DIVERSITY-MULTIPLEXING TRADEOFF IN SELECTIVE-FADING CHANNELS

Pedro E. Coronel

DIVERSITY-MULTIPLEXING TRADEOFF IN SELECTIVE-FADING CHANNELS

Hartung-Gorre Verlag Konstanz
2009

Reprint of Diss. ETH No. 17896

SERIES IN COMMUNICATION THEORY VOLUME 5

edited by Helmut Bölcskei

Bibliographic Information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available at http://dnb.ddb.de.

First edition 2009

HARTUNG-GORRE VERLAG KONSTANZ

ISSN 1865-6765
ISBN-10 3-86628-251-6
ISBN-13 978-3-86628-251-3

A mi querida familia

Abstract

Recent years have seen the proliferation of new wireless data applications that have fueled the demand for communication systems capable of delivering very high quality-of-service (QoS) in terms of data rate and reliability. At the same time, due to the surge in the number of wireless devices, the spectrum available for communication has become scarcer, and the level of interference has increased significantly. The design of systems satisfying stringent QoS requirements is therefore a particularly challenging task that requires an accurate understanding of the ultimate performance of communication over wireless channels. In this thesis, we establish information-theoretic limits on rate and reliability in wireless communication, and we provide guidelines to design systems that achieve these limits.

Rate and reliability are two essential properties of communication systems that are subject to tension: increasing data rate typically results in larger error probability and, conversely, a reduction in error probability often comes at the price of sacrificing data rate. The balance that a system is able to strike between rate and reliability provides a comprehensive view of its performance. Unfortunately, a precise characterization of the rate-reliability tradeoff is difficult in general. In fading channels, however, the *diversity-multiplexing (DM) tradeoff framework* proposed by Zheng and Tse (2003) has proved to be very helpful in understanding the interplay between rate and reliability in the high signal-to-noise ratio (SNR) regime. The analysis conducted in

this thesis adopts this framework and uses the notions of diversity and multiplexing to capture, respectively, the reliability and data rate of the system.

In many situations of practical relevance, the wireless channel exhibits time and frequency selectivity arising from temporal variations in the environment and multipath propagation. A direct characterization of the DM tradeoff in this class of channels, referred to as selective-fading channels, is in general difficult because the corresponding mutual information is a sum of correlated random variables. This thesis presents a technique that bypasses this difficulty and establishes the optimal DM tradeoff of selective-fading channels in both the point-to-point case and the multiple-access (MA) case. The essence of our approach is to study the "Jensen channel" that is associated with the original channel and that has the same behavior in the regime of high SNR relevant to the DM tradeoff framework.

In the point-to-point case, we obtain a code design criterion that guarantees optimal performance in the sense of the DM tradeoff and that ties in nicely with several code design criteria reported in prior work. The systematic construction of optimal codes is nontrivial when the transmitter is equipped with multiple antennas. We show, however, that the design problem can be separated into two simpler and independent problems: the design of an inner code, also referred to as precoder, adapted to the selectivity characteristics of the channel, and the design of an outer code which, quite remarkably, is independent of the channel characteristics.

Our investigation of selective-fading MA channels yields a detailed characterization of the limitations that multiuser interference puts on optimal performance. In particular, we find an interesting conceptual relation between the DM tradeoff framework and the notion of dominant error event, which was first introduced in additive white Gaussian noise (AWGN) channels by Gallager (1985). Studying the dominant error event as a function of the users' rates reveals the existence of operational regimes in which multiuser interference has only a negligible impact on error performance. As in the point-to-point case, we obtain a set of code design criteria that guarantees optimal performance to all users. The construction of optimal codes is nontrivial and requires the users to jointly design their codes. We finally examine a code construction which satisfies our criteria for the two-user flat-fading channel.

Résumé

La plupart des nouvelles applications sans fil exigent des systèmes de communication assurant une haute qualité de service (QoS) mesurée en termes de débit et de fiabilité. Cependant, l'augmentation rapide du nombre de ces systèmes se heurte à la difficulté de trouver des fréquences disponibles dans les bandes spectrales réglementaires. Des utilisateurs de plus en plus nombreux doivent se partager des bandes spectrales dont la largeur n'augmente pas et les systèmes de communication sous-jacents sont confrontés à une élévation du niveau d'interférence. Dès lors, la conception de systèmes de communication sans fil offrant une QoS élevée nécessite une compréhension approfondie des limites fondamentales qui régissent leur performance. Cette thèse vise tout d'abord à établir certaines de ces limites en utilisant la théorie de l'information. Elle présente aussi des critères permettant la construction de codes capables d'atteindre une performance optimale.

Le débit de transmission et la fiabilité sont deux paramètres essentiels et antagonistes d'un système de communication: une augmentation du débit entraine généralement plus d'erreurs et, inversement, une amélioration de la fiabilité se traduit souvent par une diminution du débit. La performance d'un système se caractérise intégralement par le compromis qu'il est capable de réaliser entre débit et fiabilité. Mais il est souvent difficile de déterminer ce compromis avec précision. Dans le cas des canaux à évanouissements, le cadre théorique du *compromis diversité-multiplexage* (DM) introduit par Zheng and Tse (2003) permet

de quantifier l'antagonisme entre débit et fiabilité lorsque le rapport signal sur bruit (SNR) est élevé. L'analyse développée dans cette thèse s'inscrit dans ce cadre et utilise en particulier les notions de diversité et de multiplexage pour rendre compte respectivement de la fiabilité et du débit du système.

Dans la pratique, le canal sans fil est souvent sélectif en fréquence et en temps en raison de variations temporelles de l'environnement et de la propagation par trajets multiples. En général, le compromis DM des canaux sélectifs est difficile à déterminer car l'information mutuelle correspondante est une somme de variables aléatoires corrélées. Cette thèse introduit une méthode qui contourne cette difficulté et qui permet d'établir le compromis DM optimal pour les canaux sélectifs point-à-point et à accès multiple (MA). Notre approche consiste à étudier le "canal de Jensen" qui est associé au canal sans fil original et qui se comporte de façon similaire dans le régime de grand SNR, c'est-à-dire dans le régime qui est pertinent du point de vue du compromis DM.

Dans le cas des canaux point-à-point, nous obtenons un critère pour la construction de codes optimaux dans le sense du compromis DM. De plus, nous montrons que notre critère est cohérent avec d'autres critères énoncés dans des travaux antérieurs. La construction systématique de codes optimaux devient un problème complexe lorsque le transmetteur est équipé de multiples antennes. Cependant, nos résultats démontrent qu'une telle tâche peut être décomposée en deux opérations plus simples et indépendantes: d'une part, la construction d'un code interne, ou "précodeur", développé en fonction des statistiques du canal, et d'autre part, la conception d'un code externe qui est indépendant du canal.

Notre étude des canaux sélectifs à MA fournit une caractérisation détaillée de la manière dont l'interférence entre utilisateurs affecte la performance optimale. En particulier, nous trouvons une relation conceptuelle intéressante entre le compromis DM et la notion d'erreur dominante introduite dans le cadre des canaux à bruit additif gaussien par Gallager (1985). Nos résultats révèlent qu'en fonction des débits des utilisateurs, l'interférence peut parfois n'avoir qu'une influence négligeable sur la probabilité d'erreur. Comme dans le cas des liens point-à-point, nous obtenons un ensemble de critères pour construire des codes distribués qui atteignent la performance optimale. Nous examinons finalement un exemple de code distribué qui satisfait nos critères dans le cas d'un canal non-sélectif avec deux utilisateurs.

Acknowledgments

I thank Prof. Helmut Bölcskei for his constant support during the development of this dissertation. I also thank my manager at the IBM Zurich Research Laboratory, Dr. Pierre Chevillat, for his trust and support. Furthermore, I would like to express my gratitude to Dr. Wolfgang Schott for being an exceptional mentor throughout my IBM PhD fellowship.

I am very much indebted to all my friends and colleagues at ETH Zurich and at the IBM Zurich Research Laboratory for making my research truly interesting and enjoyable. For their help and invaluable input to my work, my very special thanks go to Cemal Akçaba, Dr. Simeon Furrer, Dr. Markus Gärtner, Dr. Ateet Kapur, Patrick Kupinger, Dr. Ulrich Schuster, and Beat Weiss.

My final word of gratitude goes to my family and friends for their patience and care. They have been a priceless source of inspiration and help during all these years.

Contents

CHAPTER 1

Introduction

1.1. MOTIVATION

FUTURE WIRELESS data applications call for communication systems capable of delivering very high quality-of-service (QoS) in terms of data rate and reliability. Since bandwidth is every day scarcer and more expensive, system designers are additionally facing the challenge of optimizing spectrum efficiency. Several results established in the field of communication and information theory in recent years suggest promising approaches to develop systems that meet these requirements. In particular, the use of multiple antennas at both ends of the wireless link leads to sizable improvements in terms of system reliability (Tarokh et al., 1998) and data rate (Telatar, 1999) without requiring additional bandwidth.

In comparing and optimizing the performance of systems that exploit these newly available gains, a precise understanding of how they interplay is necessary. Experience shows that lowering the rate of a transmission improves the chances of correct reception. In practical systems, for instance, control information, which is critical for proper system operation, is usually sent at very low rate to prevent its loss. While the existence of a tension between error performance and data rate is intuitively clear, a precise characterization thereof is particularly intricate.

A common approach to study the rate-reliability tradeoff is the theory of error exponents (Gallager, 1968). However, computing the corresponding reliability function requires solving a difficult optimization problem and, in general, closed-form solutions are not available. The diversity-multiplexing (DM) tradeoff framework (Zheng and Tse, 2003), which is conjectured to have an intimate connection with the theory of error exponents, constitutes an elegant approach to study the rate-reliability tradeoff in fading channels. By analyzing the system performance in terms of the outage probability and exploiting the near-zero behavior of the fading distribution, one can obtain a characterization of the tradeoff in the high-SNR regime. The DM tradeoff framework has proved to be very useful not only to obtain theoretical performance limits but also to construct optimal codes (Yao and Wornell, 2003; Belfiore et al., 2005; Dayal and Varanasi, 2005; Tavildar and Viswanath, 2006; Elia et al., 2006).

In most situations of practical relevance, the wireless channel exhibits time and frequency selectivity arising from temporal variations in the environment and multipath propagation, respectively. Interestingly, time and frequency selectivity are capable of improving performance in terms of outage probability (Bölcskei et al., 2002a). However, the DM tradeoff corresponding to this class of channels remains unknown. The main difficulty stems from the fact that the mutual information of selective-fading channels is a sum of correlated random variables, and the corresponding outage probability is therefore difficult to compute.

The present thesis bridges this gap by establishing the optimal DM tradeoff of selective-fading multiple-antenna channels. Our approach consists in analyzing the "Jensen channel" associated with the original channel and showing that both channels have the same asymptotic behavior in SNR. We also obtain a code design criterion that guarantees DM tradeoff optimality and which ties in nicely with other criteria established in the space-time coding literature. The systematic construction of optimal codes is nontrivial when the transmitter is equipped with multiple antennas. We show, however, that the design problem can be separated into two simpler and independent problems: the design of an inner code, also referred to as precoder, adapted to the selectivity characteristics of the channel, and the design of an outer code which, quite remarkably, is independent of the channel characteristics. Our technique is also pertinent to the analysis of selective-fading multiple-access (MA) channels where an additional factor, the multiuser interference, comes into play. In addition to establishing the optimal DM tradeoff in MA channels, we find a set of code design criteria for the construction of optimal distributed space-time codes, and we prove that the construction proposed by Badr and Belfiore (2008b) is optimal with respect to our criteria. Besides the cases of point-to-point and MA selective-fading channels analyzed in this thesis, the concept of Jensen channel has also been instrumental in studying performance limits of relay channels and in obtaining corresponding optimal code constructions (Akçaba et al., 2007). Therefore, the Jensen channel constitutes an important conceptual tool in the DM tradeoff framework.

1.2. OUTLINE

This thesis is organized as follows.

The selective-fading channel (Chapter 2)

This chapter presents the channel model underlying our analysis of selective-fading channels. The channel is modeled as a linear time-varying (LTV) system which is mathematically described by an underspread operator (Kozek, 1997), and which satisfies the standard wide-sense stationary uncorrelated scattering (WSSUS) assumption (Bello, 1963). Thanks to its underspread character, the continuous channel model is approximately diagonalized by Weyl-Heinsenberg bases (Kozek, 1997). Invoking this property, one can obtain a discrete representation of the input-output relation. Throughout the thesis we shall make use of that representation.

The diversity-multiplexing tradeoff framework (Chapter 3)

The DM tradeoff framework (Zheng and Tse, 2003) is an elegant and effective tool to characterize the tradeoff between rate and reliability in fading channels. This chapter lays down the foundations of our analysis by presenting the fundamental concepts of the framework. Furthermore, we review known DM tradeoff results in the case of flat-fading channels, and we also discuss the extension of the framework to the case of flat-fading MA channels (Tse et al., 2004).

Performance limits in selective-fading channels (Chapter 4)

This chapter presents a technique to characterize the optimal DM tradeoff of selective-fading multiple-input multiple-output (MIMO) channels (Coronel and Bölcskei, 2007). In contrast to the flat-fading case, the mutual information corresponding to a selective-fading channel is difficult to analyze as the channel exhibits memory both in time and frequency. The technique presented here overcomes this difficulty by analyzing the "Jensen channel" which is associated with the original channel and happens to have the same behavior as the latter in the scale of interest in the DM tradeoff framework. Furthermore, our approach yields a code design criterion that guarantees optimality with respect to the DM tradeoff. We shall compare this criterion with other criteria reported in prior work, and examine the implications of channel selectivity on code design.

Optimal code construction (Chapter 5)

We examine the problem of constructing DM tradeoff optimal codes, and we show that the overall code design problem can be separated in two simpler and independent tasks: the design of an inner code, also referred to as “precoder”, adapted to the channel selectivity characteristics and an outer code independent of the channel statistics. The inner code can be obtained in a systematic fashion as a function of the channel statistics and the design criterion for the outer code is standard.

Multiple-access channels (Chapter 6)

In assessing the performance limits of communication over MA channels, an additional effect, unknown in point-to-point channels, has to be taken into account: multiuser interference. This chapter extends the techniques developed in the point-to-point case and establishes the ultimate performance in selective-fading MA MIMO channels (Coronel et al., 2008). Analogously to the point-to-point case presented in Chapter 4, we obtain a set of code design criteria that can be used to jointly design the users’ codebooks. Moreover, we find a conceptual relation between the DM tradeoff framework and the notion of dominant error event which was originally introduced by Gallager (1985). The analysis of the dominant error events reveals operational regimes where multiuser interference has a negligible effect on the error performance. We shall conclude this chapter by showing that the distributed code construction presented recently by Badr and Belfiore (2008b) satisfies our code design criteria, and it is hence optimal with respect to the DM tradeoff.

Finally, Chapter 7 concludes this thesis.

CHAPTER 2

The Selective-Fading Channel

AN ACCURATE UNDERSTANDING of the wireless channel is required to design efficient communication systems. Since wireless propagation occurs via electromagnetic radiation from transmitter to receiver, it is in theory possible to solve Maxwell's equations to find the electromagnetic field at the receive antenna. This approach should take into account the reflections due to multiple scatterers and the obstructions caused by objects located in the line-of-sight between transmitter and receiver. For systems operating at carrier frequencies ranging between 1 and 10 GHz, the wavelength of electromagnetic radiation lies between 3 and 30 cm. Hence, a direct computation of the electromagnetic field at the receiver would require a precise characterization of the propagation environment within centimeters. Solving the electromagnetic field equations is therefore an unrealistic approach to obtain a description of the channel.

Motivated by the fact that a precise characterization of the propagation environment is intricate in most cases of practical interest, wireless channels are modeled as stochastic processes. The engineering appeal of this approach is clear: communication systems designed on the basis of a random channel model can be expected to work reasonably well in a large variety of environments. More generally, the design of robust and efficient systems calls for a channel model capable of capturing with sufficient accuracy the transformations undergone by the signal during propagation while at the same time remaining mathematically tractable.

In this thesis, we assume that the wireless channel is a random Gaussian linear time-varying (LTV) system which satisfies the wide-sense stationary uncorrelated scattering (WSSUS) assumption (Bello, 1963). This model is valid for a fairly general class of channels, and it provides an accurate description of the propagation effects while retaining mathematical tractability.

The continuous nature of the input-ouput relation pertaining to this model renders difficult an analysis based on standard communication and information theoretic tools. We shall therefore assume that the LTV channel is *underspread*, i.e., the product of maximum delay and maximum Doppler shift is small, in which case one can get a discretized representation of the channel's input-output relation (Kozek, 1997). The underspread LTV channel is a useful model to analyze coding and modulation schemes (Bölcskei et al., 2002b; Matz et al., 2007) and to study information-theoretic performance limits (Durisi et al., 2008; Schuster et al., 2008).

In this chapter, we describe in more detail the underspread WSSUS channel model and the corresponding discretized input-output relation. We start by examining the single-input single-ouput (SISO) selective-fading channel, and we will subsequently present the general principles underlying the system models employed later in the thesis for point-to-point multiple-input multiple-output (MIMO) systems and multiple-access (MA) MIMO systems.

2.1. THE UNDERSPREAD FADING CHANNEL

A wireless channel can be described in terms of a linear operator $\mathbb{H}$ that maps a scalar input signal $x(t)$, which is assumed to be a square integrable function, into an output signal $r(t)$ that belongs to the range space of $\mathbb{H}$. The corresponding (noise-free) input-output relation is given by $r(t) = (\mathbb{H}x)(t)$. As mentioned in the outset, it is convenient to model the channel as a stochastic process and, hence, we assume that $\mathbb{H}$ is a random operator. The noise-free input-output relation of the LTV channel can be written as

$$r(t) = (\mathbb{H}x)(t) = \int_{t'} K_{\mathbb{H}}(t, t')x(t')dt' \tag{2.1}$$

where the *kernel* $K_{\mathbb{H}}(t, t')$ can be interpreted as the channel response at time t to a Dirac impulse at time t'. The *time-varying impulse response* is given by $H_{\mathbb{H}}(t, \tau) = K_{\mathbb{H}}(t, t - \tau)$, and the corresponding input-output relation follows as (Bello, 1963)

$$r(t) = \int_{\tau} H_{\mathbb{H}}(t, \tau)x(t - \tau)d\tau. \tag{2.2}$$

We shall use in the sequel two additional system functions; first, the *time-varying transfer function* of the channel given by

$$L_{\mathbb{H}}(t, f) = \int_{\tau} H_{\mathbb{H}}(t, \tau)e^{-j2\pi f\tau}d\tau \tag{2.3}$$

and the channel's *delay-Doppler spreading function*

$$S_{\mathbb{H}}(\tau, \nu) = \int_{t} H_{\mathbb{H}}(t, \tau)e^{-j2\pi\nu t}dt. \tag{2.4}$$

For every channel realization, $H_{\mathbb{H}}(t, \tau)$ is assumed to be square integrable so that (2.3) and (2.4) are well defined. The above Fourier

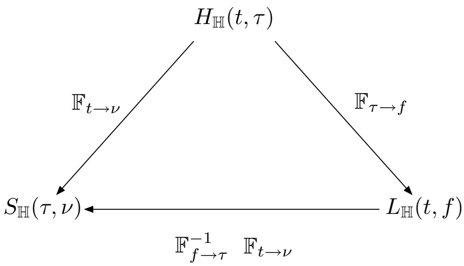

Fig. 2.1: Summary of Fourier relations between system functions

relations are summarized in Figure 2.1. An alternative representation of the input-output relation (2.2) in terms of the spreading function is

$$r(t) = \int\limits_{\tau}\int\limits_{\nu} S_{\mathbb{H}}(\tau,\nu)x(t-\tau)e^{j2\pi\nu t}d\tau d\nu \tag{2.5}$$

which essentially says that the output signal is a superposition of replicas of the input signal weighted by $S_{\mathbb{H}}(\tau,\nu)$, and shifted in time and frequency by the delay τ and the Doppler shift ν, respectively.

2.1.1. Stochastic model

We assume that $L_{\mathbb{H}}(t,f)$ is a zero-mean jointly proper Gaussian (JPG) random process in t and f, hence completely characterized by its correlation function. Furthermore, we assume that the channel process is wide-sense stationary in time t and uncorrelated in delay τ, i.e., the so-called WSSUS assumption (Bello, 1963), so that $L_{\mathbb{H}}(t,f)$ is wide-sense stationary in time t and frequency f, or, equivalently, $S_{\mathbb{H}}(\tau,\nu)$ is uncorrelated in delay τ and Doppler ν, i.e.,

$$\mathbb{E}\{L_{\mathbb{H}}(t,f)L^*_{\mathbb{H}}(t',f')\} = R_{\mathbb{H}}(t-t',f-f') \tag{2.6}$$

$$\mathbb{E}\{S_{\mathbb{H}}(\tau,\nu)S^*_{\mathbb{H}}(\tau',\nu')\} = C_{\mathbb{H}}(\tau,\nu)\,\delta(\tau-\tau')\,\delta(\nu-\nu') \tag{2.7}$$

where the scattering function $C_{\mathbb{H}}(\tau, \nu)$ and the time-frequency correlation function $R_{\mathbb{H}}(t, f)$ are related by the two-dimensional Fourier relation

$$C_{\mathbb{H}}(\tau, \nu) = \int_t \int_f R_{\mathbb{H}}(t, f) e^{-j2\pi(\nu t - \tau f)} dt\, df. \tag{2.8}$$

Since $R_{\mathbb{H}}(t, f)$ is stationary in both t and f, the scattering function is nonnegative and real-valued for all τ and ν. It can therefore be interpreted as the spectrum of the channel process.

2.1.2. The underspread assumption

Because the transmitter, receiver, and the objects in the scattering environment have finite velocities, the maximum Doppler shift ν_0 experienced by the transmitted signal is finite. Similarly, we presume that multipath propagation incurs a finite delay that is at most τ_0. For simplicity, we assume, without loss of generality, that the scattering function is compactly supported on the rectangle $[0, \tau_0] \times [0, \nu_0]$, i.e.,

$$C_{\mathbb{H}}(\tau, \nu) = 0 \quad \text{for } (\tau, \nu) \notin [0, \tau_0] \times [0, \nu_0].$$

Note that the spreading function $S_{\mathbb{H}}(\tau, \nu)$ is also supported in this rectangle with probability 1 (w.p. 1). The channel spread is given by the quantity $\Delta_{\mathbb{H}} = \nu_0 \tau_0$. The underspread assumption, which implies that the total channel spread satisfies $\Delta_{\mathbb{H}} < 1$ (Kozek, 1997), is relevant as most mobile radio channels are (highly) underspread. Moreover, underspread channels have a set of approximate eigenfunctions that yield a discretized input-output relation as discussed next.

2.1.3. Approximate diagonalization

We build our developments on the fact that underspread channels are approximately diagonalized by orthogonal Weyl-Heisenberg bases

(Kozek, 1997) which are obtained by time-frequency shifting a prototyping pulse $g(t)$ as

$$g_{m,k}(t) = g(t - mT)e^{j2\pi kFt}$$

where

$$\langle g_{m,k}, g_{n,p} \rangle = \int_t g_{m,k}(t) g^*_{n,p}(t) dt = \delta_{m,n} \delta_{k,p}. \tag{2.9}$$

In order to satisfy the orthogonality requirement on the basis $\{g_{m,k}(t)\}$, the grid parameters T and F are assumed to satisfy $TF \geq 1$ (Kozek, 1997; Kozek and Molisch, 1998). Note that large values of the product TF allow a better time-frequency localization of the prototyping pulse $g(t)$, but result in a loss of signal space dimension in comparison to the case $TF = 1$.

Since the spreading function $S_{\mathbb{H}}(\tau, \nu)$ is strictly limited in delay τ and Doppler ν, by virtue of the Fourier relation between $L_{\mathbb{H}}(t, f)$ and $S_{\mathbb{H}}(\tau, \nu)$, the time-varying transfer function $L_{\mathbb{H}}(t, f)$ is exactly characterized by its samples $L_{\mathbb{H}}(mT, kF)$ taken on a regular grid satisfying the Nyquist conditions $T \leq \frac{1}{\nu_0}$ and $F \leq \frac{1}{\tau_0}$. With this choice of grid parameters, it has been shown (Kozek, 1997) that the kernel $K_{\mathbb{H}}(t, t')$ admits the following approximate eigenvalue decomposition

$$K_{\mathbb{H}}(t, t') = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} L_{\mathbb{H}}(mT, kF)\; g_{m,k}(t) g^*_{m,k}(t'). \tag{2.10}$$

The error associated with the approximate decomposition (2.10), which is characterized for example in (Durisi et al., 2008), will be neglected in our developments. We stress, however, that the choice of prototyping function $g(t)$ is crucial in obtaining (2.10) : $g(t)$ should depend on the shape of the scattering function $C_{\mathbb{H}}(\tau, \nu)$ and be well localized in time and frequency (Kozek, 1997; Kozek and Molisch, 1998; Matz et al., 2007; Durisi et al., 2008).

2.1.4. Canonical signaling scheme

The above approximate diagonalization does not require channel information at the transmitter nor at the receiver. In particular, the transmit signal can be constructed as a linear combination of the basis functions $\{g_{m,k}(t)\}$ as

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{k=0}^{K-1} \tilde{x}_{m,k}\, g_{m,k}(t) \tag{2.11}$$

where $\tilde{x}_{m,k}$ is the information bearing data symbol corresponding to the time-frequency slot (m,k). Since the prototyping function $g(t)$ is well-localized in frequency by construction and practically realizable signals are band limited, we have assumed that the input signal has an effective bandwidth $W = KF$.

The receiver computes the inner products $y_{m,k} = \langle y, g_{m,k} \rangle$, where $y(t) = r(t) + z(t)$ and $z(t)$ is additive white Gaussian noise with $\mathbb{E}\{z(t)z^*(t')\} = \delta(t-t')$. Introducing the normalization $x_{m,k} = \tilde{x}_{m,k}/\sqrt{\mathsf{SNR}}$, with SNR denoting the average signal-to-noise ratio (SNR), the overall input-output relation is hence given by

$$\begin{aligned} y_{m,k} &= \langle \mathbb{H}\, x, g_{m,k} \rangle + \underbrace{\langle z, g_{m,k} \rangle}_{z_{m,k}} \\ &= \sqrt{\mathsf{SNR}} \sum_{n,p} x_{n,p} \, \langle \mathbb{H}\, g_{n,p}, g_{m,k} \rangle + z_{m,k} && (2.12) \\ &= \sqrt{\mathsf{SNR}}\, L_{\mathbb{H}}(mT, kF)\, x_{m,k} + z_{m,k} && (2.13) \end{aligned}$$

where (2.12) is a direct consequence of the input signal structure in (2.11), and (2.13) follows from the approximate diagonalization (2.10) and the orthonormality of the set $\{g_{m,k}(t)\}$. The projections of the noise process onto the Weyl-Heisenberg basis are independent and identically distributed across m and k, i.e., $z_{m,k} \sim \mathcal{CN}(0,1)$, for all m and k, and $\mathbb{E}\{z_{m,k} z^*_{n,p}\} = \delta_{m,n}\delta_{k,p}$.

In essence, this scheme corresponds to transmitting and receiving on the channel's approximate eigenfunctions and, hence, it yields an approximate diagonalization of the channel. The canonical signaling scheme in (2.11) can be interpreted in terms of a pulse-shaped orthogonal frequency-division multiplexing (OFDM) system, where the data symbols $\tilde{x}_{m,k}$ are modulated onto a set of orthogonal signals (Durisi et al., 2008). According to this interpretation, the error resulting in approximating $K_{\mathbb{H}}(t, t')$ as in (2.10) corresponds to intersymbol interference (ISI) and intercarrier interference (ICI). Thus, (2.13) can be thought of as the input-output relation corresponding to a pulse-shaped OFDM system in which the ISI and ICI terms are neglected. Minimizing the interference terms, and, hence, the error in the approximation (2.10), can by accomplished by appropriate design of the prototyping pulse $g(t)$ and a proper choice of the grid parameters T and F (Kozek and Molisch, 1998; Durisi et al., 2008). Throughout this thesis, we shall assume that $g(t)$, T, and F are chosen so that the approximation error in (2.10) can be neglected.

2.2. SYSTEM MODEL

The system models employed throughout the thesis are based on the approximate diagonalization of underspread WSSUS channels by Weyl-Heisenberg bases presented above. A precise description of the models corresponding to the point-point and MA cases is given at the beginning of Chapter 4 and Chapter 6, respectively. However, we shall next examine in further detail some aspects that are common to both cases due to their common underlying structure.

Time-frequency slot mapping and input-output relations

We assume that communication takes place over M time slots and K frequency slots so that, for the sake of notation, we use the bijective mapping $\mathcal{M}$, defined as

$$\begin{aligned} \mathcal{M}: \quad [0:M-1]\times[0:K-1] \quad &\longrightarrow \quad [0:N-1] \\ (m,k) \quad &\longmapsto \quad n = mK+k \end{aligned} \tag{2.14}$$

to index any time-frequency slot (m,k) in (2.13) by $n = \mathcal{M}(m,k)$. Moreover, we extend the input-output relation (2.13) to a MIMO channel with $\mathrm{M_T}$ transmit and $\mathrm{M_R}$ receive antennas, assuming for simplicity that the scalar subchannels of the $\mathrm{M_T} \times \mathrm{M_R}$ MIMO channel have statistically independent kernels with identical statistics, i.e., identical scattering functions. Consequently, all subchannels are approximately diagonalized by the same Weyl-Heisenberg basis so that, based on (2.13) and the mapping in (2.14), we get

$$\mathbf{y}_n = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}}\,\mathbf{H}_n\mathbf{x}_n + \mathbf{z}_n, \quad n = 0,1,\ldots,N-1 \tag{2.15}$$

where SNR is the average signal-to-noise ratio at each receive antenna, $\mathbf{y}_n$, $\mathbf{x}_n$ 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 $\mathrm{M_R}\times 1$ JPG noise vector satisfying $\mathbf{z}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{\mathrm{M_R}})$. The channel matrices are given by $\mathbf{H}_n(i,j) = L_{\mathbb{H}}^{(i,j)}(mT, kF)$ $(i = 1,2,\ldots,\mathrm{M_R}, j = 1,2,\ldots,\mathrm{M_T})$, where the superscript (i,j) designates the time-varying transfer function corresponding to the subchannel between transmit antenna j and receive antenna i.

In the MA case, we shall assume that channels corresponding to different users are statistically independent with identical statistics. Hence, by the same line of reasoning as before, the corresponding input-output relation follows as

$$\mathbf{y}_n = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}}\sum_{u=0}^{U-1}\mathbf{H}_{u,n}\mathbf{x}_{u,n} + \mathbf{z}_n, \quad n = 0,1,\ldots,N-1 \tag{2.16}$$

where $\mathbf{x}_{u,n}$ is the transmit signal vector for user $u \in \mathcal{U} = \{1,2,\ldots,U\}$ and the corresponding channel matrix $\mathbf{H}_{u,n}$ has entries $\mathbf{H}_{u,n}(i,j) = L_{\mathbb{H}}{}^{(i,j,u)}(mT, kF)$.

Statistics of the discrete fading process

Because the scalar subchannels are assumed to be JPG processes having statistically independent kernels with identical statistics, the channel matrices are spatially uncorrelated and the correlation across slots is related to the time-frequency correlation function (2.6) as

$$\mathbb{E}\{\mathbf{H}_n(i,j)(\mathbf{H}_{n'}(i,j))^*\} = R_{\mathbb{H}}((m-m')T,(k-k')F),$$
$$i = 1,2,\ldots,\mathrm{M_R},\ j = 1,2,\ldots,\mathrm{M_T} \quad (2.17)$$

where $n = \mathcal{M}(m,k)$ and $n' = \mathcal{M}(m',k')$, with $n, n' = 0, 1, \ldots, N-1$, are two time-frequency slots. In the MA case, we have

$$\mathbb{E}\{\mathbf{H}_{u,n}(i,j)(\mathbf{H}_{u',n'}(i,j))^*\} = R_{\mathbb{H}}((m-m')T,(k-k')F)\delta(u-u'),$$
$$i = 1,2,\ldots,\mathrm{M_R},\ j = 1,2,\ldots,\mathrm{M_T} \quad (2.18)$$

where u and u' are two arbitrary users. For later reference, we define the $N \times N$ covariance matrix $\mathbf{R}_{\mathbb{H}}$ as

$$\mathbf{R}_{\mathbb{H}}(n,n') = R_{\mathbb{H}}((m-m')T,(k-k')F). \quad (2.19)$$

Clearly, $\mathbf{R}_{\mathbb{H}}$ follows from the time-frequency correlation function $R_{\mathbb{H}}(t,f)$, the mapping (2.14), and the relation (2.17). Hence, in general, $\mathbf{R}_{\mathbb{H}}$ is a two-level Toeplitz matrix, i.e., a Toeplitz-block-Toeplitz matrix.

The two-dimensional power spectral density of the process is given by

$$S(\xi,\mu) = \sum_{m=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} R_{\mathbb{H}}(mT,kF)\, e^{-j2\pi(m\mu-k\xi)},$$
$$0 \le \xi, \mu < 1. \quad (2.20)$$

The power spectral density $S(\xi,\mu)$ can be related to the scattering function as follows

$$
\begin{aligned}
S(\xi,\mu) = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} & e^{-j2\pi(m\mu-k\xi)} \times \\
& \int_{\nu}\int_{\tau} C_{\mathbb{H}}(\tau,\nu)\, e^{j2\pi(\nu mT-\tau kF)} d\tau\, d\nu \qquad (2.21) \\
= \int_{\nu}\int_{\tau} C_{\mathbb{H}}(\tau,\nu) & \sum_{m=-\infty}^{\infty} e^{j2\pi mT(\nu-\frac{\mu}{T})} \times \\
& \sum_{k=-\infty}^{\infty} e^{-j2\pi kF(\tau-\frac{\xi}{F})}\, d\tau\, d\nu \\
= \frac{1}{TF}\int_{\nu}\int_{\tau} C_{\mathbb{H}}(\tau,\nu) & \sum_{m=-\infty}^{\infty} \delta\Big(\nu-\frac{\mu-m}{T}\Big) \times \\
& \sum_{k=-\infty}^{\infty} \delta\Big(\tau-\frac{\xi-k}{F}\Big)\, d\tau\, d\nu \qquad (2.22) \\
= \frac{1}{TF} \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} & C_{\mathbb{H}}\Big(\frac{\xi-k}{F},\frac{\mu-m}{T}\Big) \qquad (2.23)
\end{aligned}
$$

where we have used (2.8) to get (2.21), and (2.22) is a consequence of Poisson's sum formula. The variance of each channel coefficient satisfies

$$\sigma_{\mathbb{H}}^2 = \int_{\tau}\int_{\nu} C_{\mathbb{H}}(\tau,\nu)\, d\tau d\nu. \qquad (2.24)$$

We conclude this chapter with the following remarks. First, we emphasize that the input-ouput relations in (2.15) and (2.16) presume that the time-varying transfer functions are characterized by the same scattering function $C_{\mathbb{H}}(\tau,\nu)$. An important implication of this assumption is that the channel's correlation across time-frequency slots is described by the same covariance matrix $\mathbf{R}_{\mathbb{H}}$ given in (2.19) for all spatial subchannels and all users.

Second, we note that the input-ouput relations obtained here encompass many simple channel models in addition to the WSSUS model (see the discussion on linear frequency-invariant (LFI) and linear time-invariant (LTI) channels in Appendix A). The results developed in the sequel do therefore apply to these models provided one takes into account the corresponding structural differences when pertinent. We shall explicitly consider the most important instances of theses models in our exposition.

CHAPTER 3

The Diversity-Multiplexing Tradeoff

RATE AND RELIABILITY are two essential characteristics of communication systems that have been widely studied by information theorists. Shannon proved that channel capacity is the maximum rate at which reliable communication is achievable: for any rate below capacity there exist encoders and decoders that achieve arbitrarily small error probabilities and, conversely, error probability is bounded away from zero for any rate above capacity (Shannon, 1948).

Approaching error-free performance at rates below capacity requires the use of complex coding schemes with large block lengths. The theory of error exponents developed by Gallager (1968) provides tools to characterize the maximum decay, or error exponent, of error probability with block length. The error exponent is a decreasing function of the rate, which is to say, rate and reliability are subject to a fundamental tradeoff: an increase in rate comes at the expense of error performance and vice versa. Unfortunately, a precise characterization of how rate and probability of error interplay is a difficult task in general.

The *diversity-multiplexing (DM) tradeoff* framework introduced by Zheng and Tse (2003) has proved to be an efficient tool to elegantly capture the rate-reliability tradeoff in fading channels. In contrast to the error exponent approach, this new framework develops an asymptotic analysis in signal-to-noise ratio (SNR) in lieu of block length.

In this chapter, we review the fundamentals of the DM tradeoff framework as it constitutes the conceptual foundation for the results presented in this thesis. We start by discussing the framework for point-to-point flat-fading channels (Zheng and Tse, 2003), and we shall subsequently address its extension to flat-fading multiple-access (MA) channels (Tse et al., 2004). Note that the concepts developed below will be generalized to the selective-fading case in Chapter 4 and Chapter 6.

3.1. PRELIMINARIES

In flat-fading multiple-input multiple-output (MIMO) channels, the $\mathrm{M_R} \times \mathrm{M_T}$ channel matrix $\mathbf{H}_0$ remains fixed so that the input-ouput relation over a block of N slots is obtained from (2.15) as

$$\mathbf{Y} = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}}\mathbf{H}_0\,\mathbf{X} + \mathbf{Z} \tag{3.1}$$

where $\mathbf{X} = [\mathbf{x}_0\, \mathbf{x}_1\, \cdots\, \mathbf{x}_{N-1}]$ denotes the $\mathrm{M_T} \times N$ input codeword and $\mathbf{Y} = [\mathbf{y}_0\, \mathbf{y}_1\, \cdots\, \mathbf{y}_{N-1}]$ is the $\mathrm{M_R} \times N$ received matrix. The $\mathrm{M_R} \times N$ additive noise matrix $\mathbf{Z}$ satisfies $\mathbf{Z}(i,j) \sim \mathcal{CN}(0,1)$, $\forall i,j$. We are interested in the situation where the receiver has channel state information (CSI), but the transmitter has only access to the fading law. The DM tradeoff framework examines system performance for fixed block length when both the data rate and the error probability are allowed to scale with SNR. In this context, rate and reliability are respectively captured by the somewhat coarser notions of multiplexing and diversity.

3.1.1. Rate and multiplexing

Assuming a code that spans infinitely many independent realizations of the channel with input-output relation (3.1), the maximum rate that can be reliably transmitted with CSI at the receiver is given by the ergodic capacity (Telatar, 1999)

$$C(\mathsf{SNR}) = \mathbb{E}\left\{\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{H}_0\mathbf{H}_0^H\right)\right\}. \qquad (3.2)$$

Asymptotically in SNR, it can be shown that (3.2) statisfies (Foschini, 1996)

$$C(\mathsf{SNR}) = \min(\mathrm{M_T}, \mathrm{M_R}) \log \mathsf{SNR} + O(1) \qquad (3.3)$$

where the factor $\mathrm{m} \triangleq \min(\mathrm{M_T}, \mathrm{M_R})$, known as the (maximum) *multiplexing gain*, represents the number of parallel data pipes that are available for data transmission in a given bandwidth and at a certain transmit power. At high SNR, the multiplexing gain constitutes the maximal increase in data rate that can be obtained by increasing the SNR. For instance, by increasing SNR by $3\,\mathrm{dB}$, m additional bits (per second per Hertz) can be reliably conveyed through the channel.

