4.7}$$

$$\mathbf{\Gamma} = \mathrm{diag}\{\{\}\, \gamma_1, \dots, \gamma_K\}$$

$$\mathbf{V}(z) = [\boldsymbol{v}_1(z_1), \dots, \boldsymbol{v}_K(z_K)]^T$$

$$\boldsymbol{n}(z) = [n_1(z_1), \dots, n_K(z_K)]^T \, .$$

With a feasible initialization $\boldsymbol{p}^{(0)}$, the iteration (4.6) achieves the power minimum (4.2) with super-linear convergence. This shows the benefit from exploiting the special concave structure of the interference function (4.5).

The iteration (4.6) can also be used for the optimization of transmit strategies. Under certain conditions (see e.g. [43, 44]), optimal transmit strategies can be found by solving an equivalent problem involving 'virtual receive strategies'. Note, that the *joint* optimization of receive and transmit strategies—as it occurs in the context of multiuser MIMO optimization—cannot be solved within this framework. But the results can be used in order to derive near-optimal algorithms, as in [43, 44].

4.2.2 Robust Designs

In analogy to (4.5), we can define a *worst-case interference function*

$$\mathcal{I}_k(\overline{\boldsymbol{p}}) = \max_{x_k \in \mathcal{X}_k} \big[\boldsymbol{p}^T \boldsymbol{v}_k(x_k) + \sigma_n^2 n_k(x_k)\big] \, . \tag{4.8}$$

Here, the maximum is taken instead the minimum. Such an approach is typically used in the context of worst case designs. For example, x_k can stand for some channel uncertainty from a compact set $\mathcal{X}_k$ (the uncertainty region), as in [52, 28]. Another example is the multicasting scenario in [20], where a signal is received by a group of receivers, which should all fulfill a certain SINR constraint. Most of the properties discussed in the previous section can be transferred to these scenarios. In particular, the power minimization problem (4.2) can be solved with super-linear convergence [39].

Now, consider coupling coefficients $\boldsymbol{v}_k(z_k, x_k)$. That is, interference depends on a receive strategy z_k (see Section 4.2.1) *and* an uncertainty x_k. Likewise, the noise coefficient is $n_k(z_k, x_k)$. The strategy x_k is chosen so as to minimize the worst-case interference, i.e.,

$$\mathcal{I}_k(\overline{\boldsymbol{p}}) = \min_{z_k \in \mathcal{Z}_k} \max_{x_k \in \mathcal{X}_k} \left[\boldsymbol{p}^T \boldsymbol{v}_k(z_k, x_k) + \sigma_n^2 n_k(z_k, x_k) \right] . \tag{4.9}$$

The function (4.9) fulfills the properties A1–A3 so it is an interference function. However, it is neither convex nor concave, so we cannot expect to solve the power minimization problem (4.2) with super-linear convergence. But since (4.1) is fulfilled, the fixed-point iteration [56] can be applied, so linear convergence can be achieved.

4.2.3 SIR Balancing and Feasible Sets

The theory of interference functions was originally motivated by power control problems of the form (4.2). But its scope is much more general. For example, consider the SIR balancing problem

$$C(\boldsymbol{\gamma}) = \inf_{\boldsymbol{p}>0} \left(\max_{1 \le k \le K} \frac{\gamma_k \mathcal{I}_k(\boldsymbol{p})}{p_k} \right) . \tag{4.10}$$

If the interference functions $\mathcal{I}_1, \ldots \mathcal{I}_K$ are concave, like (4.4) and (4.5), and if all users are coupled by interference, then the optimum $C(\boldsymbol{\gamma})$ can be computed by the globally convergent iteration proposed in [7].

Note, that this result does not rely on the scaling property (4.1), which played an important role in the context of the power minimization problem (4.2). This justifies our approach A1–A3, which provides a common basis for different types of interference functions.

