uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Stojan Denic

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Electrical Engineering)

GRADE / DEGREE

School of Information Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Communication Subject to Normed Channel Uncertainties

TITRE DE LA THÈSE / TITLE OF THESIS

Charalambos Charalambous

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Nasir Uddin Ahmed

Frank Kschischang

Ioannis Lambadaris

Sergey Loyka

Gary W. Slater

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Communication Subject to Normed Channel Uncertainties

by

Stojan Denic

School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada

A dissertation sumbitted to the Faculty of Graduate and Postdoctoral Studies
in a partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Under the supervision of Prof. C.D. Charalambous

Your file Votre référence
ISBN: 978-0-494-15016-0
Our file Notre référence
ISBN: 978-0-494-15016-0

To
my beloved parents Zivodar and Ljiljana,
and my beloved sister Vesna.

Abstract

The transmission of information over a communication channel vastly depends on the level of knowledge that a transmitter and a receiver have about the channel and the interference. The transmission of information subject to insufficient knowledge of communication environment is called communication subject to uncertainties. The goal of this thesis is twofold: 1) To introduce new models for uncertain communication channels; 2) To define, compute, and analyze the performance of communication systems subject to introduced uncertainties from an information theoretic point of view. Various communication scenarios of compound single-input single-output and multiple-input multiple-output Gaussian channels are considered. There are three main contributions of the thesis: 1) The modeling of the channel and the noise uncertainties using H^{∞} and L_1 normed liner spaces in frequency domain; 2) In the case of single-input single-output channels, the channel uncertainty is modeled as a subset of H^{∞} space, while the noise uncertainty is modeled either by a subset of H^{∞} space or by a subset of L_1 space. Explicit formulas for the channel capacities, called robust capacities, and the optimal transmitted powers in the form of new water-filling formulas, are derived that explicitly depend on the sizes of the uncertainty sets. Moreover, when the noise uncertainty is modeled by a subset of L_1 space, the capacity formula has a game theoretical interpretation, where the transmitter tries to maximize the mutual information, while the noise tries to minimize it. It is shown that a saddle point exists and that the optimal PSD of the transmitter is proportional to the optimal PSD of the noise; 3) In the case of multiple-input multiple-output channels, two problems are considered. When the channel uncertainty is described by a subset of H^{∞} space, it is found that the transmission over the strongest singular value of the nominal channel frequency response matrix, representing the partial channel knowledge, is optimal for a large uncertainty set. When the noise uncertainty is described by a subset of L_1 space, the optimal power spectral density matrix of the noise is proportional to the optimal power spectral density matrix of the transmitted signal.

Acknowledgement

Foremost, I wish to express my deepest appreciation for two men without whose support, this thesis would not be possible. First, I would like to thank my Ph.D. advisor, Professor Charalambos D. Charalambous, for his patience, continuous guidance, encouragement and high standards throughout the last five years. I have benefited tremendously from his profound way of thinking, comprehensive knowledge, vision and enthusiasm. I thank Professor Charalambous for being a great teacher and a friend. Secondly, I would like to thank my friend and M.Sc. advisor, Professor Bane Vasic from the Department of Electrical and Computer Engineering, University of Arizona, for "infecting" me with the "diseases" called information theory and encoding and for his continuous support after my graduation. I will always remember how much fun we had during my undergraduate and M.Sc. studies.

I would like to expresses my deep gratitude to Professor Seddik M. Djouadi from the Department of Electrical and Computer Engineering, University of Tennessee, for many stimulating academic and social discussions. I am deeply indebted to his ideas and insights.

I also would like to thank Professor Ahmed Nasir from the School of Information Technology and Engineering, University of Ottawa, for interesting discussions we had during our visit to University of Cyprus.

I am deeply indebted to Professor Jean Yves Chouinard from the Department of Electrical and Computer Engineering, Laval University, for his in-depth technical views and sensibility.

I thank many of my colleagues, graduate students at the School of Information Technology and Engineering, University of Ottawa for their kindness and friendship, in particular, Alireza Farhadi, Farzad Rezai, Masud Kamel, Mohammad Pour, Ahmad Nejad, Afsaneh Sattari and Shili Lu. I am fortunate to have been a part of the Professor Charalambous' student group whose members are always fun to be with.

I would like to express a special gratitude to my friends and colleagues Natalija Vlajic, Amir Radji, Vincent Ng and Milenko Jorgic, for being truthful and devoted comrades throughout past years.

I wish to express my deepest appreciation for my friends and colleagues Dragoslav Culum, Ognjan Kabranov, Vladimir Tosic and their families for making their homes my home. I thank them for their unconditional support. I will miss all the good times we had together.

At the end, I would like to thank all members of the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, where this thesis has been written, for their hospitality, friendship, and support. In particular, I would like to thank Skevi, Irine, Filippos, Professor Andreas Kyprianou, Professor George Georghiou, Professor Christos Panayiotou, Professor Marios Polykarpou, graduate students Yiannis, Ioannis, Kiriakos, Costas, Stellios, Michalis, and researchers Christos Papageorgiou and Nick Polydorides.

Contents

List of Figures

List of Acronyms

1. a.s. - almost surely
2. AWGN - additive white Gaussian noise
3. CDI - channel distribution information
4. CSI - channel state information
5. GAVC - Gaussian arbitrarily varying channel
6. MIMO - multiple-input multiple-output
7. PSD - power spectral density
8. SISO - single-input single-output
9. SNR - signal-to-noise ratio

Notation and List of Symbols

Encoding and Decoding

1. Ω denotes the set of messages that are transmitted.
2. $\mathcal{X}$ denotes the channel input alphabet.
3. Ξ denotes the set of transmitted codewords.
4. $\mathbf{x}_k = (x_k^1, ..., x_k^l)$ denotes a codeword.
5. $\mathcal{Y}$ denotes the channel output alphabet.
6. Υ denotes the set of received words.
7. $\mathbf{y}_k = (y_k^1, ..., y_k^l)$ denotes a received word.
8. $(n, 2^{nR})$ denotes a code with a codeword length n and code rate R.
9. P_e^n denotes the probability of decoding error.
10. $I(p, Q)$ denotes the mutual information.
11. $p(x_k)$ denotes the probability mass functions of transmitted letter x_k.
12. $Q(y_k|x_k)$ denotes a conditional probability mass function of received letter y_k given transmitted letter x_k.
13. C denotes a channel capacity.
14. R denotes a code rate.
15. $Q(y|x, \theta)$ denotes a transition probability of a compound channel.
16. $D(P||Q)$ denotes the relative entropy between the probability measures P and Q.

17. (M, ϵ, T) denotes a channel code for an uncertain set of communication channels $\mathcal{B}$.

18. $I(\mathbf{x}; \mathbf{y})$ denotes the mutual information between a transmitted signal $\mathbf{x}$ and a received signal $\mathbf{y}$.

19. N_0 denotes the power spectral density of a white noise.

20. C^+ denotes the upper value or minimax capacity.

21. C^- denotes the lower value or maximin capacity.

Fields of Numbers

1. $\mathbf{C}$ denotes the set of complex numbers.

2. $\mathbf{R}$ denotes the set of real numbers.

3. $\mathbf{Z}$ denotes the set of integers.

Signals and Signal Norms

1. $\mathbf{x}$ denotes a transmitted signal.

2. $\mathbf{y}$ denotes a received signal.

3. $\mathbf{n}$ denotes a noise signal.

4. $S_{\mathbf{x}}(f)$ denotes the power spectral density of a transmitted signal $\mathbf{x}$.

5. $S_{\mathbf{n}}(f)$ denotes the power spectral density of a noise signal $\mathbf{n}$.

6. $J(S_{\mathbf{x}}, S_{\mathbf{n}}) \triangleq I(\mathbf{x}; \mathbf{y})$ denotes the mutual information between a transmitted signal $\mathbf{x}$ and a received signal $\mathbf{y}$

7. f denotes either the frequency with respect to time t or the spatial frequency with respect to delay τ for continuous-time signals

8. θ denotes the frequency with respect to time t for discrete-time signals

9. P is average transmitted power.

10. V denotes a vector space.

11. $\|.\|$ denotes the norm of a vector space V.

12. $\|.\|_p$ is a p-norm defined on the set of Lebesgue measurable functions.

13. $L_p(A)$ denotes the space of all functions having finite $\|.\|_p$ norm over set A.

14. $L_2(-\infty, +\infty)$ denotes the space of functions with finite $\|.\|_2$ norm and in the same time denotes the space of signals with finite energy.

15. $\|.\|_{\mathbf{P}}$ denotes power semi-norm.

16. $\mathbf{P}$ denotes the space of signals with finite power.

Systems

1. $G(f)$ denotes the frequency response of the linear system and $g(t)$ is a corresponding impulse response.

2. $\bar{\sigma}[G(s)]$ is a maximum singular value of a matrix $G(s)$.

3. $L_\infty(j\mathbf{R})$ denotes the space of essentially bounded functions on $j\mathbf{R}$.

4. H^∞ denotes the space of analytic and bounded functions in open right-half complex plain, while $\|.\|_\infty$ represents corresponding system norm.

5. For transfer function $H(s)$, $s = j\omega$, $arg(H)$ denotes the phase of $H(s)$.

Modeling of System Uncertainties

1. $H_{nom}(f)$ denotes the nominal frequency response.

2. $\Delta_1(f)W_1(f), \Delta_2(f)W_2(f)$ denote the perturbation of a nominal frequency response.

Chapter 1

Introduction

One of the most fundamental problems in communications is the computation of a maximal information transmission rate over a communication channel, which is known as Shannon's capacity, introduced in "A mathematical theory of communication" [64]. Channel capacity gives the ultimate upper bound on a transmission rate for a particular communication channel, for a specific decoding error (for instance, average or maximum). The channel capacity is an ultimate bound because if a transmission rate is above the channel capacity, then the reliable transmission of information is not possible, i.e., the probability of a decoding error is bounded away from zero.

This fundamental result is explained in more detail, below. Assume that a source generates messages belonging to a set $\Omega = \{\omega_1, ..., \omega_M\}$ that should be reliably transmitted over a communication channel, which is defined as a probabilistic mapping $h : \mathcal{X}^l \to \mathcal{Y}^l$, where $\mathcal{X}$ is the channel input alphabet, $\mathcal{Y}$ is the channel output alphabet, and l is the length of the channel input sequence. In order to communicate messages reliably over a communication channel subject to disturbances, one introduces a mapping $f : \Omega \to \Xi$ at the transmitter, called encoding, from the set of messages Ω, to the set of codewords $\Xi = \{\mathbf{x}_1, ..., \mathbf{x}_M\}$, $M = 2^{lR}$. A codeword $\mathbf{x}_k = (x_1^k, ..., x_l^k)$, $k = 1, ..., M$, which is transmitted through a communication channel, is a sequence of letters x_i^k, $i = 1, ..., l$, of length l from the channel input alphabet $\mathcal{X}$. Thus, $\Xi \subseteq \mathcal{X}^l$. R represents a code rate, i.e., the ratio of the number of bits needed for representation of all messages from Ω, $\log_2 M$, and the codeword length l.

The channel h maps a codeword $\mathbf{x}_k \in \Xi$ to the output word $\mathbf{y}_k \in \Upsilon$, $\Upsilon \subseteq \mathcal{Y}^l$. A word $\mathbf{y}_k = (y_1^k, ..., y_l^k)$ is a sequence of letters y_i^k, $i = 1, ..., l$, of length l from the channel output alphabet $\mathcal{Y}$. The received noisy codeword $\mathbf{y}_k$ is mapped into the original set Ω by a decoding function $g : \Upsilon \to \Omega$.

Assume that a codeword $\mathbf{x}_k = f(\omega_k)$ is transmitted and that $\mathbf{y}_k = h(\mathbf{x}_k)$ is a received word. An error in transmission occurs when $g(\mathbf{y}_k) = g(h(\mathbf{x}_k)) = g(h(f(\omega_k))) \neq \omega_k$. The probability of transmission (decoding) error when the encoding-decoding pair (f, g) is used, is denote as P_e^l. A positive constant R_a is called an attainable rate if there exists a sequence of codes $(l, 2^{lR_a})$ such that the decoding error $P_e^l \to 0$, as the codeword length l tends to infinity. The channel capacity C is equal to the supremum of all attainable rates.