3.1.2. Reliability and diversity

The drops in signal power that characterize fading channels result in high detection error probabilities. Diversity techniques can improve error performance by providing the receiver with multiple independently-faded replicas of the same transmit signal. The probability of having all replicas experiencing a deep fade simultaneously is then drastically reduced. Indeed, assuming d i.i.d. Rayleigh fading branches, the average error probability can be shown to satisfy

$$\mathbb{P}(\text{error}) = O\left(\mathsf{SNR}^{-d}\right). \qquad (3.4)$$

Time and frequency diversity can be realized using standard forward error correcting codes in conjunction with interleaving (Proakis, 2001; Biglieri, 2005). Antenna diversity can be exploited at the receiver (Jakes, 1974) and/or at the transmitter, for example, by appropriately coding across transmit antennas (Alamouti, 1998; Tarokh et al., 1998, 1999; Guey et al., 1999; Bölcskei and Paulraj, 2000), or by converting spatial diversity into time or frequency diversity and using then standard coding techniques (Hiroike et al., 1992; Wittneben, 1993; Seshadri and Winters, 1994; Kuo and Fitz, 1997).

Two systems satisfying (3.4) may have drastically different error performance at fixed SNR, e.g., because they use codes with different gains. Asymptotically in SNR, however, the error performance turns out to be completely determined by the exponent d. Since we focus on the high-SNR regime, we shall use the SNR exponent d as a measure of reliability.

3.1.3. System definitions

The central concept invoked in the DM tradeoff framework to study the interrelation between rate and reliability is that of *family of codes*. A family of codes $\mathcal{C}_r$ is a sequence of codebooks $\mathcal{C}_r(\mathsf{SNR})$ parametrized by SNR. At a given SNR, the corresponding codebook $\mathcal{C}_r(\mathsf{SNR})$ contains SNR^{Nr} codewords, implying that the data rate scales with SNR as $R(\mathsf{SNR}) = r \log \mathsf{SNR}$, where r is the multiplexing rate. More formally, we have the following definition.

Definition 3.1. A family of codes $\mathcal{C}_r$ operates at multiplexing rate $r \in [0, \mathrm{m}]$ if the rate $R(\mathsf{SNR})$ corresponding to the codebook $\mathcal{C}_r(\mathsf{SNR})$ satisfies

$$\lim_{\mathsf{SNR}\to\infty} \frac{R(\mathsf{SNR})}{\log \mathsf{SNR}} = r. \tag{3.5}$$

The multiplexing rate r is a fraction of the maximum multiplexing gain $\min(\mathrm{M_T}, \mathrm{M_R})$ obtained in (3.3), which is to say that the family of codes $\mathcal{C}_r$ achieves a non-vanishing fraction of the ergodic capacity (3.2) as SNR increases. We also note that the multiplexing rate corresponding to a family of codes with fixed rate equals zero because, however large the latter may be, the limit in (3.5) for such a family is zero.

The concept of family of codes enables us to characterize the rate-reliability tradeoff by relating the diversity achieved by $\mathcal{C}_r$, i.e., the reliability associated to $\mathcal{C}_r$ which is measured by the SNR exponent of the corresponding error probability, to its multiplexing rate r. This analysis gives rise to the DM tradeoff, which is formally defined as follows.

Definition 3.2. The DM tradeoff realized by $\mathcal{C}_r$ is given by the function

$$d(\mathcal{C}_r) = -\lim_{\mathsf{SNR}\to\infty} \frac{\log P_e(\mathcal{C}_r)}{\log \mathsf{SNR}} \tag{3.6}$$

where $P_e(\mathcal{C}_r)$ is the error probability obtained through maximum-likelihood (ML) detection. Moreover, the optimal DM tradeoff curve

$$d^\star(r) = \sup_{\mathcal{C}_r} d(\mathcal{C}_r) \tag{3.7}$$

quantifies the maximum achievable diversity gain over all families of codes that operate at multiplexing rate r.

In contrast to the traditional concept of diversity gain which is used to characterize the high-SNR performance of a single codebook, the DM tradeoff $d(\mathcal{C}_r)$ is a performance measure associated with the entire family of codes $\mathcal{C}_r$. Both the data rate and the error probability corresponding to $\mathcal{C}_r$ scale with SNR, and the function $d(\mathcal{C}_r)$ captures their tradeoff in the high-SNR limit. As we shall see in the next section, there exists a fundamental limit to the optimal DM tradeoff $d^\star(r)$ that is dictated by the fading channel.

Throughout the rest of this thesis, two functions $f(x)$ and $g(x)$ are said to be exponentially equal, denoted by $f(x) \doteq g(x)$, if and only if

$$\lim_{x\to\infty} \frac{\log f(x)}{\log x} = \lim_{x\to\infty} \frac{\log g(x)}{\log x}. \tag{3.8}$$

Exponential inequality, denoted by $\dot{\leq}$ and $\dot{\geq}$, is defined analogously.

3.2. PERFORMANCE LIMITS

In the situation where coding is performed over a single realization of the fading channel with input-output relation (3.1), the ergodic capacity has no operational meaning. Indeed, there is a nonzero probability to encounter a channel realization that does not support the rate of communication, however small it may be. In this case, we resort to the outage capacity formulation (Ozarow et al., 1994; Telatar, 1999) to determine the maximum rate at which reliable communication can be supported for some prescribed probability of outage. We shall see that outage capacity constitutes a fundamental performance limit not only for large block lengths but also asymptotically in SNR.

3.2.1. Outage probability

We assume that the receiver has perfect CSI but the transmitter has only access to the fading law. An outage event occurs when the mutual information associated with the channel realization $\mathbf{H}_0$ is below the target data rate R, i.e.,

$$\frac{1}{N}\, I(\mathbf{X};\mathbf{Y}\,|\,\mathbf{H}_0) < R. \tag{3.9}$$

The corresponding outage probability follows from optimizing the mutual information over the input distribution $f(\mathbf{X})$ as

$$P_{\text{out}}(R) = \inf_{f(\mathbf{X})}\ \mathbb{P}\left(\frac{1}{N}\, I(\mathbf{X};\mathbf{Y}|\mathbf{H}_0) < R\right). \tag{3.10}$$

The optimal distribution is Gaussian with $\text{vec}(\mathbf{X}) \sim \mathcal{CN}(\mathbf{0},\mathbf{Q})$. While Telatar (1999) conjectures that the optimal input covariance $\mathbf{Q}$ is diagonal and the available power is evenly split over some or all transmit antennas depending on the SNR and R, no general proof is available. In the high-SNR regime, however, Zheng and Tse (2003) show that a covariance matrix satisfying $\mathbf{Q} = \mathbf{I}$ incurs no loss of optimality; consequently, we can write

$$P_{\text{out}}(R) \doteq \mathbb{P}\left(\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\,\mathbf{H}_0\mathbf{H}_0^H\right) < R\right). \tag{3.11}$$

The outage event can be related to the "level of singularity" of the channel matrix. To this end, define the singularity levels (in descending order)

$$\mu_k = -\frac{\log \lambda_k(\mathbf{H}_0\mathbf{H}_0^H)}{\log \mathsf{SNR}}, \quad k = 1, 2, \ldots, \mathrm{m}, \tag{3.12}$$

where the eigenvalues of the channel matrix $\lambda_k(\mathbf{H}_0\mathbf{H}_0^H)$ are arranged in ascending order. With this notation, it has been shown in (Zheng and Tse, 2003, Th. 4) that the outage probability at rate $R(\mathsf{SNR}) = r \log \mathsf{SNR}$ satisfies the following exponential equality:

$$P_{\text{out}}(r \log \mathsf{SNR}) \doteq \mathbb{P}(\mathcal{O}_r) \tag{3.13}$$

where the outage event $\mathcal{O}_r$ is given by

$$\mathcal{O}_r = \left\{ \boldsymbol{\mu} \in \mathbb{R}_+^{\mathrm{m}} : \mu_1 \geq \mu_2 \geq \ldots \geq \mu_{\mathrm{m}}, \sum_{k=1}^{\mathrm{m}} [1-\mu_k]^+ < r \right\} \tag{3.14}$$

with $\boldsymbol{\mu} = [\mu_1 \; \mu_2 \; \ldots \; \mu_{\mathrm{m}}]$. The definition (3.14) provides a parameterization of the outage event $\mathcal{O}_r$ in terms of the singularity levels of the channel matrix. We shall see in the sequel that the outage probability (3.13) can be explicitly evaluated. First, we present a result that confers to $\mathbb{P}(\mathcal{O}_r)$ a central role in the analysis of the high-SNR performance limits of communication.

3.2.2. Performance bound

The outage probability constitutes the ultimate performance in the high-SNR regime. Next, we state this fundamental result.

Theorem 3.1 (Zheng and Tse (2003)). *For any family of codes $\mathcal{C}_r$, the error probability $P_e(\mathcal{C}_r)$ satisfies*

$$P_e(\mathcal{C}_r) \,\dot{\geq}\, \mathbb{P}(\mathcal{O}_r). \tag{3.15}$$

Proof. Fix a codebook $\mathcal{C}_r(\mathsf{SNR})$ with SNR^{Nr} codewords, and assume that the transmit codeword $\mathbf{X}$ is uniformly drawn from $\mathcal{C}_r(\mathsf{SNR})$ independently of the channel realization $\mathbf{H}_0$. By Fano's inequality (Cover and Thomas, 1991, Thm. 2.10.1), the probability of a detection error $\mathcal{E}_r$ (conditional on the channel realization) satisfies

$$1 + \mathbb{P}(\mathcal{E}_r \,|\, \mathbf{H}_0) \log |\mathcal{C}_r(\mathsf{SNR})| \geq H(\mathbf{X} \,|\, \mathbf{Y}, \mathbf{H}_0) \tag{3.16}$$

where $H(\cdot)$ denotes the entropy function. Since $\mathbf{X}$ is independent of $\mathbf{H}_0$ and uniformly drawn from $\mathcal{C}_r(\mathsf{SNR})$, we have $H(\mathbf{X} \,|\, \mathbf{H}_0) = Nr \log \mathsf{SNR}$ and, hence, adding the mutual information $I_{\mathcal{C}}(\mathbf{X}; \mathbf{Y}|\mathbf{H}_0)$ to both sides of (3.16) and performing some additional manipulations yields

$$\mathbb{P}(\mathcal{E}_r \,|\, \mathbf{H}_0) \geq 1 - \frac{I_{\mathcal{C}}(\mathbf{X}; \mathbf{Y}|\mathbf{H}_0)}{Nr \log \mathsf{SNR}} - \frac{1}{Nr \log \mathsf{SNR}}. \tag{3.17}$$

For $\epsilon > 0$, define the set

$$\mathcal{O}_r(\epsilon) = \{\mathbf{H}_0 : I_{\mathcal{C}}(\mathbf{X}; \mathbf{Y}|\mathbf{H}_0) \leq Nr(1-\epsilon) \log \mathsf{SNR}\}$$

and note that for any channel realization in $\mathcal{O}_r(\epsilon)$, it follows from (3.17) that

$$\mathbb{P}(\mathcal{E}_r \,|\, \mathbf{H}_0 \in \mathcal{O}_r(\epsilon)) \geq \epsilon - \frac{1}{Nr \log \mathsf{SNR}}. \tag{3.18}$$

Averaging the error probability over the channel distribution yields

$$\begin{aligned} P_e(\mathcal{C}_r) &= \mathbb{E}\{\mathbb{P}(\mathcal{E}_r \,|\, \mathbf{H}_0)\} \\ &\geq \mathbb{E}\Big\{\mathbb{P}(\mathcal{E}_r \,|\, \mathbf{H}_0) \,\Big|\, \mathbf{H}_0 \in \mathcal{O}_r(\epsilon)\Big\} && (3.19) \\ &\geq \left(\epsilon - \frac{1}{Nr \log \mathsf{SNR}}\right) \mathbb{P}(\mathcal{O}_r(\epsilon)) && (3.20) \\ &\dot{\geq} \; \mathbb{P}\big(\mathcal{O}_{r(1-\epsilon)}\big) && (3.21) \end{aligned}$$

where (3.20) follows from using (3.18) in (3.19), and (3.21) is a consequence of taking the exponential limit in SNR and invoking the optimality (in the sense of the outage probability) of i.i.d. Gaussian codebooks in this regime. The claim follows by taking ϵ arbitrarily close to zero. □

Since the outage probability $\mathbb{P}(\mathcal{O}_r)$ constitutes a lower bound to the error probability of any family of codes, the best DM tradeoff curve must satisfy

$$d^\star(r) \le d_{\mathcal{O}}(r) \tag{3.22}$$

where the outage curve $d_{\mathcal{O}}(r)$ is defined as

$$d_{\mathcal{O}}(r) \triangleq -\lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}(\mathcal{O}_r)}{\log \mathsf{SNR}}. \tag{3.23}$$

It is well known that the outage probability is the ultimate performance for infinite block length: for a given channel realization, any rate below mutual information can be conveyed with vanishing error probability as the block length tends to infinity. Theorem 3.1 says that outage probability is also the ultimate performance for finite block lengths, bringing thus the question of whether there exist families of codes that can achieve that performance asymptotically in SNR (instead of block length). In the case of flat-fading, Zheng and Tse (2003) have shown the existence of such codes as we detail next.

3.2.3. Achievability

The outage bound of Theorem 3.1 is actually achievable in the flat-fading case. Next, we summarize this result by Zheng and Tse (2003).

Theorem 3.2 (Zheng and Tse (2003)). *In the flat-fading channel, the optimal DM tradeoff among all family of codes $\mathcal{C}_r$ with block length $N \ge \mathrm{M_T} + \mathrm{M_R} - 1$ satisfies*

$$d^\star(r) = d_{\mathcal{O}}(r), \quad \forall r \in [0, \mathrm{m}]. \tag{3.24}$$

This result says that there exist good families of codes $\mathcal{C}_r$ that achieve the outage bound. The error probability of such schemes is exponentially equal to the outage probability, which essentially means that the typical way an error occurs is to encounter a channel realization that is in outage. While the existence of codes that approach the outage bound in the limit of infinite block length is well established, the fact that this performance bound is achievable asymptotically in SNR but with finite block lengths is remarkable.

The proof of Theorem 3.2 consists in averaging the error probability over the ensemble of Gaussian codes, and noting that for sufficiently large block lengths, i.e., larger than $\mathrm{M_T} + \mathrm{M_R} - 1$, the SNR exponent of the error probability is not larger than that of the outage probability. Although this result establishes the existence of optimal codes, it does not give any indication on how to construct them. Sparked by this result, a number of DM tradeoff optimal encoding and decoding schemes have been found during the past few years. In particular, the *non-vanishing* determinant criterion (Belfiore and Rekaya, 2003; Yao and Wornell, 2003) on codeword difference matrices has been shown to constitute a sufficient condition for DM tradeoff optimality in flat-fading MIMO channels with two transmit and two or more receive antennas (Yao and Wornell, 2003); this criterion has led to the construction of space-time codes based on constellation rotation (Yao and Wornell, 2003; Dayal and Varanasi, 2005) and cyclic division algebras (Belfiore et al., 2005). In (El Gamal et al., 2004), 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 (Tavildar and Viswanath, 2006).

3.3. EVALUATION OF THE OUTAGE PROBABILITY

In flat fading, the distribution of the Hermitian matrix $\mathbf{H}_0\mathbf{H}_0^H$ is known as the Wishart distribution, and the joint density of its eigenvalues $\lambda_k = \lambda_k(\mathbf{H}_0\mathbf{H}_0^H)$, $k = 1, 2, \ldots, \mathrm{m}$, is given by (James, 1964; Muirhead, 1982; Edelman, 1989)

$$f_{\boldsymbol{\lambda}}(\lambda_1, \lambda_2, \ldots, \lambda_{\mathrm{m}}) = K \prod_{k=1}^{\mathrm{m}} \lambda_k^{\mathrm{M}-\mathrm{m}} \prod_{j<k} (\lambda_k - \lambda_j)^2 \exp\left(-\sum_{k=1}^{\mathrm{m}} \lambda_k\right), \tag{3.25}$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{\mathrm{m}}$, $\mathrm{M} \triangleq \max(\mathrm{M_T}, \mathrm{M_R})$ and K is a normalizing constant. Using (3.25), we shall see next how to evaluate the outage probability (3.13).

The parameterization of the outage event (3.14) shows that some eigenvalues of $\mathbf{H}_0\mathbf{H}_0^H$ must decay with SNR for an outage to occur at multiplexing rate $r > 0$, or, equivalently, some singularities μ_k's must be positive, in which case the channel matrix is near singular. Motivated by this observation, we approximate the joint eigenvalue distribution (3.25) in the vicinity of zero. First, using the standard approximation of the exponential function around zero, we write

$$e^{-\lambda_k} = 1 + o(\lambda_k)\,. \tag{3.26}$$

Then, the eigenvalue ordering yields $\lambda_k - \lambda_j < \lambda_k$ for all $j < k$, and hence, we get

$$\prod_{j<k} (\lambda_k - \lambda_j)^2 = \prod_{j<k} o\big(\lambda_k^2\big) = o\Big(\lambda_k^{2(k-1)}\Big)\,. \tag{3.27}$$

Inserting (3.26) and (3.27) into (3.25) yields

$$f_{\boldsymbol{\lambda}}(\lambda_1, \lambda_2, \ldots, \lambda_{\mathrm{m}}) = K \prod_{k=1}^{\mathrm{m}} (1 + o(\lambda_k))\; o\big(\lambda_k^{2k-2+\mathrm{M}-\mathrm{m}}\big)$$

$$= \prod_{k=1}^{\mathrm{m}} o\big(\lambda_k^{2k-2+\mathrm{M}-\mathrm{m}}\big)\,. \tag{3.28}$$

From (3.12), the eigenvalues are related to the singularity levels according to $\lambda_k = \mathsf{SNR}^{-\mu_k}$, $k = 1, 2, \ldots, \mathrm{m}$. Using this variable change in (3.28) yields the following asymptotic expression for (3.25):

$$\begin{aligned} f_{\boldsymbol{\mu}}(\mu_1, \mu_2, \ldots, \mu_{\mathrm{m}}) &\doteq \prod_{k=1}^{\mathrm{m}} \mathsf{SNR}^{-(2k-1+\mathrm{M}-\mathrm{m})\mu_k} \\ &= \mathsf{SNR}^{-\sum_{k=1}^{\mathrm{m}}(2k-1+\mathrm{M}-\mathrm{m})\mu_k}. \end{aligned} \tag{3.29}$$

We can now use this approximation of the distribution of $\boldsymbol{\mu}$ around zero* to compute $\mathbb{P}(\mathcal{O}_r)$. First, the outage probability is trivially lower bounded as

$$\begin{aligned} \mathbb{P}(\mathcal{O}_r) &= \int_{\mathcal{O}_r} f_{\boldsymbol{\mu}}(\mu_1, \mu_2, \ldots, \mu_{\mathrm{m}})\, d\boldsymbol{\mu} \\ &\geq \sup_{\boldsymbol{\mu} \in \mathcal{O}_r} f_{\boldsymbol{\mu}}(\mu_1, \mu_2, \ldots, \mu_{\mathrm{m}}). \end{aligned} \tag{3.30}$$

On the other hand, we have the natural upper bound

$$\mathbb{P}(\mathcal{O}_r) \leq \left(\int_{\mathcal{O}_r} d\boldsymbol{\mu} \right) \sup_{\boldsymbol{\mu} \in \mathcal{O}_r} f_{\boldsymbol{\mu}}(\mu_1, \mu_2, \ldots, \mu_{\mathrm{m}}). \tag{3.31}$$

Note that the measure of $\mathcal{O}_r$ is independent of SNR. Therefore, combining (3.30) and (3.31), and taking the limit in SNR yields

$$\begin{aligned} \mathbb{P}(\mathcal{O}_r) &\doteq \sup_{\boldsymbol{\mu} \in \mathcal{O}_r} f_{\boldsymbol{\mu}}(\mu_1, \mu_2, \ldots, \mu_{\mathrm{m}}) \\ &\doteq \mathsf{SNR}^{-\inf_{\boldsymbol{\mu} \in \mathcal{O}_r} \sum_{k=1}^{\mathrm{m}}(2k-1+\mathrm{M}-\mathrm{m})\mu_k} \end{aligned} \tag{3.32}$$

*Note that the substitution $\lambda_k = \epsilon^{\mu_k}$ yields an expression similar to (3.29): $f_{\boldsymbol{\mu}}(\mu_1, \mu_2, \ldots, \mu_{\mathrm{m}}) \approx \epsilon^{\sum_{k=1}^{\mathrm{m}}(2k-1+\mathrm{M}-\mathrm{m})\mu_k}$ for $\epsilon \to 0$. In our framework, we set $\epsilon = \mathsf{SNR}^{-1}$ to obtain a parameterization w.r.t. SNR.

where last step follows from (3.29). This simple reasoning suggests that the outage probability can be asymptotically approximated in terms of the realization of $\boldsymbol{\mu}$ in the set $\mathcal{O}_r$ which is most likely to occur. Using large deviation techniques, both the approximation (3.29) and the asymptotic probability (3.32) have been rigorously derived by Zheng and Tse (2003). We summarize hereafter this result for later reference.

Theorem 3.3 (Zheng and Tse (2003))**.** *For the flat-fading channel, the outage probability satisfies*

$$\mathbb{P}(\mathcal{O}_r) \doteq \mathsf{SNR}^{-d_{\mathcal{O}}(r)} \tag{3.33}$$

where

$$d_{\mathcal{O}}(r) = \inf_{\boldsymbol{\mu}\in\mathcal{O}_r} \sum_{k=1}^{\mathrm{m}} (2k-1+\mathrm{M}-\mathrm{m})\mu_k. \tag{3.34}$$

Moreover, the solution to (3.34) is given by the piecewise linear curve connecting the points $(r, d_{\mathcal{O}}(r))$ *for* $r = 0, 1, \ldots, \mathrm{m}$*, and*

$$d_{\mathcal{O}}(r) = (\mathrm{M}-r)(\mathrm{m}-r). \tag{3.35}$$

There are many different realizations of the singularity vector $\boldsymbol{\mu}$ in the outage set $\mathcal{O}_r$, but Theorem 3.3 says that the outage probability is dominated (exponentially in SNR) by the probability corresponding to the $\boldsymbol{\mu}$ which minimizes the SNR exponent $\sum_{k=1}^{\mathrm{m}}(2k-1+\mathrm{M}-\mathrm{m})\mu_k$. We shall refer to the minimizing $\boldsymbol{\mu}$ as the dominant outage event $\boldsymbol{\mu}^\star$.

The solution (3.35) to the above optimization problem can be obtained as follows. Assume that the multiplexing rate r is an integer within $[0, \mathrm{m}]$. Then, $\boldsymbol{\mu}^\star$ can be shown to satisfy (Zheng and Tse, 2003)

$$\begin{aligned} \mu_k^\star &= 1, \quad k = 1, 2, \ldots, \mathrm{m}-r, \\ \mu_k^\star &= 0, \quad k = \mathrm{m}-r+1, \mathrm{m}-r+2, \ldots, \mathrm{m}, \end{aligned} \tag{3.36}$$

which is to say that $\mathrm{m}-r$ eigenvalues decay with SNR as SNR^{-1} and r eigenvalues are constant w.r.t. SNR. Intuitively, the typical way an outage occurs at multiplexing rate r is to have r constant eigenvalues while the remaining ones decay to zero with SNR. The SNR exponent corresponding to $\boldsymbol{\mu}^\star$ is now given by

$$\sum_{k=1}^{\mathrm{m}-r}(2k-1+\mathrm{M}-\mathrm{m}) = (\mathrm{M}-r)(\mathrm{m}-r)$$

which corresponds to (3.35). In the case of a non integer multiplexing rate, say $r' = r+\delta$ with $0 < \delta < 1$, $\boldsymbol{\mu}^\star$ can be shown to satisfy $\mu^\star_{\mathrm{m}-r} = 1-\delta$ and the remaining entries are still given by (3.36). In other words, for an additional δ in multiplexing rate, the $(\mathrm{m}-r)$th eigenvalue of the channel corresponding to the dominant outage event is adjusted by δ. Hence, the exponent $d_{\mathcal{O}}(r)$ varies linearly between integer values of multiplexing rate.

3.3.1. Geometric interpretation

Equation (3.35) has a nice geometric interpretation in terms of the rank of the random channel matrix $\mathbf{H}_0$. Let w.l.o.g. $\mathrm{m} = \mathrm{M_T}$ and define the set

$$\mathcal{M}_r = \left\{\mathbf{H}_0 \in \mathbb{C}^{\mathrm{M}\times\mathrm{m}} : \mathrm{rank}(\mathbf{H}_0) = r\right\}.$$

Any matrix in $\mathcal{M}_r$ has r independent columns with M entries and the remaining $\mathrm{m}-r$ columns are linear combinations thereof. Hence, such a matrix can be specified by $r\mathrm{M} + r(\mathrm{m}-r) = \mathrm{Mm} - (\mathrm{M}-r)(\mathrm{m}-r)$ parameters, which can be thought of as the dimensionality* of $\mathcal{M}_r$. Put differently, the matrix $\mathbf{H}_0$ will be close to $\mathcal{M}_r$ when the components of $\mathbf{H}_0$ lying in the $(\mathrm{M}-r)(\mathrm{m}-r)$ dimensions orthogonal to $\mathcal{M}_r$ are close to zero. Since $(\mathrm{M}-r)(\mathrm{m}-r) = d_{\mathcal{O}}(r)$ and the entries of $\mathbf{H}_0$ are i.i.d. $\mathcal{CN}(0,1)$, the probability of having $d_{\mathcal{O}}(r)$ entries in $\mathbf{H}_0$ of magnitude SNR^{-1} (i.e., close to zero for large SNR) is given by $\mathsf{SNR}^{-d_{\mathcal{O}}(r)}$. Thus, the outage probability $\mathsf{SNR}^{-d_{\mathcal{O}}(r)}$ can be interpreted as the probability that the random channel matrix is close to $\mathcal{M}_r$.

*$\mathcal{M}_r$ has a well-defined dimensionality. See Tse and Viswanath (2005) for more details.

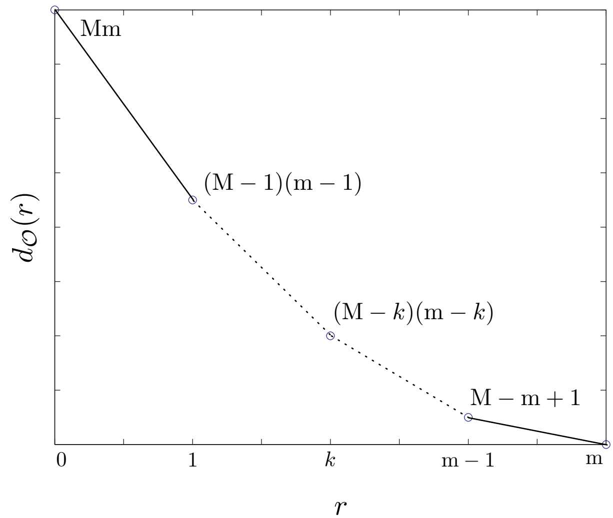

Fig. 3.1: Outage curve $d_{\mathcal{O}}(r)$ corresponding to a flat-fading MIMO channel with $\mathrm{m} = \min(\mathrm{M_T}, \mathrm{M_R})$ and $\mathrm{M} = \max(\mathrm{M_T}, \mathrm{M_R})$.

3.3.2. Visualization of the tradeoff

Figure 3.1 illustrates the optimal DM tradeoff $d_{\mathcal{O}}(r)$. The best error performance is obtained for multiplexing rate $r = 0$, i.e., $d_{\mathcal{O}}(0) = \mathrm{Mm}$, which basically says that maximum diversity gain can only to be achieved at fixed rates. On the other hand, at maximum multiplexing rate m, the system operates close to the ergodic capacity (3.2), and the optimal diversity is given by $d_{\mathcal{O}}(\mathrm{m}) = 0$, implying that an increase in SNR does not improve error performance. Between these extreme points, the outage curve $d_{\mathcal{O}}(r)$ is a strictly decreasing function of the multiplexing rate r: an increase in data rate comes at the expense of diversity and vice versa.

The relation between outage probability, rate, and SNR is illustrated

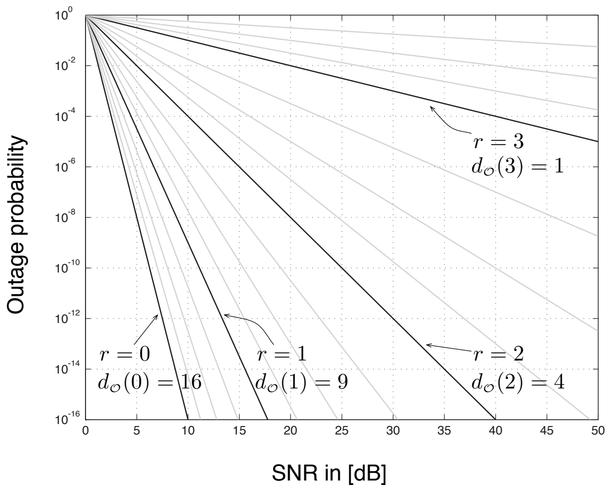

Fig. 3.2: Outage probability asymptotes for a flat-fading MIMO channel with $\mathrm{M_T} = 4$ and $\mathrm{M_R} = 4$. The asymptotes correspond to multiplexing rates separated by multiples of 0.25 and range from $r = 0$ to $r = 3.75$.

in Figure 3.2 for a system with $\mathrm{M_T} = 4$ and $\mathrm{M_R} = 4$. The asymptote corresponding to multiplexing rate $r = 0$ has the maximum diversity gain, i.e., outage probability decreases by 10^{-16} per decade; every increase in SNR is only used to improve reliability while keeping the data rate fixed. In contrast, when the multiplexing rate equals the maximum multiplexing gain $\mathrm{m} = 4$, an additional 3dB in SNR yields an increase of 4 bps/Hz in data rate but leaves the outage probability unchanged. More generally, every curve in Figure 3.2 corresponds to the asymptotic performance achieved by an optimal family of codes $\mathcal{C}_r$ with corresponding multiplexing rate. For any of the asymptotes at intermediate multiplexing rates $0 < r < \mathrm{m}$, an increase in SNR yields simultaneously an increase in data rate and an improvement in error performance. Both gains are balanced according to the DM tradeoff.

3.4. MULTIPLE-ACCESS CHANNEL

The DM tradeoff framework can be extended to analyze fundamental performance limits of communication over MA channels (Tse et al., 2004). In this section, we prepare the ground for our study of MA selective-fading channels in Chapter 6 by reviewing prior work.

We start by extending the input-output relation (3.1) to MA channels as follows:

$$\mathbf{Y} = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}} \sum_{u=1}^{U} \mathbf{H}_{u,0}\, \mathbf{X}_u + \mathbf{Z} \tag{3.37}$$

where the output matrix $\mathbf{Y}$ and the noise matrix $\mathbf{Z}$ are defined analogously to those in (3.1), but now the input consists of U codewords $\mathbf{X}_u = [\mathbf{x}_{u,0}\ \mathbf{x}_{u,1}\ \cdots\ \mathbf{x}_{u,N-1}]$, one for each user. For the sake of notation, we set

$$\mathbf{H}_{\mathcal{S},0} = [\mathbf{H}_{u_1,0}\, \mathbf{H}_{u_2,0}\ \cdots\ \mathbf{H}_{u_{|\mathcal{S}|},0}]$$
$$\mathbf{X}_{\mathcal{S}} = [\mathbf{X}_{u_1}^T\ \mathbf{X}_{u_2}^T\ \cdots\ \mathbf{X}_{u_{|\mathcal{S}|}}^T]^T$$

for any $\mathcal{S} = \{u_1, u_2, \ldots, u_{|\mathcal{S}|}\}$ with $\mathcal{S} \subseteq \mathcal{U} = \{1, 2, \ldots, U\}$. As in the point-to-point case, we focus on the case where the receiver has perfect CSI for all users, but the transmitters have only access to the fading law.

3.4.1. Achievable rate regions

In contrast to the point-to-point case where the maximum rate of reliable communication is determined by a single quantity, any rate tuple $(R_1, R_2, \ldots, R_U)$ that is achievable, i.e., for which there exist sequences of codes, one for each user, with vanishing error probabilities as block length increases, belongs to the so-called capacity region.

Capacity region

The first characterizations of the capacity region were reported for the MA additive white Gaussian noise (AWGN) channel in the early 1970's (Ahlswede, 1971; Liao, 1972; Wyner, 1974). Assuming momentarily that the channel matrix $\mathbf{H}_{\mathcal{U},0}$ is deterministic, the capacity region corresponding to the MA channel in (3.37) is given by the following theorem.

Theorem 3.4 (Cover and Thomas (1991))**.** *The capacity region of the U-user MA channel is the closure of the convex hull of the rate tuples $(R_1, R_2, \ldots, R_U)$ satisfying*

$$R(\mathcal{S}) = \sum_{u \in \mathcal{S}} R_u \leq \frac{1}{N} I(\mathbf{X}_{\mathcal{S}}; \mathbf{Y} \,|\, \mathbf{X}_{\bar{\mathcal{S}}}), \quad \forall \mathcal{S} \subseteq \mathcal{U} \tag{3.38}$$

for some product distribution $f_1(\mathbf{X}_1) \cdot f_2(\mathbf{X}_2) \cdot \cdots \cdot f_U(\mathbf{X}_U)$.

Equation (3.38) gives rise to $2^U - 1$ constraints, one for every subset of $\mathcal{U}$, and maximizing the corresponding mutual information over the input product distribution yields the boundaries of the capacity region. For a given set of users $\mathcal{S} \subseteq \mathcal{U}$, the largest boundary on $R(\mathcal{S})$ is obtained with Gaussian codebooks. In particular, let $\mathbf{x}_{u,n}$ be i.i.d. across slots n and independent across users $u \in \mathcal{S}$ with $\mathbf{x}_{u,n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_u)$, where $\mathbf{Q}_u$ is subject to the constraint $\mathrm{Tr}\,(\mathbf{Q}_u) \leq \mathrm{M_T}$. Then, the optimization problem reduces to

$$\max_{f_u(\mathbf{X}_u), u \in \mathcal{S}} \frac{1}{N} I(\mathbf{X}_{\mathcal{S}}; \mathbf{Y} \,|\, \mathbf{X}_{\bar{\mathcal{S}}}) = \max_{\substack{\mathbf{Q}_u \succeq 0, u \in \mathcal{S} \\ \mathrm{Tr}(\mathbf{Q}_u) \leq \mathrm{M_T}}} \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}} \sum_{u \in \mathcal{S}} \mathbf{H}_{u,0} \mathbf{Q}_u \mathbf{H}_{u,0}^H\right). \tag{3.39}$$

We note, however, that choosing $\{\mathbf{Q}_u\}$ so as to maximize the mutual information corresponding to the set $\mathcal{S}$ does not *necessarily* maximize the boundaries corresponding to the remaining sets of users. In general,

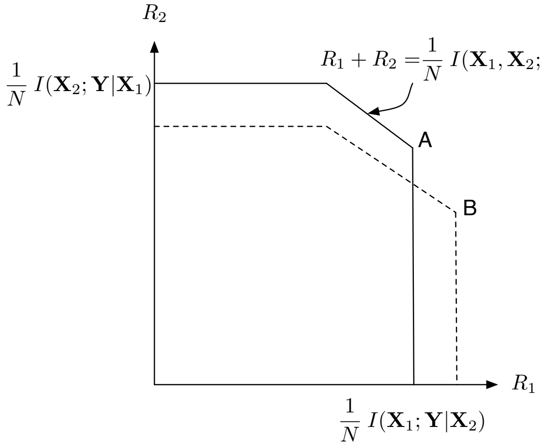

Fig. 3.3: Two achievable rate regions for the 2-user MA channel corresponding to different input distributions. Any rate tuple on the segment AB can be achieved by time-sharing.

there is no unique input distribution that optimizes all boundaries. The capacity region in Theorem 3.4 is therefore given by the convex hull of all the achievable rate regions obtained from admissible (w.r.t. an individual power constraint) input distributions. Taking the convex hull of all such regions is natural since convex combinations of rate tuples can be achieved by time-sharing. Figure 3.3 illustrates two possible rate regions for the 2-user MA channel arising from different input distributions. Rate tuples on the segment between the corner points A and B are achievable by time-sharing. In particular, if (R_1, R_2) is the rate tuple at A and (R'_1, R'_2) is that at B, then any rate tuple $(\lambda R_1 + (1-\lambda)R'_1, \lambda R_2 + (1-\lambda)R'_2)$ for $\lambda \in [0,1]$ is achievable by operating λ percent of the time at A and the rest at B.

Ergodic capacity region

When coding is performed across infinitely many realizations of the channel with input-ouput relation (3.37), the set of achievable rates tuples $(R_1, R_2, \ldots, R_U)$ is given by the ergodic capacity region with boundaries

$$R(\mathcal{S}) \leq \max_{\substack{\mathbf{Q}_u \succeq 0, u \in \mathcal{S} \\ \mathrm{Tr}(\mathbf{Q}_u) \leq \mathrm{M_T}}} \mathbb{E}\left\{ \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}} \sum_{u \in \mathcal{S}} \mathbf{H}_{u,0} \mathbf{Q}_u \mathbf{H}_{u,0}^H \right) \right\} \tag{3.40}$$

for $\mathcal{S} \subseteq \mathcal{U}$. The solution can be found by recalling the point-to-point case treated in Section 3.1.1 for which the optimal input covariance is given by $\mathbf{Q} = \mathbf{I}$. Put differently, the optimal strategy is to transmit independent signals across antennas. With such signal structure, the fact that the users cannot cooperate in the MA case becomes immaterial, and we can hence conclude that picking $\mathbf{Q}_u = \mathbf{I}$ for every $u \in \mathcal{U}$ is optimal (Telatar, 1999; Tse and Viswanath, 2005). Noteworthily, a unique input product distribution simultaneously maximizes all constraints. Upon inserting the optimal covariance matrices into (3.40), we obtain the following characterization of the ergodic capacity region in the high-SNR regime (Foschini, 1996)

$$R(\mathcal{S}) \leq \min(\mathrm{M_R}, |\mathcal{S}|\mathrm{M_T}) \log \mathsf{SNR} + O(1)\,,\ \mathcal{S} \subseteq \mathcal{U}. \tag{3.41}$$

3.4.2. Multiple-access system definitions

Paralleling Section 3.1.3, we invoke also in the MA case the concept of family of codes whereby user u employs a sequence of codebooks $\mathcal{C}_{r_u}(\mathsf{SNR})$, one for each SNR. The corresponding data rate scales with SNR as $R_u(\mathsf{SNR}) = r_u \log \mathsf{SNR}$, where $r_u \in [0, \mathrm{m}]$ is the multiplexing rate. The family of codes $\mathcal{C}_{r_u}$ is assumed to have block length N so that, at any given SNR, $\mathcal{C}_{r_u}(\mathsf{SNR})$ contains SNR^{Nr_u} codewords $\mathbf{X}_u$.