The function $C(\boldsymbol{\gamma})$ can be regarded as an indicator function for feasibility. Targets $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_K]$ are feasible if $C(\boldsymbol{\gamma}) \le 1$. So the SIR feasible region is the sub-level set

$$\mathcal{S} = \{\boldsymbol{\gamma} : C(\boldsymbol{\gamma}) \le 1\} . \tag{4.11}$$

It can be observed that $C(\boldsymbol{\gamma})$ itself fulfills the axioms A1-A3, so it can be regarded as an interference function.

Further properties of $C(\boldsymbol{\gamma})$ and the associated region $\mathcal{S}$ have been investigated in [9] under the assumption that the underlying interference functions $\mathcal{I}_1, \ldots, \mathcal{I}_K$ are *log-convex*. We say that an interference function $\mathcal{I}(\boldsymbol{p})$ is log-convex if $\log \mathcal{I}(e^{\boldsymbol{s}})$ is convex on $\mathbb{R}^K$. Note, that *every* convex interference function is log-convex when using the substitution $\boldsymbol{p} = e^{\boldsymbol{s}}$. Thus, log-convexity holds for all interference functions of the type (4.8). We will see later how this can be generalized.

It was shown in [9] that $C(\mathrm{e}^{\boldsymbol{q}})$ is log-convex with respect to the QoS vector $\boldsymbol{q} \in \mathbb{R}^K$. This means that our interference function $C(\boldsymbol{\gamma})$ has a special property: The log-convexity of the underlying interference functions $\mathcal{I}_k$ is preserved. This behavior seems to be a particularity of the min-max approach. When taking the infimum over the maximum value, i.e.,

$$c(\boldsymbol{\gamma}) = \sup_{\boldsymbol{p}>0} \Big(\min_{1\leq k\leq K} \frac{\gamma_k \mathcal{I}_k(\boldsymbol{p})}{p_k} \Big) \,, \tag{4.12}$$

it is not clear whether log-convexity is fulfilled or not. The techniques which can be used to show this property for the min-max case cannot be applied here.

Now, every log-convex function is convex. So $C(\mathrm{e}^{\boldsymbol{q}})$ is convex with respect to $\boldsymbol{q}$. Since a sub-level set of a convex function is convex, it can be concluded that the region

$$\mathcal{Q} = \{\boldsymbol{q} : C(\mathrm{e}^{\boldsymbol{q}}) \leq 1\} \tag{4.13}$$

is a convex set if the underlying interference functions $\mathcal{I}_1, \ldots, \mathcal{I}_K$ are log-convex. This result can even be extended to arbitrary bijective quality-of-service mappings QoS $= \phi(\text{SIR})$ for which the inverse mapping is log-convex. Similar observations were made in the context of linear interference functions based on a fixed coupling matrix $\mathbf{V}$ [46, 45].

This discussion shows that log-convexity is another important property which leads to efficient algorithmic designs. If $\mathcal{I}_1, \ldots, \mathcal{I}_K$ fulfill A1–A3 and if $\mathcal{I}_k(\mathrm{e}^{\boldsymbol{s}})$ is log-convex, then the resulting QoS region is convex. If, in addition, the interference functions are completely coupled (see [8]), then all points on the boundary of $\mathcal{Q}$ can be attained and a proportionally fair operating point can be computed by solving a convex optimization problem [11].

4.3 Elementary Representations and Bounds

In the previous section we have discussed the interference function $C(\boldsymbol{\gamma})$, which is composed by underlying interference functions $\mathcal{I}_1, \ldots, \mathcal{I}_K$. This has proved useful for the characterization of the region $\mathcal{Q}$. This example shows that the analysis of interference functions is closely connected with the analysis of QoS regions. This will now be specified for different types of interference functions.