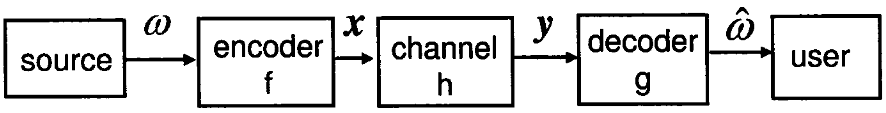

Figure 1.0.1: Model of communication system

Shannon brought forth the mathematical framework for the computation of the channel capacity C if there exists a suitable stochastic communication channel model. He showed that the capacity of a discrete memoryless channel is related to the so-called mutual information $I(p, Q)$. The mutual information is defined as the relative entropy between the joint probability $p(x_k, y_k) = p(x_k)Q(y_k|x_k)$ and the product of marginal probabilities $p(x_k)p(y_k)$ [64, 22], where x_k and y_k are the letters from the input and output channel alphabets, respectively. The channel capacity C is equal to the maximum of the mutual information $I(p, Q)$ over all possible probability mass functions of an input letter $p(x_k)$, where $Q(y_k|x_k)$ is a known conditional probability mass function of an output letter y_k given an input letter x_k [64]. Shannon's channel coding theorem states that if the code rate R is less than the capacity C defined by a maximal mutual information, then reliable transmission is possible with arbitrary small probability of error. And vice versa, the converse to the Shannon's coding theorem states that if $R > C$, then the probability of the decoding error is bounded away from zero.

In Shannon's definition of the channel capacity, it is assumed that the conditional distribution $Q(y_k|x_k)$ is completely known. The legitimate questions that can be asked are: How does the channel capacity change if the conditional distribution $Q(y_k|x_k)$ is not completely known? How can the channel capacity be computed if the conditional distribution is not completely known? What are the optimal communication strategies when the channel is uncertain? These questions do not have purely theoretical meaning. The practicality of the previous questions can be understood by realizing that any channel estimation introduces errors. This implies that the channel will never be perfectly known to the transmitter and/or receiver. This thesis tries to give the answers to the above questions for a specific class of communication channels when there is partial knowledge of the former. Communication when there is partial knowledge of a channel is often called communication subject to uncertainty.

Shannon himself realized the importance of computing the channel capacities when the channels are uncertain. For this specific reason, he introduced the notion of "channel state information" (CSI) [65]. For instance, the knowledge of attenuations, delays, and Doppler spreads at the transmitter and/or receiver represent the CSI in the case of wireless channels. Shannon computed the channel capacity of discrete memoryless communication channels when the CSI is available to the transmitter. The first paper dealing with the capacity for the class of channels is the work by Blackwell, Breiman, Thomasian [12] and Dobrushin [29], in which they computed the channel capacity of a class of channels. A comprehensive review of the topic can be found in [48].

The rest of the chapter is organized as follows. In Section 1.1, the review of related literature is given. In Section 1.2, the thesis goals are introduced and motivated. In Section 1.3, the survey of the technologies where the thesis results may be applied is given. In Section 1.4, the statements of the problems are presented. In Section 1.5, the main contributions of the thesis are outlined.

1.1 Survey of Related Research

This section summarizes the models and results on the capacity of uncertain channels which are relevant to this thesis.

1.1.1 Channel Modeling

In practical applications, one of the important issues in channel capacity computation, is the choice of an appropriate channel model. When the channel is uncertain, this is not a trivial problem. One of the main thesis contributions is the introduction of new communication channel models in the frequency domain, when the channel is subject to uncertainty. In this section, some current models are presented.

In the information theory literature, a basic model, which is often used to describe a parametric uncertainty in the channel is the one depicted in Fig. 1.1.2 [48]. A variable θ,

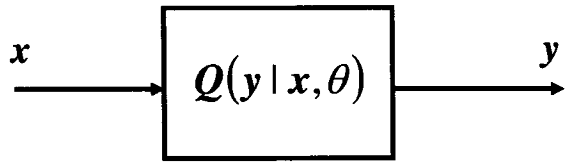

Figure 1.1.2: Probabilistic representation of a communication channel

which belongs to a certain set Θ, parameterizes a conditional distribution $Q(y|x,\theta)$. Hence, instead of dealing with a fixed known channel, one considers the capacity for the class of channels $\{Q(y|x,\theta) : \theta \in \Theta\}$. Generally speaking, there are two classes of uncertainty models; the class of *compound channels*, and the class of *arbitrarily varying channels.* Compound and arbitrarily varying channels (AVC) are further classified into discrete memoryless and finite-state channels [48].

Next, the compound and arbitrarily varying discrete memoryless channels are presented, following [48].

Discrete memoryless channels

A family of discrete memoryless channels

$$\{Q(\mathbf{y}|\mathbf{x},\theta),\ \mathbf{x} \in \mathcal{X}^n,\ \mathbf{y} \in \mathcal{Y}^n,\ \theta \in \Theta\}_{n=1}^{\infty}, \tag{1.1.1}$$

where

$$Q(\mathbf{y}|\mathbf{x},\theta) = \prod_{t=1}^{n} Q(y_t|x_t,\theta), \tag{1.1.2}$$

and $\{Q(y_t|x_t,\theta),\ x_t \in \mathcal{X},\ y_t \in \mathcal{Y},\ \theta \in \Theta\}$ is a suitable subset of the set of all stochastic matrices $\mathcal{X} \times \Theta \mapsto \mathcal{Y}$, is called a discrete memoryless compound channel. Thus, compound channels assume that the true channel $Q_{true}(\mathbf{y}|\mathbf{x},\theta)$ is unknown, though, it is assumed that $Q_{true}(\mathbf{y}|\mathbf{x},\theta)$ belongs to the family of channels (1.1.1) and *remains* unchanged during the course of a transmission.

AVC's are generalization of (1.1.1) to include time variation in θ. The difference comes from a need to model situations in which the channel changes during subsequent transmitted letters x_k. Assume that Σ is a finite set of channel states and $\Theta = \Sigma^{\infty}$. Then an AVC is determined by

$$Q(\mathbf{y}|\mathbf{x},\mathbf{s}) = \prod_{t=1}^{n} Q(y_t|x_t, s_t), \tag{1.1.3}$$

where $\mathbf{s} = (s_1, ..., s_n)$, and $\{Q(y_t|x_t, s_t),\ x_t \in \mathcal{X},\ y_t \in \mathcal{Y},\ s_t \in \Sigma\}$ is a suitable subset of the set of all stochastic matrices $\mathcal{X} \times \Sigma \mapsto \mathcal{Y}$. Hence, at each moment t, the transition matrix $Q(y_t|x_t, s_t)$ is unknown, and it is determined by the channel state $s_t \in \Sigma$.

Gaussian channels

Continuous alphabet uncertain channels received much less attention in the literature than their discrete counterparts [48]. Most of the results are related to Gaussian uncertain channels, which are briefly described below.

A Gaussian arbitrarily varying channel (GAVC) is defined by

$$\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{n}, \tag{1.1.4}$$

where $\mathbf{x}$, $\mathbf{s}$, $\mathbf{n}$, $\mathbf{y}$ are random variables in $\mathbf{R}^k$. $\mathbf{n}$ is an additive noise consisting of k independent and identically distributed (i.i.d.) zero mean Gaussian random variables. Here, $\mathbf{s}$ is a jamming sequence, whose distribution is unknown and satisfies a power constraint, $\mathbf{x}$ is a transmitted signal, and $\mathbf{y}$ is a received signal. The special case of (1.1.4), when $\mathbf{s}$ consists of k i.i.d random variables, gives rise to a compound Gaussian channel.

Another type of Gaussian compound channel is found in [63], where the channel is represented by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}. \tag{1.1.5}$$

The uncertainty is introduced by assuming that a linear transformation $\mathbf{A}$ is unknown, although it is known to belong to some pre-specified class of linear transformations. $\mathbf{x}$ and $\mathbf{y}$ are in general continuous signals, and $\mathbf{n}$ is an additive Gaussian noise.

MIMO channels

Of special interest are MIMO channel models, due to the applications of multiple-antenna systems for wireless communication. Initially, multiple-antenna systems promised considerable gain for fading channels as compared to single-antenna systems [69], [34]. However, at a later stage, it has been shown that this gain depends on the level of knowledge that the transmitter and receiver have about the channel [38]. The received signal of a flat fading channel is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1.1.6}$$

where $\mathbf{x}$ is a transmitted vector in $\mathbf{C}^{\mathbf{m}}$, $\mathbf{n}$ and $\mathbf{y}$ are random variables in $\mathbf{C}^{p}$, and $\mathbf{H}$ is a channel matrix in $\mathbf{C}^{p\times m}$. The additive noise $\mathbf{n}$ and channel matrix $\mathbf{H}$ are ergodic and stationary, and their entries are i.i.d., zero mean, circularly symmetric complex Gaussian random variables. This model is also called zero-mean spatially white (ZMSW), and it corresponds to the long-term average distribution of the channel coefficients, averaged over a number of propagation environments.

If the fading paths between different antenna pairs are correlated, then the proposed model is called a channel covariance information (CCI) model, and is given by

$$\mathbf{H_c} = (\mathbf{A^r})^{1/2}\mathbf{H}(\mathbf{A^t})^{1/2}. \tag{1.1.7}$$

Here, $\mathbf{H}$ is a zero mean circularly symmetric complex Gaussian, and $\mathbf{A^t}$ and $\mathbf{A^r}$ are called the transmit and receive fade covariance matrices, respectively.

If the channel matrix $\mathbf{H}$ is not perfectly known to the transmitter and/or receiver, the uncertainty is modeled as additive

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{E}. \tag{1.1.8}$$

Here, $\hat{\mathbf{H}}$ represents the estimation of $\mathbf{H}$, while $\mathbf{E}$ is an estimation error. When $\hat{\mathbf{H}}$ is constant, and $\mathbf{E} = \sqrt{\alpha}\,\mathbf{H^w}$ (α is constant, $\mathbf{H^w}$ is zero mean circularly symmetric complex Gaussian), this model is called channel mean information (CMI) model.

1.1.2 Literature Review

In the past, based on the previously discussed models, channel capacity problems were defined and for most of them the solutions, in terms of channel coding theorems, were found [48]. Also, universal encoding and decoding strategies were proposed that enable reliable transmission of information over uncertain channels close to channel capacities. One deficiency of universal strategies is their unacceptable complexity, due to the difficulty in implementing it in real systems, such as wireless communication systems. Hence, this is an area of active research.

This section provides a review of the channel capacity results for Gaussian uncertain channels. The review outlines some important results in the area, and motivates the work undertaken in the thesis.

A. Previous work on Gaussian uncertain channels

Historically, it appears that Blachman was the first to investigate the channel capacity subject to uncertainty using a two player game theoretical framework [9]. One participant in the game is a transmitter-receiver pair, while the other is a jammer. The role of the former is to choose a communication strategy that maximizes the channel capacity. The role of the jammer is to minimize it by choosing between different noise signals. Blachman considered the case when the transmitter and jammer can decide between finite number of strategies. Based on a game theoretic approach, he defined "pure" and "mixed" strategies. In pure strategies, the players pick up their strategies deterministically, as opposed to mixed, where the players pick up their strategies according to some probability law. Using the von Neumann theorem, he established the existence of a saddle point for the mixed strategy problem, when the pay-off function is the average capacity with respect to the probability laws of the transmitter and jammer. Moreover, the capacity of the uncertain Gaussian band-limited channel is given, when the transmitter and jammer have limited power. Blachman proved that the worst jamming strategy is a white Gaussian noise, while the optimal transmitted signal is also a white Gaussian signal. From his game solution, the well-known Shannon's formula for Gaussian band-limited channels is obtained.

The capacity of a band-limited channel perturbed by the interference which depends on a transmitted signal was discussed in [10]. Here, power constraints are imposed on both the signal and the interference. Lower and upper bounds on the channel capacity are derived by using a sphere caps packing technique. The channel capacity is computed for the case when the channel is subject to additive white noise in addition to an interference signal, finite in power. Then, the interference is expanded such that a part of the interference power is proportional but anti-parallel to the signal. The rest of the power is orthogonal to the signal, and it is used to augment the white noise. This is so-called "interference of the second kind", relative to the case when the interference is present only ("interference of the first kind"). The channel capacity is given by $C_2 = W \log[1 + (\sqrt{P} - \sqrt{aJ})^2/(N + (1-a)J)]$. Here, P is the transmitted power, J is the interference power, W is the channel bandwidth, and N is the white noise power. The coefficient $a < 1$ is chosen to minimize the signal-to-noise ratio

$(\sqrt{P}-\sqrt{aJ})^2/(N+(1-a)J)$. It was shown also that for randomly selected transmitted messages on the sphere of radius $\sqrt{P}$, the worst case interference is the interference of the first kind or the interference of the second kind, whose component orthogonal to the signal, is white Gaussian noise, i.e., it is random in direction.