Since there are multiple users, the overall family of codes is given by

$$\mathcal{C}_{\boldsymbol{r}} = \mathcal{C}_{r_1} \times \mathcal{C}_{r_2} \times \cdots \times \mathcal{C}_{r_U}$$

where $\boldsymbol{r} = (r_1, r_2, \ldots, r_U)$ denotes the multiplexing rate tuple. At a given SNR, the corresponding codebook $\mathcal{C}_{\boldsymbol{r}}(\mathsf{SNR})$ contains $\mathsf{SNR}^{Nr(\mathcal{U})}$ codewords with $r(\mathcal{U}) = \sum_{u=1}^{U} r_u$. Invoking the high-SNR characterization of the ergodic capacity boundaries in (3.41), the sum of the multiplexing rates corresponding to the users in $\mathcal{S} \subseteq \mathcal{U}$ satisfies

$$r(\mathcal{S}) \triangleq \sum_{u \in \mathcal{S}} r_u \leq \min(\mathrm{M_R}, |\mathcal{S}|\mathrm{M_T}).$$

The DM tradeoff realized by $\mathcal{C}_{\boldsymbol{r}}$ is characterized by the function

$$d(\mathcal{C}_{\boldsymbol{r}}) = - \lim_{\mathsf{SNR} \to \infty} \frac{\log P_e(\mathcal{C}_{\boldsymbol{r}})}{\log \mathsf{SNR}} \tag{3.42}$$

where $P_e(\mathcal{C}_{\boldsymbol{r}})$ is the *total* error probability (that is, the probability for the receiver to make a detection error for at least one user) obtained through ML detection. The optimal DM tradeoff curve

$$d^{\star}(\boldsymbol{r}) = \sup_{\mathcal{C}_{\boldsymbol{r}}} d(\mathcal{C}_{\boldsymbol{r}}) \tag{3.43}$$

where the supremum is taken over all possible families of codes $\mathcal{C}_{\boldsymbol{r}}$, quantifies the maximum achievable diversity gain as a function of the multiplexing rate tuple $\boldsymbol{r}$.

3.4.3. Outage formulation

Since we are interested in characterizing ultimate performance limits over a single realization of the channel with input-output relation (3.37), we shall use the concept of outage capacity (Ozarow et al., 1994; Telatar, 1999; Tse et al., 2004) as we did previously in the point-to-point case.

Conditional on a realization of the fading channel matrix $\mathbf{H}_{\mathcal{U},0}$, the set of rates that are achievable over the channel (3.37) for perfect CSI at the receiver only are characterized by Theorem 3.4. As in the point-to-point case, it can be shown that one can restrict the input to be i.i.d. Gaussian without loss of optimality (Tse et al., 2004). Hence, the set of achievable rate tuples conditional on the channel realization $\mathbf{H}_{\mathcal{U},0}$ satisfies the boundary constraints

$$R(\mathcal{S}) \leq \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}} \sum_{u\in\mathcal{S}} \mathbf{H}_{u,0}\mathbf{H}_{u,0}^H\right), \ \forall\mathcal{S} \subseteq \mathcal{U}. \quad (3.44)$$

If a given rate tuple $(R_1, R_2, \ldots, R_U)$ lies outside the region defined by (3.44), we say that the channel is in outage w.r.t. this rate tuple and denote the corresponding outage probability by $P_{\text{out}}(R_1, R_2, \ldots, R_U)$. Letting the users' rates scale with SNR as $R_u(\mathsf{SNR}) = r_u \log \mathsf{SNR}$, for all u, the outage probability satisfies

$$P_{\text{out}}(r_1 \log \mathsf{SNR}, r_2 \log \mathsf{SNR}, \ldots, r_U \log \mathsf{SNR}) \doteq \mathbb{P}(\mathcal{O}_{\boldsymbol{r}}) \quad (3.45)$$

where $\mathcal{O}_{\boldsymbol{r}}$ shall be referred to as the outage event corresponding to the multiplexing rate tuple $\boldsymbol{r}$. Before characterizing $\mathcal{O}_{\boldsymbol{r}}$ more precisely, we note that, as in the point-to-point case, the outage probability constitutes a fundamental performance bound on any family of codes.

Theorem 3.5 (Tse et al. (2004)). *Assuming an ML receiver, the detection error probability of any family of codes $\mathcal{C}_{\boldsymbol{r}}$ satisfies*

$$d^\star(\boldsymbol{r}) \leq -\lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}(\mathcal{O}_{\boldsymbol{r}})}{\log \mathsf{SNR}} \quad (3.46)$$

This result suggests that, irrespectively of the family of codes $\mathcal{C}_{\boldsymbol{r}}$ employed for communication, detection errors are very likely to occur whenever the channel is in outage with respect to the multiplexing rate tuple $\boldsymbol{r}$.

Exploiting the structure of the rate region in (3.44), the outage probability is naturally decomposed as follows

$$\mathbb{P}(\mathcal{O}_{\boldsymbol{r}}) = \mathbb{P}\left(\bigcup_{\mathcal{S}\subseteq\mathcal{U}} \mathcal{O}_{\mathcal{S}}\right) \quad (3.47)$$

where the $\mathcal{S}$-outage event $\mathcal{O}_{\mathcal{S}}$ can be rigorously defined in terms of the singularity levels of the channel matrices $\{\mathbf{H}_{u,0}\}$ just like the outage event corresponding to the point-to-point case was defined in (3.14). For our purposes, it is sufficient to note that the $\mathcal{S}$-outage probability satisfies

$$\mathbb{P}(\mathcal{O}_{\mathcal{S}}) \doteq \mathbb{P}\left(\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}} \sum_{u\in\mathcal{S}} \mathbf{H}_{u,0}\mathbf{H}_{u,0}^H\right) < R(\mathcal{S})\right) \quad (3.48)$$

$$\doteq \mathsf{SNR}^{-d_{\mathcal{S}}(r(\mathcal{S}))} \quad (3.49)$$

where last step follows upon observing that (3.48) is equivalent to the outage probability of a point-to-point flat-fading channel with $|\mathcal{S}|\mathrm{M_T}$ transmit and $\mathrm{M_R}$ receive antennas and, hence, invoking Theorem 3.3, the corresponding tradeoff curve is the piecewise linear function determined by

$$d_{\mathcal{S}}(r) = (\mathrm{M_R} - r)(|\mathcal{S}|\mathrm{M_T} - r), \qquad r = 0, 1, \ldots, \min(\mathrm{M_R}, |\mathcal{S}|\mathrm{M_T}). \quad (3.50)$$

We conclude our review of the DM tradeoff in flat-fading MA channels with the following result which characterizes the multiplexing rate region, denoted by $\mathcal{R}(d)$, where all the users are guaranteed a certain level of diversity d.

Theorem 3.6 (Tse et al. (2004))**.** *If the block length satisfies $N \geq |\mathcal{S}|\mathrm{M_T} + \mathrm{M_R} - 1$, then there exist families of codes $\mathcal{C}_{\boldsymbol{r}}$ for which the multiplexing rate region is given by*

$$\mathcal{R}(d) = \left\{\boldsymbol{r} : r(\mathcal{S}) \leq r_{\mathcal{S}}(d), \forall\mathcal{S} \subseteq \mathcal{U}\right\} \quad (3.51)$$

where $r_{\mathcal{S}}(d)$ denotes the inverse of the function $d_{\mathcal{S}}(r)$.

Theorem 3.6 says that the outage bound (3.46) is achievable. The proof technique, analogous to that employed in the point-to-point case, establishes the existence of optimal codes in the Gaussian ensemble of codes. Recent work (Nam and El Gamal, 2007; Badr and Belfiore, 2008b) provides explicit code constructions that achieve the optimal tradeoff.

3.5. BEYOND FLAT FADING

The framework presented in this chapter allows to efficiently characterize the tradeoff between rate and reliability in MIMO fading channels, and the outage bound, both in the point-to-point and multiple-access cases, constitutes the ultimate performance limit for systems with fixed block lengths. However, the proof techniques used in (Zheng and Tse, 2003) to evaluate the outage probability and to establish the existence of optimal codes are not directly applicable to the more general class of selective-fading channels. More precisely, 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 terms in the expression for the mutual information. 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. In the next chapter, we present a technique that formalizes this idea. We note that the ideas at the heart of our technique have also proved to be useful in examining performance limits for the relay channel (Akçaba et al., 2007).

CHAPTER 4

Performance Limits in Selective-Fading Channels

THE DIVERSITY-MULTIPLEXING (DM) tradeoff framework presented in Chapter 3 allows to efficiently characterize the rate-reliability tradeoff over multiple-input multiple-output (MIMO) flat-fading channels. The essence of the approach is to study the interrelation between the data rate and the error probability of a family of codes parametrized by SNR. As the SNR increases, a precise characterization of the tradeoff emerges. The outage probability, which constitutes the ultimate limit on the performance of a family of codes, depends on the distribution of the fading channel. For this reason, the results on the DM tradeoff summarized in Chapter 3 do only hold in the flat-fading case.

In environments that are subject to temporal variations and/or multipath propagation, the flat-fading channel model fails to capture the channel's inherent memory in frequency and time. In this case, the wireless channel is more accurately described in terms of the selective-fading channel model presented in Chapter 2. Since real world wireless communication systems operate over channels that are selective both in frequency and in time, identifying the corresponding performance limits is pertinent. However, a characterization of the optimal DM tradeoff of this class of channels along with a code design criterion capable of guaranteeing optimal performance remain to be found.

In this chapter, we present a technique that solves this problem by providing exponentially tight upper and lower bounds on the outage probability of coherent selective-fading MIMO channels. The idea underlying our approach is to analyze the "Jensen channel" associated with 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 is equivalent to the original channel in the sense of the DM tradeoff and lends itself nicely to analytical treatment. In addition, our technique provides an explicit code design criterion to guarantee DM tradeoff optimality in selective-fading MIMO channels. Our criterion ties in perfectly well with other criteria reported in prior work.

The rest of this chapter is organized as follows. Building on the selective-fading channel model examined in Chapter 2, we first introduce the system model that we shall adopt throughout the chapter. Section 4.2 presents a rigorous formulation of the problem at hand in terms of the DM tradeoff framework. In Section 4.3, we introduce the "Jensen channel", which is the principal tool employed to derive the optimal tradeoff, and other related concepts such as the Jensen outage event and the corresponding probability. We will then be able to derive the optimal DM tradeoff in Section 4.4 by providing a code design criterion that guarantees the error probability to be exponentially tight to the Jensen outage probability. Section 4.5 concludes this chapter by discussing the relation of our code design criterion with prior work.

4.1. SIGNAL AND CHANNEL MODEL

The system model presented below is based on the considerations made in Chapter 2. We consider a point-to-point channel with M_T transmit and M_R receive antennas. The corresponding input-ouput relation is given by

$$\mathbf{y}_n = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}}\,\mathbf{H}_n\mathbf{x}_n + \mathbf{z}_n, \quad n = 0, 1, \ldots, N-1 \tag{4.1}$$

where the index $n = \mathcal{M}(m, k)$ denotes a time, frequency, or time-frequency slot according to the mapping defined in (2.14), and SNR denotes the average signal-to-noise ratio at each receive antenna. The vectors $\mathbf{y}_n$, $\mathbf{x}_n$, and $\mathbf{z}_n$ denote, respectively, the $\mathrm{M_R} \times 1$ receive signal vector, the $\mathrm{M_T} \times 1$ transmit signal vector, and the $\mathrm{M_R} \times 1$ zero-mean jointly proper Gaussian (JPG) noise vector satisfying $\mathbb{E}\{\mathbf{z}_n\mathbf{z}_{n'}^H\} = \delta_{n,n'}\,\mathbf{I}$. We also define the matrix $\mathbf{Y} = [\mathbf{y}_0\,\mathbf{y}_1\;\cdots\;\mathbf{y}_{N-1}]$, and assume that at any SNR level the codebook $\mathcal{C}_r(\mathsf{SNR})$ contains SNR^{Nr} codewords $\mathbf{X} = [\mathbf{x}_0\,\mathbf{x}_1\;\cdots\;\mathbf{x}_{N-1}]$ that satisfy the power constraint

$$\|\mathbf{X}\|_{\mathrm{F}}^2 \le N\mathrm{M_T}, \forall\,\mathbf{X} \in \mathcal{C}_r(\mathsf{SNR}). \tag{4.2}$$

This constraint implies that the codewords of any (with respect to (w.r.t.) SNR) codebook $\mathcal{C}_r(\mathsf{SNR})$ lie inside a sphere of radius $\sqrt{N\mathrm{M_T}}$ in $\mathbb{C}^{\mathrm{M_T}\times N}$ centered at the origin. Because the sphere radius is constant w.r.t. SNR so are the entries of $\mathcal{C}_r(\mathsf{SNR})$. Yet the codeword difference matrix $\mathbf{E} = \mathbf{X} - \mathbf{X}'$, where $\mathbf{X}, \mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})$, does in general depend on SNR because the number of codewords packed within the sphere increases as $|\mathcal{C}_r(\mathsf{SNR})| = \mathsf{SNR}^{Nr}$ for the rate $R(\mathsf{SNR}) = r \log \mathsf{SNR}$ to be sustained.

In the case $N = 1$ and $\mathrm{M_T} = 1$, for example, an admissible $\mathcal{C}_r$ would be a family of quadrature amplitude modulation (QAM) constellations with $|\mathcal{C}_r(\mathsf{SNR})| = \mathsf{SNR}^r$ points satisfying the constraint $x^2 \le 1$. Note that the minimum distance in such a family of codes scales as* $d_{\min}^2 \doteq \mathsf{SNR}^{-r}$, i.e., the area of the unit disk divided by the number of constellation points in $\mathcal{C}_r(\mathsf{SNR})$.

*A discussion of the DM tradeoff properties of QAM constellations for the scalar Rayleigh fading channel can be found in (Tse and Viswanath, 2005, Section 9.1.2)

We assume that the channel is spatially uncorrelated, but correlated across slots. Specifically, we let

$$\mathbb{E}\{\mathbf{H}_n(i,j)(\mathbf{H}_{n'}(i',j'))^*\} = \mathbf{R}_{\mathbb{H}}(n,n')\,\delta_{i,i'}\,\delta_{j,j'}, \\ i,i'=1,2,\dots,\mathrm{M_R},\; j,j'=1,2,\dots,\mathrm{M_T} \\ n,n'=0,1,\dots,N-1. \quad (4.3)$$

The covariance matrix $\mathbf{R}_{\mathbb{H}}$, for which we define $\rho \triangleq \operatorname{rank}(\mathbf{R}_{\mathbb{H}})$, follows from the time-frequency correlation function in (2.17). Finally, we stack the channel matrices across slots according to

$$\mathbf{H} = [\mathbf{H}_0\,\mathbf{H}_1\;\cdots\;\mathbf{H}_{N-1}]. \quad (4.4)$$

With this notation, we have

$$\mathbb{E}\big\{\mathrm{vec}(\mathbf{H})(\mathrm{vec}(\mathbf{H}))^H\big\} = \mathbf{R}_{\mathbb{H}} \otimes \mathbf{I}_{\mathrm{M_T}\mathrm{M_R}}. \quad (4.5)$$

We note that the input-output relation (4.1) can also be used to describe communication over linear time-invariant and linear frequency-invariant channels. We refer the reader to Appendix A for a justification of this statement. Unless explicitly indicated otherwise, we shall assume that the fading channel underlying (4.1) belongs to the class of channels which are approximately diagonalized by a suitably chosen Weyl-Heisenberg basis as explained in Chapter 2.

4.2. PROBLEM FORMULATION

We consider the situation where the receiver has perfect channel state information (CSI) while the transmitter does not have CSI but knows the channel law. Given a target outage probability, the transmitter can determine the maximum rate that is supported by the channel and optimize the family of codes $\mathcal{C}_r$ accordingly. In the high-SNR regime, the relation between outage probability and data rate for selective-fading MIMO channels is obtained as follows.

Optimal input distribution

We assume without loss of optimality a Gaussian input distribution so that the input codeword $\mathbf{X}$ satisfies $\mathrm{vec}(\mathbf{X}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ with average* power constraint $\mathrm{Tr}\,(\mathbf{Q}) \leq N\mathrm{M_T}$. Then, the outage probability follows by optimizing over the input covariance as

$$P_{\text{out}}(R) = \inf_{\substack{\mathbf{Q}\succeq 0 \\ \mathrm{Tr}(\mathbf{Q})\leq N\mathrm{M_T}}} \mathbb{P}\left(\frac{1}{N}\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\,\mathbf{D_H}\mathbf{Q}\,\mathbf{D}_{\mathbf{H}}^{H}\right) < R\right) \tag{4.6}$$

where $\mathbf{D_H} = \mathrm{diag}\{\mathbf{H}_n\}_{n=0}^{N-1}$. The optimal covariance matrix can be explicitly obtained using the following arguments.

First, any $\mathbf{Q}$ that is admissible under the power constraint yields a lower bound to the maximum mutual information. In particular, by picking $\mathbf{Q} = \mathbf{I}$, we get the following upper bound to the outage probability:

$$P_{\text{out}}(R) \leq \mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1}\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\,\mathbf{H}_n\mathbf{H}_n^{H}\right) < R\right). \tag{4.7}$$

On the other hand, due to the power constraint, $\mathbf{Q}$ must satisfy the positive semidefinite ordering $\mathbf{Q} \preceq N\mathrm{M_T}\mathbf{I}$. Since the function $f(\mathbf{A}) = \log\det(\mathbf{I} + \mathbf{A})$ is increasing over the cone of positive semidefinite matrices (S. Boyd and L. Vandenberghe, 2004), we obtain the lower bound

$$P_{\text{out}}(R) \geq \mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1}\log\det\left(\mathbf{I} + \mathsf{SNR}\,N\,\mathbf{H}_n\mathbf{H}_n^{H}\right) < R\right). \tag{4.8}$$

The bounds in (4.7) and (4.8) are equal up to a factor that is independent

*This average power constraint, which is a relaxation of the per-codeword constraint (4.2), serves only the purpose of establishing a bound on performance. It will later become manifest that the bound we obtain is still achievable if we restrict our attention to the subset of codes satisfying (4.2).

of SNR and, therefore, in the high SNR limit we have

$$\begin{aligned}
&\lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}\Big(\frac{1}{N}\sum_{n=0}^{N-1} \log\det\Big(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\,\mathbf{H}_n\mathbf{H}_n^H\Big) < R\Big)}{\log \mathsf{SNR}} \\
&= \lim_{\mathsf{SNR}\to\infty} \frac{\log P_{\text{out}}(R)}{\log \mathsf{SNR}} \\
&= \lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}\Big(\frac{1}{N}\sum_{n=0}^{N-1} \log\det\big(\mathbf{I} + \mathsf{SNR}\,N\,\mathbf{H}_n\mathbf{H}_n^H\big) < R\Big)}{\log \mathsf{SNR}}.
\end{aligned}$$

We can thus conclude that an i.i.d. (across space and slots) Gaussian input distribution achieves the optimal outage performance in the scale of interest and, consequently, the outage probability satisfies the exponential equality

$$P_{\text{out}}(R) \doteq \mathbb{P}(\mathrm{I}(\mathsf{SNR}) < R) \tag{4.9}$$

where

$$\mathrm{I}(\mathsf{SNR}) = \frac{1}{N}\sum_{n=0}^{N-1} \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\,\mathbf{H}_n\mathbf{H}_n^H\right). \tag{4.10}$$

Maximum multiplexing rate

The multiplexing rate r is at most $\mathrm{m} = \min(\mathrm{M_T}, \mathrm{M_R})$ as the following argument shows. By the same line of reasoning as in flat fading (Telatar, 1999), we know that for selective fading a uniform input power allocation is optimal in the sense of the ergodic capacity (Bölcskei et al., 2002a; Tse and Viswanath, 2005). We therefore have

$$\begin{aligned}
C(\mathsf{SNR}) &= \mathbb{E}\left\{\frac{1}{N}\sum_{n=0}^{N-1} \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{H}_n\mathbf{H}_n^H\right)\right\} \\
&= \mathbb{E}\left\{\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{H}_m\mathbf{H}_m^H\right)\right\}
\end{aligned} \tag{4.11}$$

where m can be taken to be any slot as the fading process is stationary. It follows from our discussion in Section 3.1.1 that (4.11) satisfies

$$C(\mathsf{SNR}) = \mathrm{m}\ \log \mathsf{SNR} + O(1)$$

which, recalling Definition 3.1, means that the multiplexing rate r cannot exceed m. In general, the multiplexing rate satisfies $r \in [0, \mathrm{m}]$.

Outage event

As in the point-to-point case, the outage event can be characterized in terms of the "level of singularity" of the channel matrices. To this end, let the eigenvalues of the channel matrix corresponding to slot n ($n = 0, 1, \ldots, N-1$) be given in increasing order by

$$\lambda_k(\mathbf{H}_n \mathbf{H}_n^H), \quad k = 1, 2, \ldots, \mathrm{m} \tag{4.12}$$

and define the singularity levels (in descending order across index k)

$$\mu_k(n) = -\frac{\log \lambda_k(\mathbf{H}_n \mathbf{H}_n^H)}{\log \mathsf{SNR}}, \quad k = 1, 2, \ldots, \mathrm{m}. \tag{4.13}$$

The mutual information in (4.10) can now be written as

$$\begin{aligned} \mathrm{I}(\mathsf{SNR}) &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=1}^{\mathrm{m}} \log\left(1 + \frac{\mathsf{SNR}}{\mathrm{M_T}} \lambda_k(\mathbf{H}_n \mathbf{H}_n^H)\right) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \log\left(\prod_{k=1}^{\mathrm{m}} \left(1 + \frac{\mathsf{SNR}}{\mathrm{M_T}} \mathsf{SNR}^{-\mu_k(n)}\right)\right) \end{aligned} \tag{4.14}$$

where last equality follows from the variable change (4.13). At high SNR, we have

$$\left(1 + \frac{\mathsf{SNR}}{\mathrm{M_T}} \mathsf{SNR}^{-\mu_k(n)}\right) \doteq \mathsf{SNR}^{[1-\mu_k(n)]^+}$$

which, upon inserting into (4.14), yields

$$\mathrm{I}(\mathsf{SNR}) = \left(\frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=1}^{\mathrm{m}} [1 - \mu_k(n)]^+ \right) \log \mathsf{SNR} + O(1) \,. \quad (4.15)$$

The dependence of the outage probability (4.9) on the singularity levels can now be made explicit. Using (4.15) in (4.9) and letting the data rate scale with SNR as $R(\mathsf{SNR}) = r \log \mathsf{SNR}$, we get

$$P_{\text{out}}(r \log \mathsf{SNR}) \doteq \mathbb{P}(\mathcal{O}_r) \quad (4.16)$$

where the outage event* $\mathcal{O}_r$ is given by

$$\mathcal{O}_r = \Bigg\{ \boldsymbol{\mu}(n) \in \mathbb{R}_+^{\mathrm{m}}, n = 0, 1, \ldots, N-1 : \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=1}^{\mathrm{m}} [1 - \mu_k(n)]^+ < r \Bigg\}. \quad (4.17)$$

with $\boldsymbol{\mu}(n) \triangleq [\mu_1(n) \; \mu_2(n) \; \ldots \; \mu_{\mathrm{m}}(n)]$. The outage probability (4.16) constitutes a lower bound to the error probability of any family of codes† $\mathcal{C}_r$, which is to say that the optimal DM tradeoff satisfies

$$d^\star(r) \le d_{\mathcal{O}}(r) \quad (4.18)$$

where we recall that

$$d_{\mathcal{O}}(r) = - \lim_{\mathsf{SNR} \to \infty} \frac{\log \mathbb{P}(\mathcal{O}_r)}{\log \mathsf{SNR}}. \quad (4.19)$$

The outage event (4.17) depends upon the singularity levels corresponding to the individual channel matrices. In order to *directly* compute the

*We note that (4.17) is a generalization of (3.14) to the case of selective-fading channels, and it can be proved using the same arguments as in the proof of (3.14) given by (Zheng and Tse, 2003).

†The proof for this statement is omitted as it proceeds along the lines of the proof of Theorem 3.1.

outage probability, one would need to know the joint distribution of the singularity levels. Unfortunately, the latter is in general unknown for selective-fading channels. In the next sections, we develop a technique that bypasses this difficulty.

4.3. ANALYSIS TOOL

Jensen's inequality (Jensen, 1906), perhaps one of the most widely used inequalities in mathematics and information theory, states that if the real function $\varphi(x)$ is continuous and convex on some interval, then

$$\varphi\left(\frac{\sum_n a_n x_n}{\sum_n a_n}\right) \leq \frac{\sum_n a_n \varphi(x_n)}{\sum_n a_n} \tag{4.20}$$

where the real numbers $\{x_n\}$ do all belong to the interval, and the $\{a_n\}$ are arbitrary positive numbers. This simple inequality underlies the forthcoming analysis of the DM tradeoff in selective-fading MIMO channels as it gives rise to the concept of "Jensen channel".

4.3.1. The Jensen channel

Consider the mutual information in (4.10) and note that the matrix-valued arguments of the function $\log\det(\cdot)$ are positive definite. We can hence invoke the concavity of the function $\log\det(\cdot)$ over the convex cone of positive definite matrices (e.g. Horn and Johnson, 1993, 7.6.7) to get the following variant of Jensen's inequality

$$\mathrm{I}(\mathsf{SNR}) \leq \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T} N}\sum_{n=0}^{N-1}\mathbf{H}_n\mathbf{H}_n^H\right). \tag{4.21}$$

The upper bound to $\mathrm{I}(\mathsf{SNR})$ can be interpreted as the mutual information of an abstract channel that is *associated* with the original channel. This interpretation motivates the following definition.

Definition 4.1. The Jensen channel is the $\mathrm{m} \times N\mathrm{M}$ matrix given by

$$\boldsymbol{\mathcal{H}} = \begin{cases} [\mathbf{H}_0\ \mathbf{H}_1\ \cdots\ \mathbf{H}_{N-1}], & \text{if } \mathrm{M_R} \leq \mathrm{M_T}, \\ [\mathbf{H}_0^H\ \mathbf{H}_1^H\ \cdots\ \mathbf{H}_{N-1}^H], & \text{if } \mathrm{M_R} > \mathrm{M_T}. \end{cases} \tag{4.22}$$

A general upper bound to the mutual information can now be written in terms of the Jensen channel $\boldsymbol{\mathcal{H}}$ as

$$\mathrm{I}(\mathsf{SNR}) \leq \mathrm{J}(\mathsf{SNR}) = \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}N}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^H\right). \tag{4.23}$$

The operational significance of the Jensen channel will be established in Section 4.4. We shall first introduce and characterize analytically the outage probability corresponding to this channel.

4.3.2. The Jensen outage probability

In the following, we say that a Jensen outage occurs if the Jensen channel $\boldsymbol{\mathcal{H}}$ is in outage w.r.t. the target data rate R, i.e., $\boldsymbol{\mathcal{H}}$ belongs to the set $\{\boldsymbol{\mathcal{H}} : \mathrm{J}(\mathsf{SNR}) < R\}$. In analogy to the outage probability, we define the Jensen outage probability as

$$P_{\mathrm{J}}(R) \triangleq \mathbb{P}(\mathrm{J}(\mathsf{SNR}) < R)\,. \tag{4.24}$$

Since $\mathrm{J}(\mathsf{SNR})$ is an upper bound to the mutual information, the set of channel realizations in Jensen outage is a subset of the realizations that are in outage. This observation, illustrated in Figure 4.1, yields trivially

$$P_{\mathrm{J}}(R) \leq P_{\mathrm{out}}(R). \tag{4.25}$$

In the sequel, we will show that both probabilities are in fact exponentially equal, and hence the Jensen outage probability constitutes a fundamental performance limit in the high-SNR regime. Before establishing this somewhat surprising result, we shall develop an analytical characterization of the Jensen outage probability.

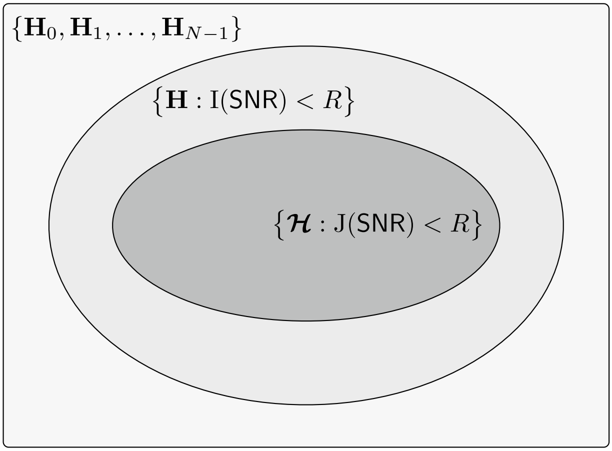

Fig. 4.1: The set of channel matrices $\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1}$ in Jensen outage is a subset of the channel matrices in outage.

High-SNR characterization

Using (4.5), we show in Appendix B that the Jensen channel satisfies

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_w(\mathbf{R}^{T/2} \otimes \mathbf{I}_{\mathrm{M}}) \tag{4.26}$$

where $\mathbf{R} = \mathbf{R}_{\mathbb{H}}$, if $\mathrm{M_R} \leq \mathrm{M_T}$, and $\mathbf{R} = \mathbf{R}_{\mathbb{H}}^T$, if $\mathrm{M_R} > \mathrm{M_T}$. The matrix $\boldsymbol{\mathcal{H}}_w$ is the i.i.d. $\mathcal{CN}(0,1)$ matrix with the same dimensions as $\boldsymbol{\mathcal{H}}$ given by

$$\boldsymbol{\mathcal{H}}_w = \begin{cases} [\mathbf{H}_{w,0}\ \mathbf{H}_{w,1}\ \ldots\ \mathbf{H}_{w,N-1}], & \text{if } \mathrm{M_R} \leq \mathrm{M_T}, \\ [\mathbf{H}_{w,0}^H\ \mathbf{H}_{w,1}^H\ \ldots\ \mathbf{H}_{w,N-1}^H], & \text{if } \mathrm{M_R} > \mathrm{M_T}. \end{cases} \tag{4.27}$$

Here, the $\mathbf{H}_{w,n}$ denote i.i.d. $\mathcal{CN}(0,1)$ matrices of dimension $\mathrm{M_R} \times \mathrm{M_T}$. Consider the eigenvalue decomposition $\mathbf{R}_{\mathbb{H}} = \mathbf{V}\boldsymbol{\Theta}\mathbf{V}^H$, where $\mathbf{V}$ is unitary and $\boldsymbol{\Theta}$ is given by

$$\boldsymbol{\Theta} = \mathrm{diag}\{\lambda_1(\mathbf{R}_{\mathbb{H}}), \lambda_2(\mathbf{R}_{\mathbb{H}}), \ldots, \lambda_\rho(\mathbf{R}_{\mathbb{H}}),\ 0, \ldots, 0\}\,.$$

Using the fact that $\boldsymbol{\mathcal{H}}_w \mathbf{U} \sim \boldsymbol{\mathcal{H}}_w$ for any unitary matrix $\mathbf{U}$ (Muirhead, 1982), and noting that $\mathbf{R}_{\mathbb{H}}$ and $\mathbf{R}_{\mathbb{H}}^T$ have the same eigenvalues, we get

$$\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^H \sim \boldsymbol{\mathcal{H}}_w(\boldsymbol{\Theta} \otimes \mathbf{I}_{\mathrm{M}})\boldsymbol{\mathcal{H}}_w^H$$

which, upon inserting into (4.23), yields

$$\mathsf{J}(\mathsf{SNR}) \sim \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}N}\boldsymbol{\mathcal{H}}_w(\boldsymbol{\Theta} \otimes \mathbf{I}_{\mathrm{M}})\,\boldsymbol{\mathcal{H}}_w^H\right).$$

Next, observe that the following positive semidefinite ordering holds

$$\begin{aligned}\lambda_1(\mathbf{R}_{\mathbb{H}})\,\mathrm{diag}\{\mathbf{I}_{\rho\mathrm{M}}, \mathbf{0}\} &\preceq \boldsymbol{\Theta} \otimes \mathbf{I}_{\mathrm{M}} \\ &\preceq \lambda_\rho(\mathbf{R}_{\mathbb{H}})\,\mathrm{diag}\{\mathbf{I}_{\rho\mathrm{M}}, \mathbf{0}\}\,. \end{aligned} \tag{4.28}$$

Since $f(\mathbf{A}) = \log\det(\mathbf{I} + \mathbf{A})$ is increasing over the cone of positive semidefinite matrices (S. Boyd and L. Vandenberghe, 2004), we get the following bounds on the Jensen outage probability

$$\begin{aligned}&\mathbb{P}\left(\log\det\left(\mathbf{I} + \lambda_\rho(\mathbf{R}_{\mathbb{H}})\frac{\mathsf{SNR}}{\mathrm{M_T}N}\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\right) < R\right) \\ &\le P_{\mathrm{J}}(R) \\ &\le \mathbb{P}\left(\log\det\left(\mathbf{I} + \lambda_1(\mathbf{R}_{\mathbb{H}})\frac{\mathsf{SNR}}{\mathrm{M_T}N}\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\right) < R\right)\end{aligned} \tag{4.29}$$

where $\overline{\boldsymbol{\mathcal{H}}}_w = \boldsymbol{\mathcal{H}}_w([1:\mathrm{m}], [1:\rho\mathrm{M}])$. Letting the data rate scale with SNR as $R(\mathsf{SNR}) = r\log\mathsf{SNR}$ and taking the exponential limit in SNR in (4.29), it readily follows that

$$P_{\mathrm{J}}(r\log\mathsf{SNR}) \doteq \mathbb{P}\left(\log\det\left(\mathbf{I} + \mathsf{SNR}\,\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\right) < r\log\mathsf{SNR}\right). \tag{4.30}$$

Analogously to (4.13), we define the singularity levels

$$\alpha_k = -\frac{\log\lambda_k(\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H)}{\log\mathsf{SNR}}, \quad k = 1, 2, \ldots, \mathrm{m} \tag{4.31}$$

and let $\boldsymbol{\alpha} \triangleq [\alpha_1 \; \alpha_2 \; \ldots \; \alpha_{\mathrm{m}}]$. With this notation, we can deduce from Theorem 3.3 that the Jensen outage probability satisfies the asymptotic equality

$$P_{\mathrm{J}}(r \log \mathsf{SNR}) \doteq \mathbb{P}(\mathcal{J}_r) \tag{4.32}$$

where $\mathcal{J}_r$ is the counterpart to the outage event $\mathcal{O}_r$ defined in (4.17) and is given by

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

It is now natural to define the Jensen DM tradeoff curve as

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

Based on the exponential equalities (4.30) and (4.32), it follows immediately that $d_{\mathcal{J}}(r)$ is nothing but the DM tradeoff curve of an effective MIMO channel with $\rho\mathrm{M}$ transmit and m receive antennas. We can therefore directly apply Theorem 3.3 to infer that the Jensen DM tradeoff curve is the piecewise linear function connecting the points $(r, d_{\mathcal{J}}(r))$, where

$$d_{\mathcal{J}}(r) = (\rho\mathrm{M} - r)(\mathrm{m} - r), r = 0, 1, \ldots, \mathrm{m}. \tag{4.33}$$

Since, as already noted, $P_{\mathrm{J}}(R) \leq P_{\mathrm{out}}(R)$, the exponential equalities (4.16) and (4.32) imply

$$\mathbb{P}(\mathcal{J}_r) \dot{\leq} \mathbb{P}(\mathcal{O}_r) \tag{4.34}$$

which, in conjunction with (4.18) and the definition in (3.7), eventually yields

$$d(\mathcal{C}_r) \leq d^{\star}(r) \leq d_{\mathcal{O}}(r) \leq d_{\mathcal{J}}(r), \quad r \in [0, \mathrm{m}] \tag{4.35}$$

for any family of codes $\mathcal{C}_r$. As evidenced by (4.35), the Jensen DM tradeoff curve $d_{\mathcal{J}}(r)$ provides an upper bound to the best achievable

DM tradeoff $d^{\star}(r)$, which is to say that at high SNR the Jensen outage probability yields a lower bound on error performance. In contrast with the outage curve $d_{\mathcal{O}}(r)$, however, $d_{\mathcal{J}}(r)$ can be easily computed for selective-fading MIMO channels.

In addition to its analytical tractability, $d_{\mathcal{J}}(r)$ is also relevant in an operational sense as it constitutes the best *achievable* performance over all possible families of codes. Indeed, the optimal DM tradeoff curve $d^{\star}(r)$ will be obtained in the next section by deriving a sufficient condition on $\mathcal{C}_r$ to guarantee that $d(\mathcal{C}_r) = d_{\mathcal{J}}(r)$ and hence, by (4.35), that $d^{\star}(r) = d_{\mathcal{J}}(r)$. Code constructions that satisfy this condition will be developed in Chapter 5.

4.4. OPTIMAL DM TRADEOFF

The goal of this section is to demonstrate that $d_{\mathcal{J}}(r)$ constitutes the optimal DM tradeoff in selective-fading MIMO channels. Our approach consists in deriving a sufficient condition for a family of codes to achieve $d(\mathcal{C}_r) = d_{\mathcal{J}}(r)$, and hence, by virtue of (4.35), to be DM tradeoff optimal. We shall show that our result is supported by an appealing geometric interpretation, and we will provide examples of how the optimal tradeoff curve depends upon the selectivity characteristics of the fading channel.

4.4.1. Code design criterion

For any given family of codes $\mathcal{C}_r$, we shall refer to the $N \times N$ matrix $\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E}$, where $\mathbf{E} = \mathbf{X} - \mathbf{X}'$ and $\mathbf{X}, \mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})$, as the *effective codeword difference matrix*. The corresponding eigenvalues

$$\lambda_k = \lambda_k(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E}),\ k = 1, 2, \ldots, \rho\mathrm{M_T} \tag{4.36}$$

are in general functions of SNR (for simplicity, this dependency is not explicit in our notation) and are assumed to satisfy $\lambda_1 \leq \lambda_2 \leq \cdots \leq$

$\lambda_{\rho \mathrm{M_T}}$ at every SNR level. Before presenting the main result of this chapter, the following remarks are now in order.

Remark 1

There are at most $\rho \mathrm{M_T}$ positive eigenvalues in (4.36); the remaining $N - \rho \mathrm{M_T}$ eigenvalues are identically zero for any effective codeword difference matrix arising from $\mathcal{C}_r(\mathsf{SNR})$ and at any SNR level. This observation simply follows from the inequality $\mathrm{rank}(\mathbf{A} \odot \mathbf{B}) \leq \mathrm{rank}(\mathbf{A})\,\mathrm{rank}(\mathbf{B})$, where $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite matrices of same dimension (Horn and Johnson, 1993, p. 458). Hence, we have $\mathrm{rank}\big(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E}\big) \leq \rho \mathrm{M_T}$, for all $\mathbf{E} = \mathbf{X} - \mathbf{X}'$, $\mathbf{X}, \mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})$ and all SNRs.