4.3.1 General Interference Functions

Consider level sets

$$\underline{L}(\mathcal{I}) = \{\hat{\boldsymbol{p}} > 0 : \mathcal{I}(\hat{\boldsymbol{p}}) \leq 1\} \tag{4.14}$$
$$\overline{L}(\mathcal{I}) = \{\hat{\boldsymbol{p}} > 0 : \mathcal{I}(\hat{\boldsymbol{p}}) \geq 1\} . \tag{4.15}$$

It can be shown that an arbitrary interference function $\mathcal{I}(\boldsymbol{p})$ (only characterized by A1–A3) can be represented as an optimum over elementary building blocks. For arbitrary $\boldsymbol{p} > 0$, the following representation holds

$$\mathcal{I}(\boldsymbol{p}) = \min_{\hat{\boldsymbol{p}} \in \underline{L}(\mathcal{I})} \max_k \frac{p_k}{\hat{p}_k} = \max_{\hat{\boldsymbol{p}} \in \overline{L}(\mathcal{I})} \min_k \frac{p_k}{\hat{p}_k} . \tag{4.16}$$

So every interference function can be expressed as a point-wise minimum over elementary interference functions $\max_k p_k/\hat{p}_k$. This reveals some basic properties of the associated level sets. For example, consider the interference function $C(\boldsymbol{\gamma})$, for which the sub-level set is the SIR feasible region $\mathcal{S}$, as defined by (4.11). Although $C(\boldsymbol{\gamma})$ is generally not convex, the result nevertheless helps to characterize the SIR feasible region. It can be observed from (4.16) that $\mathcal{S}$ has a certain monotonicity property: If $\hat{\boldsymbol{\gamma}} \in \mathcal{S}$, then all $\boldsymbol{\gamma}$ fulfilling $\boldsymbol{\gamma} \leq \hat{\boldsymbol{\gamma}}$ are also contained in $\mathcal{S}$. This holds for *all* SIR regions based on interference functions fulfilling A1–A3.

4.3.2 Convex Interference Functions

If the interference function $\mathcal{I}(\boldsymbol{p})$ is convex, then it can additionally be represented as a maximum over linear elementary functions:

$$\mathcal{I}(\boldsymbol{p}) = \max_{\boldsymbol{v} \in \mathcal{W}_0(\mathcal{I})} \sum_{k=1}^{K} v_k \cdot p_k , \tag{4.17}$$

The maximization is over the set

$$\mathcal{W}_0(\mathcal{I}) = \{\boldsymbol{w} \in \mathbb{R}_+^K : \bar{\mathcal{I}}^*(\boldsymbol{w}) = 0\} , \tag{4.18}$$

where $\bar{\mathcal{I}}^*(\boldsymbol{w})$ is the conjugate function of $\mathcal{I}(\boldsymbol{p})$.

The set $\mathcal{W}_0(\mathcal{I})$ can be shown to be closed bounded convex. So every convex interference function $\mathcal{I}$ is associated with a convex set $\mathcal{W}_0(\mathcal{I})$. The converse holds as well, i.e., every closed bounded convex set $\mathcal{V}$ leads to a convex interference function of the form (4.17). This can be exploited in order to derive a best-possible convex minorant for an arbitrary interference function. As an example, consider again the SIR feasible region $\mathcal{S}$, which is defined as the sub-level set of the interference function $C(\boldsymbol{\gamma})$. While $C(\boldsymbol{\gamma})$ is generally non-convex, the set $\mathcal{W}_0(C)$ is convex. This leads to a convex minorant $\underline{C}_x(\boldsymbol{\gamma})$, such that

$$\underline{C}_x(\boldsymbol{\gamma}) \leq C(\boldsymbol{\gamma}), \quad \forall \boldsymbol{\gamma} \,. \tag{4.19}$$

The sub-level set associated with $\underline{C}_x(\boldsymbol{\gamma})$ is convex and it contains the SIR region $\mathcal{S}$. Moreover, it can be shown that this is the smallest convex set containing $\mathcal{S}$. Such a best-possible convex approximation of an arbitrary SIR region $\mathcal{S}$ can be useful for the development of resource allocation techniques operating on the boundary of $\mathcal{S}$.