It appears that the most general framework to analyze the capacity of compound Gaussian channels is the one provided by Baker [2], [3], and Baker and Chao [5], [6]. These results are also related to the classical work of Kolmogorov [45], Gelfand and Yaglom [36], Pinsker [59], and Gallager [35].

In [2], the channel is represented by $\mathbf{y}=\mathbf{A}\mathbf{x}+\mathbf{n}$, where $\mathbf{x}$ is a stochastic process representing a transmitted signal, and $\mathbf{n}$ is a zero-mean Gaussian process representing a noise. It is assumed that almost all sample paths of $\mathbf{x}$ belong to a real separable Banach space B, and that almost all sample paths of $\mathbf{n}$ belong to a real separable Hilbert space H. $\mathbf{A}$ is assumed to be a measurable function from B into H. Thus, the classical results are extended to a non-linear $\mathbf{A}$. Necessary and sufficient conditions are determined for the existence of the mutual information between the transmitted and received signals, $I(\mathbf{x};\mathbf{y})$, in terms of constraints imposed on the probability measure of the transmitted signal $p_{\mathbf{x}}$ and the covariance operator $R_{\mathbf{n}}$ of a Gaussian noise probability measure $p_{\mathbf{n}}$. The sufficient condition can be expressed as

$$p_{\mathbf{x}}\circ\mathbf{A}^{-1}[\mathrm{range}(R_{\mathbf{n}}^{1/2})]=1, \tag{1.1.9}$$

where $p_{\mathbf{x}}\circ\mathbf{A}^{-1}[C]=p_{\mathbf{x}}\{\mathbf{x}:\mathbf{A}\mathbf{x}\in C\}$. For this kind of constraint, the supremum of $I(\mathbf{x};\mathbf{y})$ is achieved when $p_{\mathbf{x}}\circ\mathbf{A}^{-1}$ is Gaussian.

This approach is further investigated in [5]. The channel model is given by $\mathbf{y}=\mathbf{x}+\mathbf{s}+\mathbf{n}$, where $\mathbf{s}$ is a jammer signal, and $\mathbf{n}$ is a Gaussian background noise. All considered signals belong to $\mathbf{R^M}$. In this case, $\mathbf{R^M}$ denotes also an inner product space $(\mathbf{R^M},\langle\cdot,\cdot\rangle)$, $\langle u,v\rangle=\sum_{i=1}^{M}u_iv_i$. Denote the overall disturbance by $\mathbf{z}\triangleq\mathbf{s}+\mathbf{n}$. Then, the covariance matrix of $\mathbf{z}$ can be expressed as $R_{\mathbf{z}}=R_{\mathbf{n}}+R_{\mathbf{s}}=R_{\mathbf{n}}^{1/2}(I+Z)R_{\mathbf{n}}^{1/2}$, where Z is a covariance matrix. The energy constraint on the transmitted signal is given by $E\|\mathbf{x}\|_{\mathbf{n}}^2\le P_{\mathbf{x}}$, where $\|\mathbf{x}\|_{\mathbf{n}}=\|R_{\mathbf{n}}^{1/2}\mathbf{x}\|$, and $\|.\|$ is Euclidian norm. According to [2], the maximization over the transmitted signals can be limited to Gaussian measures $p_{\mathbf{x}}$ having a covariance $R_{\mathbf{x}}=R_{\mathbf{z}}^{1/2}TR_{\mathbf{z}}^{1/2}$, where T is a linear and trace class operator, with eigenvalues $\{\tau_i\}_{i\ge1}$. This is a necessary and sufficient condition for the existence of the mutual information, and it is related to (1.1.9). The transmitted signal constraint can be further reduced to $E\|\mathbf{x}\|_{\mathbf{n}}^2=\sum_i z_i\le P_{\mathbf{x}}$, $z_i=\tau_i\|(I+Z)^{1/2}U^*u_i\|^2$, where U is a unitary matrix satisfying $R_{\mathbf{z}}^{1/2} = R_{\mathbf{n}}^{1/2}(I + Z)^{1/2}U^*$, and $\{u_n\}_{n\geq 1}$ is complete orthonormal set. Then, the mutual information can be expressed as

$$I(\mathbf{x};\mathbf{y}) = \frac{1}{2}\sum_i \log(1 + z_i(1+\gamma_i)^{-1}). \tag{1.1.10}$$

The coefficients γ_i, which are related to the jammer's signal $\mathbf{s}$, are equal to $\langle Zv_i, v_i\rangle$, $v_i = U^*u_i$. Further, the expression for the mutual information (1.1.10) is used as a pay-off function to compute the channel capacity in the presence of the jammer $\mathbf{s}$.

Baker and Chao generalized previous result to the infinite dimensional channels in [6]. It was concluded that the effect of jamming is to convert the infinite-dimensional channel into a finite-dimensional channel having the same constraints in terms of the jammer's and transmitter's power constraints and the covariance of known Gaussian noise.

Instead of modeling uncertainty as additive interference with unknown statistic, the uncertainty can be imposed on the operator $\mathbf{A}$ that affects the transmitted signal as in the case of (1.1.5) [63]. This equation is general, and it is applicable to discrete-time channels as well as continuous-time. The operator $\mathbf{A}$ may represent a convolution operation. In this case, the received signal is given by

$$y(t) = \int_{-\infty}^{+\infty} h(t-\tau)x(\tau)d\tau + n(t). \tag{1.1.11}$$

The uncertain channel is defined by a set B of communication channel impulse responses $h(t)$, which satisfy certain constraints. Denote the corresponding channel frequency response by $H(f)$, the power spectral density of the transmitted signal by $S_{\mathbf{x}}(f)$ and the power spectral density of a white Gaussian noise $n(t)$ by N. The proof of coding theorem requires that $h(t) \in B$ be integrable and that B be conditionally compact. Under these conditions, the channel capacity of this compound channel is given by

$$C = \sup_{S_{\mathbf{x}}\in A}\inf_{h\in B}\int_{-\infty}^{+\infty} \log(1 + \frac{S_{\mathbf{x}}(f)|H(f)|^2}{N})df. \tag{1.1.12}$$

Thus, the channel capacity is expressed in the frequency domain. It is interesting to note that the capacity depends on the magnitude but not on the phase of the transfer function $H(f)$.

Compound Gaussian channels are also investigated in [29], [51], [28], [71], and [13].

Next, an overview of certain results associated with GAVC's is given. A typical model for GAVC is described by (1.1.4). Here, the interference $\mathbf{s}$ and transmitted signal $\mathbf{x}$ are mutually independent. Hughes and Narayan defined several problems for different types of constraints imposed on the transmitted signal $\mathbf{x}$ and unknown interference $\mathbf{s}$ [40]. It is assumed that the channel is scalar and that randomized codes are used. They considered two types of constraints, peak and average power constraint on both transmitted signal and interference. Out of four different cases investigated, only one, when both the transmitted signal and interference are subject to peak power constrain, has the Shannon's capacity. In this case, the capacity is equal to the capacity of the Gaussian channel when the power of a known Gaussian noise is augmented by the power of the interference. Thus, the interference cannot do more harm than the Gaussian noise of the same power. In other cases, only the λ capacities exist. Hence, the error probability cannot be made close to zero as the length of the codewords tends to infinity.

In [41], the vector version of the GAVC is considered. The transmitted signal and interference are subject to peak power constraints. It is found that there exist transmitter and jammer strategies that achieve a saddle point with respect to the mutual information. The optimal transmitted power is the one that water-fills the overall power of a background noise and jammer signal.

In [24], Csiszar and Narayan computed the capacity of GAVC for deterministic codes. Basar and Wu [7] employed a game theoretic approach to study uncertain channels, but in their approach the mean-square error was chosen as a pay-off function. For more detailed, and in-depth consideration of classical results on communication channels with uncertainties see [23], [48], [8], [75] and references therein.

Previous work on MIMO uncertain channels

In communications, MIMO models can be used to describe different communication problems. Most recently, MIMO models have been used for representation of multiple-antenna wireless communication. But earlier, the model was used to describe multi-pair telephone cable that includes the effect of far-end crosstalk [16]. MIMO model is also used for representation of communication between sensors in a sensor network. The literature review given here is mostly focused on the results related to the multiple-antenna communication systems.

When computing the capacity for MIMO channels, one usually specifies the type of channel information available to a transmitter and/or receiver. Two types of information patterns are used, namely, the availability of CSI, which means that the realization of the channel matrix $\mathbf{H}$ is known to the transmitter and/or receiver, and the availability of channel distribution information (CDI), which means that only the channel matrix distribution is known to the transmitter and/or receiver.

When the channel matrix has a full rank and is constant and known to both the transmitter and the receiver, the MIMO channel can be transformed into $min(p, m)$ independent channels. Here, m and p denote the numbers of transmitter and receiver antennas, respectively. The optimal transmitter's covariance matrix of the transformed channel is diagonal, and it is found by applying the water-filling argument [70].

When the CSI is available at the receiver side, and the CDI is available at the transmitter side, the main conclusions are the following. For ZMSW channels, the optimal transmitter's covariance matrix is a scaled identity matrix [70]. For large number of transmit antennas m, the channel capacity grows linearly with the number of receiver antennas p. It was this result that sparked the interest in MIMO communication systems. Regarding the outage probability, it is conjectured that the optimal covariance matrix is diagonal with the power equally distributed among a subset of the transmitter antennas. The higher the rate, the higher the outage probability, and the smaller the number of active antennas.

For the CCI model, the optimal transmitter's covariance matrix has the same eigenvectors as the transmit fade covariance matrix $\mathbf{A^t}$ [43]. For CMI model, the optimal transmitter's covariance matrix has the same principal eigenvector as the channel mean matrix, and the eigenvalues of the remaining eigenvectors are equal [44], [68]. For both CCI and CMI models, the necessary and sufficient conditions for beamforming are determined [44]. The conclusion is that the additional information regarding the channel, in the form of the channel mean or covariance, increases the capacity and helps determine the dominant mode (beamforming); moreover, the capacity is achieved using scalar codes [38].

When changes in the channel are fast enough so that the receiver is not able to track them, there is no channel state information. In this case, the receiver might be able to track the short-term distribution of the channel [38]. The main conclusions are as follows. For a ZMSW model with CDI at the transmitter and receiver, if block fading is assumed, the capacity is achieved when the transmitter's signal $T \times m$ matrix is equal to the product of two statistically independent matrices: a $T \times T$ isotropically distributed unitary matrix and a certain $T \times m$ random matrix that is diagonal, real, and nonnegative [50]. T is a positive integer that represents the length of the coherence interval in symbol periods, and m is the number of transmit antennas. Here, the channel capacity does not grow if the number of transmit antennas is increased beyond T. This result is generalized in [79]. While it is assumed that the channel model is the same as in [50], a geometric approach in computing the channel capacity is taken. It is found that the channel capacity grows linearly with $m^*(1-\frac{m^*}{T})$ for high SNR, where $m^* \triangleq \{m, p, \lfloor T/2 \rfloor\}$. Also, it is shown that the optimal number of transmit antennas is m^*, i.e., the use of larger number of antennas does not give larger capacity gain. If the block fading assumption is removed, then for high SNR, it is proved that the channel capacity grows only double logarithmically as a function of SNR [49].

For CCI models, the results are not so pessimistic [38]. Similarly to the ZMSW case, the capacity is achieved when the transmitter's signal $T \times m$ matrix is equal to the product of statistically independent $T \times T$ isotropically distributed unitary matrix and certain $T \times m$ random matrix that is diagonal, real, and nonnegative, and the matrix of eigenvectors of the transmit fade covariance matrix $\mathbf{A^t}$. Contrary to the ZMSW model, the channel capacity increases for $m > T$, as long as the transmitter's antenna channel fading coefficients are spatially correlated.

The effect of the channel estimation error on the channel capacity and outage probability is studied in [77], and the optimal spatial and temporal power adaptation strategies are provided. It is concluded that the spatial adaptation improves the ergodic capacity and reduces the outage probability. The temporal adaptation is useful in reducing the outage probability, while the effect on the ergodic capacity is negligible. Additional references that consider the capacity subject to uncertainties, when the channel is known to the receiver but not to the transmitter, are [58] and [53].