Remark 2

Due to the power constraint, all positive eigenvalues are strictly non-increasing functions of SNR, i.e.

$$\lambda_k(\mathsf{SNR}) \dot{\leq} 1, \; k = 1, 2, \ldots, \rho \mathrm{M_T}. \tag{4.37}$$

This statement can be proved as follows. Since the largest positive eigenvalue is smaller than the sum of eigenvalues, we get

$$\begin{aligned} \lambda_{\rho \mathrm{M_T}}(\mathsf{SNR}) &\leq \mathrm{Tr}\left(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E}\right) \\ &\leq \sigma_{\mathbb{H}}^2 \mathrm{Tr}\left(\mathbf{E}^H\mathbf{E}\right) &(4.38)\\ &\leq 4\sigma_{\mathbb{H}}^2 \mathrm{M_T} N &(4.39) \end{aligned}$$

where (4.38) is a consequence of $\mathbf{R}_{\mathbb{H}} \preceq \sigma_{\mathbb{H}}^2 \mathbf{I}$ and the fact that for positive semidefinite matrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$, if $\mathbf{A} \preceq \mathbf{B}$ then $\mathbf{A} \odot \mathbf{C} \preceq \mathbf{B} \odot \mathbf{C}$ so that $\mathrm{Tr}(\mathbf{A} \odot \mathbf{C}) \leq \mathrm{Tr}(\mathbf{B} \odot \mathbf{C})$. Finally, (4.39) follows from (4.2) and $\mathbf{E} = \mathbf{X} - \mathbf{X}'$. Since (4.39) is exponentially equal to $\mathsf{SNR}^0 \doteq 1$, we have (4.37).

We are now ready to present the following Theorem.

Theorem 4.1. *Consider a family of codes $\mathcal{C}_r$ with block length $N \geq \rho \mathrm{M_T}$ that operates over the channel with input-ouput relation (4.1). Define*

$$\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}(\mathsf{SNR}) \triangleq \min_{\substack{\mathbf{E}=\mathbf{X}-\mathbf{X}' \\ \mathbf{X},\mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})}} \prod_{k=1}^{\mathrm{m}} \lambda_k. \tag{4.40}$$

If there exists an $\epsilon > 0$ such that

$$\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}(\mathsf{SNR}) \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)} \tag{4.41}$$

then the error probability (for maximum-likelihood (ML) decoding) satisfies

$$P_e(\mathcal{C}_r) \doteq \mathsf{SNR}^{-d_{\mathcal{J}}(r)}. \tag{4.42}$$

Proof. We start by deriving an upper bound on the average (w.r.t. the random channel) pairwise error probability (PEP). Under the assumptions of perfect CSI at the receiver and additive white Gaussian noise, ML decoding amounts to choosing the codeword $\mathbf{X} \in \mathcal{C}_r(\mathsf{SNR})$, with $\mathbf{X} = [\mathbf{x}_0 \, \mathbf{x}_1 \, \ldots \, \mathbf{x}_{N-1}]$, that minimizes the quantity

$$\eta(\mathbf{X}) = \sum_{n=0}^{N-1} \left\| \mathbf{y}_n - \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}} \mathbf{H}_n \mathbf{x}_n \right\|^2.$$

Assuming that $\mathbf{X}$ was transmitted, the probability of the ML decoder mistakenly deciding in favor of codeword $\mathbf{X}' = [\mathbf{x}'_0 \, \mathbf{x}'_1 \, \ldots \, \mathbf{x}'_{N-1}]$ is given by

$$\begin{aligned} \mathbb{P}(\mathbf{X} \to \mathbf{X}') &= \mathbb{E}_{\{\mathbf{H}_n\}}\Big\{ \mathbb{P}(\eta(\mathbf{X}') < \eta(\mathbf{X})) \Big\} \\ &= \mathbb{E}_{\{\mathbf{H}_n\}}\left\{ \mathbb{P}\left(\frac{\mathsf{SNR}}{\mathrm{M_T}} \sum_{n=0}^{N-1} \|\mathbf{H}_n \mathbf{e}_n\|^2 + 2\zeta < 0 \right) \right\} \end{aligned} \tag{4.43}$$

where $\mathbf{e}_n = \mathbf{x}_n - \mathbf{x}'_n$, $n = 0, 1, \ldots, N-1$, and the random variable $\zeta = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}} \sum_{n=0}^{N-1} \mathfrak{Re}\big(\mathbf{z}_n^H \mathbf{H}_n \mathbf{e}_n\big)$ can be shown to satisfy

$$\zeta \sim \mathcal{N}\left(0, \frac{\mathsf{SNR}}{2\mathrm{M_T}} \sum_{n=0}^{N-1} \|\mathbf{H}_n \mathbf{e}_n\|^2\right).$$

Consequently, (4.43) can be written in terms of the Gaussian tail function $\mathcal{Q}(\cdot)$ as

$$\mathbb{P}(\mathbf{X} \to \mathbf{X}') = \mathbb{E}_{\{\mathbf{H}_n\}}\left\{\mathcal{Q}\left(\sqrt{\frac{\mathsf{SNR}}{2\mathrm{M_T}} \sum_{n=0}^{N-1} \|\mathbf{H}_n \mathbf{e}_n\|^2}\right)\right\}$$

which, using the standard approximation $\mathcal{Q}(x) \le e^{-x^2/2}$, yields

$$\begin{aligned} \mathbb{P}(\mathbf{X} \to \mathbf{X}') &\le \mathbb{E}_{\{\mathbf{H}_n\}}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}} \sum_{n=0}^{N-1} \|\mathbf{H}_n \mathbf{e}_n\|^2\right)\right\} \\ &= \mathbb{E}_{\mathbf{H}_w}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}} \mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H \mathbf{H}_w^H\right)\right)\right\} \end{aligned} \qquad (4.44)$$

where

$$\boldsymbol{\Upsilon} = (\mathbf{R}_{\mathbb{H}}^{T/2} \otimes \mathbf{I}_{\mathrm{M_T}})\,\mathrm{diag}\{\mathbf{e}_n\}_{n=0}^{N-1} \qquad (4.45)$$

and $\mathbf{H}_w$ denotes an $\mathrm{M_R} \times \mathrm{M_T} N$ i.i.d. $\mathcal{CN}(0,1)$ matrix. Noting that

$$\boldsymbol{\Upsilon}^H \boldsymbol{\Upsilon} = \mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H \mathbf{E} \qquad (4.46)$$

where $\mathbf{E} = [\mathbf{e}_0\ \mathbf{e}_1\ \ldots\ \mathbf{e}_{N-1}]$ is the codeword difference matrix, and using the fact that the nonzero eigenvalues of $\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}$ equal the nonzero eigenvalues of $\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H$, it follows, by assumption, that $\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H$ has $\rho\mathrm{M_T}$ nonzero eigenvalues. Then, employing the eigenvalue decomposition $\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where $\boldsymbol{\Lambda} = \mathrm{diag}\{\bar{\boldsymbol{\Lambda}}, \mathbf{0}\}$ and $\bar{\boldsymbol{\Lambda}} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_{\rho\mathrm{M_T}}\}$, we have

$$\mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H \mathbf{H}_w^H\right) \sim \mathrm{Tr}\left(\mathbf{H}_w \boldsymbol{\Lambda} \mathbf{H}_w^H\right)$$

where equality in distribution follows from $\mathbf{H}_w\mathbf{U} \sim \mathbf{H}_w$ (Muirhead, 1982). Thus, setting $\overline{\mathbf{H}}_w = \mathbf{H}_w([1{:}\mathrm{M_R}], [1{:}\rho\mathrm{M_T}])$, it clearly follows that

$$\mathbb{P}(\mathbf{X} \to \mathbf{X}') \le \mathbb{E}_{\overline{\mathbf{H}}_w}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}}\,\mathrm{Tr}\left(\overline{\mathbf{H}}_w\bar{\mathbf{\Lambda}}\overline{\mathbf{H}}_w^H\right)\right)\right\}. \tag{4.47}$$

Next, we upper bound (4.47) in terms of the Jensen channel $\boldsymbol{\mathcal{H}}$. Recalling (4.26), we have $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_w(\mathbf{R}^{T/2} \otimes \mathbf{I}_\mathrm{M})$, where $\boldsymbol{\mathcal{H}}_w$ is given in (4.27), and $\mathbf{R} = \mathbf{R}_\mathbb{H}$ if $\mathrm{M_R} \le \mathrm{M_T}$ and $\mathbf{R} = \mathbf{R}_\mathbb{H}^T$ if $\mathrm{M_R} > \mathrm{M_T}$. We treat these two cases separately.

In the case $\mathrm{M_R} \le \mathrm{M_T}$, it follows from (4.27) that $\overline{\mathbf{H}}_w = \overline{\boldsymbol{\mathcal{H}}}_w$, with $\overline{\boldsymbol{\mathcal{H}}}_w = \boldsymbol{\mathcal{H}}_w([1 : \mathrm{M_R}], [1 : \rho\mathrm{M_T}])$ and, hence, we have

$$\begin{aligned}\mathrm{Tr}\left(\overline{\mathbf{H}}_w\bar{\mathbf{\Lambda}}\overline{\mathbf{H}}_w^H\right) &= \mathrm{Tr}\left(\overline{\boldsymbol{\mathcal{H}}}_w\bar{\mathbf{\Lambda}}\overline{\boldsymbol{\mathcal{H}}}_w^H\right)\\ &\ge \sum_{k=1}^{\mathrm{M_R}} \lambda_k(\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H)\,\lambda_{\mathrm{M_R}+1-k}.\end{aligned} \tag{4.48}$$

where, recalling that $\lambda_k(\mathbf{A})$ are the nonzero eigenvalues of the Hermitian matrix $\mathbf{A}$ in ascending order, we have used Theorem C.1 to get the lower bound (4.48).

In the case $\mathrm{M_R} > \mathrm{M_T}$, we use $\mathbf{H}_w = [\mathbf{H}_{w,0}\,\mathbf{H}_{w,1}\,\cdots\,\mathbf{H}_{w,N-1}]$ and the partition $\bar{\mathbf{\Lambda}} = \mathrm{diag}\{\bar{\mathbf{\Lambda}}_0, \bar{\mathbf{\Lambda}}_1, \ldots, \bar{\mathbf{\Lambda}}_{\rho-1}\}$, where $\bar{\mathbf{\Lambda}}_n$ has dimension $\mathrm{M_T} \times \mathrm{M_T}$, to write

$$\begin{aligned}\mathrm{Tr}\left(\overline{\mathbf{H}}_w\bar{\mathbf{\Lambda}}\overline{\mathbf{H}}_w^H\right) &= \sum_{n=0}^{\rho-1}\mathrm{Tr}\left(\mathbf{H}_{w,n}\bar{\mathbf{\Lambda}}_n\mathbf{H}_{w,n}^H\right)\\ &\ge \sum_{n=0}^{\rho-1}\mathrm{Tr}\left(\mathbf{H}_{w,n}\bar{\mathbf{\Lambda}}_0\mathbf{H}_{w,n}^H\right) \qquad (4.49)\\ &= \sum_{n=0}^{\rho-1}\mathrm{Tr}\left(\bar{\mathbf{\Lambda}}_0^{1/2}\mathbf{H}_{w,n}^H\mathbf{H}_{w,n}\bar{\mathbf{\Lambda}}_0^{1/2}\right)\\ &= \mathrm{Tr}\left(\bar{\mathbf{\Lambda}}_0^{1/2}\left(\sum_{n=0}^{\rho-1}\mathbf{H}_{w,n}^H\mathbf{H}_{w,n}\right)\bar{\mathbf{\Lambda}}_0^{1/2}\right)\end{aligned}$$

$$= \mathrm{Tr}\left(\bar{\mathbf{\Lambda}}_0^{1/2}\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\bar{\mathbf{\Lambda}}_0^{1/2}\right) \tag{4.50}$$

$$\geq \sum_{k=1}^{\mathrm{M_T}} \lambda_k(\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H)\,\lambda_{\mathrm{M_T}+1-k} \tag{4.51}$$

where (4.49) follows from the positive semidefinite ordering $\bar{\mathbf{\Lambda}}_0 \preceq \bar{\mathbf{\Lambda}}_n$ due to the eigenvalues of $\bar{\mathbf{\Lambda}}$ being given in ascending order, (4.50) follows from (4.27) and $\overline{\boldsymbol{\mathcal{H}}}_w = \boldsymbol{\mathcal{H}}_w([1:\mathrm{M_T}],[1:\rho\mathrm{M_R}])$, and (4.51) is a direct consequence of applying Theorem C.1 in Appendix C to (4.50).

Combining (4.48) and (4.51), we have in general

$$\mathrm{Tr}\left(\overline{\mathbf{H}}_w\bar{\mathbf{\Lambda}}\overline{\mathbf{H}}_w^H\right) \geq \sum_{k=1}^{\mathrm{m}} \lambda_k(\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H)\,\lambda_{\mathrm{m}+1-k}$$

$$= \sum_{k=1}^{\mathrm{m}} \mathsf{SNR}^{-\alpha_k}\,\lambda_{\mathrm{m}+1-k} \tag{4.52}$$

where we have used (4.31). Using (4.52) in (4.47), the PEP upper bound is now given in terms of the singularity levels α_k ($k=1,2,\ldots,\mathrm{m}$) characterizing the Jensen outage event

$$\mathbb{P}(\mathbf{X}\to\mathbf{X}') \leq \mathbb{E}_{\boldsymbol{\alpha}}\left\{\exp\left(-\frac{1}{4\mathrm{M_T}}\sum_{k=1}^{\mathrm{m}}\mathsf{SNR}^{1-\alpha_k}\,\lambda_{\mathrm{m}+1-k}\right)\right\}. \tag{4.53}$$

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

$$\sum_{k\in\mathcal{S}_{\boldsymbol{\alpha}}}(1-\alpha_k) = \sum_{k=1}^{\mathrm{m}}[1-\alpha_k]^+$$

which, in conjunction with the arithmetic-geometric mean inequality, yields

$$\sum_{k=1}^{\mathrm{m}} \mathsf{SNR}^{1-\alpha_k}\,\lambda_{\mathrm{m}+1-k} \geq \sum_{k\in\mathcal{S}_{\boldsymbol{\alpha}}} \mathsf{SNR}^{1-\alpha_k}\,\lambda_{\mathrm{m}+1-k} \geq |\mathcal{S}_{\boldsymbol{\alpha}}| \left(\mathsf{SNR}^{\sum_{k=1}^{\mathrm{m}}[1-\alpha_k]^+} \prod_{k\in\mathcal{S}_{\boldsymbol{\alpha}}} \lambda_{\mathrm{m}+1-k}\right)^{\frac{1}{|\mathcal{S}_{\boldsymbol{\alpha}}|}}. \quad (4.54)$$

Using (4.54) in (4.53), we get

$$\mathbb{P}(\mathbf{X}\to\mathbf{X}') \leq \mathbb{E}_{\boldsymbol{\alpha}}\left\{\exp\left(-\frac{|\mathcal{S}_{\boldsymbol{\alpha}}|}{4\mathrm{M_T}}\left(\mathsf{SNR}^{\sum_{k=1}^{\mathrm{m}}[1-\alpha_k]^+}\prod_{k\in\mathcal{S}_{\boldsymbol{\alpha}}}\lambda_{\mathrm{m}+1-k}\right)^{\frac{1}{|\mathcal{S}_{\boldsymbol{\alpha}}|}}\right)\right\}. \quad (4.55)$$

The dependence of the PEP upper bound (4.55) on the singularity levels characterizing the Jensen outage event suggests to split up the overall error probability according to

$$\begin{aligned} P_e(\mathcal{C}_r) &= \mathbb{E}_{\boldsymbol{\alpha}}\{\mathbb{P}(\mathcal{E}_r)\} \\ &= \mathbb{P}(\mathcal{E}_r, \boldsymbol{\alpha}\in\mathcal{J}_r) + \mathbb{P}(\mathcal{E}_r, \boldsymbol{\alpha}\notin\mathcal{J}_r) \\ &= \mathbb{P}(\boldsymbol{\alpha}\in\mathcal{J}_r)\,\mathbb{P}(\mathcal{E}_r|\boldsymbol{\alpha}\in\mathcal{J}_r) \\ &\qquad + \mathbb{P}(\boldsymbol{\alpha}\notin\mathcal{J}_r)\,\mathbb{P}(\mathcal{E}_r|\boldsymbol{\alpha}\notin\mathcal{J}_r) \\ &\leq \mathbb{P}(\boldsymbol{\alpha}\in\mathcal{J}_r) \\ &\qquad + \mathbb{P}(\boldsymbol{\alpha}\notin\mathcal{J}_r)\,\mathbb{P}(\mathcal{E}_r|\boldsymbol{\alpha}\notin\mathcal{J}_r)\,. \end{aligned} \quad (4.56)$$

For any $\boldsymbol{\alpha}\notin\mathcal{J}_r$, we have, by definition, $\sum_{k=1}^{\mathrm{m}}[1-\alpha_k]^+ \geq r$ and consequently $|\mathcal{S}_{\boldsymbol{\alpha}}| \geq 1$, which upon noting that $|\mathcal{C}_r(\mathsf{SNR})| = \mathsf{SNR}^{Nr}$, yields the following union bound based on the PEP in (4.55)

$$\mathbb{P}(\mathcal{E}_r|\boldsymbol{\alpha} \notin \mathcal{J}_r) \leq \mathsf{SNR}^{Nr} \exp\left(-\frac{1}{4\mathrm{M_T}}\left(\mathsf{SNR}^r \prod_{k\in\mathcal{S}_{\boldsymbol{\alpha}}} \lambda_{\mathrm{m}+1-k}\right)^{\frac{1}{\mathrm{m}}}\right) \tag{4.57}$$

where we used $|\mathcal{S}_{\boldsymbol{\alpha}}| \leq \mathrm{m}$. Next, we note that the code design criterion in (4.41) implies that $\prod_{k=1}^{\mathrm{m}} \lambda_k \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)}$ for some $\epsilon > 0$ so that, recalling (4.37), we necessarily have

$$\prod_{k\in\mathcal{S}_{\boldsymbol{\alpha}}} \lambda_{\mathrm{m}+1-k} \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)} \tag{4.58}$$

for any $\mathcal{S}_{\boldsymbol{\alpha}} \subseteq \{1, 2, \ldots, \mathrm{m}\}$. Inserting (4.58) into (4.57) yields

$$\begin{aligned}\mathbb{P}(\mathcal{E}_r, \boldsymbol{\alpha} \notin \mathcal{J}_r) &= \underbrace{\mathbb{P}(\boldsymbol{\alpha} \notin \mathcal{J}_r)}_{\leq 1} \mathbb{P}(\mathcal{E}_r|\boldsymbol{\alpha} \notin \mathcal{J}_r) \\ &\dot{\leq} \mathsf{SNR}^{Nr} \exp\left(-\frac{\mathsf{SNR}^{\epsilon/\mathrm{m}}}{4\mathrm{M_T}}\right).\end{aligned} \tag{4.59}$$

In contrast to the Jensen outage probability which decays polynomially with SNR, (4.59) decays exponentially with SNR. Consequently, noting that $\mathbb{P}(\boldsymbol{\alpha} \in \mathcal{J}_r) \doteq \mathbb{P}(\mathcal{J}_r)$ and using (4.59) in (4.56), we get

$$P_e(\mathcal{C}_r) \dot{\leq} \mathbb{P}(\mathcal{J}_r) \tag{4.60}$$

for $r > 0$. Recalling (4.34), we also have $\mathbb{P}(\mathcal{J}_r) \dot{\leq} \mathbb{P}(\mathcal{O}_r)$. In addition, for a specific family of codes $\mathcal{C}_r$, $\mathbb{P}(\mathcal{O}_r) \dot{\leq} P_e(\mathcal{C}_r)$ holds true. Putting the pieces together, we obtain that for any $r > 0$

$$\mathbb{P}(\mathcal{O}_r) \dot{\leq} P_e(\mathcal{C}_r) \dot{\leq} \mathbb{P}(\mathcal{J}_r) \dot{\leq} \mathbb{P}(\mathcal{O}_r)$$

which implies

$$P_e(\mathcal{C}_r) \doteq \mathbb{P}(\mathcal{J}_r)$$

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

$$P_e(\mathcal{C}_r) \doteq \mathsf{SNR}^{-d_{\mathcal{J}}(r)}. \tag{4.61}$$

Finally, by the continuity of $d_{\mathcal{J}}(r)$, (4.61) also holds for $r \to 0$. □

As a direct consequence of Theorem 4.1, a family of codes $\mathcal{C}_r$ that satisfies (4.41) for any (w.r.t. SNR) 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 (4.35), we obtain

$$d^{\star}(r) = d_{\mathcal{J}}(r). \tag{4.62}$$

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 4.1 shows that, even though the channels in Jensen outage form by definition a subset of the channels in outage, 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.

4.4.2. Geometric interpretation

In the following, we provide a geometric interpretation of the optimal DM tradeoff in selective fading. Our presentation assumes an OFDM system operating over an intersymbol interference (ISI) channel. This discussion extends the geometric interpretation in the flat-fading case provided by Zheng and Tse (2003) and briefly summarized in Section 3.3.1.

We start by noting that in an OFDM system, the mutual information (after discarding the cyclic prefix) is given by $\mathrm{I}(\mathsf{SNR}) = \frac{1}{N}\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{M_T}\mathbf{D_H}\mathbf{D_H}^H\right)$, where $\mathbf{D_H} = \mathrm{diag}\{\mathbf{H}_n\}_{n=0}^{N-1}$. Following the geometric argument in Zheng and Tse (2003), we wish to relate the outage probability at multiplexing rate r to the rank of the matrix $\mathbf{D_H}$. Unfortunately, the quantity $\mathrm{rank}(\mathbf{D_H})$ is difficult to study in general because the different diagonal blocks are correlated, i.e.,

$$\mathbf{H}_n = \sum_{l=0}^{L-1} \mathbf{H}(l) e^{-j\frac{2\pi}{N}ln}$$

where $\mathbf{H}(l)$, $l = 0, 1, \ldots, L-1$, are the i.i.d. matrix-valued channel taps with $\mathcal{CN}(0,1)$ entries. In OFDM, however, the matrix $\mathbf{D_H}\mathbf{D_H}^H$ can be readily shown to be unitary equivalent to $\mathbf{C_H}\mathbf{C_H}^H$ (see e.g. Paulraj et al., 2003, Ch. 9), where $\mathbf{C_H}$ is the following $N\mathrm{M_R} \times N\mathrm{M_T}$ block-circulant matrix

$$\mathbf{C_H} = \begin{bmatrix} \mathbf{H}(0) & \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{H}(L-1) & \cdots & \mathbf{H}(1) \\ \mathbf{H}(1) & \mathbf{H}(0) & \ddots & & \vdots & \mathbf{0} & \ddots & \vdots \\ \vdots & \mathbf{H}(1) & \ddots & \ddots & \vdots & \vdots & \ddots & \mathbf{H}(L-1) \\ \mathbf{H}(L-1) & \vdots & \ddots & \ddots & \mathbf{0} & \vdots & & \mathbf{0} \\ \mathbf{0} & \mathbf{H}(L-1) & & \ddots & \mathbf{H}(0) & \mathbf{0} & & \vdots \\ \vdots & \mathbf{0} & \ddots & & \mathbf{H}(1) & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{H}(L-1) & \cdots & \mathbf{H}(1) & \mathbf{H}(0) \end{bmatrix}.$$

For $N > L$, the structure of $\mathbf{C_H}$ implies that for any channel realization

$$\mathrm{rank}(\mathbf{C_H}) = N\mathrm{rank}(\mathbf{C}_w) \tag{4.63}$$

where we have defined $\mathbf{C}_w = \mathbf{C_H}([1 : L\mathrm{M_R}], [1 : \mathrm{M_T}])^H$, if $\mathrm{M_T} \leq \mathrm{M_R}$, and $\mathbf{C}_w = \mathbf{C_H}([(N-1)\mathrm{M_R} + 1 : N\mathrm{M_R}], [(N-L)\mathrm{M_T} + 1 : N\mathrm{M_T}])$, if $\mathrm{M_T} > \mathrm{M_R}$. Note that with this definition $\mathbf{C}_w$ is a $\mathrm{m} \times L\mathrm{M}$ matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. In order to characterize $\mathrm{rank}(\mathbf{C_H})$, it suffices to consider $\mathrm{rank}(\mathbf{C}_w)$ only and determine the number of parameters required to specify r linearly independent rows in $\mathbf{C}_w$. This number is obtained as follows: $L\mathrm{M}r$ parameters are required to specify

r linearly independent rows in $\mathbf{C}_w$. The remaining $\mathrm{m} - r$ rows are then given by linear combinations of the r linearly independent rows. Specifying these linearly dependent rows requires r parameters per row and hence $(\mathrm{m} - r)r$ parameters in total. The number of parameters specifying a matrix $\mathbf{C_H}$ with rank Nr is therefore obtained as

$$LMr + (\mathrm{m} - r)r = LM\mathrm{m} - (LM - r)(\mathrm{m} - r). \tag{4.64}$$

Summarizing the results, again invoking the arguments in Zheng and Tse (2003) and noting that $\mathbf{C}_w$ lies in a $LM\mathrm{m}$-dimensional space, we obtain that $\mathrm{rank}(\mathbf{C_H})$ is close to Nr when $\mathbf{C}_w$ is close to the subspace of all rank-r matrices. This requires a collapse in the components of $\mathbf{C}_w$ in all the dimensions orthogonal to that subspace; the number of such dimensions is given by $(LM - r)(\mathrm{m} - r)$, which, recalling that $\rho = L$, is precisely the SNR exponent given in (4.33).

4.4.3. Selectivity and optimal performance

In this section, we apply the concept of Jensen channel to different selective-fading channels and obtain the corresponding optimal DM tradeoff curve.

A. Time-frequency selectivity

For time-frequency selective channels, the covariance matrix $\mathbf{R}_{\mathbb{H}}$ is Toeplitz-block-Toeplitz, and its entries are given by the sampled time-frequency correlation

$$\mathbf{R}_{\mathbb{H}}(n, n') = R_{\mathbb{H}}((m - m')T, (k - k')F)$$

where the slot index $n = 0, 1, \ldots, N - 1$ is related to the time-frequency couple (m, k) by the mapping $n = \mathcal{M}(m, k)$ defined in (2.14). From standard results on the eigenvalue distribution of Toeplitz-block-Toeplitz matrices (Gray, 2006; Voois, 1996; Tyrtyshnikov, 1996), the eigenvalues of $\mathbf{R}_{\mathbb{H}}$ can be shown to be asymptotically ($M, K \to \infty$) distributed as*

$$S(\xi,\mu) = \frac{1}{TF} C_{\mathbb{H}}\left(\frac{\xi}{F}, \frac{\mu}{T}\right),\ 0 \le \mu, \xi < 1.$$

The eigenvalues of $\mathbf{R}_{\mathbb{H}}$ are well approximated by uniformly sampling $S(\xi,\mu)$ (Gray, 2006; Voois, 1996), i.e.,

$$\lambda_{\mathcal{M}(m,k)}(\mathbf{R}_{\mathbb{H}}) = S\left(\frac{k}{K}, \frac{m}{M}\right),$$
$$\text{for} \quad m = 0, 1, \ldots, M-1,\ k = 0, 1, \ldots, K-1.$$

Now, since the scattering function is compactly supported in the rectangle $[0, \tau_0] \times [0, \nu_0]$, we can infer that $\mathbf{R}_{\mathbb{H}}$ has $v \cdot t$ nonzero eigenvalues, where

$$v \triangleq \lfloor \nu_0 T M \rfloor \quad \text{and} \quad t \triangleq \lfloor \tau_0 F K \rfloor.$$

Hence, the rank of $\mathbf{R}_{\mathbb{H}}$ is approximately given by $\rho = v \cdot t$. From (4.33) and (4.62), the optimal DM tradeoff curve follows as the piecewise linear function connecting the points $(r, d^\star(r))$ for $r = 0, 1, \ldots, \mathrm{m}$, where

$$d^\star(r) = (vt\,\mathrm{M} - r)(\mathrm{m} - r). \tag{4.65}$$

Interestingly, this expression relates the optimal performance in the sense of the DM tradeoff to system properties like output signal duration TM and signal bandwith FK, and also to quantities that depend on the channel selectivity characteristics such as the Doppler spread ν_0 and delay spread τ_0.

B. Circulant covariance matrix

Next, we apply our results on the DM tradeoff to the case of linear time-invariant (LTI) and linear frequency-invariant (LFI) systems. Under

*This relation was derived in (2.23).

the assumptions of Appendix A, the channel covariance matrix is unitarily diagonalizable by the Fourier matrix $\boldsymbol{\Psi}$. By (A.2), we have $\mathrm{rank}(\mathbf{R}_{\mathbb{H}}) = L$ and, after inserting $\rho = L$ into (4.33) and using (4.62), we get the optimal DM tradeoff curve as the piecewise linear function connecting the points $(r, d^\star(r))$ for $r = 0, 1, \ldots, \mathrm{m}$, with*

$$d^\star(r) = (L\,\mathrm{M} - r)(\mathrm{m} - r). \tag{4.66}$$

This is the optimal DM tradeoff curve for frequency-selective fading MIMO channels reported previously in Medles and Slock (2005). Specializing (4.66) to the case single-antenna 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 result reported in Grokop and Tse (2004). We note that the proof techniques employed in (Grokop and Tse, 2004; Medles and Slock, 2005) are different from the approach taken here and seem to be tailored to the frequency-selective case. In addition, since (4.41) can be guaranteed for any $N \geq L\mathrm{M_T}$, our approach is not limited to large code lengths as required in (Grokop and Tse, 2004; Medles and Slock, 2005).

4.5. RELATION TO OTHER DESIGN CRITERIA

This section examines how our code design criterion ties in with other results on space-time code design that were previously reported in literature.

4.5.1. Non-vanishing determinant

The non-vanishing determinant criterion, which is well-known in flat-fading MIMO channels, can be recovered from the code design criterion of Theorem 4.1. Recall that in flat fading, the channel remains constant

*Compare (4.66) with (4.65) and note that L is analogous to t in the LTI case and to v in the LFI case.

for the codeword duration, i.e., $\mathbf{H}_n = \mathbf{H}_0$ for all n. Therefore, the channel covariance matrix satisfies $\mathbf{R}_{\mathbb{H}} = \mathbb{1}_N$ and, hence, $\rho = 1$, and $\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E} = \mathbf{E}^H\mathbf{E}$ for all possible $\mathbf{E} = \mathbf{X} - \mathbf{X}'$. We note also that the quantity $\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}(\mathsf{SNR})$ in Theorem 4.1 lower bounds the product of the m smallest nonzero eigenvalues of any matrix $\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E}$ arising from the codebook $\mathcal{C}_r(\mathsf{SNR})$. From these observations, we can conclude that, in the flat-fading case, the code design criterion of Theorem 4.1 specializes to

$$\Xi_{\mathrm{m}}^{\mathrm{M_T}}(\mathsf{SNR}) = \min_{\substack{\mathbf{E}=\mathbf{X}-\mathbf{X}' \\ \mathbf{X},\mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})}} \prod_{k=1}^{\mathrm{m}} \lambda_k(\mathbf{E}\mathbf{E}^H) \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)} \quad (4.67)$$

for some $\epsilon > 0$. For $\mathrm{M_T} \leq \mathrm{M_R}$, i.e., $\mathrm{m} = \mathrm{M_T}$, we clearly have

$$\Xi_{\mathrm{m}}^{\mathrm{m}}(\mathsf{SNR}) = \min_{\mathbf{E}} \det(\mathbf{E}\mathbf{E}^H)$$

and condition (4.67) turns out to be the non-vanishing* determinant criterion proposed for the first time in (Belfiore and Rekaya, 2003). While many space-time code constructions have a nonzero minimum determinant, it does often vanish with increasing data rates and, hence, it fails to satisfy (4.67). In contrast, the constructions based on the non-vanishing determinant criterion (e.g., Yao and Wornell, 2003; Belfiore et al., 2005; Dayal and Varanasi, 2005; Elia et al., 2006) have a minimum determinant that is bounded away from zero for increasing rates. This property is a sufficient condition for optimality (in the sense of the DM tradeoff) over the i.i.d. Rayleigh flat-fading MIMO channel (Elia et al., 2006).

4.5.2. Approximate universality

The code design criterion of Theorem 4.1 subsumes as a special case the criterion for approximate universality (Tavildar and Viswanath,

*This terminology is to be understood prior to SNR normalization, i.e, it applies to codewords $\tilde{\mathbf{X}} = \sqrt{\mathsf{SNR}}\,\mathbf{X}$.

2006) over flat-fading MIMO channels. This observation follows easily by comparing (4.67) to the criterion given in (Tavildar and Viswanath, 2006, Theorem 3.1). The coincidence of both criteria in the flat-fading case is noteworthy as they are obtained using different approaches: while our proof technique makes explicit assumptions on the fading law, approximately universal codes are designed to perform optimally over any channel that is not in outage irrespectively of the fading distribution. In the context of approximate universality, the code design criterion aims at protecting the codewords against the worst-case channel, i.e., the channel that aligns itself in the weakest directions of the codeword difference matrix. Our approach does not explicitly consider the worst-case channel, but it turns out that any family $\mathcal{C}_r$ satisfying (4.41) achieves the optimal DM tradeoff in flat fading irrespectively of the distribution of the channel matrix $\mathbf{H}_0$.

4.5.3. Classical space-time code design criteria

Next, we show that the classical rank criterion (Tarokh et al., 1998, 1999; Bölcskei and Paulraj, 2000; Bölcskei et al., 2002b, 2003a; Ma et al., 2005) coincides with the sufficient condition of Theorem 4.1 at multiplexing rate $r = 0$ and that the determinant criterion has no impact on error performance in the sense of the DM tradeoff.

We start by deriving the classical space-time code design criteria. Consider a codebook $\mathcal{C}_0$ that operates at multiplexing rate $r = 0$. Assuming that codeword $\mathbf{X}$ was transmitted, the probability of the ML decoder mistakenly deciding in favor of codeword $\mathbf{X}'$ was upper bounded in (4.44) in terms of the codeword difference matrix $\mathbf{E} = [\mathbf{e}_0\, \mathbf{e}_1\, \ldots\, \mathbf{e}_{N-1}]$, with $\mathbf{e}_n = \mathbf{x}_n - \mathbf{x}'_n$, as

$$\mathbb{P}(\mathbf{X} \to \mathbf{X}') \leq \mathbb{E}_{\mathbf{H}}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}}\sum_{n=0}^{N-1}||\mathbf{H}_n\mathbf{e}_n||^2\right)\right\}. \qquad (4.68)$$

The key point in the proof of Theorem 4.1 is to guarantee that (4.68) decays exponentially for every channel that is not in the set $\mathcal{J}_r$. In contrast, the traditional approach summarized next focuses on obtaining an upper bound on the error probability by averaging (4.68) over all channel realizations. We note that

$$\sum_{n=0}^{N-1} ||\mathbf{H}_n \mathbf{e}_n||^2 \sim \mathbf{h}_w^H \left(\boldsymbol{\Upsilon}^* \boldsymbol{\Upsilon}^T \otimes \mathbf{I}_{\mathrm{M_R}}\right) \mathbf{h}_w$$

where $\mathbf{h}_w$ is a $N\mathrm{M_T}\mathrm{M_R}$-dimensional Gaussian random vector with i.i.d. $\mathcal{CN}(0,1)$ entries and $\boldsymbol{\Upsilon}$ is given by

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

Using the fact that (Turin, 1960)

$$\mathbb{E}\Big\{\exp\Big(-v \left\|\mathbf{h}\right\|^2\Big)\Big\} = \det(\mathbf{I} + v\mathbf{C})^{-1}$$

for $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{C})$, we take the expectation in (4.68) to get

$$\begin{aligned} \mathbb{P}(\mathbf{X} \to \mathbf{X}') \leq \; & \det\left(\mathbf{I} + \frac{\mathsf{SNR}}{4\mathrm{M_T}} \boldsymbol{\Upsilon}^* \boldsymbol{\Upsilon}^T\right)^{-\mathrm{M_R}} \\ & \leq \Delta_{\mathbf{E}}^{-\mathrm{M_R}} \left(\frac{\mathsf{SNR}}{4\mathrm{M_T}}\right)^{-\kappa_{\mathbf{E}}\mathrm{M_R}} \end{aligned} \tag{4.69}$$

where, using the identity $\operatorname{rank}\big(\boldsymbol{\Upsilon}^* \boldsymbol{\Upsilon}^T\big) = \operatorname{rank}\big(\boldsymbol{\Upsilon}^H \boldsymbol{\Upsilon}\big)$ and noting that $\boldsymbol{\Upsilon}^H \boldsymbol{\Upsilon} = \mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H \mathbf{E}$ (see also (4.45), (4.46)), we have used

$$\kappa_{\mathbf{E}} = \operatorname{rank}\big(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H \mathbf{E}\big) \tag{4.70}$$

and $\Delta_{\mathbf{E}}$ denotes the product of the nonzero eigenvalues of $\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H \mathbf{E}$. The total probability of error can be decomposed as

$$P_e(\mathcal{C}_0) = \sum_{\mathbf{X}} \mathbb{P}(\mathbf{X}) \sum_{\mathbf{X}' \neq \mathbf{X}} \mathbb{P}(\mathbf{X} \to \mathbf{X}') \tag{4.71}$$

and hence, by (4.69) and the fact that $\kappa_{\mathbf{E}} \leq \rho \mathrm{M_T}$, the SNR exponent of the error probability corresponding to $\mathcal{C}_0$ satisfies

$$\min_{\mathbf{E}} \kappa_{\mathbf{E}} \, \mathrm{M_R} \leq - \lim_{\mathsf{SNR} \to \infty} \frac{\log P_e(\mathcal{C}_0)}{\log \mathsf{SNR}} \leq \rho \mathrm{M_T} \mathrm{M_R} \tag{4.72}$$

where the minimum is over all $\mathbf{E}$ arising from $\mathcal{C}_0$. The rank and determinant criteria consist, respectively, in maximizing $\Delta_{\mathbf{E}}$ and $\kappa_{\mathbf{E}}$ over all possible effective codeword difference matrices so as to minimize (4.71).

Since the lower bound to the SNR exponent of $P_e(\mathcal{C}_0)$ in (4.72) is independent of $\Delta_{\mathbf{E}}$, maximizing the determinant plays no role in the DM tradeoff framework. However, by maximizing $\kappa_{\mathbf{E}}$ over all possible $\mathbf{E}$, the rank criterion yields

$$\kappa_{\mathbf{E}} = \rho \mathrm{M_T}, \ \forall\, \mathbf{E} = \mathbf{X} - \mathbf{X}', \mathbf{X}, \mathbf{X}' \in \mathcal{C}_0 \tag{4.73}$$

which, upon inserting into the left-hand side of (4.72), readily shows that the SNR exponent of $P_e(\mathcal{C}_0)$ is $\rho \mathrm{M_T M_R}$. Since this is the maximal diversity gain $d_{\mathcal{J}}(0)$, we can conclude that the rank criterion realizes the optimal DM tradeoff at multiplexing rate $r = 0$.

4.5.4. A geometric view on code design

The optimal code design criterion of Theorem 4.1 depends explicitly on the selectivity characteristics of the fading channel through the covariance matrix $\mathbf{R}_{\mathbb{H}}$. The effective codeword difference matrix can be stated in a form that yields geometric insight into the implications of this dependency and nicely recovers the code design criterion reported in Bölcskei and Paulraj (2000) for frequency-selective MIMO channels.