Note that $\underline{C}_x(\boldsymbol{\gamma}) = 0$ can occur. In this case, $\mathbb{R}^K_+$ is the only convex set containing $\mathcal{S}$.

In analogy, there exists a representation for concave interference functions. Every concave $\mathcal{I}(\boldsymbol{p})$ can be represented as a minimum of linear elementary functions, optimized over a closed convex set. This representation, which is not discussed here, leads to a best-possible concave majorant.

Finally, it is interesting that the coefficients v_k in the maximization problem (4.17) can be interpreted as utilities, which are optimized over an "utility" region $\mathcal{W}_0(\mathcal{I})$, with weighting factors p_k. This shows a direct link to problems in resource allocation, where such an optimization strategy is commonly used. This aspect will be further discussed in Section 4.4 (see also [42, 10]).

4.3.3 Log-Convex Interference Functions

It was shown in the previous section that even if the SIR region $\mathcal{S}$ is not convex, efficient optimization is possible if the log-SIR region is convex. This motivates the construction of log-convex approximations.

To this end, we use another representation result: Let $\mathcal{I}(\boldsymbol{p})$ be an arbitrary log-convex interference function, then

$$\mathcal{I}(\boldsymbol{p}) = \max_{\boldsymbol{w} \in \mathcal{L}(\mathcal{I})} \Big(f_{\mathcal{I}}(\boldsymbol{w}) \cdot \prod_{l=1}^{K} (p_l)^{w_l} \Big) \tag{4.20}$$

$$\text{where} \quad f_{\mathcal{I}}(\boldsymbol{w}) = \inf_{\boldsymbol{p}>0} \frac{\mathcal{I}(\boldsymbol{p})}{\prod_{l=1}^{K} (p_l)^{w_l}} \,, \quad \boldsymbol{w} \in \mathbb{R}^K_+ \,. \tag{4.21}$$

$$\mathcal{L}(\mathcal{I}) = \big\{ \boldsymbol{w} \in \mathbb{R}^K_+ : f_{\mathcal{I}}(\boldsymbol{w}) > 0 \big\} \,.$$

The set $\mathcal{L}(\mathcal{I})$ is convex. Conversely, starting with an arbitrary closed convex set, we can use (4.20) in order to synthesize a log-convex interference function.

As in the previous section, we consider again the example of the feasible region $\mathcal{S}$, with $C(\boldsymbol{\gamma})$ as defined by (4.11). With $\mathcal{L}(C)$ and (4.20), we obtain a log-convex minorant $\underline{C}_l(\boldsymbol{\gamma})$ of the original function $C(\boldsymbol{\gamma})$. This minorant is the largest log-convex minorant, and it is tighter than the convex minorant $\underline{C}_x(\boldsymbol{\gamma})$ studied in the last section, i.e.,

$$\underline{C}_x(\boldsymbol{\gamma}) \leq \underline{C}_l(\boldsymbol{\gamma}) \leq C(\boldsymbol{\gamma}), \quad \forall \boldsymbol{\gamma} \; . \tag{4.22}$$

The first inequality follows from the fact that the convex interference function $\underline{C}_x(\boldsymbol{\gamma})$ is necessarily log-convex (with the substitution $\mathrm{e}^{\boldsymbol{q}}$), but the converse is not true. That is, $\underline{C}_x(\boldsymbol{\gamma})$ also provides a log-convex minorant, but $\underline{C}_l(\boldsymbol{\gamma})$ is the *best-possible* log-convex minorant. So the associated sub-level set is the smallest log-convex region containing the SIR region $\mathcal{S}$. Again, this region can be used as an approximation of the original non-convex region $\mathcal{S}$, which is more difficult to handle.

4.4 Proportional Fairness and Bargaining

Proportional fairness was introduced in [27] in the context of stability and fairness of rate control algorithms. The proportional fair equilibrium was defined as the point at which the relative difference to any other QoS vector is non-positive. This corresponds to a resource allocation strategy which favors users with good channel conditions over those with bad channels. This is desirable in the presence of elastic traffic, which is able to adapt to channel fluctuations, as in a wireless communication system.