The previous discussion applies to the MIMO flat fading wireless channels. The other type of wireless channels is frequency selective fading channel. It is used to model broadband wireless channels.

A frequency selective fading channel with L significant scatterer clusters is given by

$$y(t) = \sum_{j=0}^{L-1} h(j)x(t-j) + n(t), \tag{1.1.13}$$

where $h(t)$ represents the t^{th} tap of the discrete-time MIMO fading channel impulse response [14].

Brandenburg and Wyner [16] computed the MIMO capacity for the case when the channel frequency response matrix $H(e^{j\theta})$ (corresponding to the discrete Fourier transform of

$\{h(t)\}_{t=-\infty}^{+\infty}$) and the power spectral density of the noise $n(t)$, $W_{\mathbf{n}}(\theta)$, are known to the transmitter and receiver. A similar result to this one is obtained by Raleigh and Cioffi [62]. The capacity result is given in terms of singular values of $H(e^{j\theta})$ and water-filling in the frequency domain over the singular values. Another interesting result is [14], where the channel capacity of the orthogonal frequency division multiplexing (OFDM)-based spatial multiplexing system is derived. It is shown that the use of MIMO frequency selective fading channels is beneficial in terms of the capacity gain relative to the flat fading MIMO channels in contrast to the SISO fading channel case.

1.2 Thesis Motivation

The focus of this thesis is on Gaussian compound channels, which are described by the following equation

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}. \tag{1.2.14}$$

Here, $\mathbf{x}$ and $\mathbf{y}$ represent discrete-time or continuous-time signals, $\mathbf{A}$ is a convolution operator, and $\mathbf{n}$ is an additive Gaussian noise. A convolutional operator is defined by

$$\mathbf{A}x(t) \triangleq \int_{-\infty}^{\infty} a(t-\tau)x(\tau)d\tau \tag{1.2.15}$$

in continuous-time case, and by

$$\mathbf{A}x(t) \triangleq \sum_{\tau=-\infty}^{\infty} a(t-\tau)x(\tau) \tag{1.2.16}$$

in discrete-time case. The main goal is to compute the channel capacities and determine the optimal transmission strategies for two basic models: 1) When there is uncertainty in the channel frequency response, and 2) When there is uncertainty in additive noise, which is not correlated with the transmitted signal. Thus, a clear distinction is made between the two models of uncertainty. To achieve these goals, the thesis introduces a modeling of uncertainty in the frequency domain.

From the previous literature review, it can be seen that current uncertainty models are often given in the time domain. On the other hand, modeling in the frequency domain is natural for both continuous-time and discrete-time stationary Gaussian channels, because the mutual information and random coding exponents are expressed in terms of channel frequency responses and noise power spectral densities as in (1.1.12) (see [35], [63]). The goal of this thesis is to fill in this gap.

The uncertainty modeling in frequency domain can be applied to frequency selective channels, in which the channel capacity, for the perfect CSI at the receiver, is given by [57]

$$C = E \int_{-\infty}^{+\infty} \log(1 + \frac{P|G(t,f)|^2}{N}) df, \tag{1.2.17}$$

where $G(t, f)$ is a frequency response at time t,

$$G(t,f) = \int_{-\infty}^{+\infty} g(\tau, t) \mathrm{e}^{-j2\pi f \tau} d\tau, \tag{1.2.18}$$

and E denotes the expectation taken over the statistics of the random process $G(t, f)$. Here, $g(\tau, t)$ is the impulse response of the wireless channel, N is the power spectral density of white noise, and P defines the power constraint.

The models introduced in the thesis bring some advantages and flexibility that previously introduced models do not have. One advantage is practicality, because it is possible to extract the uncertainty model from channel measurements, based on the Bode and Nyquist plots. Also, the capacity expressions and optimal transmitters' strategies (in terms of optimal power spectral densities) are presented in the frequency domain, which is more suitable representation from the engineering point of view as opposed to the capacity formulas given for instance in work by Baker and Chao [5] and [6], (1.1.10). By analyzing the optimal strategies in the frequency domain, one can obtain the information about the change of the optimal bandwidth subject to uncertainties in the channel frequency response or in the power spectral density of the noise.

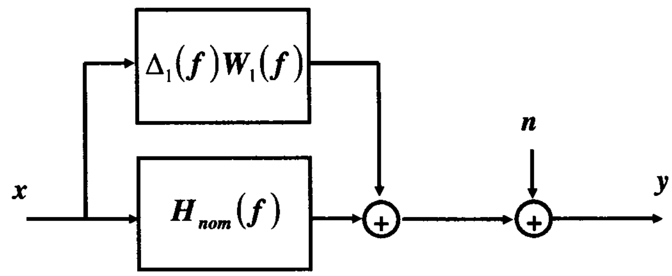

Figure 1.2.3: Additive description of uncertain Gaussian channel

1.2.1 Modeling in Frequency Domain

To described the uncertainty of a communication channel, H^∞ methods [80], which model the uncertainty of a system frequency response by using the H^∞ normed linear space (the space of analytic and bounded frequency responses in the open right half plane) are employed. If $H(f)$ is a transfer function matrix, then the H^∞ norm, $\|.\|_\infty$, is equivalent to

$$\|H\|_\infty = \sup_{f \in \mathbf{R}} \bar{\sigma}[H(f)], \; H \in H^\infty, \tag{1.2.19}$$

where $\bar{\sigma}[H(f)]$ is a maximal singular value of $H(f)$. For instance, an additive uncertainty description of a channel frequency response $\tilde{H}(f)$ in the presence of additive noise $\mathbf{n}$, shown in Fig. 1.2.3, is given by the following set

$$\{\tilde{H}(f) \in H^\infty : \tilde{H} = H_{nom} + \Delta_1 W_1, H_{nom} \in H^\infty, W_1 \in H^\infty, \Delta_1 \in H^\infty, \|\Delta_1\|_\infty \leq 1\} \tag{1.2.20}$$

$H_{nom}(f)$ is the so-called nominal channel frequency response. It is the result of the previous channel measurements or belief regarding the channel. $\Delta_1(f)W_1(f)$ represents the perturbation that determines the size of the uncertainty set. $\Delta_1(f)$ is a variable stable transfer function that accounts for the uncertainty in the phase, and acts like a scaling factor on the magnitude of the perturbation. $W_1(f)$ is a fixed stable transfer function, the weight. From the definition of the additive uncertainty description, it follows that the uncertainty set is a ball in the frequency domain centered at $H_{nom}(f)$, with the radius $|W_1(f)|$. Therefore, the uncertainty set can contain all kinds of transfer functions which are located at distance $|W_1(f)|$ from $H_{nom}(f)$.

H^∞ models capture both parametric and non-parametric uncertainties. To understand a difference between parametric and non-parametric uncertainty descriptions, note the following example. Take a second order transfer function

$$H(s) = \frac{1}{s^2 + as + b}, \tag{1.2.21}$$

$s = j2\pi f$. A parametric uncertainty description of $H(s)$ is obtained by letting the parameters a and b to take values in certain interval. Hence, a set of second order transfer functions is generated by varying a and b in their corresponding sets.

On the other hand, a low structured uncertainty description can be obtained by identifying the nominal transfer function $H_{nom}(s) = H(s)$ and the radius $W_1(s)$. In this way, the uncertainty description is richer than the parametric description, as the uncertainty set can contain transfer functions that are different from second order transfer functions.

In addition, by randomizing $H_{nom}(f)$, the proposed H^{∞} models can be made even more general (see [57]). Randomized H^{∞} models can be used to compute the capacity of time varying wireless fading channels. The short term uncertainty (within a coherence time [60]) can be modeled by H^{∞} methods, while the long term uncertainty can be captured by the distribution of $H_{nom}(f)$.

The previously discussed H^{∞} uncertainty models are used to describe the uncertainties with respect to the channel frequency response as well as with respect to the power spectral density of the noise. The uncertainty in the power spectral density of the noise can also be represented by a subset of L_1 function space, by limiting the total power of the noise. For some applications, as for instance, in communication subject to jamming, it is a reasonable to assume that the transmitter and the receiver do not have the information about the nominal power spectral density of the noise as required by H^{∞} uncertainty models. Rather, they may have just a rough information about the total transmitted power of the jammer. Thus, in this thesis, H^{∞} and L_1 uncertainty models will be used to model the uncertainty in the power spectral density of the noise.

1.3 Survey of Related Technologies

This section describes several examples of communication systems in which channel uncertainty is an intrinsic part of the systems.

1.3.1 Wireless Systems

Because of the mobility of users, the time variations of the environment and the multipath character of the wireless signals, the wireless channel continuously changes with time. In order to provide a required quality of service throughout the transmission, mobile and base stations estimate the channel parameters and adapt the encoding/decoding schemes and the transmitted power accordingly. Because the estimation introduces errors, this leads to the uncertainty about true channel parameters. Also, since the cellular system is a communication network, the users in addition to the targeted signal, receive interference from other users which work in the same bandwidth. Therefore, the interference from other users in a communication network represents another source of uncertainty. This problem is more prominent in the systems that use unlicensed bands (such as IEEE 802.11 standard), where many services operate in the same band without any regulations. Thus, it is necessary to compute the channel capacity under uncertain conditions and provide optimal transmission schemes. In Section 3.3, the capacity of a wireless multi-path channel is computed for two cases: 1) when only the receiver has a partial knowledge of the channel, and 2) when both, the receiver and transmitter have partial knowledge of the channel.

1.3.2 Jamming Systems

The communication systems in the presence of jamming are the classical representatives of the communication systems that deal with uncertainties. In this communication scenario, two or more parties communicate among each other. The adversary party tries to disturb the communication in the opponent network by sending jamming signals. Because the jamming signal and its statistics are usually unknown to the parties that communicate, the jamming signal may be viewed as a source of uncertainty for the users of the communication network. Section 3.4 provide the formula for the capacity of the channel subject to jamming.

1.4 Thesis Objective

The objective of this thesis is to define and compute the channel capacities and give optimal transmission strategies for uncertain stationary Gaussian channels when the channel uncertainty is described using H^∞ and L_1 normed linear spaces, in frequency domain. The problems, when the channel frequency response and/or the power spectral density (PSD) of the noise are uncertain, are considered for both single-input single-output (SISO) and multiple-input multiple-output (MIMO) communication channels. Thus, the main focus of the thesis is on a specific class of compound Gaussian channels. In this thesis, the capacity of the uncertain channel is referred to as the robust capacity.

In Chapter 3, SISO communication channels subject to uncertainties are considered. The channel model is given by the following equation

$$y(t) = \int_{-\infty}^{+\infty} \tilde{h}(t-\tau)x(\tau)d\tau + \int_{-\infty}^{+\infty} \tilde{w}(t-\tau)n(\tau)d\tau. \tag{1.4.22}$$

Here, $\mathbf{x} \triangleq \{x(t); -\infty < t < +\infty\}$ is a wide sense stationary process with finite power representing a transmitted signal, $\mathbf{n} \triangleq \{n(t); -\infty < t < +\infty\}$ is a Gaussian noise, $\mathbf{y} \triangleq \{y(t); -\infty < t < +\infty\}$ is a received signal, and $\tilde{h}(t)$ is the impulse response of the channel. The filter $\tilde{w}(t)$ shapes the PSD of the noise $\mathbf{n}$. It is assumed that $\tilde{h}(t)$, $\tilde{w}(t) \in L_2$. Moreover, $S_{\mathbf{x}}(f)$ and $S_{\mathbf{n}}(f)$ are the PSD's of $\mathbf{x}$ and $\mathbf{n}$, respectively, and $\tilde{H}(f)$ is a channel frequency response that corresponds to $\tilde{h}(t)$. The channel capacities of uncertain Gaussian channels are defined as the optimization problems by using the mutual information rate

$$J(S_{\mathbf{x}}, S_{\mathbf{n}}) = \int_{-\infty}^{+\infty} \log\Big(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}\Big) df, \tag{1.4.23}$$

as a pay-off function. Here and throughout the thesis, the base of the logarithm is e. The condition for the existence of $J(S_{\mathbf{x}}, S_{\mathbf{n}})$ (see chapter 8 of [35]) is given by $\int_{-\infty}^{+\infty} \frac{|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2} df < +\infty$.