Lemma 4.2. Let $\mathbf{R}_{\mathbb{H}} = \sum_{n=0}^{\rho-1} \theta_n\, \mathbf{v}_n \mathbf{v}_n^H$ be the eigenvalue decomposition of the channel covariance matrix. Then,

$$\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H \mathbf{E} = \tilde{\mathbf{E}}^H \tilde{\mathbf{E}} \tag{4.74}$$

where $\tilde{\mathbf{E}}$ is the $\rho \mathrm{M_T} \times N$ matrix given by

$$\tilde{\mathbf{E}}^H = \begin{bmatrix} \sqrt{\theta_0}\, \mathbf{D}_{\mathbf{v}_0^*} \mathbf{E}^H & \sqrt{\theta_1}\, \mathbf{D}_{\mathbf{v}_1^*} \mathbf{E}^H & \cdots & \sqrt{\theta_{\rho-1}}\, \mathbf{D}_{\mathbf{v}_{\rho-1}^*} \mathbf{E}^H \end{bmatrix}. \tag{4.75}$$

Proof. Inserting the eigenvalue decomposition of $\mathbf{R}_{\mathbb{H}}$ into the LHS of (4.74), we get

$$\begin{aligned}\mathbf{R}_{\mathbb{H}}^{T} \odot \mathbf{E}^{H}\mathbf{E} &= \left(\sum_{n=0}^{\rho-1} \theta_n \, \mathbf{v}_n^{*}\mathbf{v}_n^{T}\right) \odot \mathbf{E}^{H}\mathbf{E} \\ &= \sum_{n=0}^{\rho-1} \theta_n \, \mathbf{D}_{\mathbf{v}_n^{*}} \mathbf{E}^{H}\mathbf{E}\, \mathbf{D}_{\mathbf{v}_n^{T}} \\ &= \tilde{\mathbf{E}}^{H}\tilde{\mathbf{E}}.\end{aligned}$$

□

Note that the pseudo codeword $\tilde{\mathbf{E}} = \tilde{\mathbf{X}} - \tilde{\mathbf{X}}'$ has dimension $\rho \mathrm{M_T} \times N$ and so do the effective codewords $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{X}}'$. This suggests that, from the receiver's perspective, the channel has $\rho \mathrm{M_T}$ virtual antennas. When $\mathcal{C}_r$ is designed so that any effective codeword difference matrix $\tilde{\mathbf{E}}$ has the maximal number of linearly independent columns, i.e., its rank is $\rho \mathrm{M_T}$, the corresponding codewords are as distinguishable as possible at the receiver, which in turn means that a pairwise error event is less likely to occur.

Example: cyclic signaling

We consider an orthogonal frequency-division multiplexing (OFDM) system operating over a LTI channel as described in Appendix A. The implications of channel selectivity on code design are as follows. Since the covariance matrix $\mathbf{R}_{\mathbb{H}}$ is a L-rank circulant, it admits the columns of the Fourier matrix $\boldsymbol{\Psi}$ as eigenvectors. Consequently, we can use $\mathbf{D}_{\mathbf{v}_n^{*}} = \mathbf{D}^{n}$, where $\mathbf{D} = \mathrm{diag}\{e^{j\frac{2\pi}{N}k}\}_{k=0}^{N-1}$, in (4.75) to obtain the effective codeword difference matrix

$$\tilde{\mathbf{E}}^{H} = \left[\sqrt{\theta_0}\,\mathbf{E}^{H} \quad \sqrt{\theta_1}\,\mathbf{D}\mathbf{E}^{H} \quad \ldots \quad \sqrt{\theta_{L-1}}\,\mathbf{D}^{L-1}\mathbf{E}^{H}\right].$$

Next, denote by $\mathbf{E}_t$ the time-domain counterpart to the codeword difference matrix $\mathbf{E}$, i.e., $\mathbf{E}^H = \mathbf{\Psi}^* \, \mathbf{E}_t^H$, and note that

$$\mathbf{\Psi}^T \mathbf{D}^n \mathbf{E}^H = \mathbf{\Pi}^n \, \mathbf{E}_t^H$$

where $\mathbf{\Pi} = [\boldsymbol{\pi}_1 \, \cdots \, \boldsymbol{\pi}_{N-1} \, \boldsymbol{\pi}_0]$ is the basic circulant permutation matrix, and $\boldsymbol{\pi}_n = [0 \, \cdots \, 0 \; 1 \; 0 \, \cdots \, 0]^T$ contains a 1 in its nth position. The time-domain counterpart to $\tilde{\mathbf{E}}^H$ follows now as

$$\tilde{\mathbf{E}}_t^H = \left[\sqrt{\theta_0} \, \mathbf{E}_t^H \quad \sqrt{\theta_1} \, \mathbf{\Pi} \mathbf{E}_t^H \quad \ldots \quad \sqrt{\theta_{L-1}} \, \mathbf{\Pi}^{L-1} \mathbf{E}_t^H \right].$$

Since the frequency- and time-domain versions of the codeword difference matrices are related by a unitary transformation, the rank of $\tilde{\mathbf{E}}_t$ is the same as the rank of $\tilde{\mathbf{E}}$. It follows that $\mathcal{C}_r$ should be constructed so that the matrices $\mathbf{\Pi}^l \mathbf{E}_t^H$ and $\mathbf{\Pi}^{l'} \mathbf{E}_t^H$ have linearly independent columns for any $l \neq l'$ and any possible codeword difference matrix $\mathbf{E}_t$. This is the code design criterion for OFDM systems over frequency-selective MIMO channels reported for the first time in (Bölcskei and Paulraj, 2000).

CHAPTER 5

Optimal Code Construction

In previous chapters, we have established the optimal diversity-multiplexing (DM) tradeoff for the general class of selective-fading channels and provided a code design criterion for DM tradeoff optimality. The goal of this chapter is to demonstrate the existence of codes satisfying this criterion by providing corresponding design procedures. In addition, we want to ensure that the proposed code designs are practicable in the sense of being DM tradeoff optimal, ideally independently of the channel covariance matrix (i.e., of the selectivity characteristics). We shall see that in the single-transmit-antenna case this is rather straightforward to accomplish. In the case of multiple transmit antennas, we propose a procedure that decouples the problem into the design of a precoder, which can be obtained systematically for a given channel covariance matrix, and an outer code, which has to satisfy a design criterion that is independent of the covariance matrix.

The remainder of this chapter is organized as follows. In Section 5.1, we examine the single-transmit-antenna channel and give an example of an optimal code. Section 5.2 presents a new code construction procedure for multiple-transmit-antenna channels which is based on precoding. Finally, Section 5.3 introduces a method to construct the precoder as a function of the channel selectivity characteristics.

5.1. SINGLE TRANSMIT ANTENNA

We consider a family of codes $\mathcal{C}_r$ of block length N. Since we consider a single-transmit-antenna channel, the codewords of $\mathcal{C}_r(\mathsf{SNR})$ are row vectors $\bar{\mathbf{x}}$ of length N and satisfy the power constraint

$$\|\bar{\mathbf{x}}\|^2 \leq N, \forall \bar{\mathbf{x}} \in \mathcal{C}_r(\mathsf{SNR}). \tag{5.1}$$

The effective codeword difference matrix is now given by

$$\mathbf{R}_{\mathbb{H}}^T \odot \bar{\mathbf{e}}^H \bar{\mathbf{e}} = \mathbf{D}_{\bar{\mathbf{e}}}^H \, \mathbf{R}_{\mathbb{H}}^T \, \mathbf{D}_{\bar{\mathbf{e}}}$$

where $\bar{\mathbf{e}} = \bar{\mathbf{x}} - \bar{\mathbf{x}}'$, with $\bar{\mathbf{x}}, \bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})$. Hence, the quantity $\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}$ defined in (4.40) specializes to

$$\Xi_1^{\rho}(\mathsf{SNR}) = \min_{\substack{\bar{\mathbf{e}} = \bar{\mathbf{x}} - \bar{\mathbf{x}}' \\ \bar{\mathbf{x}}, \bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})}} \lambda_1(\mathbf{D}_{\bar{\mathbf{e}}}^H \mathbf{R}_{\mathbb{H}}^T \mathbf{D}_{\bar{\mathbf{e}}}). \tag{5.2}$$

Recall that the notation Ξ_1^{ρ} implies that the effective codeword difference matrix has ρ nonzero eigenvalues. The dependence of (5.2) on $\mathbf{R}_{\mathbb{H}}$ leads to different design criteria depending on the channel selectivity characteristics.

In a flat-fading channel, where $\mathbf{R}_{\mathbb{H}} = \mathbb{1}_N$ and $\rho = 1$, it follows from (5.2) that

$$\Xi_1^{1}(\mathsf{SNR}) = \min_{\substack{\bar{\mathbf{e}} = \bar{\mathbf{x}} - \bar{\mathbf{x}}' \\ \bar{\mathbf{x}}, \bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})}} ||\bar{\mathbf{e}}||^2$$

which is simply the minimum Euclidian distance between codewords. Hence, by Theorem 4.1, we can conclude that the decay rate of the Euclidian distance constitutes the criterion for optimal performance over single-transmit-antenna flat-fading channels.

In fast fading, the channel covariance matrix satisfies $\mathbf{R}_{\mathbb{H}} = \mathbf{I}$ with $\rho = N$; hence, the sufficient condition for optimality follows from (5.2) and Theorem 4.1 as

$$\Xi_1^{N}(\mathsf{SNR}) = \min_{\substack{\bar{\mathbf{e}} = \bar{\mathbf{x}} - \bar{\mathbf{x}}' \\ \bar{\mathbf{x}}, \bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})}} \min_n |\bar{\mathbf{e}}(n)|^2 \,\dot{\geq}\, \mathsf{SNR}^{-(r-\epsilon)}$$

which implies that no entry in the codeword difference vector $\bar{\mathbf{e}}$ can be equal to zero. We show next that the class of codes that satisfy this condition is of general interest because they perform optimally irrespectively of the covariance matrix $\mathbf{R}_{\mathbb{H}}$.

Lemma 5.1. If there exists an $\epsilon > 0$ such that

$$\min_{\substack{\bar{\mathbf{e}}=\bar{\mathbf{x}}-\bar{\mathbf{x}}' \\ \bar{\mathbf{x}},\bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})}} \min_n |\bar{\mathbf{e}}(n)|^2 \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)} \tag{5.3}$$

then $\mathcal{C}_r$ is optimal over the single-transmit-antenna channel irrespectively of $\mathbf{R}_{\mathbb{H}}$.

Proof. Applying Ostrowski's Theorem (Horn and Johnson, 1993, 4.5.9) to the effective codeword difference matrix yields

$$\lambda_n(\mathbf{D}_{\bar{\mathbf{e}}}^H \mathbf{R}_{\mathbb{H}}^T \mathbf{D}_{\bar{\mathbf{e}}}) = \theta\, \lambda_n(\mathbf{R}_{\mathbb{H}}),\ n = 0, 1, \ldots, N-1,$$

where $\theta \in [\min_n |\bar{\mathbf{e}}(n)|^2, \max_n |\bar{\mathbf{e}}(n)|^2]$. Hence, by (5.3), every $\bar{\mathbf{e}}$ arising form $\mathcal{C}_r$ satisfies

$$\lambda_k(\mathbf{D}_{\bar{\mathbf{e}}}^H \mathbf{R}_{\mathbb{H}}^T \mathbf{D}_{\bar{\mathbf{e}}}) \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)}\, \lambda_k(\mathbf{R}_{\mathbb{H}}), \quad k = 1, 2 \ldots, \rho. \tag{5.4}$$

Since the eigenvalues of $\mathbf{R}_{\mathbb{H}}$ are not functions of SNR, we can conclude from (5.4) that $\Xi_1^{\rho}(\mathsf{SNR}) \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)}$, implying that $\mathcal{C}_r$ is optimal. □

Noteworthily, uncoded* quadrature amplitude modulation (QAM) constellations satisfy the condition of Lemma 5.1. Indeed, the minimum distance in a QAM constellation with SNR^r points scales as

*By "uncoded" we mean that the entries of $\mathbf{x}$ have no dependency across slots, but, still, any two codewords differ in all their entries. More precisely, if $\mathcal{A}$ is the underlying QAM constellation with SNR^r points, then, for every $n = 0, 1, \ldots, N-1$, the equality $\mathcal{A} = \{\mathbf{x}(n) : \mathbf{x} \in \mathcal{C}_r(\mathsf{SNR})\}$ holds true. In contrast, in (Tse and Viswanath, 2005, Sec. 9.2.2), uncoded QAM can result in two codewords having the same symbol at the same position.

$d^2_{\text{min}} \approx \mathsf{SNR}^{-r}$ (Tse and Viswanath, 2005, Sec. 9.1.2). With one such constellation for each slot, the system operates at data rate $R(\mathsf{SNR}) = r \log \mathsf{SNR}$ and the codeword difference vectors $\bar{\mathbf{e}}$ satisfy (5.3) as $\epsilon \to 0$. Hence, the optimality (in the sense of the DM tradeoff) of uncoded QAM over the single-transmit-antenna flat-fading Rayleigh channel (Tse and Viswanath, 2005, Sec. 9.1.2) also holds for selective-fading channels.

Condition (5.3) can be related to the product distance (Divsalar and Simon, 1988) of $\mathcal{C}_r$ defined as

$$\Pi_{\mathcal{C}_r}(\mathsf{SNR}) \triangleq \min_{\substack{\bar{\mathbf{e}}=\bar{\mathbf{x}}-\bar{\mathbf{x}}' \\ \bar{\mathbf{x}},\bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})}} \prod_{n=0}^{N-1} |\bar{\mathbf{e}}(n)|^2. \tag{5.5}$$

Note that the power constraint implies that $|\bar{\mathbf{e}}(n)|^2 \dot{\leq} 1$ for any n. We hence have

$$\Pi_{\mathcal{C}_r}(\mathsf{SNR}) \dot{\leq} \min_{\bar{\mathbf{e}},n} |\bar{\mathbf{e}}(n)|^2$$

and, consequently, if the product distance of $\mathcal{C}_r$ satisfies

$$\Pi_{\mathcal{C}_r}(\mathsf{SNR}) \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)}, \ \epsilon > 0 \tag{5.6}$$

then, by Lemma 5.1, $\mathcal{C}_r$ is DM tradeoff optimal. The criterion of maximizing the product distance of $\mathcal{C}_r$ is known to be optimal (for fixed rates w.r.t. SNR) in i.i.d. Rayleigh fading scalar channels (Divsalar and Simon, 1988; Boutros and Viterbo, 1998) and has led to the construction of a variety of codes.

Note that the minimum product distance in transmitting uncoded QAM constellations scales approximately as SNR^{-Nr}, i.e., the worst case product distance is obtained for two codewords separated by the minimum distance SNR^{-r} in each constellation. Condition (5.6) can only be guaranteed by coding across slots, and hence, it is not a useful criterion for the single-transmit-antenna channel where uncoded transmissions are optimal. In contrast, in the multiple-transmit-antenna case, we will see that (5.6) is more useful.

5.2. MULTIPLE TRANSMIT ANTENNAS

In contrast to the single-transmit-antenna case, optimal code constructions for the multiple-transmit-antenna channel has to deal with interference between the signals transmitted from different antennas. In this section, we develop a new procedure to construct codes achieving optimal performance.

5.2.1. Preliminaries

Precoding is an effective technique to simplify data transmission and reception over wireless channels, and the task of designing good codes is often made easier by the use of an appropriate precoder. In essence, the method consists in processing the transmit signal in a manner that takes advantage of the knowledge that the transmitter has on the channel. Depending on the channel characteristics and the degree of channel knowledge at the transmitter, a variety of approaches to precoding have been developed in the past.

With perfect transmit channel state information (CSI), Tomlinson-Harashima precoding (Tomlinson, 1971; Harashima and Miyakawa, 1972) pre-processes the data symbols transmitted over a single-input single-ouput (SISO) frequency-selective channel so that no intersymbol interference is present at the receiver. In fact, this technique is the transmitter dual to receiver-based decision-feedback equalization. In the multiple-input multiple-output (MIMO) case, full CSI at both the transmitter and at the receiver can be used to design the transmit filter, i.e., the precoder, and the receive filter so as to diagonalize the channel matrix and obtain a set of parallel non-interfering spatial channels for communication. The filters can be optimized according to different performance measures, e.g., to maximize rate, or to reduce error probability (Sampath et al., 2001; Scaglione et al., 2002) and (Paulraj et al., 2003, Chapter 8). Because these methods assume up-to-date CSI at the transmitter, their applicability is restricted to slowly-varying channels.

In a second, drastically different, approach to precoding, the transmitter is only required to know the delay spread τ_0 and the Doppler spread ν_0 of the channel. This type of precoding*, which aims at exploiting transmit antenna diversity, is often referred to as delay diversity, or phase-roll diversity (e.g., Adachi, 1979; Hiroike et al., 1992; Wittneben, 1993; Weerackody, 1993; Seshadri and Winters, 1994; Kuo and Fitz, 1997; Wornell and Trott, 1997; Kaiser, 2000). The basic principle shared by these techniques is to convert transmit antenna diversity into selective diversity (either in time, in frequency, or in time-frequency), and exploit it through standard forward error correction (FEC) coding and interleaving. Thanks to the precoder, the design of the outer code becomes simpler as it is not affected by the presence of multiple transmit antennas.

In this section, we investigate the latter type of precoding. Based on the code design criterion for DM tradeoff optimality presented in Chapter 4, we show that the design problem can be separated in two simpler and independent problems: the design of an inner code, or precoder, adapted to the channel statistics and an outer code independent of the channel statistics. The inner code can be obtained in a systematic fashion as a function of the selectivity characteristics of the channel and the design criterion for the outer code is standard.

5.2.2. Precoding setup

Consider an *outer family of codes* $\mathcal{C}_r$ of block length N and an $\mathrm{M_T} \times N$ precoder $\mathbf{P}$. The transmitted codeword $\tilde{\mathbf{X}}$ is obtained as the element-wise multiplication of the original codeword and the precoder according to

$$\tilde{\mathbf{X}} = \mathbf{P} \odot \mathbf{X}. \tag{5.7}$$

We shall denote the rows of $\mathbf{X}$ by $\bar{\mathbf{x}}_l$, and the rows of $\mathbf{P}$ by $\bar{\mathbf{p}}_l$. The transmitter structure is depicted in Figure 5.1 part (i). For two code-

*The terminology "linear precoding" to refer to these techniques is used in Wornell and Trott (1997).

words $\mathbf{X}$ and $\mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})$, the precoded codeword difference matrix is given by $\tilde{\mathbf{E}} = \mathbf{P} \odot \mathbf{E}$, where $\mathbf{E} = \mathbf{X} - \mathbf{X}'$. Denoting the rows of $\mathbf{E}$ by $\bar{\mathbf{e}}_l$, we have

$$\begin{aligned}\tilde{\mathbf{E}}^H\tilde{\mathbf{E}} &= \sum_{l=1}^{\mathrm{M_T}} (\bar{\mathbf{p}}_l \odot \bar{\mathbf{e}}_l)^H(\bar{\mathbf{p}}_l \odot \bar{\mathbf{e}}_l) \\ &= \sum_{l=1}^{\mathrm{M_T}} \mathbf{D}_{\bar{\mathbf{e}}_l}^H \, \bar{\mathbf{p}}_l^H\bar{\mathbf{p}}_l \, \mathbf{D}_{\bar{\mathbf{e}}_l} \qquad (5.8) \\ &= \sum_{l=1}^{\mathrm{M_T}} \bar{\mathbf{p}}_l^H\bar{\mathbf{p}}_l \odot \bar{\mathbf{e}}_l^H\bar{\mathbf{e}}_l\end{aligned}$$

where we have used $\bar{\mathbf{p}}_l \odot \bar{\mathbf{e}}_l = \bar{\mathbf{p}}_l \, \mathbf{D}_{\bar{\mathbf{e}}_l}$ to get (5.8). For the sake of notation, let us introduce

$$\mathbf{R}_l = \mathbf{D}_{\bar{\mathbf{p}}_l}^H \, \mathbf{R}_{\mathbb{H}}^T \, \mathbf{D}_{\bar{\mathbf{p}}_l}, \quad l = 1, 2, \ldots, \mathrm{M_T}. \qquad (5.9)$$

Since, trivially, $\mathbf{R}_l \odot \bar{\mathbf{e}}_l^H\bar{\mathbf{e}}_l = \mathbf{D}_{\bar{\mathbf{e}}_l}^H \, \mathbf{R}_l \, \mathbf{D}_{\bar{\mathbf{e}}_l}$ ($l = 1, 2, \ldots, \mathrm{M_T}$), the effective codeword difference matrix can be written as

$$\mathbf{R}_{\mathbb{H}}^T \odot \tilde{\mathbf{E}}^H\tilde{\mathbf{E}} = \sum_{l=1}^{\mathrm{M_T}} \mathbf{D}_{\bar{\mathbf{e}}_l}^H \, \mathbf{R}_l \, \mathbf{D}_{\bar{\mathbf{e}}_l}. \qquad (5.10)$$

The optimal code design criterion follows from (4.41) as

$$\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}(\mathsf{SNR}) \,\dot{\geq}\, \mathsf{SNR}^{-(r-\epsilon)}, \ \epsilon > 0 \qquad (5.11)$$

where the quantity $\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}$ specializes to

$$\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}(\mathsf{SNR}) = \min_{\substack{\mathbf{E}=\mathbf{X}-\mathbf{X}' \\ \mathbf{X},\mathbf{X}' \in \mathcal{C}_r(\mathsf{SNR})}} \prod_{k=1}^{\mathrm{m}} \lambda_k \left(\sum_{l=1}^{\mathrm{M_T}} \mathbf{D}_{\bar{\mathbf{e}}_l}^H \, \mathbf{R}_l \, \mathbf{D}_{\bar{\mathbf{e}}_l} \right). \qquad (5.12)$$

The matrix $\mathbf{R}_l$ can be thought of as an effective covariance matrix induced by $\bar{\mathbf{p}}_l$ on transmit antenna l. Put differently, signals transmitted

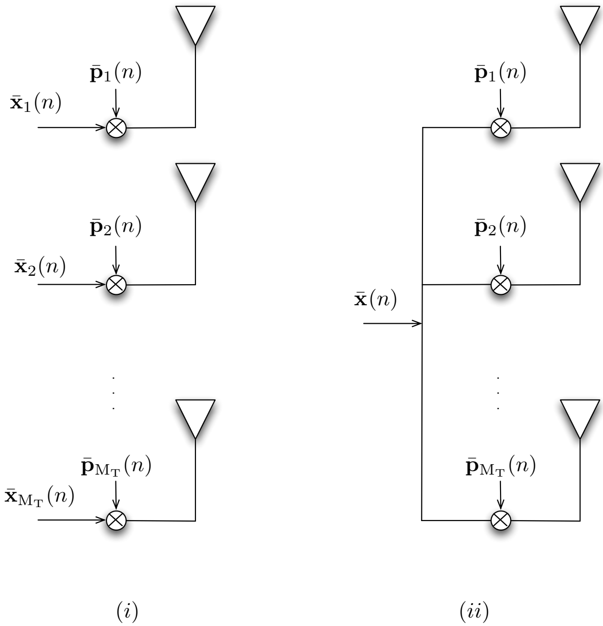

Fig. 5.1: Depiction of the transmitter with precoding for slot n. The general structure is depicted in (i), and part (ii) shows the structure assumed in Theorem 5.2.

from different antennas undergo different effective channels which can be controlled by precoder design. Our goal is to find a precoder that induces more structure in the effective codeword matrix so as to simplify the design of the outer code $\mathcal{C}_r$.

5.2.3. Design criteria with precoding

In conjunction with the precoding scheme defined above, we consider an outer family of codes $\mathcal{C}_r$ for which the same $1 \times N$ codeword vector

$\bar{\mathbf{x}} = [x_0\, x_1\, \ldots\, x_{N-1}]$ is transmitted over all antennas. The codeword difference vector is given by $\bar{\mathbf{e}} = \bar{\mathbf{x}} - \bar{\mathbf{x}}'$, where $\bar{\mathbf{x}}, \bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})$. The corresponding transmitter structure is illustrated in Figure 5.1-(ii). With this notation in place, next theorem provides code design criteria to achieve the optimal DM tradeoff in selective-fading multiple-transmit-antenna channels.

Theorem 5.2. *Let the family of codes $\mathcal{C}_r$, $r \in [0, \mathrm{m}]$, have block length $N \geq \rho\mathrm{M_T}$, and suppose that the signal transmitted on antenna l is given by*

$$\bar{\mathbf{p}}_l \odot \bar{\mathbf{x}},\ l = 1, 2, \ldots, \mathrm{M_T}, \tag{5.13}$$

where $\bar{\mathbf{x}} = [x_0\, x_1\, \cdots\, x_{N-1}] \in \mathcal{C}_r(\mathsf{SNR})$ and $\bar{\mathbf{p}}_l$ is the lth row of the precoder $\mathbf{P}$ ($\mathrm{M_T} \times N$). If there exists $\epsilon > 0$ such that $\mathcal{C}_r$ satisfies

$$\min_{\substack{\bar{\mathbf{e}}=\bar{\mathbf{x}}-\bar{\mathbf{x}}' \\ \bar{\mathbf{x}},\bar{\mathbf{x}}' \in \mathcal{C}_r(\mathsf{SNR})}} \prod_{n=0}^{\mathrm{m}-1} |\bar{\mathbf{e}}(\pi(n))|^2 \dot{\geq} \mathsf{SNR}^{-(r-\epsilon)} \tag{5.14}$$

where the permutation π sorts the entries of $\mathbf{e}$ in ascending magnitude, and $\mathbf{P}$ is such that

$$\mathrm{rank}\big(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\big) = \rho\mathrm{M_T} \tag{5.15}$$

then the pair of inner and outer codes $(\mathbf{P}, \mathcal{C}_r)$ achieves the optimal DM tradeoff.

Proof. We start by noting that since the same $1 \times N$ codeword $\mathbf{x}$ is transmitted over all antennas, we have $\bar{\mathbf{e}}_l = \bar{\mathbf{e}}$, which, upon insertion into (5.10), yields

$$\mathbf{R}_{\mathbb{H}}^T \odot \tilde{\mathbf{E}}^H\tilde{\mathbf{E}} = \sum_{l=1}^{\mathrm{M_T}} \mathbf{D}_{\bar{\mathbf{e}}_l}^H \mathbf{R}_l \mathbf{D}_{\bar{\mathbf{e}}_l} = \mathbf{D}_{\bar{\mathbf{e}}}^H \left(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\right) \mathbf{D}_{\bar{\mathbf{e}}}. \tag{5.16}$$

Next, condition (5.15) implies that exactly $\rho\mathrm{M_T}$ eigenvalues of $\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}$ are nonzero. With the eigenvalue decomposition $\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P} =$

$\mathbf{V\Sigma V}^H$, where we have defined $\mathbf{\Sigma} = \operatorname{diag}\left\{\tilde{\mathbf{\Sigma}}, 0, \ldots, 0\right\}$ with $\tilde{\mathbf{\Sigma}} = \operatorname{diag}\{\sigma_0, \sigma_1, \ldots, \sigma_{\rho \mathrm{M_T}-1}\}$ and the nonzero eigenvalues σ_i are sorted in ascending order, we get $\mathbf{R}_{\mathbb{H}}^T \odot \tilde{\mathbf{E}}^H \tilde{\mathbf{E}} = \mathbf{D}_{\bar{\mathbf{e}}}^H \mathbf{V\Sigma V}^H \mathbf{D}_{\bar{\mathbf{e}}}$. Using the fact that $\lambda_n(\mathbf{M}\mathbf{M}^H) = \lambda_n(\mathbf{M}^H\mathbf{M})$, $\forall n$, for a square matrix $\mathbf{M}$, we obtain

$$\begin{aligned}
\lambda_n(\mathbf{R}_{\mathbb{H}}^T \odot \tilde{\mathbf{E}}^H \tilde{\mathbf{E}}) &= \lambda_n(\mathbf{\Sigma}^{1/2} \underbrace{\mathbf{V}^H \mathbf{D}_{\bar{\mathbf{e}}} \mathbf{D}_{\bar{\mathbf{e}}}^H \mathbf{V}}_{\triangleq \mathbf{B}} \mathbf{\Sigma}^{1/2}) \\
&= \lambda_n(\tilde{\mathbf{\Sigma}}^{1/2} \tilde{\mathbf{B}} \tilde{\mathbf{\Sigma}}^{1/2}) \qquad (5.17) \\
&\geq \sigma_0 \, \lambda_n(\tilde{\mathbf{B}}) \qquad (5.18)
\end{aligned}$$

where $\tilde{\mathbf{B}}$ has dimension $\rho \mathrm{M_T} \times \rho \mathrm{M_T}$, (5.17) holds only for the nonzero eigenvalues of $\mathbf{B}$, and we have applied Ostrowski's Theorem (Horn and Johnson, 1993, 4.5.9) to get (5.18). Since $\mathbf{B}$ is Hermitian and $\tilde{\mathbf{B}}$ is its principal submatrix obtained by deleting the $N - \rho \mathrm{M_T}$ last rows and the corresponding columns from $\mathbf{B}$, we can invoke (Horn and Johnson, 1993, 4.3.15) to conclude that

$$\lambda_k(\tilde{\mathbf{B}}) \geq \lambda_k(\mathbf{B}) = |\bar{\mathbf{e}}(\pi(k))|^2, \; k = 0, 1, \ldots, \rho \mathrm{M_T} - 1 \qquad (5.19)$$

where π is an index permutation that sorts the entries of $\bar{\mathbf{e}}$ in ascending magnitude for every SNR level. Next, combining (5.18) with (5.19), we find that the nonzero eigenvalues of $\mathbf{R}_{\mathbb{H}}^T \odot \tilde{\mathbf{E}}^H \tilde{\mathbf{E}}$ satisfy

$$\lambda_k(\mathbf{R}_{\mathbb{H}}^T \odot \tilde{\mathbf{E}}^H \tilde{\mathbf{E}}) \geq \sigma_0 \, |\bar{\mathbf{e}}(\pi(k))|^2, \; k = 0, 1, \ldots, \rho \mathrm{M_T} - 1. \quad (5.20)$$

By (5.14), we can therefore conclude that

$$\Xi_{\mathrm{m}}^{\rho \mathrm{M_T}}(\mathsf{SNR}) \geq (\sigma_0)^{\mathrm{m}} \; \min_{\bar{\mathbf{e}}} \prod_{n=0}^{\mathrm{m}-1} |\bar{\mathbf{e}}(\pi(n))|^2 \,\dot{\geq}\, \mathsf{SNR}^{-(r-\epsilon)}.$$

□

Theorem 5.2 says that it is possible to split the code design problem into the design of a precoder adapted to $\mathbf{R}_{\mathbb{H}}$ (the precoder effectively decorrelates the channel into its independent diversity branches) and the design of a single-antenna outer code $\mathcal{C}_r$ which is independent of the channel selectivity characteristics. The design of the precoder shall be examined in detail in the next section. We shall consider before an example of outer code that meets the requirement (5.14).

5.2.4. Outer family of codes

The QAM-based permutation codes proposed in Tavildar and Viswanath (2006) in the context of the parallel channel satisfy (5.14). To see this, we start by recalling that the problem addressed in (Tavildar and Viswanath, 2006, V.B) is the construction of space-only codes (i.e., unit-block length) that are approximately universal over the parallel channel with L flat-fading subchannels. Let x_l be the lth entry of such a code, with $l = 1, 2, \ldots, L$, denoting the subchannel index, and assume it belongs to a QAM constellation with SNR^r points. Denote the symbol difference corresponding to subchannel l by $e_l = x_l - x'_l$. Then, a sufficient condition for approximate universality is given by (Tavildar and Viswanath, 2006, Th. 5.1) as

$$|e_1|^2 \cdot |e_2|^2 \cdots |e_L|^2 \dot{>} 2^{-R(\mathsf{SNR})} \quad \text{for all } e_1, e_2, \ldots, e_L \qquad (5.21)$$

where the data rate satisfies $R(\mathsf{SNR}) = r \log \mathsf{SNR}$, $r \in [0, L]$. The existence of codes satisfying this condition is guaranteed by (Tavildar and Viswanath, 2006, Th. 5.2). The proof of this result consists in averaging the product distance $\Pi_{\mathcal{C}_r}$ over the ensemble of random permutations codes (i.e., codes obtained by randomly permuting the points labels in successive QAM constellations), and showing that there must be codes within the ensemble that satisfy (5.21).

This code design criterion is not fundamentally different from the condition on the outer family of codes given in Theorem 5.2. In fact, mapping the spatial dimension in (5.21) to the temporal dimension and setting $L = N$, we can conclude from (Tavildar and Viswanath, 2006, Th. 5.2) that there exist family of codes $\mathcal{C}_r$, whose codewords

$\bar{\mathbf{x}} = [x_0\, x_1\, \cdots\, x_{N-1}]$ are transmitted across slots, that satisfy

$$|\bar{\mathbf{e}}(0)|^2 \cdot |\bar{\mathbf{e}}(1)|^2 \cdots |\bar{\mathbf{e}}(N-1)|^2 \dot{>} 2^{-R(\mathsf{SNR})}$$
$$\text{for all } \bar{\mathbf{e}}(0), \bar{\mathbf{e}}(1), \ldots, \bar{\mathbf{e}}(N-1) \quad (5.22)$$

and $R(\mathsf{SNR}) = r \log \mathsf{SNR}$, $r \in [0, N]$. Due to the power constraint (5.1) on the codewords of $\mathcal{C}_r$, we necessarily have $|\bar{\mathbf{e}}(n)|^2 \dot{\leq} 1$ for every n and, therefore, the product of the m smallest $|\bar{\mathbf{e}}(n)|^2$ does also satisfy the decay rate of (5.22). Hence, we can conclude that the permutation codes of Tavildar and Viswanath (2006) satisfy (5.14). The development of practical codes based on QAM permutations with appropriate product distance seems to remain largely unexplored.

5.3. PRECODER DESIGN

The essence of the following precoder construction consists in approximating the (two-level*) Toeplitz matrix $\mathbf{R}_{\mathbb{H}}$ by a (two-level) circulant matrix $\mathbf{C}_{\mathbb{H}}$. Indeed, the latter matrix admits a precoder $\mathbf{P}$ for which

$$\operatorname{rank}\big(\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\big) = \mathrm{M_T} \cdot \operatorname{rank}(\mathbf{C}_{\mathbb{H}})$$

holds true for any block length $N \geq \mathrm{M_T} \cdot \operatorname{rank}(\mathbf{C}_{\mathbb{H}})$. By standard results on the asymptotic eigenvalue distribution of $\mathbf{C}_{\mathbb{H}}$ and $\mathbf{R}_{\mathbb{H}}$ (Gray, 2006; Voois, 1996; Tyrtyshnikov, 1996), we will argue that for reasonably large block lengths $\mathbf{P}$ also satisfies

$$\operatorname{rank}\big(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\big) = \mathrm{M_T} \cdot \operatorname{rank}(\mathbf{R}_{\mathbb{H}})$$

which is to say that $\mathbf{P}$ meets the condition of Theorem 5.2. We shall show that such a precoder is a generalized version of what is commonly known as delay and phase-roll diversity (Adachi, 1979; Hiroike et al.,

*A two-level Toeplitz matrix is a Toeplitz-block-Toeplitz matrix. Similarly, a two-level circulant matrix is a circulant-block-circulant matrix.

1992; Wittneben, 1993; Weerackody, 1993; Seshadri and Winters, 1994; Kuo and Fitz, 1997; Wornell and Trott, 1997; Kaiser, 2000).

We start our presentation by considering the general case, i.e., $\mathbf{R}_{\mathbb{H}}$ is two-level Toeplitz, and then show how our result applies to the particular cases of Toeplitz and circulant covariance matrices.

5.3.1. Two-level Toeplitz covariance matrix

Since the time-varying transfer function $L_{\mathbb{H}}(t, f)$ is assumed to be wide-sense stationary both in t and f, the $N \times N$ covariance matrix $\mathbf{R}_{\mathbb{H}}$ is a two-level Toeplitz matrix with entries given by the sampled time-frequency correlation as

$$\mathbf{R}_{\mathbb{H}}(n, n') = R_{\mathbb{H}}((m - m')T, (k - k')F) \tag{5.23}$$

where $n = \mathcal{M}(m, k)$, $n' = \mathcal{M}(m', k')$, and $\mathcal{M}(\cdot, \cdot)$ is defined in (2.14). From standard results on the eigenvalue distribution of (two-level) Toeplitz matrices (Gray, 2006; Voois, 1996; Tyrtyshnikov, 1996), the eigenvalues of $\mathbf{R}_{\mathbb{H}}$ are asymptotically ($M, K \to \infty$) distributed as the 2-dimensional power spectral density given by

$$S(\xi, \mu) = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} R_{\mathbb{H}}(mT, kF)\, e^{-j2\pi(\mu m - \xi k)}, \quad 0 \le \mu, \xi < 1$$

where, recalling (2.23), $S(\xi, \mu)$ can be related to the channel scattering function as

$$S(\xi, \mu) = \frac{1}{TF} C_{\mathbb{H}}\left(\frac{\xi}{F}, \frac{\mu}{T}\right), \; 0 \le \xi, \mu < 1.$$

Motivated by the fact that uniformly sampling $S(\xi, \mu)$ yields good approximations of the eigenvalues of $\mathbf{R}_{\mathbb{H}}$ (Gray, 2006; Voois, 1996), we define a $N \times N$ two-level circulant matrix $\mathbf{C}_{\mathbb{H}}$ by setting

$$\lambda_n(\mathbf{C}_{\mathbb{H}}) \triangleq S\left(\frac{k}{K}, \frac{m}{M}\right) \tag{5.24}$$

for $m = 0, 1, \ldots, M-1$ and $k = 0, 1, \ldots, K-1$. Note that (5.24) completely determines the matrix $\mathbf{C}_{\mathbb{H}}$ because two-level circulant matrices are diagonalized by two-level Fourier matrices (Voois, 1996; Tyrtyshnikov, 1996). Hence, setting $\mathbf{F} = \mathbf{\Psi} \otimes \mathbf{\Phi}$, where $\mathbf{\Psi}$ and $\mathbf{\Phi}$ denote, respectively, the Fourier matrices of size M and K, we have

$$\mathbf{C}_{\mathbb{H}} = \mathbf{F}\mathbf{\Lambda}\mathbf{F}^H \tag{5.25}$$

where $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_n(\mathbf{C}_{\mathbb{H}})\}_{n=0}^{N-1}$ is given by (5.24). We recall that the scattering function is compactly supported in the rectangle $[0, \tau_0] \times [0, \nu_0]$ and, hence, it follows from (5.24) that the nonzero eigenvalues of $\mathbf{C}_{\mathbb{H}}$ are indexed by

$$(m, k) \in \{0, 1, \ldots, v-1\} \times \{0, 1, \ldots, t-1\} \tag{5.26}$$

where

$$v \triangleq \lfloor \nu_0 T M \rfloor \quad \text{and} \quad t \triangleq \lfloor \tau_0 F K \rfloor. \tag{5.27}$$

We recall that the grid parameters satisfy the Nyquist condition $T \leq \frac{1}{\nu_0}$ and $F \leq \frac{1}{\tau_0}$ (we exclude therefore the cases $\nu_0 = 0$ and $\tau_0 = 0$). For grid parameters satisfying $\nu_0 T = \tau_0 F = 1$, i.e., critical channel sampling, we have $v = M$ and $t = K$, implying that $\mathbf{C}_{\mathbb{H}}$ is full rank. In general, the rank of $\mathbf{C}_{\mathbb{H}}$ is given by the product $v \cdot t$, which is the number of nonzero elements in the set $\{S\big(\frac{k}{K}, \frac{m}{M}\big)\}$.