Another approach to resource allocation is the *Nash bargaining theory* [26, 34]. A bargaining solution can be seen as the unanimous agreement on some QoS point $\boldsymbol{q}$ from a convex feasible set $\mathcal{Q}$ (e.g. rates from a capacity region). The K users (denoted as players) cooperate in order to achieve a solution outcome which is better than their minimum requirements (*disagreement point*).

It was already observed in [27] (see also [41]) that the symmetric Nash bargaining solution is achieved by proportional fairness, provided that the underlying QoS region is convex. With the results of the previous sections, we are now able to characterize systems for which this is fulfilled. If $\mathrm{QoS}_k = \log \mathrm{SIR}_k$, and the QoS region depends on log-convex interference functions, then the region is convex and equivalence holds.

Next, we can exploit the structure of log-convex interference functions in order to characterize the *asymmetric weighted Nash bargaining game*

$$\mathcal{N}(\boldsymbol{w}, \boldsymbol{d}) = \max_{\boldsymbol{q} \in \mathcal{Q} : \boldsymbol{q} > \boldsymbol{d}} \prod_{k=1}^{K} (q_k - d_k)^{w_k}, \quad \boldsymbol{d} > 0 \; , \tag{4.23}$$

where $\boldsymbol{d} = [d_1, \ldots, d_K]$ denote minimum QoS requirements. The weights $\boldsymbol{w} = [w_1, \ldots, w_K]^T$, with $\sum_k w_k = 1$, are chosen according to individual user priorities.

This problem is closely connected with the elementary building blocks of log-convex interference functions. In particular, we can use the function $f_{\mathcal{I}_1}$, as defined by (4.21), in order to show that the weighted Nash bargaining (4.23) is equivalent to a resource allocation strategy, known as *weighted proportional fairness*

$$\max_{\boldsymbol{q} \in \mathcal{Q}: \boldsymbol{q} > \boldsymbol{d}} \sum_{k=1}^{K} w_k \log(q_k - d_k) \ . \tag{4.24}$$

This shows an interesting connection between two areas of research with independent backgrounds and motivations. Bargaining theory was derived in the context of economics, while proportional fair resource allocation has a background in communications and networking. This result demonstrates that the theory of interference functions has a wide range of applications and is not restricted to power control.

Bibliography

[1] S. Alamouti. A simple transmitter diversity scheme for wireless communications. *IEEE Transactions on Information Theory.*

[2] J. C. Belfiore and G. Rekaya. Quaternionic lattices for space-time coding. In *Proc. IEEE Int. Symp. on Information Theory Workshop ITW.*

[3] J. C. Belfiore, G. Rekaya, and E. Viterbo. The golden code: A 2×2 full rate space-time code with nonvanishing determinants. *IEEE Transactions on Information Theory*, 51(4):1432–1436, April 2005.

[4] P. A. Bello. Characterization of randomly time varying linear channels. *IEEE Transactions commun System.*

[5] M. Bengtsson and B. Ottersten. *Handbook of Antennas in Wireless Communications*, chapter 18: Optimal and Suboptimal Transmit Beamforming.

[6] H. Boche and M. Schubert. Convergence behavior of matrix-based iterative transceiver optimization. In *Proc. IEEE Int. Workshop on Signal Processing Advances for Wireless Comm. SPAWC.*

[7] H. Boche and M. Schubert. Multiuser interference balancing for general interference functions – a convergence analysis. In *Proc. IEEE Int. Conf. on Comm. (ICC)*, Glasgow, Scotland, June 2007. to appear.

[8] H. Boche and M. Schubert. On the existence of a proportionally fair operating point for wireless communication systems. In *Signal Processing Advances in Wireless Communications (SPAWC)*, Helsinki, Finland, June 2007.