The uncertainty of a channel frequency response $\tilde{H}(f)$ is modeled by using two low structured uncertainty models in H^∞ space; namely, additive and multiplicative models. The additive model is given by

$$B_1 \triangleq \{\tilde{H}(f) \in H^\infty : \tilde{H} = H_{nom} + \Delta_1 W_1, H_{nom}, W_1, \Delta_1 \in H^\infty, \|\Delta_1\|_\infty \le 1\}, \tag{1.4.24}$$

where $|W_1(f)|$ determines the size of the uncertainty set. The multiplicative model will be presented in Chapter 2. The uncertainty in the overall PSD of a noise, $S_{\mathbf{n}}(f)|\tilde{W}(f)|^2$, is modeled either through the uncertainty of the filter $\tilde{W}(f)$, as a subset of the H^∞ space

$$B_2 \triangleq \{\tilde{W}(f) \in H^\infty : \tilde{W} = W_{nom} + \Delta_2 W_2, W_{nom}, W_2, \Delta_2 \in H^\infty, \|\Delta_2\|_\infty \le 1\}, \tag{1.4.25}$$

or through the uncertainty of $S_{\mathbf{n}}(f)$, as the subset of the L_1 space

$$B_3 \triangleq \Big\{S_{\mathbf{n}}(f) : \int_{-\infty}^{+\infty} S_{\mathbf{n}}(f) df \le P_{\mathbf{n}}\Big\}. \tag{1.4.26}$$

Here, the constant $P_{\mathbf{n}}$ determines the size of the uncertainty set. A number of different problems are defined depending on whether the channel, the noise, or both are uncertain. The channel capacities and the transmitters' optimal PSD's are obtained as the solutions of corresponding maximin optimization problems, when $J(S_{\mathbf{x}}, S_{\mathbf{n}})$ is a pay-off function. In all problems, the maximum is taken over the set of all possible PSD's of a transmitted signal having bounded power

$$A \triangleq \Big\{S_{\mathbf{x}}(f) : \int_{-\infty}^{+\infty} S_{\mathbf{x}}(f) df \le P\Big\}. \tag{1.4.27}$$

If the channel frequency response is uncertain, the minimum is taken with respect to all channel frequency responses which belong to the uncertainty set B_1. If the power spectral density of the noise is uncertain, the minimum is taken with respect to all power spectral densities of the noise which belong to the uncertainty sets defined by B_2 or B_3. The most general information capacity problems are defined by

$$\sup_{S_{\mathbf{x}}\in A} \inf_{\tilde{W}\in B_2} \inf_{\tilde{H}\in B_1} J(S_{\mathbf{x}}, S_{\mathbf{n}}), \tag{1.4.28}$$

and

$$\sup_{S_{\mathbf{x}}\in A} \inf_{S_{\mathbf{n}}\in B_3} \inf_{\tilde{H}\in B_1} J(S_{\mathbf{x}}, S_{\mathbf{n}}). \tag{1.4.29}$$

The term "information" capacity is introduced in [22] (see page 184) to denote the the maximum of the mutual information. In this thesis, information capacity is used to denote the maximin of the mutual information in order to distinguish between the maximin mutual information and the channel capacity, here referred to as operational capacity.

The definition of the information capacity as a maximin of the mutual information rate may seem too conservative. However, the level of conservatism depends on the size of the uncertainty sets, and smaller uncertainty sets yield less conservative capacity. In the thesis, it will be shown that these information channel capacities represent the operational capacities. The maximin definition of the capacity formulas have a game theoretic explanation. One player is a transmitter that wishes to maximize the mutual information rate over the set A, while the other player is the noise (jammer), that wishes to minimize it over the set B_3. However, it should be noted that the channel capacities of compound channels are not always equal to the maximin of the mutual information. For an example see [48].

In Chapter 4, MIMO communication channels subject to uncertainties are considered. The channel is defined as follows

$$y(t) = \sum_{j=-\infty}^{+\infty} h(t-j)x(j) + n(t), \tag{1.4.30}$$

where $\mathbf{x} \triangleq \{x(t) : t \in \mathbf{Z}\}$ is a m-component complex stationary stochastic process representing a transmitted signal, $\mathbf{n} \triangleq \{n(t) : t \in \mathbf{Z}\}$ is a p-component Gaussian stochastic process representing an additive noise, $\mathbf{y} \triangleq \{y(t) : t \in \mathbf{Z}\}$ is a p-component stationary stochastic process representing a received signal and $\mathbf{h} \triangleq \{h(t) : t \in \mathbf{Z}\}$ is the sequence of complex $p \times m$ matrices representing the impulse response of the MIMO communication channel.

The information channel capacity problems are defined as maximin optimization problems, in which the pay-off function is the mutual information rate given by

$$J(W_{\mathbf{x}}, W_{\mathbf{n}}) = \frac{1}{4\pi}\int_0^{2\pi} \log\det(I + H(e^{j\theta})W_{\mathbf{x}}(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^{-1}(\theta))d\theta, \tag{1.4.31}$$

where $W_{\mathbf{x}}(\theta)$ is the PSD matrix of $\mathbf{x}$, $W_{\mathbf{n}}(\theta)$ is the PSD matrix of $\mathbf{n}$ and $H(e^{j\theta})$ is the channel frequency response matrix, which is the discrete Fourier transform of $\mathbf{h}$, $\theta \in [0, 2\pi]$.

The solutions of the following two problems are derived:

1) When the noise PSD matrix is completely known, while the channel frequency response matrix uncertainty is modeled by an additive model

$$B_1 \triangleq \left\{H \in H^{\infty} : H = H_{nom} + W_1 \Delta W_2, H, H_{nom}, W_1, W_2, \Delta \in H^{\infty}, \|\Delta\|_{\infty} \leq 1\right\}, \quad (1.4.32)$$

2) When the channel frequency response matrix is completely known, while the uncertainty in the noise PSD matrix is modeled by a subset of L_1 space,

$$B_2 \triangleq \{W_{\mathbf{n}}(\theta) : \int_0^{2\pi} \mathrm{Trace}(W_{\mathbf{n}}(\theta))d\theta \leq P_{\mathbf{n}}\}. \quad (1.4.33)$$

The maximum is taken over the set of transmitter's PSD matrices subject to the power constraint

$$A \triangleq \{W_{\mathbf{x}}(\theta) : \int_0^{2\pi} \mathrm{Trace}(W_{\mathbf{x}}(\theta))d\theta \leq P_{\mathbf{x}}\}. \quad (1.4.34)$$

In the case of uncertain channel frequency response matrix, the minimum is taken over all channel frequency response matrices, which belong to the set B_1. The information capacity is defined by

$$\sup_{W_{\mathbf{x}} \in A} \inf_{H \in B_1} J(W_{\mathbf{x}}, W_{\mathbf{n}}). \quad (1.4.35)$$

In the case of the uncertainty in the noise PSD matrix, the minimum is taken over the set B_2. Thus, the information capacity is defined by

$$\sup_{W_{\mathbf{x}} \in A} \inf_{W_{\mathbf{n}} \in B_2} J(W_{\mathbf{x}}, W_{\mathbf{n}}). \quad (1.4.36)$$

For certain special cases of the problems defined by the sets B_1 and B_2, it is shown that the computed MIMO information capacities (1.4.35) and (1.4.36) are equal to the operational capacities.

Chapter 5 contains the main points of the thesis and suggests directions for future research.

1.5 Contributions

The main contributions of Chapter 3 are the following:

1. The channel capacity formulas are derived which show how the capacities decrease with the size of uncertainty sets;

2. The transmitters' optimal PSD's are obtained in the form of water-filling equations which depend on the size of uncertainty sets as well. When the size of an uncertainty set increases, the transmitter tends to shrink the optimal bandwidth of a transmitted signal and to regroup the power towards lower frequencies;

3. The derived capacity formulas are employed to compute the capacity of ergodic channels;

4. When the noise uncertainty is modeled by the set B_3, it is found that the optimal PSD of the noise $S^o_{\mathbf{n}}(f)$, i.e. jammer, is proportional to the optimal PSD of the transmitter, $S^o_{\mathbf{x}}(f)$;

5. It is shown that all derived capacity formulas are equal to the operational capacities of corresponding communication problems.

The main contributions of Chapter 4 are the following:

1. The information channel capacities decrease as the size of uncertainty sets increase;

2. When the channel is uncertain, the information capacity formula suggests that the transmission over the strongest mode of the nominal channel frequency response matrix, representing the partial channel knowledge, is the optimal transmission strategy for high uncertainties ;

3. When the noise is uncertain, the optimal PSD matrix of the noise $W^o_{\mathbf{n}}(\theta)$ is, in a sense, proportional to the optimal PSD matrix of the transmitter $W^o_{\mathbf{x}}(\theta)$;

4. It is proved for certain special cases of problems defined by the sets B_1 and B_2 that computed information capacities are equal to the operational capacities.

Chapter 2

Modeling of Uncertainty

The objective of Information Theory is to provide the fundamental limits for reliable transmission of information subject to different assumptions on transmitted signals, disturbances and channel models. In this section, signal normed linear spaces and system normed linear spaces are introduced. System normed linear spaces are useful for the characterization of channel uncertainty. In this thesis, the uncertainty of channel frequency response or the uncertainty of power spectral density of the noise will be described by subsets of normed linear spaces. The sizes of those uncertainty sets are quantified by their norms.

2.1 Definition of Signal and System Norms

Next, the definition of normed linear spaces used in the thesis are given. All definitions presented here may be found in [80].

Definition 2.1.1 *Given a vector space $(V, \mathbf{C})$, where V is a set of vectors and $\mathbf{C}$ is the field of complex numbers, a real valued function $\|\cdot\|$ is called a norm on $(V, \mathbf{C})$ if and only if*

1. $\|x\| > 0$, $\forall x \in V$, $x \neq 0$,
2. $\|x\| = 0$ *if and only if* $x = 0$,
3. $\|\alpha x\| = |\alpha| \|x\|$, $\forall \alpha \in \mathbf{C}$, $\forall x \in V$,
4. $\|x + y\| \leq \|x\| + \|y\|$.

Definition 2.1.2 *A vector space $(V, \mathbf{C})$, accompanied with a norm $\|.\|$, is called a normed linear space.*

Remark 2.1.3 *A distance between $x, y \in V$ may be defined as $d(x,y) = \|x - y\|$, which is so-called "induced" matric, because it is defined by using a norm.*

2.1.1 Signal Norms

The signals, which are used in the thesis are finite energy or finite power signals. In this section, the formal mathematical definitions of those signals are given. Although here the definitions of real-valued signals are given, the complex-valued signals are defined in similar manner.

Definition 2.1.4 *The $\|\cdot\|_p$ norm of Lebesgue measurable functions $x : A \rightarrow \mathbf{R}$ over a set $A \subseteq \mathbf{R}$ is defined as*

$$\left(\int_A |x(t)|^p dt\right)^{1/p}, \; 1 \leq p < +\infty, \tag{2.1.1}$$

and

$$\|x(t)\|_\infty \triangleq ess \sup_{t\in A} |x(t)|. \tag{2.1.2}$$

Here, esssup $x(t)$ stands for essential supremum, and it is defined as the smallest number α such that the measure of the set $\{x : f(x) > \alpha\}$ is zero.

Definition 2.1.5 *The set of all functions on A having finite $\|\cdot\|_p$ norm is called $L_p(A)$ space.*

Definition 2.1.6 *Assume that a signal $x : (-\infty, +\infty) \rightarrow \mathbf{R}$ has autocorrelation function $R_{\mathbf{x}}(\tau)$, which is finite for any $\tau \in \mathbf{R}$, and has a corresponding PSD $S_{\mathbf{x}}(f)$. Then, the power semi-norm is defined by*

$$\|x\|_{\mathbf{P}}^2 \triangleq \lim_{T\rightarrow\infty} \frac{1}{2T} \int_{-T}^{+T} |x(t)|^2 dt \tag{2.1.3}$$

$$= \int_{-\infty}^{+\infty} S_{\mathbf{x}}(f) df. \tag{2.1.4}$$

The norm $\|.\|_{\mathbf{P}}$ is a semi-norm because there exist signals different from zero having $\|.\|_{\mathbf{P}}$ norm equal to zero. For instance, those are all signals that have finite duration. The norm $\|.\|_{\mathbf{P}}$ is also called the root-mean-square value of the signal.