The following theorem presents a precoder $\mathbf{P}$ that satisfies the condition $\mathrm{rank}\big(\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\big) = \rho \mathrm{M_T}$. The idea is to apply different time-frequency shifts to the channel spectra corresponding to different antennas. As mentioned in the outset, our construction is a general version of delay and phase-roll diversity.

Theorem 5.3. *Consider the $N \times N$ 2-level circulant matrix $\mathbf{C}_{\mathbb{H}}$ with rank vt defined by (5.24) and (5.25). If $N \geq vt\mathrm{M_T}$ and the rows of $\mathbf{P}$ satisfy*

$$\bar{\mathbf{p}}_l = \boldsymbol{\psi}_{p_l v}^T \otimes \boldsymbol{\phi}_{q_l t}^T, \;\; \textit{for} \;\; l = 1, 2, \ldots, \mathrm{M_T} \tag{5.28}$$

where ψ_m *and* ϕ_k *are, respectively, the* m*th and* k*th columns of the Fourier matrices* $\mathbf{\Psi}$ *and* $\mathbf{\Phi}$*, and*

$$(p_l, q_l) \in \left\{0, 1, \ldots, \left\lfloor \frac{1}{\nu_0 T} \right\rfloor - 1\right\} \times \left\{0, 1, \ldots, \left\lfloor \frac{1}{\tau_0 F} \right\rfloor - 1\right\}, \\ (p_l, q_l) \neq (p_{l'}, q_{l'}) \textit{ for } l \neq l' \quad \textit{(5.29)}$$

then $\operatorname{rank}\left(\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H \mathbf{P}\right) = \rho \mathrm{M_T}$.

Proof. We start by noting that $\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H \mathbf{P}$ can be written as

$$\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H \mathbf{P} = \sum_{l=1}^{\mathrm{M_T}} \underbrace{\mathbf{D}_{\bar{\mathbf{p}}_l^H} \mathbf{C}_{\mathbb{H}}^T \mathbf{D}_{\bar{\mathbf{p}}_l}}_{\triangleq \mathbf{C}_l}. \tag{5.30}$$

Next, consider the following similarity transformation

$$\mathbf{F}^T \mathbf{C}_l \mathbf{F}^* = \mathbf{F}^T \mathbf{D}_{\bar{\mathbf{p}}_l^H} \mathbf{F}^* \mathbf{\Lambda} \mathbf{F}^T \mathbf{D}_{\bar{\mathbf{p}}_l} \mathbf{F}^* \tag{5.31}$$

where we have used $\mathbf{C}_{\mathbb{H}} = \mathbf{F} \mathbf{\Lambda} \mathbf{F}^H$. With (5.28) and $\mathbf{F} = \mathbf{\Psi} \otimes \mathbf{\Phi}$, we get

$$\begin{aligned} \mathbf{F}^T \mathbf{D}_{\bar{\mathbf{p}}_l^H} \mathbf{F}^* &= \left(\mathbf{\Psi}^T \mathbf{D}_{\boldsymbol{\psi}^*_{p_l v}} \mathbf{\Psi}^*\right) \otimes \left(\mathbf{\Phi}^T \mathbf{D}_{\boldsymbol{\phi}^*_{q_l t}} \mathbf{\Phi}^*\right) \\ &= \mathbf{\Pi}^{p_l v} \otimes \mathbf{\Pi}^{q_l t} \end{aligned} \tag{5.32}$$

where $\mathbf{\Pi} = [\boldsymbol{\pi}_1 \ \cdots \ \boldsymbol{\pi}_{N-1} \ \boldsymbol{\pi}_0]$ is the circulant permutation matrix with $\boldsymbol{\pi}_n = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$ containing a 1 in its nth position. Using (5.32) in (5.31), we obtain

$$\mathbf{F}^T \mathbf{C}_l \mathbf{F}^* = \left(\mathbf{\Pi}^{p_l v} \otimes \mathbf{\Pi}^{q_l t}\right) \mathbf{\Lambda} \left(\mathbf{\Pi}^{p_l v} \otimes \mathbf{\Pi}^{q_l t}\right)^T \tag{5.33}$$

and consequently

$$\mathbf{F}^T \left(\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H \mathbf{P}\right) \mathbf{F}^* = \sum_{l=1}^{\mathrm{M_T}} \left(\mathbf{\Pi}^{p_l v} \otimes \mathbf{\Pi}^{q_l t}\right) \mathbf{\Lambda} \left(\mathbf{\Pi}^{p_l v} \otimes \mathbf{\Pi}^{q_l t}\right)^T. \tag{5.34}$$

Since $\left(\mathbf{\Pi}^k \otimes \mathbf{\Pi}^l\right) \mathbf{\Lambda}\left(\mathbf{\Pi}^k \otimes \mathbf{\Pi}^l\right)^T$ simply permutes the entries of $\mathbf{\Lambda}$ along the main diagonal, the rank of $\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}$ is trivially upper bounded by $\rho \mathrm{M_T}$. To achieve this maximum, we need to ensure that the different shifts in (5.34) distribute the ρ eigenvalues of $\mathbf{C}_{\mathbb{H}}$ in mutually orthogonal subspaces. This can be accomplished as follows. With (5.26) and (5.33), we find that the indices (m, k) corresponding to the nonzero eigenvalues of $\mathbf{C}_l$ are given by the set

$$\mathcal{I}_l = \{p_l v, p_l v + 1, \ldots, (p_l + 1)v - 1\} \times \\ \{q_l t, q_l t + 1, \ldots, (q_l + 1)t - 1\}$$

that is, the nonzero eigenvalues of $\mathbf{C}_l$ are obtained by cyclically shifting the eigenvalues of $\mathbf{C}_{\mathbb{H}}$ by $p_l v$ positions along index m and $q_l t$ positions along index k. The condition in (5.29) guarantees that $\mathcal{I}_l \cap \mathcal{I}_{l'} = \emptyset$ for $l \neq l'$, which together with $\rho = vt$ in turn ensures that $\operatorname{rank}\left(\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\right) = \rho \mathrm{M_T}$. □

From (5.30) and the fact that $\mathbf{D}_{\bar{\mathbf{p}}_l}^H \mathbf{D}_{\bar{\mathbf{p}}_l} = \mathbf{I}$, it follows that the precoder of Theorem 5.3 performs a unitary transformation – a different one for each transmit antenna – on the channel covariance matrix. Hence, the original covariance matrix and the transformed covariance matrix have the same eigenvalues. The transformation, however, shifts the eigenvalues of $\mathbf{C}_l$ to different eigenspaces. The choice of shifts proposed in Theorem 5.3 guarantees that the positive eigenvalues of any two $\mathbf{C}_l$ are placed in mutually orthogonal subspaces so that $\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}$ has full $\rho \mathrm{M_T}$-rank. This idea is illustrated in Figure 5.2 where the original channel spectrum gets shifted in the $(\tau F, \nu T)$-plane by $\bar{\mathbf{p}}_l$. The precoder $\mathbf{P}$ is designed so that the spectra corresponding to different l do not overlap in the $(\tau F, \nu T)$-plane; the figure illustrates this aspect for the spectra of $\mathbf{C}_{\mathbb{H}}$ and $\mathbf{C}_l$.

We emphasize that the precoder of Theorem 5.3 does not guarantee $\operatorname{rank}\left(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\right) = \rho \mathrm{M_T}$ for every block length N. However, since the eigenvalues of $\mathbf{C}_{\mathbb{H}}$ are good approximations of the eigenvalues of

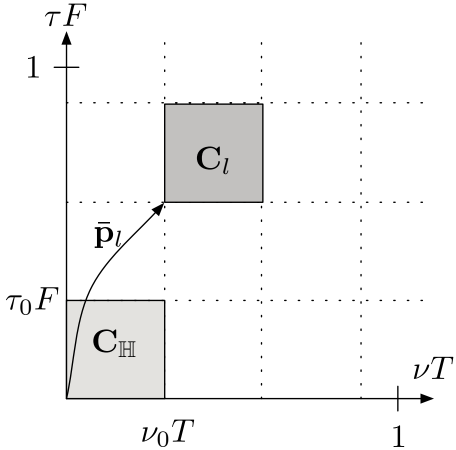

Fig. 5.2: Location of the (positive) eigenvalue spectra of $\mathbf{C}_{\mathbb{H}}$ and $\mathbf{C}_l$. The matrix $\mathbf{C}_l$, which is obtained using the precoder $\bar{\mathbf{p}}_l$ of Theorem 5.3, has the same eigenvalues as $\mathbf{C}_{\mathbb{H}}$ but time-frequency shifted (here by $\nu_0 T$ along the normalized Doppler frequency axis and $2\tau_0 F$ along the normalized delay axis).

$\mathbf{R}_{\mathbb{H}}$ for reasonably large N (Gray, 2006; Voois, 1996), we expect the precoder to yield good results for $N \geq \rho \mathrm{M_T}$.

5.3.2. Single-level covariance matrix

The technique of Theorem 5.3 can readily be applied to single-level Toeplitz matrices. For instance, suppose that $\tau_0 F = 1$ and $K = 1$ so that $t = 1$ in (5.27), and the $N \times N$ circulant $\mathbf{C}_{\mathbb{H}} = \mathbf{\Psi \Lambda \Psi}^H$ has rank v. Then, using the same arguments as in the two-level case, it can readily be seen that a precoder $\mathbf{P}$ satisfying

$$\bar{\mathbf{p}}_l = \boldsymbol{\psi}^T_{(l-1)v}, \quad l = 1, 2, \ldots, \mathrm{M_T} \tag{5.35}$$

where $\boldsymbol{\psi}_n$ is the nth column of the $N \times N$ Fourier matrix $\boldsymbol{\Psi}$, yields $\operatorname{rank}\big(\mathbf{C}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\big) = v\mathrm{M_T}$. As in the two-level case, this construction is expected to yield $\operatorname{rank}\big(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{P}^H\mathbf{P}\big) = \rho\mathrm{M_T}$ for $N \geq \rho\mathrm{M_T}$ due to the asymptotic relation between circulant and Toeplitz matrices.

When $\mathbf{R}_{\mathbb{H}}$ is circulant, like in the linear frequency-invariant (LFI) and linear time-invariant (LTI) systems considered in Appendix A, the precoder of Theorem 5.3 yields exact results because $\mathbf{C}_{\mathbb{H}} = \mathbf{R}_{\mathbb{H}}$. In this case, the construction of Theorem 5.3 is a special case of the precoder proposed in (Bölcskei et al., 2003b) which is defined by $\mathbf{P}^T = \boldsymbol{\Psi}\mathbf{A}$. In general, the $N \times \mathrm{M_T}$ matrix $\mathbf{A}$ can be constructed using the method proposed in (Bölcskei et al., 2003b, Th.1). The construction that we propose corresponds to the choice $\mathbf{A} = [\boldsymbol{\pi}_0\ \boldsymbol{\pi}_\rho\ \ldots\ \boldsymbol{\pi}_{(\mathrm{M_T}-1)\rho}]$.

5.3.3. Relation to delay and phase-roll diversity

As mentioned in the outset, the precoder of Theorem 5.3 is a generalization of well-known transmit diversity techniques that convert spatial diversity into time or frequency diversity and exploit it through forward error correcting codes and interleaving. In order to see this, note from (5.28) that the signal transmitted from the lth antenna is multiplied by

$$\bar{\mathbf{p}}_l(n) = \exp\left(-j2\pi\left(p_l v\frac{m}{M} + q_l t\frac{k}{K}\right)\right) \text{ for } n = 0, 1, \ldots, N-1 \tag{5.36}$$

where the pair (m, k) is related to the slot index n by $\mathcal{M}(m, k) = n$. Along the time index m, (5.36) introduces a frequency offset across transmit antennas (Adachi, 1979; Hiroike et al., 1992; Weerackody, 1993; Kuo and Fitz, 1997; Wornell and Trott, 1997; Kaiser, 2000), where, recalling that $v/M \approx \nu_0 T$, the offset on each antenna is an integer multiple of $\nu_0 T$. Along the frequency index k, (5.36) induces a time offset, i.e. a delay, across transmit antennas (Wittneben, 1993; Seshadri and Winters, 1994; Wornell and Trott, 1997; Kaiser, 2000), where, since $t/K \approx \tau_0 F$, the offset is a integer multiple of $\tau_0 F$. According to Theorem 5.2, these transmit diversity techniques can achieve the optimal DM tradeoff when used in conjunction with a good outer code. While phase and delay diversity are known to get good performance in systems with $\mathrm{M_R} = 1$ and, hence, their optimality in the sense of the DM tradeoff is not surprising in that case, our results show that they can also be employed in MIMO systems without loss of optimality.

CHAPTER 6

Multiple-Access Channel

How does multiuser interference affect the ultimate performance of communication over selective-fading channels? The present chapter addresses this issue in the case where multiple independent transmitters interfere in communicating with a single receiver. This type of communication channel, known as multiple-access (MA) channel, has attracted considerable attention, especially over the past few years. In selective-fading MA channels, however, a characterization of the diversity-multiplexing (DM) tradeoff remains to date unknown. Moreover, appropriate code construction guidelines to guarantee optimal performance are still missing.

Building on the techniques introduced in Chapter 4, we shall bridge this gap by establishing the optimal DM tradeoff of selective-fading MA multiple-input multiple-output (MIMO) channels. Our results provide a detailed analysis of the limitations that multiuser interference puts on rate and reliability. On the conceptual level, we find an interesting relation between the DM tradeoff framework and the notion of error event regions which was first introduced in additive white Gaussian noise (AWGN) channels by Gallager (1985). This relation leads to an accurate characterization of the error mechanisms in fading MA MIMO channels, which, quiet remarkably, follow a similar pattern as in the AWGN case. In particular, our analysis based on the notion of dominant error event, which characterizes the total error probability also in selective MA channels, reveals the existence of operational regimes in which multiuser interference has only a negligible impact on performance.

As a byproduct of our analysis, we obtain a set of code design criteria that guarantees optimal performance in the sense of the DM tradeoff. Despite the users' inability to cooperate at the time of communication, our criteria can be used to realize optimal performance by designing the codebooks beforehand. This is in sharp contrast with the prevalent approach to multiuser communication according to which users are assigned orthogonal resources, either in time or frequency, to minimize interference. The latter method is usually wasteful of communication resources since it ignores the structure of the underlying MA channel by reducing it to a set of point-to-point links. Instead, multiuser codes that satisfy our criteria are capable of achieving the ultimate performance limits promised by the DM tradeoff. The construction of distributed space-time codes that are optimal is nontrivial, and employing algebraic tools for this purpose seems to be a promising approach. We shall examine a particular example of such a construction for the 2-user flat-fading MA channel.

The rest of this chapter is organized as follows. In the next section, we introduce the signal and channel models in use throughout the chapter. Section 6.2 builds on Section 3.4 and examines how the channel selectivity affects fundamental performance measures like the ergodic capacity region and the outage probability. Based on these considerations, Section 6.3 tackles the problem of characterizing the optimal DM tradeoff of the selective-fading MA MIMO channel by extending the technique introduced for the point-to-point case in Chapter 4. Then, we examine in Section 6.4 the conceptual relation between the DM tradeoff framework and the notion of dominant error events. Section 6.5 concludes this chapter by analyzing an example of code construction which is DM tradeoff optimal over the 2-user flat-fading MA channel.

6.1. SYSTEM MODEL

The system model for the MA channel presented below builds on the considerations developed in Chapter 2. In particular, the signal model is based on the premise that the fading processes corresponding to all users are characterized by the same scattering function $C_{\mathbb{H}}(\tau,\nu)$ so that all channels can be diagonalized by the same Weyl-Heisenberg basis.

6.1.1. Signal model

We consider a selective-fading MA channel where U users, with $\mathrm{M_T}$ antennas each, communicate with a single receiver with $\mathrm{M_R}$ antennas. The corresponding input-output relation is given by

$$\mathbf{y}_n = \sqrt{\frac{\mathsf{SNR}}{\mathrm{M_T}}} \sum_{u=1}^{U} \mathbf{H}_{u,n}\, \mathbf{x}_{u,n} + \mathbf{z}_n,\ n = 0, 1, \ldots, N-1, \qquad (6.1)$$

where the index $n = \mathcal{M}(m,k)$ corresponds to a time, frequency or time-frequency slot in accordance with the mapping defined in (2.14) and SNR denotes the average per-user signal-to-noise ratio at each receive antenna. This input-ouput relation constitutes a natural extension of the point-to-point input-output relation in (4.1).

We denote the received signal matrix by $\mathbf{Y} = [\mathbf{y}_0\, \mathbf{y}_1\, \cdots\, \mathbf{y}_{N-1}]$ and, for every user $u \in \mathcal{U} \triangleq \{1, 2, \ldots, U\}$, the codebook corresponding to a given SNR level $\mathcal{C}_{r_u}(\mathsf{SNR})$ contains SNR^{Nr_u} codewords $\mathbf{X}_u = [\mathbf{x}_{u,0}\, \mathbf{x}_{u,1}\, \cdots\, \mathbf{x}_{u,N-1}]$ that satisfy the per-user power constraint

$$\|\mathbf{X}_u\|_{\mathrm{F}}^2 \le N\mathrm{M_T}, \forall\, \mathbf{X}_u \in \mathcal{C}_{r_u}(\mathsf{SNR}),\ \forall u \in \mathcal{U}. \qquad (6.2)$$

Moreover, for any subset of users $\mathcal{S} = \{u_1, u_2, \ldots, u_{|\mathcal{S}|}\} \subseteq \mathcal{U}$, we define the $\mathrm{M_T}|\mathcal{S}| \times N$ codeword corresponding to the users in $\mathcal{S}$ as

$$\mathbf{X}_{\mathcal{S}} = [\mathbf{X}_{u_1}^T\, \mathbf{X}_{u_2}^T\, \cdots\, \mathbf{X}_{u_{|\mathcal{S}|}}^T]^T. \qquad (6.3)$$

The overall family of codes $\mathcal{C}_{\boldsymbol{r}} = \mathcal{C}_{r_1} \times \mathcal{C}_{r_2} \times \cdots \times \mathcal{C}_{r_U}$ is characterized by the multiplexing rate tuple $\boldsymbol{r} = (r_1, r_2, \ldots, r_U)$, which is related to the rate tuple $(R_1, R_2, \ldots, R_U)$ through $R_u(\mathsf{SNR}) = r_u \log \mathsf{SNR}$ for all users u. For any $\mathcal{S} \subseteq \mathcal{U}$, we shall use the definitions

$$R(\mathcal{S}) = \sum_{u \in \mathcal{S}} R_u \quad \text{and} \quad r(\mathcal{S}) = \sum_{u \in \mathcal{S}} r_u. \tag{6.4}$$

6.1.2. Channel model

We refer to Section 2.2 and restrict our analysis to spatially uncorrelated Rayleigh fading channels so that, for a given n, $\mathbf{H}_{u,n}$ has i.i.d. $\mathcal{CN}(0,1)$ entries. The channels corresponding to different users are assumed to be statistically independent. We do, however, allow for correlation across n for a given u, and assume, for simplicity, that all scalar subchannels have the same correlation function so that (see (2.18))

$$\begin{aligned}\mathbb{E}\{\mathbf{H}_{u,n}(i,j)\,(\mathbf{H}_{u',n'}(i',j'))^*\} = \mathbf{R}_{\mathbb{H}}(n,n')\,\delta_{u,u'}\,\delta_{i,i'}\,\delta_{j,j'},&\\ j,j' = 1,2,\ldots,\mathrm{M_T},\ i,i' = 1,2,\ldots,\mathrm{M_R},&\\ n,n' = 0,1,\ldots,N-1,&\\ u,u' = 1,2,\ldots,U.&\end{aligned}$$

The covariance matrix $\mathbf{R}_{\mathbb{H}}$, with $\rho \triangleq \operatorname{rank}(\mathbf{R}_{\mathbb{H}})$, is obtained from the channel's time-frequency correlation function in (2.19). For any subset of users $\mathcal{S} = \{u_1, u_2, \ldots, u_{|\mathcal{S}|}\}$, we define the matrices $\mathbf{H}_{\mathcal{S},n}$ and $\mathbf{H}_{\mathcal{S}}$ according to

$$\mathbf{H}_{\mathcal{S},n} = [\mathbf{H}_{u_1,n}\ \mathbf{H}_{u_2,n}\ \cdots\ \mathbf{H}_{u_{|\mathcal{S}|},n}] \tag{6.5}$$

$$\mathbf{H}_{\mathcal{S}} = [\mathbf{H}_{\mathcal{S},0}\ \mathbf{H}_{\mathcal{S},1}\ \cdots\ \mathbf{H}_{\mathcal{S},N-1}]. \tag{6.6}$$

With this notation, it follows that

$$\mathbb{E}\Big\{\operatorname{vec}(\mathbf{H}_{\mathcal{S}})(\operatorname{vec}(\mathbf{H}_{\mathcal{S}}))^H\Big\} = \mathbf{R}_{\mathbb{H}} \otimes \mathbf{I}_{|\mathcal{S}|\mathrm{M_T M_R}}. \tag{6.7}$$

6.2. PRELIMINARIES

A distinctive aspect of the MA channel is that users are not allowed to cooperate in encoding their messages. This is reflected by the fact that every user u employs its own family of codes $\mathcal{C}_{r_u}$ to communicate with the receiver. While users do not cooperate at the time of encoding, they can still jointly design their codebooks in advance. We assume that the receiver has perfect channel state information (CSI) and the transmitters do not have CSI but are aware of the fading law of all users. Given a target outage probability, the transmitters can therefore jointly optimize their respective codebooks to achieve the corresponding maximum reliable rate. In fact, we shall see that jointly designing the users' codebooks is crucial to realize optimal performance in the sense of the DM tradeoff.

6.2.1. Multiplexing gain region

We briefly review the concept of multiplexing gain region (Tse et al., 2004; Visuri and Bölcskei, 2006). Because the boundaries of the capacity region are maximized with Gaussian inputs (see Section 3.4.1 for more details), we assume that the users' codebooks are independent and satisfy $\mathrm{vec}\,(\mathbf{X}_u) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_u)$ for all u. It follows that $\mathrm{vec}(\mathbf{X}_\mathcal{S}) \sim \mathcal{CN}(0, \mathbf{Q}_\mathcal{S})$, where $\mathbf{Q}_\mathcal{S}$ can be uniquely determined from the input covariance matrices $\{\mathbf{Q}_u\}$. The corresponding ergodic capacity region is given by the constraints

$$R(\mathcal{S}) \leq \max_{\substack{\mathbf{Q}_\mathcal{S} \succeq 0 \\ \mathrm{Tr}(\mathbf{Q}_\mathcal{S}) \leq |\mathcal{S}| N \mathrm{M_T}}} \mathbb{E}_{\mathbf{H}_\mathcal{S}} \left\{ \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}} \mathbf{D}_{\mathbf{H}_\mathcal{S}} \mathbf{Q}_\mathcal{S} \mathbf{D}_{\mathbf{H}_\mathcal{S}}^H \right) \right\} \quad \forall \mathcal{S} \subseteq \mathcal{U} \tag{6.8}$$

where $\mathbf{D}_{\mathbf{H}_\mathcal{S}} \triangleq \mathrm{diag}\{\mathbf{D}_{\mathbf{H}_{\mathcal{S},n}}\}_{n=0}^{N-1}$. By the same line of reasoning as in Section 3.4.1, the solution to the above optimization problem is $\mathbf{Q}_\mathcal{S} = \mathbf{I}$ and, hence, the choice $\mathbf{Q}_u = \mathbf{I}$ for every $u = 1, 2, \ldots, U$ is optimal (Telatar, 1999; Tse and Viswanath, 2005; Visuri and Bölcskei, 2006). Upon inserting the solution into (6.8), the ergodic capacity region follows as the set of rate tuple satisfying the constraints

$$\begin{aligned} R(\mathcal{S}) &\leq \frac{1}{N}\sum_{n=0}^{N-1}\mathbb{E}_{\mathbf{H}_{\mathcal{S},n}}\left\{\log\det\left(\mathbf{I}+\frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{H}_{\mathcal{S},n}\mathbf{H}_{\mathcal{S},n}^{H}\right)\right\} \\ &= \mathbb{E}_{\mathbf{H}_{\mathcal{S},n}}\left\{\log\det\left(\mathbf{I}+\frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{H}_{\mathcal{S},n}\mathbf{H}_{\mathcal{S},n}^{H}\right)\right\} \end{aligned} \tag{6.9}$$

for all $\mathcal{S} \subseteq \mathcal{U}$. Equation (6.9) is a consequence of the fact that the fading process is stationary. The matrix $\mathbf{H}_{\mathcal{S},n}$ has dimension $\mathrm{M_R} \times |\mathcal{S}|\mathrm{M_T}$ so that, asymptotically in SNR, the boundary in (6.9) scales as $\min(\mathrm{M_R}, |\mathcal{S}|\mathrm{M_T}) \log \mathsf{SNR}$ (Foschini, 1996), which is to say that the achievable multiplexing rate tuples belong to the set

$$\left\{\boldsymbol{r} = (r_1, r_2, \ldots, r_U) : r(\mathcal{S}) \leq \min(\mathrm{M_R}, |\mathcal{S}|\mathrm{M_T}), \forall \mathcal{S} \subseteq \mathcal{U}\right\}. \tag{6.10}$$

6.2.2. Outage formulation

Since we are interested in characterizing ultimate performance limits over a single realization of the channel with input-output relation (6.1), we shall resort to the concept of outage capacity (Ozarow et al., 1994; Telatar, 1999; Shamai (Shitz) and Wyner, 1997a,b; Visuri and Bölcskei, 2006) as we did previously in the point-to-point case, and use the DM tradeoff framework (Zheng and Tse, 2003; Tse et al., 2004) to conduct our analysis.

Conditional on a realization of the fading channel matrix $\mathbf{H}_{\mathcal{U}}$, the set of rates that are achievable over the channel (6.1) for perfect channel state information (CSI) at the receiver only are characterized by Theorem 3.4. Assuming without loss of optimality that $\mathrm{vec}(\mathbf{X}_{\mathcal{S}}) \sim \mathcal{CN}(0, \mathbf{Q}_{\mathcal{S}})$, the rate constraint on the users in $\mathcal{S}$ follows as

$$R(\mathcal{S}) \leq \frac{1}{N}\log\det\left(\mathbf{I}+\frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{D}_{\mathbf{H}_{\mathcal{S}}}\mathbf{Q}_{\mathcal{S}}\mathbf{D}_{\mathbf{H}_{\mathcal{S}}}^{H}\right). \tag{6.11}$$

By the same reasoning as in the point-to-point case treated in Section 4.2, the optimal input covariance matrix in the high-SNR regime can be shown to be $\mathbf{Q}_{\mathcal{S}} = \mathbf{I}$ (note that this is the same input covariance as in the ergodic capacity setting). This immediately implies $\mathbf{Q}_u = \mathbf{I}$ for all $u \in \mathcal{S}$. Hence, the set of achievable rate tuples $(R_1, R_2, \ldots, R_U)$ for the channel realization $\mathbf{H}_{\mathcal{U}}$ satisfies the constraints

$$R(\mathcal{S}) \leq \frac{1}{N}\sum_{n=0}^{N-1} \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\,\mathbf{H}_{\mathcal{S},n}\mathbf{H}_{\mathcal{S},n}^H\right), \forall \mathcal{S} \subseteq \mathcal{U}. \quad (6.12)$$

We recall from Theorem 3.5 that, asymptotically in SNR, the outage probability $\mathbb{P}(\mathcal{O}_{\boldsymbol{r}})$ constitutes a fundamental barrier on the error performance, i.e.,

$$d^\star(\boldsymbol{r}) \leq -\lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}(\mathcal{O}_{\boldsymbol{r}})}{\log \mathsf{SNR}} \quad (6.13)$$

where, after letting the users' data rates scale with SNR as $R_u(\mathsf{SNR}) = r_u \log \mathsf{SNR}$, the probability of outage with respect to (w.r.t.) the rate tuple $\boldsymbol{r}$ is given by

$$\mathbb{P}(\mathcal{O}_{\boldsymbol{r}}) = \mathbb{P}\left(\bigcup_{\mathcal{S}\subseteq\mathcal{U}} \mathcal{O}_{\mathcal{S}}\right) \quad (6.14)$$

where $\mathcal{O}_{\mathcal{S}}$ denotes the $\mathcal{S}$-outage event, i.e., the outage event with respect to the boundary corresponding to $\mathcal{S}$ in (6.12), for a given rate tuple $(R_1, R_2, \ldots, R_U)$. The $\mathcal{S}$-outage probability* is therefore given by

$$\mathbb{P}(\mathcal{O}_{\mathcal{S}}) \doteq \mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1} \log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\mathrm{M_T}}\,\mathbf{H}_{\mathcal{S},n}\mathbf{H}_{\mathcal{S},n}^H\right) < R(\mathcal{S})\right). \quad (6.15)$$

Our goal will be to characterize (6.14) as a function of the multiplexing rate tuple $\boldsymbol{r}$ in the high-SNR regime and to find criteria on the users' codebooks guaranteeing that the corresponding error probability

*Note that (6.14) and (6.15) extend (3.47) and (3.48) to the selective-fading case.

behaves exponentially in SNR like $\mathbb{P}(\mathcal{O}_{\boldsymbol{r}})$. However, just like in the case of point-to-point channels, a direct characterization of the right-hand side of (6.15) for the selective-fading case seems analytically intractable since one has to deal with the sum of correlated (recall that the $\mathbf{H}_{u,n}$ are correlated across n) terms. In the next section, we show how the technique introduced in Chapter 4 for characterizing the DM tradeoff of point-to-point selective-fading MIMO channels can be extended to the MA case.

6.3. OPTIMAL DIVERSITY-MULTIPLEXING TRADEOFF

In this section, we derive the optimal DM tradoff of selective-fading MA MIMO channels, and we obtain code design criteria that guarantee optimality of the family of codes $\mathcal{C}_{\boldsymbol{r}}$. As in the point-to-point case, our approach builds upon the Jensen channel associated with the selective-fading MA MIMO channel. We start by deriving a lower bound on the $\mathcal{S}$-outage probability that will be instrumental in establishing the aforementioned results.

6.3.1. Lower bound on the $\mathcal{S}$-outage probability

For any set $\mathcal{S} \subseteq \mathcal{U}$, Jensen's inequality yields the following upper bound on the corresponding mutual information in (6.12):

$$\frac{1}{N}\sum_{n=0}^{N-1}\log\det\left(\mathbf{I}+\frac{\mathsf{SNR}}{\mathrm{M_T}}\mathbf{H}_{\mathcal{S},n}\mathbf{H}_{\mathcal{S},n}^{H}\right)$$
$$\leq \log\det\left(\mathbf{I}+\frac{\mathsf{SNR}}{\mathrm{M_T}N}\boldsymbol{\mathcal{H}}_{\mathcal{S}}\boldsymbol{\mathcal{H}}_{\mathcal{S}}{}^{H}\right) \triangleq \mathrm{J}_{\mathcal{S}}(\mathsf{SNR}) \quad (6.16)$$

where, adapting Definition 4.1 to the MA case, the Jensen channel is given by

$$\boldsymbol{\mathcal{H}}_{\mathcal{S}} = \begin{cases} [\mathbf{H}_{\mathcal{S},0}\ \mathbf{H}_{\mathcal{S},1}\ \ldots\ \mathbf{H}_{\mathcal{S},N-1}], & \text{if } \mathrm{M_R} \le |\mathcal{S}|\mathrm{M_T}, \\ [\mathbf{H}^H_{\mathcal{S},0}\ \mathbf{H}^H_{\mathcal{S},1}\ \ldots\ \mathbf{H}^H_{\mathcal{S},N-1}], & \text{if } \mathrm{M_R} > |\mathcal{S}|\mathrm{M_T}. \end{cases} \tag{6.17}$$

Consequently, $\boldsymbol{\mathcal{H}}_{\mathcal{S}}$ has dimension $\mathrm{m}(\mathcal{S}) \times N\mathrm{M}(\mathcal{S})$, where

$$\mathrm{m}(\mathcal{S}) \triangleq \min(\mathrm{M_R}, |\mathcal{S}|\mathrm{M_T}) \tag{6.18}$$

$$\mathrm{M}(\mathcal{S}) \triangleq \max(\mathrm{M_R}, |\mathcal{S}|\mathrm{M_T}). \tag{6.19}$$

We say that a Jensen outage event occurs if the Jensen channel $\boldsymbol{\mathcal{H}}_{\mathcal{S}}$ is in outage w.r.t. the rate $R(\mathcal{S})$, and the corresponding probability is given by

$$P_{\mathrm{J}}(R(\mathcal{S})) = \mathbb{P}(\mathrm{J}_{\mathcal{S}}(\mathsf{SNR}) < R(\mathcal{S}))\,. \tag{6.20}$$

The Jensen outage probability can be characterized analytically as in the point-to-point case. In particular, recalling (6.7), the Jensen channel satisfies $\boldsymbol{\mathcal{H}}_{\mathcal{S}} = \boldsymbol{\mathcal{H}}_w(\mathbf{R}^{T/2} \otimes \mathbf{I}_{\mathrm{M}(\mathcal{S})})$, where $\mathbf{R} = \mathbf{R}_{\mathbb{H}}$, if $\mathrm{M_R} \le |\mathcal{S}|\mathrm{M_T}$, and $\mathbf{R} = \mathbf{R}_{\mathbb{H}}^T$, if $\mathrm{M_R} > |\mathcal{S}|\mathrm{M_T}$, and $\boldsymbol{\mathcal{H}}_w$ is the i.i.d. $\mathcal{CN}(0,1)$ matrix with the same dimensions as $\boldsymbol{\mathcal{H}}_{\mathcal{S}}$ given by

$$\boldsymbol{\mathcal{H}}_w = \begin{cases} [\mathbf{H}_{w,0}\ \mathbf{H}_{w,1}\ \ldots\ \mathbf{H}_{w,N-1}], & \text{if } \mathrm{M_R} \le |\mathcal{S}|\mathrm{M_T}, \\ [\mathbf{H}^H_{w,0}\ \mathbf{H}^H_{w,1}\ \ldots\ \mathbf{H}^H_{w,N-1}], & \text{if } \mathrm{M_R} > |\mathcal{S}|\mathrm{M_T}. \end{cases}$$

Here, $\mathbf{H}_{w,n}$ denotes i.i.d. $\mathcal{CN}(0,1)$ matrices of dimension $\mathrm{M_R} \times |\mathcal{S}|\mathrm{M_T}$. Since $\boldsymbol{\mathcal{H}}_w\mathbf{U} \sim \boldsymbol{\mathcal{H}}_w$, for any unitary $\mathbf{U}$ (Muirhead, 1982), and $\mathbf{R}_{\mathbb{H}}$ and $\mathbf{R}_{\mathbb{H}}^T$ have the same eigenvalues, we get

$$\boldsymbol{\mathcal{H}}_{\mathcal{S}}\boldsymbol{\mathcal{H}}_{\mathcal{S}}{}^H \sim \boldsymbol{\mathcal{H}}_w(\boldsymbol{\Theta} \otimes \mathbf{I}_{\mathrm{M}(\mathcal{S})})\boldsymbol{\mathcal{H}}_w^H \tag{6.21}$$

where $\boldsymbol{\Theta} = \mathrm{diag}\{\lambda_1(\mathbf{R}_{\mathbb{H}}), \lambda_2(\mathbf{R}_{\mathbb{H}}), \ldots, \lambda_\rho(\mathbf{R}_{\mathbb{H}}), 0, \ldots, 0\}$. Letting the users' data rates scale with SNR as $R_u(\mathsf{SNR}) = r_u \log \mathsf{SNR}$ and setting $\overline{\boldsymbol{\mathcal{H}}}_w = \boldsymbol{\mathcal{H}}_w([1:\mathrm{m}(\mathcal{S})], [1:\rho\mathrm{M}(\mathcal{S})])$, the same line of argumentation that we used in Section 4.3.2 shows that $P_{\mathrm{J}}(r(\mathcal{S}) \log \mathsf{SNR}) \doteq$

$\mathbb{P}(\mathcal{J}_{\mathcal{S}})$, where, although omitted here, $\mathcal{J}_{\mathcal{S}}$ can be rigorously defined in terms of the singularity levels of $\overline{\boldsymbol{\mathcal{H}}}_w$. Asymptotically in SNR, $\mathbb{P}(\mathcal{J}_{\mathcal{S}})$ is nothing but the outage probability of an effective MIMO channel with $\rho\mathrm{M}(\mathcal{S})$ transmit and $\mathrm{m}(\mathcal{S})$ receive antennas, that is

$$\begin{aligned}\mathbb{P}(\mathcal{J}_{\mathcal{S}}) &\doteq \mathbb{P}\Big(\log\det\Big(\mathbf{I} + \mathsf{SNR}\,\overline{\boldsymbol{\mathcal{H}}}_w\overline{\boldsymbol{\mathcal{H}}}_w^H\Big) < r(\mathcal{S})\log\mathsf{SNR}\Big) \\ &\doteq \mathsf{SNR}^{-d_{\mathcal{S}}(r(\mathcal{S}))} \end{aligned} \tag{6.22}$$

where we infer from Theorem 3.3 that $d_{\mathcal{S}}(r)$ is the piecewise linear function connecting the points $(r, d_{\mathcal{S}}(r))$ for $r = 0, 1, \ldots, \mathrm{m}(\mathcal{S})$, with

$$d_{\mathcal{S}}(r) = (\mathrm{m}(\mathcal{S}) - r)(\rho\mathrm{M}(\mathcal{S}) - r). \tag{6.23}$$

As an immediate consequence of (6.16), we have $\mathbb{P}(\mathcal{O}_{\mathcal{S}}) \dot{\geq} \mathbb{P}(\mathcal{J}_{\mathcal{S}})$, which, using (6.22), yields

$$\mathbb{P}(\mathcal{O}_{\mathcal{S}}) \dot{\geq} \mathsf{SNR}^{-d_{\mathcal{S}}(r(\mathcal{S}))}. \tag{6.24}$$

In the next subsection, we shall show that the right-hand side of (6.24) is also an upper bound, in the $\dot{\leq}$ sense, on the probability of erroneously decoding all the users in the set $\mathcal{S}$ while correctly decoding the users in its complement $\bar{\mathcal{S}}$. This result paves the way for establishing the optimal DM tradeoff.

6.3.2. Error event analysis

Following (Gallager, 1985; Tse et al., 2004), we decompose the total error probability into $2^U - 1$ disjoint error events according to

$$P_e(\mathcal{C}_{\boldsymbol{r}}) = \sum_{\mathcal{S}\subseteq\mathcal{U}} \mathbb{P}(\mathcal{E}_{\mathcal{S}}) \tag{6.25}$$

where the $\mathcal{S}$-error event $\mathcal{E}_{\mathcal{S}}$ corresponds to all the users in $\mathcal{S}$ being decoded incorrectly and the remaining users being decoded correctly. More precisely, we have

$$\mathcal{E}_{\mathcal{S}} \triangleq \Big\{(\hat{\mathbf{X}}_u \neq \mathbf{X}_u, \forall u \in \mathcal{S}) \;\wedge\; (\hat{\mathbf{X}}_u = \mathbf{X}_u, \forall u \in \bar{\mathcal{S}})\Big\} \tag{6.26}$$

where $\mathbf{X}_u$ and $\hat{\mathbf{X}}_u$ are, respectively, the transmitted and ML-decoded codewords corresponding to user u. The following result establishes a DM tradeoff optimal code design criterion for a specific error event $\mathcal{E}_{\mathcal{S}}$.