[9] H. Boche and M. Schubert. The supportable QoS region of a multiuser system with log-convex interference functions. In *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, Honolulu, Hawaii, USA, April 2007. to appear.

[10] H. Boche, M. Schubert, E. Jorswieck, and A. Sezgin. A general framework for concave/convex interference coordination problems and network utility optimization. In *ITG/IEEE Workshop on Smart Antennas (WSA), Vienna, Austria*, 2007. submitted.

[11] H. Boche, M. Schubert, S. Stanczak, and M. Wiczanowski. An axiomatic approach to resource allocation and interference balancing. In *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, Philadelphia, USA, 2005.

[12] H. Bölcskei, M. Borgmann, and A. J. Paulraj. Impact of the propagation environment on the performance of space-frequency coded MIMO-ofdm. *IEEE J. Sel. Ar. Comm.*, 21(3):427–439, April 2003.

[13] H. Bölcskei, R. Koetter, and S. Mallik. Coding and modulation for underspread fading channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 358, Lausanne, Switzerland, 2002.

[14] H. Bölcskei and A. J. Paulraj. Space frequency coded broadband ofdm systems. In *Proc. IEEE Wireless Commun. Net. Conf. (WCNC)*, pages 1–6, Chicago, USA, 2000.

[15] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK.

[16] P. Coronel and H. Bölcskei. Diversity multiplexing tradeoff in selective fading MIMO channels.

[17] P. Dayal and M. K. Varanasi. An optimal two transmit antenna space-time code and its stacked extensions. *IEEE Transactions on Information Theory*, 51(12):4348–4355, December 2005.

[18] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, 1(2):41–59.

[19] H. El Gamal, G. Caire, and M. O. Damen. Lattice coding and decoding achieves the optimal diversity multiplexing tradeoff of MIMO channels. *IEEE Transactions on Information Theory*, 50(9):968–985, 2004.

[20] Y. Gao and M. Schubert. Power allocation for multi-group multicasting with beamforming. In *ITG/IEEE Workshop on Smart Antennas (WSA)*.

[21] L. Grokop and D. N. C. Tse. Diversity/multiplexing tradeoff in ISI channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Chicago, USA, June/July 2004.

[22] D. Hammarwall, M. Bengtsson, and B. Ottersten. On downlink beamforming with indefinite shaping constraints. *IEEE Trans. Sign. Proc.*, 54.

[23] S. Hanly. Congestion measures in DS-CDMA networks. *IEEE Trans. Comm.*, 47(3).

[24] S. Hanly. An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity. *IEEE J. Sel. Ar. Comm.*, 13(7):1332–1340, 1995.

[25] C. Huang and R. Yates. Rate of convergence for minimum power assignment algorithms in cellular radio systems. *Baltzer/ACM Wireless Networks*, 4.

[26] J. F. Nash Jr. The bargaining problem. *Econometrica, Econometric Soc.*, 18.

[27] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *Journal of Operations Research Society*, 49(3):237–252, March 1998.

[28] R. G. Lorenz and S. Boyd. Robust minimum variance beamforming. *IEEE Trans. Sign. Proc.*, 53(5):1684–1696.

[29] Z-Q. Luo and W. Yu. An introduction to convex optimization for communications and signal processing. *IEEE J. Sel. Ar. Comm.*, 24(8):1426–1438, August 2006.

[30] X. Ma, G. Leus, and G. B. Giannakis. Space-time doppler block coding for correlated time selective fading channels. *IEEE Trans. Sign. Proc.*, 53(6):2167–2181, 2005.

[31] A. Medles and D. T. M. Slock. Optimal diversity vs multiplexing tradeoff for frequency selective MIMO channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, 2005.

[32] 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. Sign. Proc.*, 51(9):2381–2401, 2003.

[33] A. Peled and A. Ruiz. Frequency domain data transmission using reduced computational complexity algorithms. In *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, pages 964–967, Denver,USA, April 1980.

[34] H. J. M. Peters. *Axiomatic Bargaining Game Theory*. Kluwer Academic Publishers, Dordrecht.