Definition 2.1.7 *The set of Lebesgue measurable functions $x : \mathbf{R} \rightarrow \mathbf{R}$ with finite $\|.\|_2$ norm, which is denoted by $L_2(-\infty, +\infty)$, represents the signals with finite energy.*

Definition 2.1.8 *The set of functions $x : \mathbf{R} \rightarrow \mathbf{R}$ with finite $\|.\|_{\mathbf{P}}$ norm, which is denoted as $\mathbf{P}$, represents the signals with finite power.*

2.1.2 System Norms

This thesis deals with frequency responses that belong to so-called H^∞ space. Further, the definitions of $\|.\|_\infty$ norm and H^∞ space as well as related $\|.\|_2$ norm and H^2 space are given.

Definition 2.1.9 *A complex-valued function $g(s)$, defined on an open set $S \subset \mathbf{C}$ is an analytic function at point $z_0 \in S$ if it is differentiable at z_0 and at each point in some neighborhood of z_0. A function $g(s)$ is analytic in S if it is analytic at each point in S.*

Definition 2.1.10 *$L_2(j\mathbf{R})$ is a Hilbert space of matrix-valued functions on $j\mathbf{R}$ and consists of all complex matrix-valued functions $G(j2\pi f)$ ($G(j2\pi f)$ is a matrix with a fixed size) such that*

$$\int_{-\infty}^{+\infty} \mathit{Trace}[G^*(j2\pi f)G(j2\pi f)]df < +\infty. \tag{2.1.5}$$

The inner product of $L_2(j\mathbf{R})$ is defined as

$$(G_1, G_2) \triangleq \int_{-\infty}^{+\infty} \mathit{Trace}[G_1^*(j2\pi f)G_2(j2\pi f)]df, \tag{2.1.6}$$

for $G_1, G_2 \in L_2(j\mathbf{R})$. The inner product induced norm is defined as

$$\|G\|_2 \triangleq \sqrt{(G,G)}. \tag{2.1.7}$$

Definition 2.1.11 *H^2 is a closed subspace of $L_2(j\mathbf{R})$ with matrix functions $G(s)$ ($s = \sigma + j2\pi f$ is a complex frequency) that are analytic in the open right half plane. The norm of the H^2 space is defined as*

$$\|G\|_2^2 \triangleq \sup_{\sigma>0} \int_{-\infty}^{+\infty} \mathit{Trace}[G^*(\sigma + j2\pi f)G(\sigma + j2\pi f)]df \tag{2.1.8}$$

$$= \int_{-\infty}^{+\infty} \mathit{Trace}[G^*(j2\pi f)G(j2\pi f)]df. \tag{2.1.9}$$

Remark 2.1.12 *If $\mathbf{x}$, as an input to a system $G(j2\pi f)$, is a white noise such that $S_{\mathbf{x}}(f) = 1$ for all frequencies, the output PSD is given by $S_{\mathbf{y}}(f) = S_{\mathbf{x}}(f)|G(j2\pi f)|^2 = |G(j2\pi f)|^2$. Therefore,*

$$\|\mathbf{y}\|_{\mathbf{P}} = \|G\|_2, \tag{2.1.10}$$

which follows from (2.1.4). Thus, the H^2 norm of a transfer function measures the root-mean-square response of its output when it is driven by a white noise excitation.

Definition 2.1.13 *$L_\infty(j\mathbf{R})$ is the Banach space of matrix-valued functions $G(j2\pi f)$ that are essentially bounded on $j\mathbf{R}$, with a norm*

$$\|G\|_\infty \triangleq ess \sup_{f\in\mathbf{R}} \bar{\sigma}[G(j2\pi f)], \tag{2.1.11}$$

where $\bar{\sigma}[G(j2\pi f)]$ is a maximal singular value of $G(j2\pi f)$.

Definition 2.1.14 *H^∞ is a closed subspace of $L_\infty(j\mathbf{R})$ with matrix functions that are analytic and bounded in the open right half plain. The H^∞ norm is defined as*

$$\|G\|_\infty \triangleq \sup_{\sigma>0} \bar{\sigma}[G(\sigma + j2\pi f)] = \sup_{f\in\mathbf{R}} \bar{\sigma}[G(j2\pi f)]. \tag{2.1.12}$$

In the rest of the thesis, the shorthand notation for a transfer function will be used, $G(f)$, $f \in \mathbf{R}$, instead of $G(j2\pi f)$.

Remark 2.1.15 *When a communication channel is a SISO channel, the definition reduces to*

$$\|G\|_\infty \triangleq \sup_{f\in\mathbf{R}} |G(f)|. \tag{2.1.13}$$

Hence, the H^∞ norm, in the SISO case, represents the peak of $|G(f)|$.

Definition 2.1.16 *The frequency response $G(f)$, which belongs to the H^∞ space is called a stable frequency response.*

Remark 2.1.17 *If $\mathbf{x}$, as an input to a system $G(f)$, has a PSD $S_\mathbf{x}(f)$, then the root-mean-square value of the output $\mathbf{y}$ satisfies*

$$\begin{aligned}
\|\mathbf{y}\|_\mathbf{P}^2 &= \int_{-\infty}^{+\infty} S_\mathbf{y}(f)df & (2.1.14)\\
&= \int_{-\infty}^{+\infty} |G(f)|^2 S_\mathbf{x}(f)df & (2.1.15)\\
&\leq \sup_f |G(f)|^2 \int_{-\infty}^{+\infty} S_\mathbf{x}(f)df & (2.1.16)\\
&= \|G\|_\infty^2 \|\mathbf{x}\|_\mathbf{P}^2, & (2.1.17)
\end{aligned}$$

implying the connection between root-mean-square and H^∞ norm,

$$\frac{\|G\mathbf{x}\|_\mathbf{P}}{\|\mathbf{x}\|_\mathbf{P}} \leq \|G\|_\infty. \tag{2.1.18}$$

Previous remark points out on an alternative definition of the H^∞ norm. The H^∞ norm is defined by

$$\sup_{\|\mathbf{x}\|_{\mathbf{P}} \neq 0} \frac{\|G\mathbf{x}\|_{\mathbf{P}}}{\|\mathbf{x}\|_{\mathbf{P}}} \triangleq \|G\|_\infty, \tag{2.1.19}$$

which explains why the H^∞ norm is called induced norm. Next, H^2 and H^∞ norms will be compared to show the advantage of the H^∞ norm, when the two norms are used in defining the performance of robust systems.

In the case of the H^2 norm, the norm-bound specification $\|G\|_2 \leq M$, is equivalent to $\|G\mathbf{x}\|_{\mathbf{P}} \leq M$ for $\mathbf{x}$ a white noise.

However, in the case of the H^∞ norm, from (2.1.18), it follows that the norm-bound specification $\|G\|_\infty \leq M$ is equivalent to $\|G\mathbf{x}\|_{\mathbf{P}} \leq M$, for all $\mathbf{x}$, $\|\mathbf{x}\|_{\mathbf{P}} \leq 1$.

Thus, if $\mathbf{x}$ represents noise or interference signal, and if $G(f)$ is a transfer function from the noise $\mathbf{x}$ to the decoding error $\mathbf{e}$, the specification of the performance in terms of the bounded $\|G\|_\infty$, will imply that the decoding error $\mathbf{e}$ is bounded for all noises $\mathbf{x}$, $\|\mathbf{x}\|_{\mathbf{P}} \leq 1$. In contrast, the boundedness of $\|G\|_2$ implies the boundedness of the decoding error $\mathbf{e}$ only for $\mathbf{x}$ a white noise.

2.2 Modeling of Uncertainty in H^∞ Space

This section provides two examples of uncertainty set descriptions in H^∞ space, additive and multiplicative. These two descriptions are general enough to illustrate the concept. The choice of the uncertainty description depends on the problem at hand. This type of uncertainty description can be used to model uncertainty in the channel frequency response and the PSD of the noise.

2.2.1 Additive Uncertainty

The additive uncertainty model of the frequency response $\tilde{H}(f) = H_{nom}(f) + \Delta_1(f)W_1(f)$ is the sum of two terms. The first is the so-called nominal frequency response $H_{nom}(f)$ that represents the known part of $\tilde{H}(f)$. The other term, $\Delta_1(f)W_1(f)$, represents the perturbation. The choice of the nominal frequency response $H_{nom}(f)$ is based on the previous experience or belief that one has regarding the physical laws that govern that channel behavior. The transfer functions $H_{nom}(f)$, $W_1(f)$, and $\Delta_1(f)$ belong to the H^∞ space. $W_1(f)$ is a known stable transfer function, and $\Delta_1(f)$ is a variable stable transfer function with $\|\Delta_1(f)\|_\infty \leq 1$. The frequency responses $\tilde{H}(f)$ and $H_{nom}(f)$ have the same number of unstable poles. Thus, no unstable pole of $H_{nom}(f)$ is canceled in forming $\tilde{H}(f)$. Since $\tilde{H}(f) = H_{nom}(f)+\Delta_1(f)W_1(f)$, it follows that $|\tilde{H}(f) - H_{nom}(f)| \leq |W_1(f)|$, i.e., the uncertainty set is the set of all $\tilde{H}(f)$ that belong to the ball centered at $H_{nom}(f)$ with radius determined by the magnitude of a fixed known frequency response $W_1(f)$. Thus, the size of uncertainty set depends on the frequency and it is determined by $|W_1(f)|$. The smaller the magnitude $|W_1(f)|$, the smaller the uncertainty set. The transfer function $\Delta_1(f)$ accounts for the phase uncertainty and behaves as a scaling factor on the magnitude of the perturbation because its magnitude is between 0 and 1. The block diagram of additive perturbation is shown in Fig. 2.2.1.

In the case of discrete-time MIMO systems, the additive uncertainty description is of the form $H(e^{j\theta}) = H_{nom}(e^{j\theta}) + W_1(e^{j\theta})\Delta(e^{j\theta})W_2(e^{j\theta})$, where $H_{nom}(e^{j\theta})$ is a nominal frequency response and $W_1(e^{j\theta})\Delta(e^{j\theta})W_2(e^{j\theta})$ is a perturbation. $W_1(e^{j\theta})$ and $W_2(e^{j\theta})$ are known stable transfer functions. $\Delta(e^{j\theta})$ is an unknown stable transfer function such that $\|\Delta\|_\infty \leq 1$. If $W_1(e^{j\theta}) = I$ and $W_2(e^{j\theta}) = w(e^{j\theta})I$, where $w(e^{j\theta})$ is a scalar fixed function, $H(e^{j\theta})$ describes the disk centered at $H_{nom}(e^{j\theta})$ with radius $w(e^{j\theta})$.

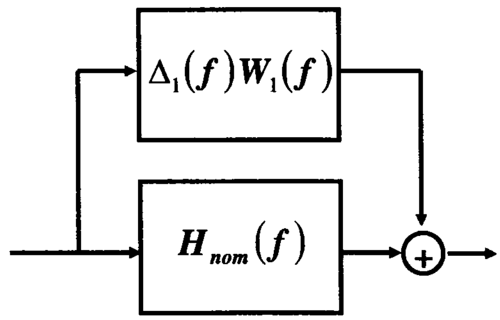

Figure 2.2.1: Additive uncertainty description

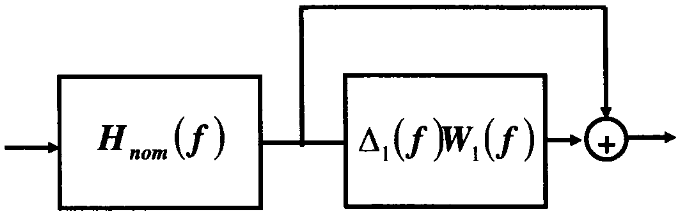

Figure 2.2.2: Multiplicative uncertainty description

2.2.2 Multiplicative Uncertainty

The multiplicative uncertainty model is described by $\tilde{H}(f) = H_{nom}(f)(1 + \Delta_1(f)W_1(f))$, where the frequency responses $\tilde{H}(f)$, $H_{nom}(f)$, $\Delta_1(f)$ and $W_1(f))$ satisfy the same conditions as in the additive uncertainty description. This model implies that $|\frac{\tilde{H}(f)}{H_{nom}(f)} - 1| \leq |W_1(f)|$. It means that the ratio $\frac{\tilde{H}(f)}{H_{nom}(f)}$ at each frequency belongs to the ball centered at 1, with radius $|W_1(f)|$. The size of uncertainty set is again determined by $|W_1(f)|$, while $\Delta_1(f)$ accounts for the phase uncertainty and behaves as a scaling factor on the magnitude of the perturbation $|W_1(f)|$. The block diagram of the multiplicative uncertainty description is depicted in Fig. 2.2.2. The corresponding MIMO model is given by $H(e^{j\theta}) = (I + W_1(e^{j\theta})\Delta(e^{j\theta})W_2(e^{j\theta}))H_{nom}(e^{j\theta})$.