Theorem 6.1. *For every $u \in \mathcal{S}$, let $\mathcal{C}_{r_u}$ have block length $N \geq \rho|\mathcal{S}|\mathrm{M_T}$, and set*

$$\lambda_n = \lambda_n\left(\mathbf{R}_{\mathbb{H}}^T \odot \sum_{u\in\mathcal{S}} \mathbf{E}_u^H \mathbf{E}_u\right),\ n = 1, 2, \ldots, \rho|\mathcal{S}|\mathrm{M_T} \qquad (6.27)$$

where $\mathbf{E}_u = \mathbf{X}_u - \mathbf{X}'_u$ and $\mathbf{X}_u$, $\mathbf{X}'_u \in \mathcal{C}_{r_u}(\mathsf{SNR})$. Furthermore, define

$$\Xi_{\mathrm{m}(\mathcal{S})}^{\rho|\mathcal{S}|\mathrm{M_T}}(\mathsf{SNR}) \triangleq \min_{\substack{\mathbf{E}_u = \mathbf{X}_u - \mathbf{X}'_u, \forall u\in\mathcal{S} \\ \mathbf{X}_u, \mathbf{X}'_u \in \mathcal{C}_{r_u}(\mathsf{SNR})}} \prod_{k=1}^{\mathrm{m}(\mathcal{S})} \lambda_k. \qquad (6.28)$$

If there exists an $\epsilon > 0$ such that

$$\Xi_{\mathrm{m}(\mathcal{S})}^{\rho|\mathcal{S}|\mathrm{M_T}}(\mathsf{SNR}) \dot{\geq} \mathsf{SNR}^{-(r(\mathcal{S})-\epsilon)}, \qquad (6.29)$$

then, under ML decoding, the $\mathcal{S}$-error probability satisfies

$$\mathbb{P}(\mathcal{E}_{\mathcal{S}}) \dot{\leq} \mathsf{SNR}^{-d_{\mathcal{S}}(r(\mathcal{S}))}. \qquad (6.30)$$

Proof. We adapt the upper bound on the average (w.r.t. the random channel) pairwise error probability (PEP) in (4.44) to the $\mathcal{S}$-error event. Based on (6.26), we note that $\mathbf{E}_u = \mathbf{X}_u - \mathbf{X}'_u$ is nonzero for $u \in \mathcal{S}$ and $\mathbf{E}_u = \mathbf{0}$ for $u \in \bar{\mathcal{S}}$. Assuming, without loss of generality, that $\mathcal{S} = \{1, 2, \ldots, |\mathcal{S}|\}$, the probability of the ML decoder mistakenly deciding in favor of the codeword $\mathbf{X}'$ when $\mathbf{X}$ was actually transmitted can be upper bounded in terms of $\mathbf{X} - \mathbf{X}' = [\mathbf{E}_1^T\, \mathbf{E}_2^T\, \ldots\, \mathbf{E}_{|\mathcal{S}|}^T\, \mathbf{0} \ldots \mathbf{0}]^T$ as

$$\begin{aligned} &\mathbb{P}(\mathbf{X} \to \mathbf{X}') \\ &\leq \mathbb{E}_{\{\mathbf{H}_{\mathcal{S},n}\}}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}} \sum_{n=0}^{N-1} \mathrm{Tr}\left(\mathbf{H}_{\mathcal{S},n}\mathbf{e}_n\mathbf{e}_n^H\mathbf{H}_{\mathcal{S},n}^H\right)\right)\right\} \end{aligned} \qquad (6.31)$$

where $\mathrm{Tr}\left(\mathbf{H}_{\mathcal{S},n}\mathbf{e}_n\mathbf{e}_n^H\mathbf{H}_{\mathcal{S},n}^H\right) = \left\|\sum_{u\in\mathcal{S}}\mathbf{H}_{u,n}\mathbf{e}_{u,n}\right\|^2$ with $\mathbf{H}_{\mathcal{S},n}$ defined in (6.5) and $\mathbf{e}_n = [\mathbf{e}_{u_1,n}^T \ \mathbf{e}_{u_2,n}^T \ \cdots \ \mathbf{e}_{u_{|\mathcal{S}|},n}^T]^T$, where $\mathbf{e}_{u,n} = \mathbf{x}_{u,n} - \mathbf{x}'_{u,n}$. Recalling $\mathbf{H}_{\mathcal{S}}$ in (6.6), we get from (6.31)

$$\begin{aligned}
&\mathbb{P}(\mathbf{X}\to\mathbf{X}') \\
&\leq \mathbb{E}_{\mathbf{H}_{\mathcal{S}}}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}}\mathrm{Tr}\left(\mathbf{H}_{\mathcal{S}}\,\mathrm{diag}\{\mathbf{e}_n\mathbf{e}_n^H\}_{n=0}^{N-1}\mathbf{H}_{\mathcal{S}}^H\right)\right)\right\} \\
&= \mathbb{E}_{\mathbf{H}_w}\left\{\exp\left(-\frac{\mathsf{SNR}}{4\mathrm{M_T}}\mathrm{Tr}\left(\mathbf{H}_w\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H\mathbf{H}_w^H\right)\right)\right\} \qquad (6.32)
\end{aligned}$$

where we used $\mathbf{H}_{\mathcal{S}} = \mathbf{H}_w(\mathbf{R}_{\mathbb{H}}^{T/2}\otimes\mathbf{I}_{|\mathcal{S}|\mathrm{M_T}})$ with $\mathbf{H}_w$ an $\mathrm{M_R}\times N|\mathcal{S}|\mathrm{M_T}$ matrix with i.i.d. $\mathcal{CN}(0,1)$ entries and

$$\boldsymbol{\Upsilon} = (\mathbf{R}_{\mathbb{H}}^{T/2}\otimes\mathbf{I}_{|\mathcal{S}|\mathrm{M_T}})\,\mathrm{diag}\{\mathbf{e}_n\}_{n=0}^{N-1}. \qquad (6.33)$$

Noting that $\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon} = \mathbf{R}_{\mathbb{H}}^T\odot\sum_{u\in\mathcal{S}}\mathbf{E}_u^H\mathbf{E}_u$ and using the fact that the nonzero eigenvalues of $\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H$ in (6.32) equal the nonzero eigenvalues of $\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}$, it follows, by assumption, that $\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H$ has precisely $\rho|\mathcal{S}|\mathrm{M_T}$ nonzero eigenvalues. The remainder of the proof proceeds along the lines of the proof of Theorem 4.1. In particular, we split and subsequently bound the $\mathcal{S}$-error probability as

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{\mathcal{S}}) &= \mathbb{P}(\mathcal{E}_{\mathcal{S}},\mathcal{J}_{\mathcal{S}}) + \mathbb{P}\big(\mathcal{E}_{\mathcal{S}},\bar{\mathcal{J}}_{\mathcal{S}}\big) \\
&= \mathbb{P}(\mathcal{J}_{\mathcal{S}})\underbrace{\mathbb{P}(\mathcal{E}_{\mathcal{S}}|\mathcal{J}_{\mathcal{S}})}_{\leq 1} + \underbrace{\mathbb{P}\big(\bar{\mathcal{J}}_{\mathcal{S}}\big)}_{\leq 1}\mathbb{P}\big(\mathcal{E}_{\mathcal{S}}|\bar{\mathcal{J}}_{\mathcal{S}}\big) \\
&\leq \mathbb{P}(\mathcal{J}_{\mathcal{S}}) + \mathbb{P}\big(\mathcal{E}_{\mathcal{S}}|\bar{\mathcal{J}}_{\mathcal{S}}\big). \qquad (6.34)
\end{aligned}$$

As detailed in the point-to-point case, the code design criterion (6.29) yields the following upper bound on the second term in (6.34):

$$\mathbb{P}\big(\mathcal{E}_{\mathcal{S}}|\bar{\mathcal{J}}_{\mathcal{S}}\big) \leq \mathsf{SNR}^{Nr(\mathcal{S})}\exp\left(-\frac{\mathsf{SNR}^{\epsilon/\mathrm{m}(\mathcal{S})}}{4\mathrm{M_T}}\right). \qquad (6.35)$$

In contrast to the Jensen outage probability which satisfies $\mathbb{P}(\mathcal{J}_{\mathcal{S}}) \doteq \mathsf{SNR}^{-d_{\mathcal{S}}(r(\mathcal{S}))}$, (6.35) decays exponentially in SNR. Hence, upon inserting (6.35) into (6.34), we get $\mathbb{P}(\mathcal{E}_{\mathcal{S}}) \dot{\leq} \mathbb{P}(\mathcal{J}_{\mathcal{S}})$, and can therefore conclude that $\mathbb{P}(\mathcal{E}_{\mathcal{S}}) \dot{\leq} \mathsf{SNR}^{-d_{\mathcal{S}}(r(\mathcal{S}))}$. □

In summary, for every $\mathcal{E}_{\mathcal{S}}$, we have a sufficient condition on $\{\mathcal{C}_{r_u} : u \in \mathcal{S}\}$ for $\mathbb{P}(\mathcal{E}_{\mathcal{S}})$ to be exponentially upper bounded by $\mathbb{P}(\mathcal{J}_{\mathcal{S}})$. Noteworthily, condition (6.29) is nothing else but the code design criterion of Theorem 4.1 applied to a system operating at multiplexing rate $r(\mathcal{S})$ and using $|\mathcal{S}|\mathrm{M_T}$ *collocated* antennas. However, while in the point-to-point case encoding is performed across antennas, the antennas corresponding to different users are not collocated and the corresponding messages have to be encoded independently. In spite of this lack of user cooperation at the time of communication, the families of codes $\mathcal{C}_{r_u}$, for all $u \in \mathcal{S}$, can be designed in advance so as to satisfy (6.29). This task is challenging but doable as we shall see later.

Building on Theorem 6.1, we will first establish the optimal DM tradeoff of selective-fading MA channels and provide a set of code design criteria that guarantees optimal performance.

6.3.3. Optimal code design

We start by noting that (6.14) implies $\mathbb{P}(\mathcal{O}_{\boldsymbol{r}}) \geq \mathbb{P}(\mathcal{O}_{\mathcal{S}})$ for any $\mathcal{S} \subseteq \mathcal{U}$, which combined with (6.24) gives rise to $2^U - 1$ lower bounds on $\mathbb{P}(\mathcal{O}_{\boldsymbol{r}})$. For a given multiplexing rate tuple $\boldsymbol{r}$, the tightest lower bound (exponentially in SNR) corresponds to the set $\mathcal{S}$ that yields the smallest SNR exponent $d_{\mathcal{S}}(r(\mathcal{S}))$. More precisely, the tightest lower bound is characterized by

$$\mathbb{P}(\mathcal{O}_{\boldsymbol{r}}) \dot{\geq} \mathsf{SNR}^{-d_{\mathcal{S}^\star}(r(\mathcal{S}^\star))} \tag{6.36}$$

where the dominant outage event corresponds to the set

$$\mathcal{S}^\star = \underset{\mathcal{S} \subseteq \mathcal{U}}{\arg\min}\ d_{\mathcal{S}}(r(\mathcal{S})). \tag{6.37}$$

Next, we show that, for any multiplexing rate tuple, the total error probability $P_e(\mathcal{C}_{\boldsymbol{r}})$ can be made exponentially equal to the lower bound in (6.36) by appropriate design of the users' codebooks. As a direct consequence thereof, recalling from Theorem 3.5 that $P_e(\mathcal{C}_{\boldsymbol{r}}) \dot{\geq} \mathbb{P}(\mathcal{O}_{\boldsymbol{r}})$, we then obtain that $d_{\mathcal{S}^\star}(r(\mathcal{S}^\star))$ constitutes the optimal DM tradeoff of the selective-fading MA MIMO channel.

Theorem 6.2. *The optimal DM tradeoff of the selective-fading MA MIMO channel in (6.1) is given by* $d^\star(\boldsymbol{r}) = d_{\mathcal{S}^\star}(r(\mathcal{S}^\star))$*, that is*

$$d^\star(\boldsymbol{r}) = (\mathrm{m}(\mathcal{S}^\star) - r(\mathcal{S}^\star))(\rho \mathrm{M}(\mathcal{S}^\star) - r(\mathcal{S}^\star)). \tag{6.38}$$

Moreover, if the family of codes $\mathcal{C}_{\boldsymbol{r}}$ *satisfies (6.29) for every* $\mathcal{S} \subseteq \mathcal{U}$*, then*

$$d(\mathcal{C}_{\boldsymbol{r}}) = d^\star(\boldsymbol{r}). \tag{6.39}$$

Proof. Upon inserting the upper bound (6.34) into (6.25), we get

$$\begin{aligned} P_e(\mathcal{C}_{\boldsymbol{r}}) &\leq \sum_{\mathcal{S} \subseteq \mathcal{U}} \big(\mathbb{P}(\mathcal{J}_{\mathcal{S}}) + \mathbb{P}\big(\mathcal{E}_{\mathcal{S}} | \bar{\mathcal{J}}_{r\mathcal{S}}\big)\big) \\ &\dot{\leq} \sum_{\mathcal{S} \subseteq \mathcal{U}} \mathbb{P}(\mathcal{J}_{\mathcal{S}}) &(6.40) \\ &\doteq \mathsf{SNR}^{-d_{\mathcal{S}^\star}(r(\mathcal{S}^\star))} &(6.41) \end{aligned}$$

where (6.40) is a consequence of the assumption that $\mathcal{C}_{\boldsymbol{r}}$ satisfies (6.29) for every $\mathcal{S} \subseteq \mathcal{U}$ and (6.41) follows from (6.22) together with the definition (6.37). By Theorem 3.5, we have $P_e(\mathcal{C}_{\boldsymbol{r}}) \dot{\geq} \mathbb{P}(\mathcal{O}_{\boldsymbol{r}})$ and hence, combining (6.36) and (6.41) yields

$$P_e(\mathcal{C}_{\boldsymbol{r}}) \doteq \mathbb{P}(\mathcal{O}_{\boldsymbol{r}}) \doteq \mathsf{SNR}^{-d_{\mathcal{S}^\star}(r(\mathcal{S}^\star))}. \tag{6.42}$$

Since, by definition, $d(\mathcal{C}_{\boldsymbol{r}}) \leq d^\star(\boldsymbol{r})$, using (6.13), we can conclude from (6.42) that

$$d(\mathcal{C}_{\boldsymbol{r}}) = d^\star(\boldsymbol{r}) = d_{\mathcal{S}^\star}(r(\mathcal{S}^\star)). \tag{6.43}$$

□

According to Theorem 6.2, the optimal DM tradeoff is determined by the tradeoff curve $d_{\mathcal{S}^\star}(r(\mathcal{S}^\star))$, which is simply the SNR exponent of the Jensen outage probability $\mathbb{P}(\mathcal{J}_{\mathcal{S}^\star})$ corresponding to the $\mathcal{S}^\star$-outage event. By virtue of (6.42) and the relations between the events $\mathcal{O}_{\mathcal{S}^\star}$, $\mathcal{J}_{\mathcal{S}^\star}$, and $\mathcal{O}_{\boldsymbol{r}}$, we get

$$\mathbb{P}(\mathcal{O}_{\mathcal{S}^\star}) \dot{\leq} \mathbb{P}(\mathcal{O}_{\boldsymbol{r}}) \doteq \mathbb{P}(\mathcal{J}_{\mathcal{S}^\star}) \dot{\leq} \mathbb{P}(\mathcal{O}_{\mathcal{S}^\star}) \tag{6.44}$$

which is to say

$$\mathbb{P}(\mathcal{O}_{\mathcal{S}^\star}) \doteq \mathsf{SNR}^{-d_{\mathcal{S}^\star}(r(\mathcal{S}^\star))}. \tag{6.45}$$

Hence, the Jensen upper bound on mutual information yields a bound on the $\mathcal{S}^\star$-outage probability which is tight exponentially in SNR and constitutes the optimal DM tradeoff. However, in order to achieve it, the families of codes $\{\mathcal{C}_{r_u}, u \in \mathcal{U}\}$ are required to satisfy (6.29) for every possible subset of users $\mathcal{S}$: a requirement that seems difficult to meet in general. We shall later consider an example of a code construction which actually satisfies this condition in the 2-user case.

The dominant outage event determines the maximal diversity order as a function of the multiplexing rate tuple $\boldsymbol{r}$. Conversely, one may also be interested in characterizing the region of achievable multiplexing rates at a minimum diversity gain d. To this end, let us denote by $r_{\mathcal{S}}(d)$ the inverse of the tradeoff curve $d_{\mathcal{S}}(r)$, i.e., $d = d_{\mathcal{S}}\big(r_{\mathcal{S}}(d)\big)$. If the decay of the $\mathcal{S}$-error probability is at least d over all $\mathcal{S} \subseteq \mathcal{U}$ (this implies that all users realize a diversity requirement of d), it follows from Theorem 6.2 that

$$d_{\mathcal{S}^\star}(r(\mathcal{S}^\star)) = d.$$

As a consequence of the definition of the dominant outage event in (6.37), we also get

$$d_{\mathcal{S}}\big(r(\mathcal{S})\big) \geq d, \quad \forall \mathcal{S} \subseteq \mathcal{U}.$$

And applying $r_{\mathcal{S}}(\cdot)$ to the both sides of the inequality yields

$$r(\mathcal{S}) \leq r_{\mathcal{S}}(d), \quad \forall \mathcal{S} \subseteq \mathcal{U}.$$

We have just proved the following extension of Theorem 3.6 to selective-fading MA MIMO channels.

Corollary 6.3. Consider an overall family of codes $\mathcal{C}_{\boldsymbol{r}}$ that achieves the optimal DM tradeoff in the sense of Theorem 6.2. Then, the region of multiplexing rates for which a minimum diversity gain d is guaranteed to all users is characterized by

$$\mathcal{R}(d) \triangleq \left\{ \boldsymbol{r} : r(\mathcal{S}) \leq r_{\mathcal{S}}(d), \forall \mathcal{S} \subseteq \mathcal{U} \right\} \tag{6.46}$$

where $r_{\mathcal{S}}(d)$ is the inverse function of $d_{\mathcal{S}}(r)$.

This explicit characterization of $\mathcal{R}(d)$ generalizes the concept of multiplexing gain region given in (6.10) to positive diversity gains. In order to illustrate this point, consider the 2-user case with $\mathrm{M_T} = 3$, $\mathrm{M_R} = 4$ and $\rho = 2$. Figure 6.1 shows the multiplexing rate regions $\mathcal{R}(d)$ corresponding to several diversity levels, i.e., $d \in \{0, 2, 4, 8, 16\}$. The region $\mathcal{R}(0)$, equivalent to the multiplexing gain region (6.10), is the pentagon described by the constraints $r_1 \leq 3$, $r_2 \leq 3$ and $r_1 + r_2 \leq \min(4, 6)$ within which $d \geq 0$. Larger diversity requirement can be achieved at the expense of tighter constraints on r_1 and r_2. For instance, for a diversity requirement $d \geq 8$, the multiplexing rate region is given by the pentagon $0\mathrm{ABCD}$. Increasing the guaranteed amount of diversity results in multiplexing rate regions that shrink towards the origin. Note that to achieve a diversity requirement $d \geq 16$, the users are required to operate at very low multiplexing rates. In this regime, the multiplexing rate region is a square: performance is not affected (in the sense of the DM tradeoff) by the presence of a second user.

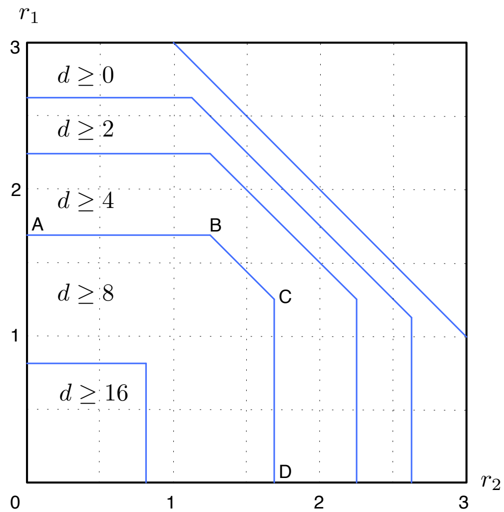

Fig. 6.1: Multiplexing rate regions as a function of the minimum diversity requirement $d \in \{0, 2, 4, 8, 16\}$ for both users ($\mathrm{M_T} = 3$, $\mathrm{M_R} = 4$ and $\rho = 2$). The diversity gain d corresponding to any multiplexing rate tuple inside the pentagon $0\mathrm{ABCD}$ satisfies $d \geq 8$.

6.4. ERROR MECHANISMS

The error exponent analysis conducted by Gallager (1985) for the AWGN MA channel can be extended to selective-fading MA MIMO channels, e.g., see (Gärtner and Bölcskei, 2006) in the context of fixed data rates. In this section, we review the concept of dominant error event, which is central to the error exponent framework, and show how it carries over into the DM tradeoff framework. We shall also see that the dominant error event leads to an accurate characterization of the mechanisms of error in selective-fading MA MIMO channels.

6.4.1. Preliminaries

The capacity region contains all the rate tuples $(R_1, R_2, \ldots, R_U)$ that can be transmitted through the channel with vanishing probability of error as the block length grows to infinity. Similar to the study of error exponents in the point-to-point case (Gallager, 1968), the error exponent analysis conducted by Gallager (1985) for the AWGN MA channel aims at quantifying the decay of the error probability with block length as a function of the rate tuple $(R_1, R_2, \ldots, R_U)$ in the capacity region. Gallager focuses on the 2-user case, and finds that, depending on the tuple (R_1, R_2), the total error probability is either dominated by single-user errors, i.e., only one of the two users is decoded incorrectly, or by a joint error, i.e., both users are simultaneously decoded in error. This distinction gives rise to the notion of error events types, which, in our analysis, are referred to as $\mathcal{S}$-error events, and justifies the following decomposition of the error probability

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) + \mathbb{P}(\mathcal{E}_3)$$

where $\mathcal{E}_1$ corresponds to user 1 only being in error, $\mathcal{E}_2$ to user 2 only being in error, and $\mathcal{E}_3$ is the event of both users being simultaneously in error (note that we have used a similar decomposition of the error probability in (6.25)). The findings of Gallager show that the dominant type of error, or dominant error event, corresponds to the term with the smallest random coding exponent, i.e., the term with the slowest exponential decay with block length. Moreover, the achievable rate region can be separated into three distinct areas, each one dominated by a different error event.

6.4.2. Dominant error event

In the context of the DM tradeoff the problem is conceptually similar to that addressed by Gallager: we aim at identifying the $\mathcal{S}$-error event that dominates the total error probability as the SNR (and not the block length) grows to infinity. We summarize our result as follows.

Theorem 6.4. *Under the assumptions of Theorem 6.2, the total error probability satisfies*

$$P_e(\mathcal{C}_{\boldsymbol{r}}) \doteq \mathbb{P}(\mathcal{E}_{\mathcal{S}^\star})\,. \tag{6.47}$$

where the set $\mathcal{S}^\star$, defined in (6.37), corresponds to the dominant $\mathcal{S}$-outage event for the multiplexing rate tuple $\boldsymbol{r}$.

Proof. Let $\mathcal{E}_{\mathcal{S}'}$ denote the $\mathcal{S}$-error event that dominates the overall error probability so that, based on (6.25),

$$P_e(\mathcal{C}_{\boldsymbol{r}}) \doteq \mathbb{P}(\mathcal{E}_{\mathcal{S}'})\,.$$

By (6.42), we necessarily have $d_{\mathcal{S}^\star}(r(\mathcal{S}^\star)) = d(\mathcal{E}_{\mathcal{S}'})$, where

$$d(\mathcal{E}_{\mathcal{S}'}) = -\lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}(\mathcal{E}_{\mathcal{S}'})}{\log \mathsf{SNR}}.$$

Since $\mathcal{C}_{\boldsymbol{r}}$ satisfies (6.29) $\forall \mathcal{S} \subseteq \mathcal{U}$, Theorem 6.1 yields $d(\mathcal{E}_{\mathcal{S}'}) \geq d_{\mathcal{S}'}(r(\mathcal{S}'))$ and, hence, we get

$$d_{\mathcal{S}^\star}(r(\mathcal{S}^\star)) \geq d_{\mathcal{S}'}(r(\mathcal{S}')).$$

However, by the definition of $\mathcal{S}^\star$, we also have $d_{\mathcal{S}^\star}(r(\mathcal{S}^\star)) \leq d_{\mathcal{S}'}(r(\mathcal{S}'))$ which implies $d_{\mathcal{S}^\star}(r(\mathcal{S}^\star)) = d_{\mathcal{S}'}(r(\mathcal{S}'))$. We can hence conclude that $P_e(\mathcal{C}_{\boldsymbol{r}}) \doteq \mathbb{P}(\mathcal{E}_{\mathcal{S}^\star})$. □

The optimal DM tradeoff is therefore given by the SNR exponent corresponding to the dominant error event, which is to say that the concept of dominant error event carries over in the context of the DM tradeoff. This fact is remarkable since, except for interference, errors over the fading MA channel are mainly due to channel outage while in the original AWGN MA channel errors were exclusively due to atypically large noise realizations.

The following example illustrates the application of Theorem 6.2 to the two-user case, and reveals multiplexing rate regions dominated by different error events. Noteworthily, despite being obtained in a drastically different setting, our regions resemble Gallager's dominating error regions for the AWGN MA channel case.

Example

We assume $M_T = 3$, $M_R = 4$, and $\rho = 2$. For $U = 2$, the $2^2 - 1 = 3$ possible error events are denoted by $\mathcal{E}_1$ (user 1 only is in error), $\mathcal{E}_2$ (user 2 only is in error) and $\mathcal{E}_3$ (both users are in error). The SNR exponents of the corresponding error probabilities are obtained from (6.23) as

$$\begin{aligned} d_u(r_u) &= (3 - r_u)(8 - r_u), \quad u = 1, 2, \\ d_3(r_1 + r_2) &= \big(4 - (r_1 + r_2)\big)\big(12 - (r_1 + r_2)\big). \end{aligned} \tag{6.48}$$

Based on (6.48), we can now explicitly determine the dominant error event for every multiplexing rate tuple $\boldsymbol{r} = (r_1, r_2)$. In Figure 6.2, we plot the rate regions dominated by the different error events. Note that the SNR exponent of the error probability is zero whenever $r_1 > 3$, $r_2 > 3$ or $r_1 + r_2 > 4$. In the rate region dominated by $\mathcal{E}_1$, we have $d_1(r_1) < d_2(r_2)$ and $d_1(r_1) < d_3(r_1 + r_2)$, implying that the SNR exponent of the total error probability equals $d_1(r_1)$, i.e., the SNR exponent that would be obtained in a point-to-point selective-fading MIMO channel with $M_T = 3$, $M_R = 4$, and $\rho = 2$. The same reasoning applies to the rate region dominated by $\mathcal{E}_2$ and, hence, we can conclude that, in the sense of the DM tradeoff, the performance in regions $\mathcal{E}_1$ and $\mathcal{E}_2$ is not affected by the presence of the second user. In contrast, in the area dominated by $\mathcal{E}_3$, we have $d_3(r_1 + r_2) < d_u(r_u)$, $u = 1, 2$, which is to say that the multiuser interference does have an impact on the DM tradeoff and reduces the diversity gain that would be obtained if only one user were present.

Figure 6.3 shows the dominant error event regions for the same system parameters but with one additional receive antenna, i.e., $M_R = 5$. The benefits from having an additional antenna are noticeable: not only larger sum multiplexing rates are achievable, i.e., $r_1 + r_2 \leq 5$, but we also observe that the area where $\mathcal{E}_3$ dominates the error probability, and hence where interference reduces diversity gain, is significantly smaller relative to the area dominated by the single user errors $\mathcal{E}_1$ and $\mathcal{E}_2$. This effect, also pointed out in the finite-rate analysis

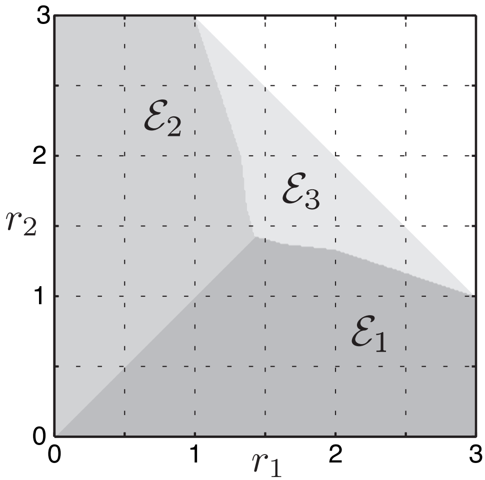

Fig. 6.2: Dominant error event regions for the two-user MA MIMO channel with $\mathrm{M_T} = 3$, $\mathrm{M_R} = 4$, and $\rho = 2$.

of (Gärtner and Bölcskei, 2006), can be intuitively explained by the fact that increasing $\mathrm{M_R}$ yields more spatial degrees of freedom at the receiver and, consequently, it alleviates the task of discriminating the users' signals in a larger portion of the multiplexing rate region. For a small set of high multiplexing rates, however, the joint error still dominates the error probability.

6.5. OPTIMAL DISTRIBUTED SPACE-TIME CODING

Satisfying the code design criterion (6.29) for every $\mathcal{S} \subseteq \mathcal{U}$ is non-trivial and systematic procedures for designing DM tradeoff optimal codes are an important open problem. In this section, we show that the

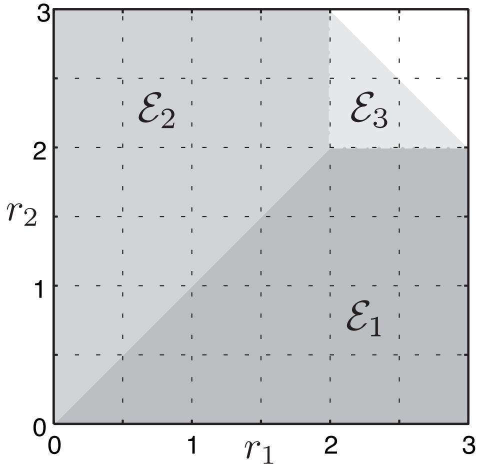

Fig. 6.3: Dominant error event regions for the two-user MA MIMO channel with $\mathrm{M_T} = 3$, $\mathrm{M_R} = 5$, and $\rho = 2$.

algebraic code construction proposed recently by Badr and Belfiore (2008b) for flat-fading MA channels with single-antenna users satisfies (6.29) for every $\mathcal{S} \subseteq \mathcal{U}$ and any multiplexing rate tuple*.

We start by briefly reviewing the code construction described in Badr and Belfiore (2008b) for a system with $\mathrm{M_T} = 1$, $\mathrm{M_R} = 2$, $U = 2$, $N = 2$, and $\rho = 1$. For each user u, let $\mathcal{A}_u$ denote a QAM constellation with $2^{R_u(\mathsf{SNR})}$ points carved from $\mathbb{Z}[i] = \{k + il : k, l \in \mathbb{Z}\}$, where $i = \sqrt{-1}$. The proposed code spans two slots so that the vector of information symbols corresponding to user u is given by $\mathbf{s}_u = [s_{u,0}\ s_{u,1}]$, where $s_{u,0}, s_{u,1} \in \mathcal{A}_u$. Using the unitary transformation matrix $\mathbf{U}$ underlying the Golden Code (Belfiore et al., 2005), the 1×2

*In (Badr and Belfiore, 2008b), the DM tradeoff optimality of the proposed code is shown for $r_1 = r_2$.

codeword $\mathbf{X}_u$ is obtained as

$$\mathbf{X}_u^T = \mathbf{U}\,\mathbf{s}_u^T = \begin{bmatrix} x_u \\ \sigma(x_u) \end{bmatrix} \tag{6.49}$$

with the matrix $\mathbf{U}$ given by

$$\mathbf{U} = \frac{1}{\sqrt{5}} \begin{bmatrix} \alpha & \alpha\varphi \\ \bar{\alpha} & \bar{\alpha}\bar{\varphi} \end{bmatrix} \tag{6.50}$$

where $\varphi = \frac{1+\sqrt{5}}{2}$ denotes the Golden number with corresponding conjugate $\bar{\varphi} = \frac{1-\sqrt{5}}{2}$, $\alpha = 1 + i - i\varphi$ and $\bar{\alpha} = 1 + i - i\bar{\varphi}$. Here, σ denotes the generator of the Galois group of the quadratic extension $\mathbb{Q}(i, \sqrt{5})$ over $\mathbb{Q}(i) = \{k + il : k, l \in \mathbb{Q}\}$ given by

$$\begin{array}{rccc} \sigma : & \mathbb{Q}(i, \sqrt{5}) & \rightarrow & \mathbb{Q}(i, \sqrt{5}) \\ & a + b\sqrt{5} & \mapsto & a - b\sqrt{5}. \end{array} \tag{6.51}$$

Moreover, one of the users, say user 2, multiplies the symbol corresponding to the first slot by a constant $\gamma \in \mathbb{Q}(i)$, resulting in the overall 2×2 codeword

$$\mathbf{X} = \begin{bmatrix} x_1 & \sigma(x_1) \\ \gamma x_2 & \sigma(x_2) \end{bmatrix}. \tag{6.52}$$

As shown in (Badr and Belfiore, 2008b), for any $\gamma \neq \pm 1$ and any two $\mathbf{X}, \mathbf{X}'$ according to (6.52), it holds that $\det(\boldsymbol{\Delta}) \neq 0$, where $\boldsymbol{\Delta} = \mathbf{X} - \mathbf{X}'$. For the so-defined construction we have the following result.

Theorem 6.5. *For any multiplexing rate tuple $\boldsymbol{r}$, the algebraic code construction in (Badr and Belfiore, 2008b) satisfies (6.29) for any $\mathcal{S} \subseteq \mathcal{U}$.*

Proof. For completeness, we start by proving that the determinant corresponding to any codeword $\mathbf{X}$ in (6.52) is nonzero for any $\gamma \neq \pm 1$, and hence, by the linearity of the mapping σ over $\mathbb{Q}(i, \sqrt{5})$, the determinant of any codeword difference matrix is also nonzero. Note that

$$\det(\mathbf{X}) = x_1\sigma(x_2) - \gamma x_2\sigma(x_1) \\ = x - \gamma\sigma(x) \tag{6.53}$$

where last step follows from setting $x = x_1\sigma(x_2)$ and using the property $\sigma(x \cdot y) = \sigma(x)\cdot\sigma(y)$ for every $x, y \in \mathbb{Q}(i, \sqrt{5})$. Hence, the determinant is zero if and only if γ satsifies $\gamma = x/\sigma(x)$. In this case, recalling that $\gamma \in \mathbb{Q}(i)$, we must have $x \in \mathbb{Q}(i)$, or $x \in \sqrt{5}\mathbb{Q}(i)$. These constraints yield, respectively, $\gamma = x/\sigma(x) = 1$ and $\gamma = x/\sigma(x) = -1$, from which we can conclude that $\det(\mathbf{X}) = 0 \iff \gamma = \pm 1$. Hence, any $\gamma \in \mathbb{Q}(i)\backslash\{\pm 1\}$ yields a nonzero determinant.

We consider now the second part of the proof. Assume that at any given SNR, user u carves out $2^{R_u(\mathsf{SNR})-\epsilon \log \mathsf{SNR}}$ points from $\mathbb{Z}[i]$ for some $\epsilon > 0$, i.e., $|\mathcal{A}_u| = \mathsf{SNR}^{r_u-\epsilon}$. In order to satisfy the power constraint (6.2), we scale $\mathcal{A}_u$ by $\mathsf{SNR}^{-(r_u-\epsilon)/2}$ so that, due to the linearity (over $\mathbb{C}$) of the transformation in (6.49), the codeword corresponding to user u is given by $\mathsf{SNR}^{-(r_u-\epsilon)/2}\,\mathbf{X}_u$. From (6.52) and the linearity of the mapping σ over $\mathbb{Q}(i, \sqrt{5})$, the codeword difference matrix is obtained as

$$\mathbf{E} = \begin{bmatrix} \mathsf{SNR}^{-(r_1-\epsilon)/2} e_1 & \mathsf{SNR}^{-(r_1-\epsilon)/2}\sigma(e_1) \\ \mathsf{SNR}^{-(r_2-\epsilon)/2}\gamma e_2 & \mathsf{SNR}^{-(r_2-\epsilon)/2}\sigma(e_2) \end{bmatrix} \tag{6.54}$$

where $e_u = x_u - x'_u$, $u = 1, 2$. Next, we note that in the flat-fading case $\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E} = \mathbf{E}^H\mathbf{E}$. In particular, considering user 1, i.e., $\mathcal{S} = \{1\}$, we have $|\mathcal{S}| = 1$ and $\mathrm{m}(\mathcal{S}) = 1$ so that, from (6.28), we obtain

$$\Xi_1^1(\mathsf{SNR}) = \mathsf{SNR}^{-(r_1-\epsilon)} \min_{e_1}(|e_1|^2 + |\sigma(e_1)|^2). \tag{6.55}$$

Letting $\mathbf{X}_1^T = \mathbf{U}\mathbf{s}_1^T$ and $(\mathbf{X}'_1)^T = \mathbf{U}(\mathbf{s}'_1)^T$ and using the fact that $\mathbf{U}$ is unitary, we get

$$(|e_1|^2 + |\sigma(e_1)|^2) = ||\mathbf{s}_1 - \mathbf{s}'_1||^2 \geq 2d_{\min}^2 \tag{6.56}$$

where $d_{\min}$ is the (nonzero and SNR-independent) minimum distance in $\mathcal{A}_1$. We therefore conclude that $\Xi_1^1(\mathsf{SNR}) \doteq \mathsf{SNR}^{-(r_1-\epsilon)}$. For user 2, a similar argument shows that $\Xi_1^1(\mathsf{SNR}) \doteq \mathsf{SNR}^{-(r_2-\epsilon)}$ and, hence, the construction satisfies the criteria arising from (6.29) for $\mathcal{S} = \{1\}$ and $\mathcal{S} = \{2\}$.

For $\mathcal{S} = \{1, 2\}$, note that $|\mathcal{S}| = 2$ and $\mathrm{m}(\mathcal{S}) = 2$ so that

$$\Xi_2^2(\mathsf{SNR}) = \min_{\mathbf{E}} |\det(\mathbf{E})|^2 . \tag{6.57}$$

From (6.54), we get

$$|\det(\mathbf{E})|^2 = \mathsf{SNR}^{-(r_1+r_2-2\epsilon)} |\det(\boldsymbol{\Delta})|^2 . \tag{6.58}$$

Recalling that for any $\gamma \neq \pm 1$, $\det(\boldsymbol{\Delta})$ is nonzero and independent of SNR, it follows that $|\det(\mathbf{E})|^2 \doteq \mathsf{SNR}^{-(r_1+r_2-2\epsilon)}$ and, consequently, we obtain

$$\Xi_2^2(\mathsf{SNR}) \doteq \mathsf{SNR}^{-(r_1+r_2-2\epsilon)} \tag{6.59}$$

from which we can conclude that (6.29) is also satisfied for $\mathcal{S} = \{1, 2\}$. The proof is concluded by taking ϵ to be arbitrarily close to zero, implying that both users operate arbitrarily close to their target multiplexing rates. □

Other code constructions that achieve the optimal DM tradeoff of flat-fading channels were recently reported in (Nam and El Gamal, 2007; Badr and Belfiore, 2008a). In Nam and El Gamal (2007), it is shown that lattice-based space-time codes achieve the optimal DM tradeoff with lattice decoding. Badr and Belfiore (2008a), on the other hand, extend the construction considered in this section to the case of multiple antennas at the users' transmitters. However, we note that the problem of systematically constructing DM tradeoff optimal codes seems to remain largely unexplored in the case of selective-fading MA MIMO channels.