[35] L. Qi. Convergence analysis of some algorithm for solving non-smooth equations. *Mathematics of Operations Research*, 18(1):227–244, February 1993.

[36] F. Rashid-Farrokhi, L. Tassiulas, and K.J. Liu. Joint optimal power control and beamforming in wireless networks using antenna arrays. *IEEE Trans. Comm.*, 46(10):1313–1323, October 1998.

[37] M. Schubert and H. Boche. A generic approach to qos-based transceiver optimization. *IEEE Trans. Comm. to appear*.

[38] M. Schubert and H. Boche. Qos-based resource allocation and transceiver optimization. *Foundation and Trends in Communications and Information Theory*, 2(6).

[39] M. Schubert and H. Boche. Robust resource allocation. In *Proc. IEEE Int. Symp. on Information Theory Workshop ITW*.

[40] M. Schubert and H. Boche. Solution of the multi-user downlink beamforming problem with individual SINR constraints. *IEEE Trans. Veh. Tech.*, 53(1):18–28, January 2004.

[41] M. Schubert and H. Boche. Properties and operational characterization of proportionally fair resource allocation. In *Signal Processing Advances in Wireless Communications (SPAWC)*, Helsinki, Finland, June 2007.

[42] M. Schubert, H. Boche, and S. Stanczak. Strict convexity of the qos feasible region for log-convex interference functions. In *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove (CA), USA, November 2006.

[43] M. Schubert, S. Shi, and H. Boche. Iterative transceiver optimization for linear multiuser MIMO channels with per-user MMSE requirements. In *Proc. European Signal Processing Conference (EUSIPCO)*.

[44] S. Shi, M. Schubert, and H. Boche. Downlink MMSE transceiver optimization for multi-user MIMO systems: Duality and sum-MMSE minimization. *IEEE Trans. Sign. Proc. to appear.*

[45] S. Stancszak and H. Boche. On the convexity of feasible qos regions. *IEEE Trans. Inf. Th.*, 53(2):779–783, February 2007.

[46] C. W. Sung. Log-convexity property of the feasible sir region in power controlled cellular systems. *IEEE Communications Letters*, 6(6):248–249, June 2002.

[47] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communications: Performance criteria in the presence of channel estimation errors, mobility and multiple paths. *IEEE Transactions on Communications*, 44.

[48] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communications: Performance criteria and code construction. *IEEE Transactions on Information Theory*, 44(2):744–765, March 1998.

[49] S. Tavildar and P. Viswanath. Approximately universal codes over slow fading channels. *IEEE Transactions on Information Theory*, 52(7):3233–3258, July 2006.

[50] E. Telatar. Capacity of multi-antenna gaussian channels. *Europ. Trans. on Telecomm*, 10(6):585–596, 1999.

[51] E. Visotsky and U. Madhow. Optimum beamforming using transmit antenna arrays. In *IEEE Veh. Tech. Conf.,* Spring.

[52] S. Vorobyov, A. Gershman, and Z-Q. Luo. Robust adaptive beamforming using worst-cast performance optimization: a solution to the signal mismatch problem. *IEEE Trans. Sign. Proc.*, 51(2):313–324, February 2003.

[53] A. Wiesel, Y. C. Eldar, and S. Shamai(Shitz). Linear precoding via conic optimization for fixed MIMO receivers. *IEEE Trans. Sign. Proc.*, 54(1):161–176.

[54] H. Yao and G. W. Wornell. Achieving the full MIMO diversity multiplexing frontier with rotation based space-time codes. In *Proc. Allerton Conference on Communication, Control, and Computing*, October 2003.

[55] R. Yates and H. Ching-Yao. Integrated power control and base station assignment. *IEEE Trans. Veh. Tech.*, 44(3):638–644, August 1995.

[56] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE J. Sel. Areas Comm.*, 53(1):18–28, January 2004.

[57] L. Zheng and D. N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, May 2003.