2.3 Modeling of Uncertainty in L_1 Space

The uncertainty in the PSD of the noise is described by a subset of the L_1 space. For SISO channels, the description is given by $\{S_{\mathbf{n}}(f) : \int_{-\infty}^{+\infty} S_{\mathbf{n}}(f)df \leq P_{\mathbf{n}}\}$, while for MIMO channels the description is given by $\{W_{\mathbf{n}}(\theta) : \int_0^{2\pi} \text{Trace}(W_{\mathbf{n}}(\theta))d\theta \leq P_{\mathbf{n}}\}$. The difference between H^∞ and L_1 descriptions is the following. In the case of H^∞ uncertainty description, it is assumed certain a priori knowledge regarding a channel or a noise in the form of nominal frequency response $H_{nom}(f)$. In the case of L_1 models, this a priori knowledge does not exist. That is why the L_1 uncertainty description is a good model for the communication subject to jamming, where it is reasonable to assume that the communicator could have just a rough knowledge regarding the power constraint of the jammer.

Remark 2.3.1 *It may be seen that the use of white Gaussian noise has been ruled out by the definitions of previous sets that define all possible noise PSD's. Because one application of the result is jamming, the reason for doing this is to exclude the process that does not exist in reality. Further as discussed in [35] (Chapter 8, Section 3), the white noise assumption has certain deficiency in terms that obtained results (probability of decoding error) depend on what occurs at very large frequencies.*

2.4 Summary

The normed linear spaces, which are used to represent the signals, noises, and channels, are introduced. More specifically, the H^∞ and L_1 uncertainty descriptions are explained, which are employed to represent the uncertainties in the channel frequency response and the PSD of the noise.

Chapter 3

Robust Capacity of SISO Gaussian Channels

3.1 Introduction

In this chapter, the capacities of SISO Gaussian channels subject to the uncertainties defined in the frequency domain are derived. These channels belong to the class of so-called Gaussian compound channels. Although, in a realistic environment, one deals with communication networks, the results obtained for SISO communication channels are the first step in studying reliable communication over uncertain channels. The SISO results will provide insight into the MIMO channels considered in Chapter 4. Moreover, the channel coding theorem is derived, which guarantees the existence of a single code for corresponding compound channel that enables the reliable transmission, provided the code rate is less than the robust capacity. This is the reminiscent of the channel coding theorem derived in [63].

The chapter is organized as follows. In Section 3.2, the channel capacity formulas are presented when the channel and the noise uncertainties are described by subsets of H^{∞} space. In Section 3.3, the results obtained in Section 3.2 are applied for derivation of the capacities of ergodic and non-ergodic uncertain wireless fading channels. In Section 3.4, the capacity formula is derived when the channel frequency response uncertainty is described by a subset of H^{∞} space, while the noise uncertainty is described by a subset of L_1 space. In Sections 3.5 and 3.6, it is shown that the derived information capacities are equal to the operational capacities.

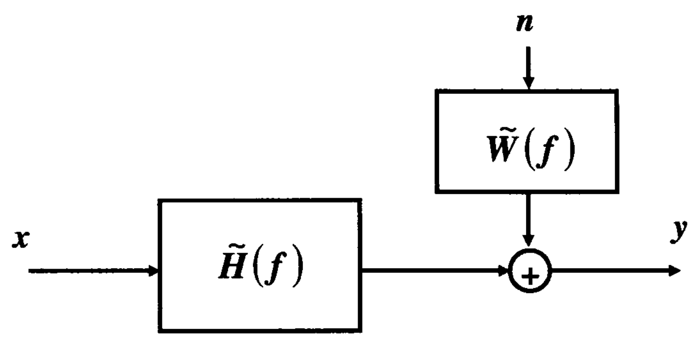

Figure 3.1.1: Uncertain Gaussian continuous-time channel

3.2 Robust Capacity with H^∞ Channel and/or Noise Uncertainty

The basic model of the compound Gaussian channel considered in this chapter is shown in Fig. 3.1.1, based on (1.4.22). The channel frequency response uncertainty and the uncertainty in the PSD of the noise are modeled using the mathematical tools described in Chapter 2. Three different problems are defined and then solved. The uncertainty sets are defined as follows

$$A_1 \triangleq \left\{ S_{\mathbf{x}}(f) : \int_{-\infty}^{+\infty} S_{\mathbf{x}}(f) df \leq P \right\}, \tag{3.2.1}$$

$$A_2 \triangleq \left\{ \tilde{H} \in H^\infty : \tilde{H} = H_{nom} + \Delta_1 W_1, H_{nom}, W_1, \Delta_1 \in H^\infty, \|\Delta_1\|_\infty \leq 1 \right\}, \tag{3.2.2}$$

$$A_3 \triangleq \left\{ \tilde{W} \in H^\infty : \tilde{W} = W_{nom} + \Delta_2 W_2, W_{nom}, W_2, \Delta_2 \in H^\infty, \|\Delta_2\|_\infty \leq 1 \right\}. \tag{3.2.3}$$

The precise definitions of the problems are given below.

First problem. (Channel unknown, noise known) Suppose the channel uncertainty is described by the set A_2, while the PSD of the noise, $S_{\mathbf{n}}(f)$, is known, and $\|\Delta_2\|_\infty = 0$. The channel capacity is defined by

$$C = \sup_{S_{\mathbf{x}} \in A_1} \inf_{\tilde{H} \in A_2} \frac{1}{2} \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|W_{nom}(f)|^2}\right) df. \tag{3.2.4}$$

Second problem. (Channel known, noise unknown) Suppose the noise uncertainty is described by the set A_3, while $S_{\mathbf{n}}(f)$ is known, and $\|\Delta_1\|_\infty = 0$. The channel capacity is defined by

$$C = \sup_{S_{\mathbf{x}} \in A_1} \inf_{\tilde{W} \in A_3} \frac{1}{2} \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)|H_{nom}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}\right) df. \tag{3.2.5}$$

Third problem. (Channel unknown, noise unknown) Suppose the noise uncertainty is described by the set A_3, the channel uncertainty by the set A_2, and $S_{\mathbf{n}}(f)$ is known. The channel capacity is defined by

$$C = \sup_{S_{\mathbf{x}} \in A_1} \inf_{\tilde{H} \in A_2} \inf_{\tilde{W} \in A_3} \frac{1}{2} \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}\right) df. \tag{3.2.6}$$

Clearly, the first and the second problem are special cases of the third problem. Therefore, only the the solution for the third problem will be given, while the solution of the other two will be stated as corollaries.

3.2.1 Robust Transmission in the Presence of Channel and Noise Uncertainty

In the next theorem, the solution of the third problem is presented.

Theorem 3.2.1 *Suppose* $\frac{(|H_{nom}(f)|+|W_1(f)|)^2}{S_{\mathbf{n}}(|W_{nom}(f)|-|W_2(f)|)^2}$ *is bounded and integrable, and* $|W_{nom}(f)| \neq |W_2(f)|$, $\forall f \in (-\infty, +\infty)$. *Then the following hold*

1. *The robust capacity defined by (3.2.6) is given by*

$$C = \frac{1}{2} \int_{\mathcal{S}} \log\left(\frac{\nu^o(|H_{nom}(f)| - |W_1(f)|)^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}\right) df, \tag{3.2.7}$$

where ν^o *is a Lagrange multiplier found via*

$$\int_{\mathcal{S}} \left(\nu^o - \frac{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}{(|H_{nom}(f)| - |W_1(f)|)^2}\right) df = P, \tag{3.2.8}$$

in which the integrations are over the set

$$\mathcal{S} = \{f : \nu^o - \frac{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}{(|H_{nom}(f)| - |W_1(f)|)^2} > 0,\ \nu^o > 0\}. \tag{3.2.9}$$

2. *The infimum over the noise uncertainty in (3.2.6) is achieved at*

$$\Delta_2^o(f) = \exp[-j\arg(W_2(f)) + j\arg(W_{nom}(f))],\ \|\Delta_2^o\|_\infty = 1, \tag{3.2.10}$$

and the resulting mutual information rate after the minimization is given by

$$\inf \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|W_{nom}(f) + \Delta_2(f)W_2(f)|^2}\right) df \tag{3.2.11}$$

$$= \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}\right) df, \tag{3.2.12}$$

where the infimum is over $\|\Delta_2(f)\|_\infty \leq 1$.

3. *The infimum over the channel uncertainty in (3.2.6) is achieved at*

$$\Delta_1^o(f) = \exp[-j\arg(W_1(f)) + j\arg(H_{nom}(f)) + j\pi],\ \|\Delta_1^o\|_\infty = 1, \tag{3.2.13}$$

and the resulting mutual information rate after minimization is given by

$$\inf \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)|H_{nom}(f) + \Delta_1(f)W_1(f)|^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}\right) df \tag{3.2.14}$$

$$= \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)(|H_{nom}(f)| - |W_1(f)|)^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}\right) df, \tag{3.2.15}$$

where the infimum is over $\|\Delta_1(f)\|_\infty \leq 1$.

4. *The supremum over* A_1 *yields the water-filling equation*

$$S_{\mathbf{x}}^o(f) + \frac{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}{(|H_{nom}(f)| - |W_1(f)|)^2} = \nu^o. \tag{3.2.16}$$

Proof. The proof is given in Appendix A.

Remark 3.2.2 *From the definition of the set* $\mathcal{S}$, *it follows that the integration in (3.2.7) and (3.2.8) is only over the frequencies for which* $|H_{nom}(f)| > |W_1(f)|$.

From the above theorem, it follows that the capacity C depends on the two fixed and known transfer functions $W_1(f)$ and $W_2(f)$, which determine the size of the channel uncertainty set and the size of the noise uncertainty set, respectively. Therefore, for smaller channel and noise measurement uncertainties, i.e., for smaller $|W_1(f)|$ and $|W_2(f)|$, the robust capacity is larger. Moreover, $W_1(f)$ and $W_2(f)$ affect the water-filling formula (3.2.16) that defines the optimal transmitted power in the presence of uncertainty. Finally, (3.2.7) shows that the robust capacity is equal to the capacity of the worst case channel, $|H_{nom}(f)| - |W_1(f)|$, and the worst case noise, $|W_{nom}(f)| + |W_2(f)|$. This, is a consequence of the norm employed to describe the uncertainty, which measures only magnitude and the fact that the mutual information is given in terms of the magnitude of the channel and the PSD of the noise. Clearly, if one employs a different norm to describe the uncertainty, then the resulting formula will be different. Nevertheless, it appears from the obtained results, (3.2.7), (3.2.8), that the H^∞ norm is a natural norm to define the uncertainty in the frequency response.

Corollary 3.2.3 *The solution of the first problem is obtained by setting $|W_2(f)|$ to zero for all $f \in (-\infty, +\infty)$ in (3.2.7) and (3.2.8), which implies that the PSD of the noise is perfectly known.*

Corollary 3.2.4 *The solution of the second problem is obtained by setting $|W_1(f)|$ to zero for all $f \in (-\infty, +\infty)$ in (3.2.7) and (3.2.8), which implies that the channel frequency response is perfectly known.*

Corollary 3.2.5 *The solution of the classical capacity formula is obtained by setting $|W_1(f)|$ and $|W_2(f)|$ to zero for all $f \in (-\infty, +\infty)$ in (3.2.7) and (3.2.8) [35].*

Remark 3.2.6 *The channel capacity of an uncertain discrete-time channel is obtained from Theorem 3.2.1 by formally removing the integrals in (3.2.7) and (3.2.8)*

$$C = \frac{1}{2}\log\left(\frac{\nu^o(|H_{nom}| - |W_1|)^2}{S_{\mathbf{n}}(|W_{nom}| + |W_2|)^2}\right), \tag{3.2.17}$$

$$\nu^o - \frac{S_{\mathbf{n}}(|W_{nom}| + |W_2|)^2}{(|H_{nom}| - |W_1|)^2} = P. \tag{3.2.18}$$

Substituting ν^o into C gives

$$C = \frac{1}{2}\log\left(1 + \frac{P(|H_{nom}| - |W_1|)^2}{S_{\mathbf{n}}(|W_{nom}| + |W_2|)^2}\right). \tag{3.2.19}$$

It should be noted that in (3.2.19), the channel capacity is measured in nats/channel use, while in (3.2.7), the unit is nats/s. Also, for the discrete-time case, $S_{\mathbf{n}}$ represents the noise variance, $S_{\mathbf{x}}$ represents the transmitted signal variance such that $Var(\mathbf{x}) = S_{\mathbf{x}} \leq P$. $|H_{nom}|$ is a constant representing the nominal channel (the proof of this claim is found in Appendix A).