CHAPTER 7

Conclusion

THIS DISSERTATION provides a high-SNR characterization of the optimal rate-reliability tradeoff for communication systems operating over selective-fading wireless channels. Our analysis and results are presented in the diversity-multiplexing (DM) tradeoff framework (Zheng and Tse, 2003), in which rate and reliability are respectively captured by the concepts of multiplexing and diversity.

The principal difficulty we overcome to establish the optimal DM tradeoff for selective-fading channels is of technical nature: the mutual information for this class of channels is in general a sum of correlated random variables whose distribution is unknown. Nevertheless, we are able to carry our analysis through by introducing the "Jensen channel", that is, a conceptual channel associated with the original channel and which behaves likewise in the high-SNR regime. The Jensen channel turns out to be a very useful tool not only in the context of point-to-point channels and multiple-access (MA) channels addressed in this thesis, but also, as evidenced by the work of Akçaba et al. (2007), in analyzing ultimate communication performance in the relay channel.

The proof techniques employed in this thesis reveal code design criteria that guarantee optimal performance in both the point-to-point case and the MA case. We also establish the existence of codes satisfying these criteria by providing examples of optimal constructions.

In the point-to-point case, we show that the problem of constructing optimal codes can be split into two simpler and independent tasks: the construction of a precoder, or inner code, which takes explicitly into account the channel's selectivity characteristics, and the construction of an outer code which is independent of the channel. This thesis presents a simple procedure, which is a generalization of the well-established phase- and delay-diversity techniques, to construct the precoder based on the channel's covariance matrix. We show that the permutation codes over quadrature amplitude modulation (QAM) constellations introduced by Tavildar and Viswanath (2006) can be used in conjunction with our precoder to achieve optimal performance. Identifying additional outer codes that could be used with our precoder would be an interesting aspect to consider in future research.

We also show that the code construction for the two-user MA channel presented in (Badr and Belfiore, 2008b) satisfies our code design criteria. However, more work is still needed to address the problem of constructing optimal codes for the general MA channel. A possible avenue to explore would be to extend our systematic code construction procedure to the MA channel, i.e., to jointly develop the users' precoders and corresponding outer codes in order to satisfy the design criteria. While the precoder design technique presented in this thesis seems to be readily extendable to multiple users, it is still unclear how to design the users' outer codes.

Finally, this thesis also presents an interesting relation between the DM tradeoff framework in MA channels and the concept of dominant error event which was first introduced by Gallager (1985) for additive white Gaussian noise (AWGN) MA channels. The idea is that a decoding error can be decomposed into several disjoint error events of which only one dominates the error probability asymptotically in SNR. Extending this concept to additional communication channels in future work seems promising, and it might lead to a more comprehensive understanding of the error mechanisms in wireless communication.

APPENDIX A

Linear Frequency-Invariant and Linear Time-Invariant Systems

This appendix examines the input-output relation over linear frequency-invariant (LFI) and linear time-invariant (LTI) channels. These channels can be thought of as limiting cases of a linear time-varying (LTV) system with vanishing delay spread or vanishing Doppler spread, respectively. Although LFI and LTI channel operators have a well-structured set of eigenfunctions (complex sinusoids for LTI systems and Dirac functions parametrized by time for LFI systems), their diagonalization cannot be casted in terms of the general underspread LTV framework presented in Chapter 2 (Durisi et al., 2008, Sec. II.C).

LFI and LTI systems are rather popular models to describe wireless channels in the communication engineering literature (e.g., Proakis, 2001; Tse and Viswanath, 2005). It is therefore interesting to note that the input-output relation used throughout this thesis, i.e.,

$$y_n = \sqrt{\mathsf{SNR}}\, h_n\, x_n + z_n,\; n = 0, 1, \ldots, N-1, \tag{A.1}$$

is also applicable to describe communication over this class of channels. We consider here the case of scalar channels, but the discussion readily extends to multiple-antenna channels.

In the case of an LTI channel, the input-output relation (A.1) can be obtained by using an orthogonal frequency-division multiplexing (OFDM) system (Peled and Ruiz, 1980) with N tones. Assuming that the OFDM system employs a cyclic prefix (guard interval) exceeding the channel's delay spread, the effect of the LTI channel on the transmit signal is described in terms of a cyclic (instead of linear) convolution. Then, applying appropriate Fourier transforms to the signal at the transmitter and at the receiver, the original frequency-selective channel is transformed into N non-interfering subchannels as in (A.1). In the case of an LFI channel, the input-ouput relation (A.1) can be obtained by employing a basis expansion model (BEM) (Ma et al., 2005).

The channel covariance matrix corresponding to (A.1) is obtained as follows. Suppose that $\{h(l)\}_{l=0}^{L-1}$ are L channel samples taken either in the delay domain for an LTI system, or in the Doppler frequency domain for an LFI system. Note that the resolution L depends upon the input signal bandwidth (LTI system), or upon the output signal duration (LFI system). For the two approaches considered above, the channel is described by

$$h_n = \sum_{l=0}^{L-1} h(l)\, e^{-j\frac{2\pi}{N}ln},\ n = 0, 1, \ldots, N-1,$$

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

$$\mathbb{E}\{h(l)h^*(l')\} = \sigma_l^2\, \delta(l - l').$$

The samples are uncorrelated due to the uncorrelated scattering (US) assumption for the LTI channel, and due to the wide-sense stationary (WSS) assumption for the LFI channel. Then, the channel's covariance matrix clearly follows as

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

where $\mathbf{\Psi}$, $\mathbf{\Psi}(i, j) = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}ij}$ $(i, j = 0, 1, \ldots, N-1)$, denotes the $N \times N$ Fourier matrix.

APPENDIX B

Distribution of the Jensen Channel

We show that the Jensen channel satisfies $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_w(\mathbf{R}^{T/2} \otimes \mathbf{I}_{\mathrm{M}})$, where

$$\begin{cases} \mathbf{R} = \mathbf{R}_{\mathbb{H}}, \boldsymbol{\mathcal{H}}_w = [\mathbf{H}_{w,0}\ \mathbf{H}_{w,1}\ \cdots\ \mathbf{H}_{w,N-1}] & \text{if } \mathrm{M_R} \leq \mathrm{M_T}, \\ \mathbf{R} = \mathbf{R}_{\mathbb{H}}^T, \boldsymbol{\mathcal{H}}_w = [\mathbf{H}_{w,0}^H\ \mathbf{H}_{w,1}^H\ \cdots\ \mathbf{H}_{w,N-1}^H] & \text{if } \mathrm{M_R} > \mathrm{M_T}. \end{cases} \tag{B.1}$$

We start by considering the case $\mathrm{M_R} \leq \mathrm{M_T}$, where, recalling (4.22), the Jensen channel is simply given by $\boldsymbol{\mathcal{H}} = [\mathbf{H}_0\ \mathbf{H}_1\ \cdots\ \mathbf{H}_{N-1}]$. By (4.5), it follows that $\mathbb{E}\{\mathrm{vec}\,(\boldsymbol{\mathcal{H}})\,\mathrm{vec}\,(\boldsymbol{\mathcal{H}})^H\} = \mathbf{R}_{\mathbb{H}} \otimes \mathbf{I}_{\mathrm{M_T M_R}}$ so that, using $\mathrm{M} = \mathrm{M_T}$, we may write $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_w(\mathbf{R}_{\mathbb{H}}^{T/2} \otimes \mathbf{I}_{\mathrm{M}})$. Partitioning $\boldsymbol{\mathcal{H}}_w$ into blocks of size $\mathrm{M_R} \times \mathrm{M_T}$ readily yields (B.1) for the case $\mathrm{M_R} \leq \mathrm{M_T}$.

In the case $\mathrm{M_R} > \mathrm{M_T}$, we set $\mathbf{R}_{\mathbb{H}}^{T/2}(n, n') = r_{n,n'}$ for the sake of notation, and note that the following equality holds for every realization of the channel:

$$\begin{aligned} [\mathbf{H}_0\ \mathbf{H}_1\ \cdots\ \mathbf{H}_{N-1}] &= [\mathbf{H}_{w,0}\ \mathbf{H}_{w,1}\ \cdots\ \mathbf{H}_{w,N-1}]\big(\mathbf{R}_{\mathbb{H}}^{T/2} \otimes \mathbf{I}_{\mathrm{M_T}}\big) \\ &= \sum_{n=0}^{N-1} [r_{n,0}\mathbf{H}_{w,n}\ r_{n,1}\mathbf{H}_{w,n}\ \cdots\ r_{n,N-1}\mathbf{H}_{w,n}]. \end{aligned}$$

Next, we use this observation to write

$$\begin{aligned}
\boldsymbol{\mathcal{H}} &= [\mathbf{H}_0^H \; \mathbf{H}_1^H \; \cdots \; \mathbf{H}_{N-1}^H] \\
&= \sum_{n=0}^{N-1} [r_{n,0}^* \mathbf{H}_{w,n}^H \; r_{n,1}^* \mathbf{H}_{w,n}^H \; \cdots \; r_{n,N-1}^* \mathbf{H}_{w,n}^H] \\
&= [\mathbf{H}_{w,0}^H \; \mathbf{H}_{w,1}^H \; \cdots \; \mathbf{H}_{w,N-1}^H] \big((\mathbf{R}_{\mathbb{H}}^{T/2})^* \otimes \mathbf{I}_{\mathrm{M_R}} \big).
\end{aligned} \tag{B.2}$$

Since $\mathbf{R}_{\mathbb{H}}^{1/2}$ is the positive semidefinite square root of $\mathbf{R}_{\mathbb{H}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, we have

$$\begin{aligned}
(\mathbf{R}_{\mathbb{H}}^{T/2})^* &= \big((\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^H)^T \big)^* \\
&= (\mathbf{U}^* \boldsymbol{\Lambda}^{1/2} \mathbf{U}^T)^* \\
&= \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^H = \mathbf{R}_{\mathbb{H}}^{1/2}
\end{aligned}$$

which, combined with (B.2) and $\mathrm{M} = \mathrm{M_R}$, yields

$$\boldsymbol{\mathcal{H}} = [\mathbf{H}_{w,0}^H \; \mathbf{H}_{w,1}^H \; \cdots \; \mathbf{H}_{w,N-1}^H](\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{\mathrm{M}})$$

hence establishing (B.1) for the case $\mathrm{M_R} > \mathrm{M_T}$.

APPENDIX C

Worst Pairwise Distance

The pairwise distance between two possible codewords takes the form $\mathrm{Tr}\left(\mathbf{B}\mathbf{A}\mathbf{B}^H\right)$, where $\mathbf{A}$ is Hermitian and $\mathbf{B}$ diagonalizable. This appendix provides a lower bound to $\mathrm{Tr}\left(\mathbf{B}\mathbf{A}\mathbf{B}^H\right)$ in terms of the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$. This result is proved in (Köse and Wesel, 2003, Theorem 2) for the case of a square matrix $\mathbf{B}$.

Suppose that $\mathbf{A}$ is a $n \times n$ Hermitian matrix and let its eigenvalue decomposition be given by

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Theta}\mathbf{U}^H \tag{C.1}$$

with $\boldsymbol{\Theta} = \mathrm{diag}\{\theta_l\}_{l=1}^n$, $\theta_1 \le \theta_2 \le \ldots \le \theta_n$, and $\mathbf{U} \in \mathcal{T}_n$, where $\mathcal{T}_n$ denotes the set of unitary matrices of size n. Moreover, let the singular value decomposition of the $m \times n$ matrix $\mathbf{B}$ be given by

$$\mathbf{B} = \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{W}^H \tag{C.2}$$

where $\mathbf{V} \in \mathcal{T}_m$, $\mathbf{W} \in \mathcal{T}_n$, and $\boldsymbol{\Lambda}$ has zeros in its off-diagonal entries but $[\boldsymbol{\Lambda}]_{k,k} = \lambda_k$ for $k = 1, 2, \ldots, m$, with $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_p$ and $p \triangleq \min(m, n)$. Then, we have the following theorem.

Theorem C.1. *For $\mathbf{A}$ and $\mathbf{B}$ given respectively by (C.1) and (C.2), we have*

$$\min_{\mathbf{U},\mathbf{W} \in \mathcal{T}_n} \mathrm{Tr}\left(\mathbf{B}\mathbf{A}\mathbf{B}^H\right) = \sum_{k=1}^{p} \lambda_k \, \theta_{p+1-k}.$$

Proof. Set $\mathbf{Q} = \mathbf{W}^H\mathbf{U}$, with $\mathbf{Q} \in \mathcal{T}_n$, and note that using (C.1) and (C.2), we get

$$\begin{aligned}
\min_{\mathbf{U},\mathbf{W}\in\mathcal{T}_n} \operatorname{Tr}\left(\mathbf{BAB}^H\right) &= \min_{\mathbf{Q}\in\mathcal{T}_n} \operatorname{Tr}\left(\mathbf{\Lambda}^{1/2}\,\mathbf{Q}\,\mathbf{\Theta}\,\mathbf{Q}^H\mathbf{\Lambda}^{1/2}\right) \\
&= \min_{\mathbf{Q}\in\mathcal{T}_n} \sum_{k=1}^{p} \lambda_k \sum_{l=1}^{n} |\mathbf{Q}(k,l)|^2\,\theta_l \\
&\geq \min_{\mathbf{D}\in\mathcal{D}_n} \sum_{k=1}^{p} \lambda_k \sum_{l=1}^{n} \mathbf{D}(k,l)\,\theta_l \qquad \text{(C.3)}
\end{aligned}$$

where $\mathbf{D}$, defined by $\mathbf{D}(k,l) = |\mathbf{Q}(k,l)|^2$, is doubly stochastic, $\mathcal{D}_n \supset \mathcal{T}_n$ denotes the set of doubly stochastic matrices of size n, and (C.3) is direct consequence of enlarging the set of admissible matrices to include doubly stochastic matrices. Since $\mathcal{D}_n$ is a convex compact set and (C.3) is a linear function of $\mathbf{D}(k,l)$, the minimum is attained at an extreme point of the set. By Birkhoff's Theorem (Horn and Johnson, 1993, 8.7.1), which says that the extreme points of $\mathcal{D}_n$ are permutation matrices $\mathbf{P}$, we have

$$\begin{aligned}
\min_{\mathbf{D}\in\mathcal{D}_n} \sum_{k=1}^{p} \lambda_k \sum_{l=1}^{n} \mathbf{D}(k,l)\,\theta_l &= \min_{\mathbf{P}} \sum_{k=1}^{p} \lambda_k \sum_{l=1}^{n} \mathbf{P}(l,k)\,\theta_l \\
&= \min_{\pi} \sum_{k=1}^{p} \lambda_k\,\theta_{\pi(k)} \qquad \text{(C.4)} \\
&= \sum_{k=1}^{p} \lambda_k\,\theta_{p+1-k} \qquad \text{(C.5)}
\end{aligned}$$

where (C.4) follows from the structure of $\mathbf{P}$, i.e., one entry per row is equal to one and the rest are zeros, π denotes a permutation on $\{1, 2, \ldots, n\}$, and (C.5) is a direct consequence of the ordering assumed for the λ_k's and θ_l's. The proof is concluded by noting that permutation matrices do also belong to $\mathcal{T}_n$, i.e., the initial set of admissible matrices, and, hence, the minimum in (C.5) is attained with equality. □

APPENDIX D

Notation

D.1. SYSTEM

$\mathrm{M_T}, \mathrm{M_R}$	number of transmit and receive antennas
m, M	$\min(\mathrm{M_T}, \mathrm{M_R})$ and $\max(\mathrm{M_T}, \mathrm{M_R})$
$R, R(\mathsf{SNR})$	data rate and SNR-dependent data rate
$\mathcal{U}$	set of users $\mathcal{U} = \{1, 2, \ldots, U\}$
$\mathcal{S}$	subset of users in $\mathcal{U}$
$\lvert\mathcal{S}\rvert$	cardinality of $\mathcal{S}$
$\bar{\mathcal{S}}$	complement of $\mathcal{S}$ in $\mathcal{U}$
$\mathrm{m}(\mathcal{S})$	$\min(\lvert\mathcal{S}\rvert\mathrm{M_T}, \mathrm{M_R})$
$\mathrm{M}(\mathcal{S})$	$\max(\lvert\mathcal{S}\rvert\mathrm{M_T}, \mathrm{M_R})$
$(R_1, R_2, \ldots, R_U)$	data rate tuple

D.2. DIVERSITY-MULTIPLEXING TRADEOFF

Point-to-point channel

r	multiplexing rate
$\mathcal{E}_r$	error event at multiplexing rate r
$\mathcal{O}_r$	outage event at multiplexing rate r
$\mathcal{J}_r$	Jensen outage event at multiplexing rate r
$\mathcal{C}_r$	family of codes operating at multiplexing rate r
$\mathcal{C}_r(\mathsf{SNR})$	codebook in $\mathcal{C}_r$ corresponding to SNR
$P_e(\mathcal{C}_r)$	error probability corresponding to $\mathcal{C}_r$
$d_{\mathcal{O}}$	SNR exponent of the outage probability
$d_{\mathcal{J}}$	SNR exponent of the Jensen outage probability
$d^{\star}$	optimal diversity-multiplexing tradeoff over all $\mathcal{C}_r$
$\mathbf{H}$	$[\mathbf{H}_0\,\mathbf{H}_1\,\cdots\,\mathbf{H}_{N-1}]$
$\boldsymbol{\mathcal{H}}$	Jensen channel
$\mathrm{I}(\mathsf{SNR})$	mutual information for i.i.d. Gaussian input
$\mathrm{J}(\mathsf{SNR})$	Jensen upper bound on $\mathrm{I}(\mathsf{SNR})$

Multiple-access channel

$\boldsymbol{r}$	multiplexing rate tuple $(r_1, r_2, \ldots, r_U)$
$r(\mathcal{S})$	total multiplexing rate corresponding to the users in $\mathcal{S} \subseteq \mathcal{U}$, i.e., $r(\mathcal{S}) = \sum_{u\in\mathcal{S}} r_u$
$\mathcal{E}_\mathcal{S}$	$\mathcal{S}$-error event
$\mathcal{O}_\mathcal{S}$	$\mathcal{S}$-outage event
$\mathcal{J}_\mathcal{S}$	Jensen outage event corresponding to $\mathcal{S}$
$\mathcal{O}_{\boldsymbol{r}}$	outage set w.r.t. the multiplexing rate tuple $\boldsymbol{r}$
$\mathcal{C}_{\boldsymbol{r}}$	overall codebook $\mathcal{C}_{r_1} \times \mathcal{C}_{r_2} \times \cdots \times \mathcal{C}_{r_U}$
$P_e(\mathcal{C}_{\boldsymbol{r}})$	error probability of $\mathcal{C}_{\boldsymbol{r}}$
$d_\mathcal{S}$	SNR exponent of $\mathbb{P}(\mathcal{J}_\mathcal{S})$
$d^\star$	optimal diversity-multiplexing tradeoff over all $\mathcal{C}_{\boldsymbol{r}}$
$\boldsymbol{\mathcal{H}}_\mathcal{S}$	Jensen channel of $\mathcal{S} = \{u_1, u_2, \ldots, u_{\lvert\mathcal{S}\rvert}\}$
$\mathbf{H}_{\mathcal{S},n}$	$[\mathbf{H}_{u_1,n}\ \mathbf{H}_{u_2,n}\ \cdots\ \mathbf{H}_{u_{\lvert\mathcal{S}\rvert},n}]$
$\mathbf{H}_\mathcal{S}$	$[\mathbf{H}_{\mathcal{S},0}\ \mathbf{H}_{\mathcal{S},1}\ \cdots\ \mathbf{H}_{\mathcal{S},N-1}]$
$\mathrm{I}_\mathcal{S}(\mathsf{SNR})$	sum mutual information for users in $\mathcal{S}$
$\mathrm{J}_\mathcal{S}(\mathsf{SNR})$	Jensen upper bound on $\mathrm{I}_\mathcal{S}(\mathsf{SNR})$

General

SNR	(per-user) average SNR at each receive antenna
$\doteq$	exponential equality: two functions $f(x)$ and $g(x)$ are said to be exponentially equal if and only if $\lim_{x\to\infty} \frac{\log f(x)}{\log x} = \lim_{x\to\infty} \frac{\log g(x)}{\log x}$.
$\dot{\leq}, \dot{\geq}$	exponential inequality
$\dot{<}, \dot{>}$	strict exponential inequality
$P_{\text{out}}(R)$	outage probability at rate R
$P_{\text{J}}(R)$	Jensen outage probability at rate R
λ_n	nonzero eigenvalues of the effective codeword difference matrix, i.e., $\lambda_k = \lambda_k(\mathbf{R}_{\mathbb{H}}^T \odot \mathbf{E}^H\mathbf{E})$
Ξ_n^m	product of the n smallest eigenvalues of m nonzero eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$

D.3. LINEAR ANALYSIS

$\mathbf{A}, \mathbf{a}, \bar{\mathbf{a}}$	matrix, column vector, and row vector
$\mathbf{A}(i,j), \mathbf{a}(i)$	(i,j)th entry of $\mathbf{A}$, ith entry of $\mathbf{a}$
$\mathbf{A}^T, \mathbf{A}^H$	transposition and conjugate transposition of $\mathbf{A}$
$\mathbf{A}^*$	element-wise conjugation
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} \odot \mathbf{B}$	Hadamard product of $\mathbf{A}$ and $\mathbf{B}$
$\mathrm{vec}(\mathbf{A})$	if $\mathbf{a}_k$ ($k{=}1, 2, \ldots, m$) are the columns of $\mathbf{A}$, then $\mathrm{vec}(\mathbf{A}) = [\mathbf{a}_1^T \; \mathbf{a}_2^T \; \ldots \; \mathbf{a}_m^T]^T$
$\mathrm{Tr}\,(\mathbf{A})$	trace of $\mathbf{A}$, i.e., $\mathrm{Tr}\,(\mathbf{A}) = \sum_i \mathbf{A}(i,i)$
$\\|\mathbf{A}\\|_{\mathrm{F}}^2$	squared Frobenius norm, i.e., $\\|\mathbf{A}\\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(\mathbf{A}\mathbf{A}^H\right)$
$\\|\mathbf{x}\\|$	norm of the vector $\mathbf{x}$
$\det(\mathbf{A})$	determinant of a nonsingular matrix, i.e., $\prod_i \lambda_i(\mathbf{A})$
$\mathrm{rank}(\mathbf{A})$	rank of $\mathbf{A}$
$\mathrm{diag}\{\mathbf{A}_n\}_{n=0}^{N-1}$	block diagonal matrix obtained by diagonally stacking $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1}$
$\mathbf{D}_{\mathbf{a}}$	$\mathrm{diag}\{\mathbf{a}(n)\}_{n=0}^{N-1}$
$\mathbf{A}^{1/2}$	positive semidefinite square root of an Hermitian matrix $\mathbf{A}$
$\mathbf{A}(\mathcal{I}_1, \mathcal{I}_2)$	(sub)matrix consisting of the rows of $\mathbf{A}$ indexed by $\mathcal{I}_1$ and the columns of $\mathbf{A}$ indexed by $\mathcal{I}_2$
$\lambda_k(\mathbf{A})$	nonzero eigenvalues of the $n \times n$ Hermitian matrix $\mathbf{A}$, sorted in ascending order ($k = 1, 2, \ldots, \mathrm{rank}(\mathbf{A})$)
$\mathbf{F}, \mathbf{\Psi}, \mathbf{\Phi}$	Fourier matrices ($\mathbf{F} = \mathbf{\Psi} \otimes \mathbf{\Phi}$ is two-level)
$\mathbb{1}_N$	$N \times N$ all-ones matrix
$\mathbf{\Pi}$	basic circulant permutation $[\boldsymbol{\pi}_1 \; \cdots \; \boldsymbol{\pi}_{N-1} \; \boldsymbol{\pi}_0]$, where $\boldsymbol{\pi}_n = [0 \; \cdots \; 0 \; 1 \; 0 \; \cdots \; 0]^T$ contains a 1 in its nth position.

D.4. PROBABILITY AND STATISTICS

$\mathbb{P}(\mathcal{A})$	probability of the event $\mathcal{A}$
$\mathbb{P}(\mathcal{A}, \mathcal{B})$	joint probability of $\mathcal{A}$ and $\mathcal{B}$
$\mathbb{P}(\mathcal{A} \mid \mathcal{B})$	conditional probability of $\mathcal{A}$ given $\mathcal{B}$
$f_X(x)$	probabiltiy density function (pdf) of X
$X \sim Y$	equality in distribution of X and Y
$f_{\mathbf{x}}, f(\mathbf{X})$	pdf of the random vector $\mathbf{x}$ or matrix $\mathbf{X}$
$\mathbb{E}\{\cdot\}$	expectation operator
$\mathbb{E}_X\{\cdot\}$	expectation operator with respect to X
$\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$	jointly proper Gaussian (JPG) random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$
$x \sim \mathcal{N}\big(\mu, \sigma^2\big)$	real Gaussian random variable with mean μ and variance σ^2

D.5. MISCELLANEOUS

$\mathbb{R}$	set of real numbers
$\mathbb{Q}$	set of rational numbers
$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of naturals
$\mathbb{C}$	set of complex numbers
$\mathfrak{Re}(x)$	real part of complex number x
$\mathfrak{Im}(x)$	imaginary part of complex number x
$\delta_{n,m}$	Kronecker delta function, i.e., $\delta_{n,m} = 1$ for $n = m$ and zero otherwise
$\delta(t)$	Dirac distribution
$\mathbb{F}$	Fourier operator
$\lvert x \rvert$	absolute value of $x \in \mathbb{R}$, or magnitude of $x \in \mathbb{C}$
$\mathcal{Q}(x)$	Gaussian tail function $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$

APPENDIX E

Acronyms and abbreviations

AWGN	additive white Gaussian noise
CSI	channel state information
DM	diversity-multiplexing
FEC	forward error correction
ICI	intercarrier interference
ISI	intersymbol interference
JPG	jointly proper Gaussian
LFI	linear frequency-invariant
LTI	linear time-invariant
LTV	linear time-varying
ML	maximum-likelihood
MA	multiple-access
MIMO	multiple-input multiple-output
OFDM	orthogonal frequency-division multiplexing
QAM	quadrature amplitude modulation
QoS	quality-of-service
SNR	signal-to-noise ratio
SISO	single-input single-ouput
WSSUS	wide-sense stationary uncorrelated scattering
w.p. 1	with probability 1
w.r.t.	with respect to

References

Adachi, F. (1979), "Transmitter diversity for a digital FM paging system," *IEEE Trans. Veh. Technol.*, vol. 28, pp. 333–338.

Ahlswede, R. (1971), "Multi-way communication channels," in *2nd Int. Symp. Inf. Theory (Tsahkadsor, Armenian S.S.R.)*, pp. 23–52, Budapest: Hungarian Academy of Sciences.

Akçaba, C., Kuppinger, P., and Bölcskei, H. (2007), "Distributed transmit diversity in relay networks," in *IEEE Information Theory Workshop (ITW)*, pp. 233–237, Bergen, Norway.

Alamouti, S. M. (1998), "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458.

Badr, M. and Belfiore, J.-C. (2008a), "Distributed space-time block codes for the MIMO multiple access channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2553–2557, Toronto, ON, Canada.

——— (2008b), "Distributed space-time block codes for the non cooperative multiple access channel," in *Proc. Int. Zurich Seminar on Commun.*, pp. 132–135.

Belfiore, J.-C. and Rekaya, G. (2003), "Quaternionic lattices for space-time coding," in *Proc. IEEE Inf. Theory Workshop*, pp. 267–270, Paris, France.

Belfiore, J.-C., Rekaya, G., and Viterbo, E. (2005), "The Golden code: A 2×2 full rate space-time code with nonvanishing determinants," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1432–1436.

Bello, P. A. (1963), "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. COM-11, pp. 360–393.

Biglieri, E. (2005), *Coding for Wireless Channels*, New York, NY: Springer.

Bölcskei, H., Borgmann, M., and Paulraj, A. J. (2003a), "Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 427–439.

——— (2003b), "Space-frequency coded MIMO-OFDM with variable multiplexing-diversity tradeoff," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 2837–2841, Anchorage, AK.

Bölcskei, H., Gesbert, D., and Paulraj, A. J. (2002a), "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 225–234.

Bölcskei, H., Koetter, R., and Mallik, S. (2002b), "Coding and modulation for underspread fading channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, p. 358, Lausanne, Switzerland.

Bölcskei, H. and Paulraj, A. J. (2000), "Space-frequency coded broadband OFDM systems," in *Proc. IEEE Wireless Commun. Net. Conf. (WCNC)*, pp. 1–6, Chicago, IL.

Boutros, J. and Viterbo, E. (1998), "Signal space diversity: A power- and bandwidth-efficient diversity technique for the rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1453–1467.

Coronel, P. and Bölcskei, H. (2007), "Diversity-multiplexing tradeoff in selective-fading MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2841–2845, Nice, France.

Coronel, P., Gärtner, M., and Bölcskei, H. (2008), "Diversity-multiplexing tradeoff in selective-fading multiple-access MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. pp. 915–919, Toronto, ON, Canada.

Cover, T. M. and Thomas, J. A. (1991), *Elements of Information Theory*, New York, NY: Wiley, 2nd ed.

Dayal, P. and Varanasi, M. K. (2005), "An optimal two transmit antenna space-time code and its stacked extensions," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4348–4355.

Divsalar, D. and Simon, M. K. (1988), "The design of trellis coded MPSK for fading channels: Performance criteria," *IEEE Trans. Commun.*, vol. 36, no. 9, pp. 1004–1012.

Durisi, G., Schuster, U. G., Bölcskei, H., and Shamai (Shitz), S. (2008), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*. Submitted. [Online]. Available: http://arxiv.org/abs/0804.1748.

Edelman, A. (1989), *Eigenvalues and condition number of random matrices*, Ph.D. thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA.

El Gamal, H., Caire, G., and Damen, M. O. (2004), "Lattice coding and decoding achieves the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 968–985.

Elia, P., Kumar, K. R., Pawar, S. A., Kumar, P. V., and Lu, H.-F. (2006), "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3869–3884.

Foschini, G. J. (1996), "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59.

Gallager, R. G. (1968), *Information Theory and Reliable Communication*, New York, NY: Wiley.

——— (1985), "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142.

Gärtner, M. and Bölcskei, H. (2006), "Multiuser space-time/frequency code design," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2819–2823, Seattle, WA.

Gray, R. M. (2006), "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239.

Grokop, L. and Tse, D. N. C. (2004), "Diversity/multiplexing tradeoff in ISI channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, p. 96, Chicago, USA.

Guey, J., Fitz, M., Bell, M., and Kuo, W. (1999), "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527–537.

Harashima, H. and Miyakawa, H. (1972), "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774–780.

Hiroike, A., Adachi, F., and Nakajima, N. (1992), "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. Veh. Technol.*, vol. 41, no. 2, pp. 170–176.

Horn, R. A. and Johnson, C. R. (1993), *Matrix Analysis*, Cambridge University Press, 2nd ed.

Jakes, W. (1974), *Microwave mobile communications*, New York, NY: Wiley.

James, A. T. (1964), "Distribution of matrix variates and latent roots derived from normal samples," *Annals of Mathematical Statistics*, vol. 35, pp. 475–501.

Jensen, J. L. W. V. (1906), "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Math.*, vol. 30, no. 1, pp. 175–193.

Kaiser, S. (2000), "Spatial transmit diversity techniques for broadband OFDM systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'00)*, vol. 3, pp. 1824–1828.

Köse, C. and Wesel, R. D. (2003), "Universal space-time trellis codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2717–2727.

Kozek, W. (1997), *Matched Weyl-Heisenberg expansions of nonstationary environments*, Ph.D. thesis, Vienna University of Technology.

Kozek, W. and Molisch, A. F. (1998), "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1579–1589.

Kuo, W. Y. and Fitz, M. (1997), "Design and analysis of transmitter diversity using intentional frequency offset for wireless communications," *IEEE Trans. Veh. Technol.*, vol. 46, no. 4, pp. 871–881.

Liao, H. H. J. (1972), *Multiple Access Channels*, Ph.D. thesis, University of Hawaii, Honolulu.

Ma, X., Leus, G., and Giannakis, G. B. (2005), "Space-time-Doppler block coding for correlated time-selective fading channels," *IEEE Trans. Sig. Proc.*, vol. 53, no. 6, pp. 2167–2181.

Matz, G., Schafhuber, D., Gröchenig, K., Hartmann, M., and Hlawatsch, F. (2007), "Analysis, optimization, and implementation of low-interference wireless multicarrier systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1921–1931.

Medles, A. and Slock, D. T. M. (2005), "Optimal diversity vs. multiplexing tradeoff for frequency selective MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1813–1817, Adelaide, Australia.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, New York, NY: Wiley.

Nam, Y. and El Gamal, H. (2007), "On the optimality of lattice coding and decoding in multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 211–215, Nice, France.

Ozarow, L. H., Shamai (Shitz), S., and Wyner, A. D. (1994), "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378.

Paulraj, A., Nabar, R., and Gore, D. (2003), *Introduction to space-time wireless communications*, Cambridge, UK: Cambridge University Press.

Peled, A. and Ruiz, A. (1980), "Frequency domain data transmission using reduced computational complexity algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 964–967, Denver, CO.

Proakis, J. G. (2001), *Digital Communications*, New York, NY: McGraw-Hill, 4th ed.

S. Boyd and L. Vandenberghe (2004), *Convex Optimization*, Cambridge, United Kingdom: Cambridge University Press.

Sampath, H., Stoica, P., and Paulraj, A. (2001), "Generalized linear precoder and decoder design for MIMO channels using hte weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206.

Sayeed, A. M. and Aazhang, B. (1999), "Joint multipath-doppler diversity in mobile wireless communications," *IEEE Trans. Commun.*, vol. 47, no. 1, pp. 123–132.

Scaglione, A., Stoica, P., Barbarossa, S., Giannakis, G. B., and Sampath, H. (2002), "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Sig. Proc.*, vol. 50, no. 5, pp. 1051–1064.

Schuster, U. G., Durisi, G., Bölcskei, H., and Poor, H. V. (2008), "Capacity bounds for peak-constrained multiantenna wideband channels," *IEEE Trans. Comm.* Submitted. [Online]. Available: http://arxiv.org/abs/0801.1002.

Seshadri, N. and Winters, J. H. (1994), "Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitted antenna diversity," *Int. J. Wireless Information Networks*, vol. 1, no. 1, pp. 49–60.

Shamai (Shitz), S. and Wyner, A. D. (1997a), "Information theoretic considerations for symmetric, cellular, multiple-access fading channels-Part I," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1877–1894.

——— (1997b), "Information theoretic considerations for symmetric, cellular, multiple-access fading channels-Part II," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1894–1911.

Shannon, C. E. (1948), "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656.

Tarokh, V., Naguib, A., Seshadri, N., and Calderbank, A. R. (1999), "Space-time codes for high data rate wireless communications: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Comm.*, vol. 47, pp. 199–207.

Tarokh, V., Seshadri, N., and Calderbank, A. R. (1998), "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765.

Tavildar, S. and Viswanath, P. (2006), "Approximately universal codes over slow-fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3233–3258.

Telatar, I. E. (1999), "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595.

Tomlinson, M. (1971), "New automatic equaliser employing modulo arithmetic," *IEE Electronics Letters*, vol. 7, no. 5/6, pp. 138–139.

Tse, D. and Viswanath, P. (2005), *Fundamentals of Wireless Communications*, Cambridge, UK: Cambridge University Press.

Tse, D. N. C., Viswanath, P., and Zheng, L. (2004), "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874.

Turin, G. L. (1960), "The characteristic function of Hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, pp. 199–201.

Tyrtyshnikov, E. E. (1996), "A unifying approach to some old and new Theorems on distribution and clustering," *Linear Algebra and its Applications*, vol. 232, pp. 1–43.

Visuri, S. and Bölcskei, H. (2006), "Multiple-access strategies for frequency-selective MIMO channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3980–3993.

Voois, P. A. (1996), "A theorem on the asymptotic eigenvalue distribution of Toeplitz-block-Toeplitz matrices," *IEEE Trans. Sig. Proc.*, vol. 44, no. 7, pp. 1837–1841.

Weerackody, V. (1993), "Diversity for the direct-sequence spread spectrum system using multiple transmit antennas," in *Proc. IEEE Int. Conf. Commun.*, pp. 1775–1779.

Wittneben, A. (1993), "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, pp. 1630–1634, Geneva, Switzerland.

Wornell, G. W. and Trott, M. D. (1997), "Efficient signal processing techniques for exploiting transmit antenna diversity on fading channels," *IEEE Trans. Sig. Proc.*, vol. 45, no. 1, pp. 191–205.

Wyner, A. D. (1974), "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10.

Yao, H. and Wornell, G. W. (2003), "Achieving the full MIMO diversity-multiplexing frontier with rotation based space-time codes," in *Proc.*

Allerton Conf. on Commun., Control and Computing, pp. 400–409, Monticello, IL.

Zheng, L. and Tse, D. N. C. (2003), "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096.

Curriculum Vitae

Pedro E. Coronel
Born May 13, 1979 in Morges (Switzerland)

Education

2003–2008	**Doctor of Science, ETH Zurich**
1997–2002	**Master of Science, EPF Lausanne** - Eurécom Institute, Sophia Antipolis, France (2001). - Royal Institute of Technology, Stockholm, Sweden: Exchange student (1999/2000).
1994–1997	**Swiss Federal Maturity Certificate** (maths/science) Gymnase Cantonal de la Cité, Lausanne.

Professional Experience

2006–2008	**Research assistant**, Communication Theory group Communication Technology Laboratory, ETH Zurich
2003–2006	**Pre-doctoral fellow**, Sensor Networks group IBM Zurich Research Laboratory
2002	**Intern**, Wireless Communication Systems group IBM Zurich Research Laboratory
2000-2001	**Teaching assitant**, Institute for Computer Communications and Applications, EPF Lausanne

Series in Communication Theory

Edited by Helmut Bölcskei

Vol. 1 Markus E. Gärtner, *Space-Time Coding and Multiple Access in MIMO Fading Channels*. ISBN 3-86628-181-1

Vol. 2 Moritz Borgmann, *Noncoherent MIMO Wideband Communications*. ISBN 3-86628-246-X

Vol. 3 Daniel S. Baum, *Information-Theoretic Analysis of a Class of MIMO Channel Measurement Devices*. ISBN 3-86628-247-8

Vol. 4 Ulrich G. Schuster, *Wireless Communication over Wideband Channels*. ISBN 3-86628-245-1

Vol. 5 Pedro E. Coronel, *Diversity-Multiplexing Tradeoff in Selective-Fading Channels*. ISBN 3-86628-251-6

More information at http://www.Hartung-Gorre.de
Hartung-Gorre Verlag Konstanz