By using a similar approach, one can derive the channel capacity formulas when other H^∞ uncertainty descriptions are employed. For instance, for the multiplicative uncertainty description

$$A_2' \triangleq \left\{\tilde{H} \in H^\infty : \tilde{H} = H_{nom}(1 + \Delta W_1), H_{nom} \in H^\infty, W_1 \in H^\infty, \Delta \in H^\infty, \|\Delta\|_\infty \leq 1\right\} \tag{3.2.20}$$

the capacity is found by replacing $W_1(f)$ with $H_{nom}(f)W_1(f)$ in (3.2.7) and (3.2.8). The optimal $\Delta(f)$ is given by $\Delta^o(f) = \exp[-j\arg(W_1(f)) + j\pi]$, $\|\Delta^o\|_\infty = 1$. In the special case, when $|W_2(f)| = 0$ for all $f \in (-\infty, +\infty)$, the robust capacity formula is given by

$$C = \frac{1}{2}\int_{\mathcal{S}'} \log\left(\frac{\nu^o[|H_{nom}(f)|(1-|W_1(f)|)]^2}{S_{\mathbf{n}}(f)|W_{nom}(f)|^2}\right)df, \tag{3.2.21}$$

where ν^o is a Lagrange multiplier found via

$$\int_{\mathcal{S}'}\left(\nu^o - \frac{S_{\mathbf{n}}(f)|W_{nom}(f)|^2}{[|H_{nom}(f)|(1-|W_1(f)|)]^2}\right)df = P, \tag{3.2.22}$$

in which the integrations are over the set

$$\mathcal{S}' = \{f : \nu^o - \frac{S_{\mathbf{n}}(f)|W_{nom}(f)|^2}{[|H_{nom}(f)|(1-|W_1(f)|)]^2} > 0,\ \nu^o > 0\}. \tag{3.2.23}$$

3.2.2 Uncertainty and Optimal Bandwidth

Theorem 3.2.1 implies that the the optimal bandwidth depends on the frequencies over which $\nu^o - \frac{S_{\mathbf{n}}(f)(|W_{nom}(f)|+|W_2(f)|)^2}{(|H_{nom}(f)|-|W_1(f)|)^2}$ is positive. The optimal PSD, $S_{\mathbf{x}}^o(f)$, is found by pouring power, constrained by P, into the well $\frac{S_{\mathbf{n}}(f)(|W_{nom}(f)|+|W_2(f)|)^2}{(|H_{nom}(f)|-|W_1(f)|)^2}$, till the level ν^o.

When the channel and noise are uncertain, the shape of the well depends on the sizes of the uncertainty sets, $|W_1(f)|$ and $|W_2(f)|$, for each frequency f. An interesting point here is that for physical systems more uncertainty is found at high frequencies rather than at low frequencies. Thus, $|W_1(f)|$ and $|W_2(f)|$ should be assigned larger values at high frequencies.

This aspect is explained by considering a wireless fading channel. The low-pass representation of a time-varying impulse response of a wireless fading channel is given by [60]

$$c(\tau, t) = \sum_n \alpha_n(t)\exp(-j2\pi f_c\tau_n(t))\delta(\tau - \tau_n(t)), \tag{3.2.24}$$

where $\alpha_n(t)$ and $\tau_n(t)$ are attenuations and delays, respectively, and f_c is a carrier frequency. The phase of the signal $\theta_n(t)$ is determined by $2\pi f_c\tau_n(t)$. This means that the change in $\tau_n(t)$ by $1/f_c$ will result in the change of $\theta_n(t)$ by 2π. Therefore, for large f_c, $1/f_c$ is a small value such that a small motion in the transmission medium can cause a change in $\theta_n(t)$ by 2π.

Hence, as frequency increases the distance between a perturbed well $\frac{S_{\mathbf{n}}(f)(|W_{nom}(f)|+|W_2(f)|)^2}{(|H_{nom}(f)|-|W_1(f)|)^2}$ and the nominal $\frac{S_{\mathbf{n}}(f)|W_{nom}(f)|^2}{|H_{nom}(f)|^2}$ will increase. Consequently, the perturbed well is narrower than the nominal one. From here, we may deduce that the optimal bandwidth for an uncertain channel will be smaller than the optimal bandwidth for a channel which is completely known. One has to have in mind that the two wells have to be filled by the same power P.

3.2.3 Examples

Next, four examples are presented to illustrate the effect of the uncertainty on the robust capacity formula and the optimal transmitted PSD.

A. Uncertain Channel and White Gaussian Noise

Suppose the nominal system is given by a first order low pass filter

$$H_{nom}(f) = \frac{\alpha\beta}{j2\pi f + \beta} \tag{3.2.25}$$

where $\alpha,\ \beta > 0$, and hence, for $f = 0$, $H(0) = \alpha$. The uncertain channel is then represented by the following multiplicative model

$$\tilde{H}(f) = \frac{\alpha_p(f)}{j2\pi f + \beta}, \tag{3.2.26}$$

where $\alpha_p(f) = \alpha(1 + \Delta_1(f)\delta)$, and $|\Delta_1(f)| < 1$, $\forall f \in (-\infty, +\infty)$, $0 \leq \delta < 1$. The transfer function $W_1(f)$ that determines the size of uncertainty set is obtained from

$$\left|\frac{\tilde{H}(f)}{H_{nom}(f)} - 1\right| = |\Delta_1(f)W_1(f)| \leq |\delta| = |W_1(f)| < 1. \tag{3.2.27}$$

If $\delta = 0$, the channel is perfectly known, and, hence, there is no uncertainty. The larger δ, the larger the uncertainty set. It should be noted that $\frac{[|H_{nom}(f)|(1+|W_1(f)|)]^2}{N}$ is an integrable function of frequency f, which is a necessary condition for the existence of the robust capacity formula. N represents the PSD of the noise, which is equal to $N_0/2$. The uncertainty set is depicted by the Nyquist plot in Fig. 3.2.2. The Nyquist plot of the nominal frequency response corresponds to the solid circle, while the Nyquist plot of the true frequency response corresponds to a Nyquist plot that falls between the dashed circles. Hence, one candidate for the true channel is depicted by the "irregular" Nyquist plot, shown in Fig. 3.2.2.

Using Theorem 3.2.1, the formula for the robust capacity is computed by substituting $|H_{nom}(f)|$ and $|W_1(f)|$ into (3.2.21) and (3.2.22). The constant ν^o is computed by solving the integral constraint equation (3.2.22). The value of the constant ν^o is substituted in (3.2.21) to give the robust capacity C. The formulae for the robust capacity accompanied with the optimal PSD is given parametrically by

$$C = \frac{1}{\pi}\left\{\sqrt{\frac{\nu^o}{c} - \beta^2} - \beta\tan^{-1}\frac{\sqrt{\frac{\nu^o}{c} - \beta^2}}{\beta}\right\} \tag{3.2.28}$$

$$S_{\mathbf{x}}^o(f) = \begin{cases} \nu^o - c((2\pi f)^2 + \beta^2), & |f| \leq \frac{1}{2\pi}\sqrt{\frac{\nu^o}{c} - \beta^2} \\ 0, & \text{otherwise} \end{cases} \tag{3.2.29}$$

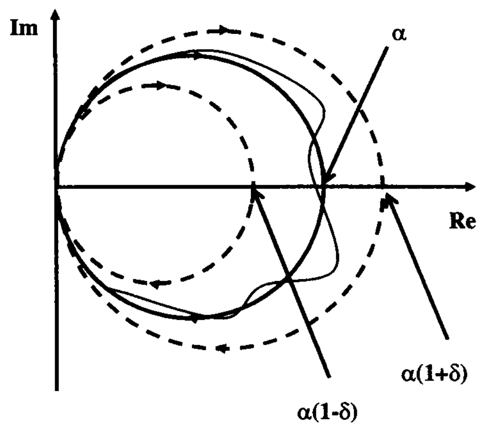

Figure 3.2.2: Nyquist plot of the first order system uncertainty set

where ν^o is the solution of

$$\{\nu^o\}^3 - 3c\beta^2\{\nu^o\}^2 + 3c^2\beta^4\nu^o - c^3\beta^6 - \frac{9}{4}c(\pi P)^2 = 0, \ c = \frac{N_0}{2\alpha^2\beta^2(1-\delta)^2}. \quad (3.2.30)$$

Here, (3.2.29) implies that the transmitted signal will be different than zero only if $\nu^o \geq c((2\pi f)^2 + \beta^2)$. Since $(2\pi f)^2 + \beta^2$ represents a parabola, then based on (3.2.29), there should be two intersections of the constant ν^o and $c((2\pi f)^2 + \beta^2)$, for a fixed power constraint P. The integral power constraint (3.2.22) gives (3.2.30), which has one real solution and two complex conjugated solutions; the real solution of (3.2.30) is unique and equal to the optimal ν^o. Hence, the points of intersections of ν^o and the parabola $c((2\pi f)^2 + \beta^2)$ determines the true bandwidth of the optimal transmitted signal $BW = \frac{2}{2\pi}\sqrt{\frac{\nu^o}{c} - \beta^2}$. This observation will be verified shortly via the closed form solution of the cubic equation (3.2.30). The real solution of (3.2.30) has the form $\nu^o = c\beta^2 + \left(\frac{9}{4}c(\pi P)^2\right)^{1/3}$. The closed form solution for the capacity is then

$$C = \frac{1}{\pi}\left\{\left(\frac{9}{4}\right)^{1/6}\left(\frac{\pi P}{c}\right)^{1/3} - \beta\tan^{-1}\frac{\left(\frac{9}{4}\right)^{1/6}\left(\frac{\pi P}{c}\right)^{1/3}}{\beta}\right\} \quad (3.2.31)$$

$$S_{\mathrm{x}}^o(f) = \begin{cases} \left(\frac{9}{4}c(\pi P)^2\right)^{1/3} - c(2\pi f)^2, & |f| \leq \frac{1}{2\pi}\left(\frac{9}{4}\right)^{1/6}\left(\frac{\pi P}{c}\right)^{1/3} \\ 0, & \text{otherwise} \end{cases} \quad (3.2.32)$$

The optimal bandwidth of the transmission is given by $BW = \frac{1}{\pi}\left(\frac{9}{4}\right)^{1/6}\left(\frac{\pi P}{c}\right)^{1/3}$. Because BW and $S_{\mathrm{x}}^o(f)$ depend on the constant c, they also depend on the parameter δ that defines the uncertainty. Thus, the larger the channel uncertainty δ, the larger the constant c, the smaller the bandwidth BW. This quantifies the importance of the presented approach, because it shows how the uncertainty affects the optimal transmitted bandwidth as well.

The previous calculation is employed to construct Fig. 3.2.3, 3.2.4, 3.2.5. Fig. 3.2.3 shows how the robust capacity decreases as a function of the channel uncertainty δ, for different values of the channel parameter α (which determines the channel gain), for a fixed cut-off frequency $\beta = 2\pi 10^4$ rad/s, $P = 0.01$ W, $N_0 = 10^{-7}$ W/Hz. As expected, the robust capacity decreases as the uncertainty increases. Moreover, it is noted that the difference between the capacities for different values of α are larger for smaller values of δ, and as δ increases, the value of the robust capacity tends to zero. Also, the larger α, the larger the slope of the capacity curve.

In Fig. 3.2.4, the robust capacity is shown versus the cut-off frequency β, for different values of the parameter δ. The capacity increases as the function of the cut-off frequency β. It can be noted that the difference between the channel capacities for different values of δ changes slowly with β. The parameters are chosen as follows: $P = 0.01$ W, $N_0 = 10^{-7}$ W/Hz, $\alpha = 1$.

Fig. 3.2.5 shows the optimal transmitter's PSD for different values of the parameter δ, which determines the size of the uncertainty set. It is shown how the optimal bandwidth shrinks as the uncertainty δ increases. The other parameters are chosen as follows: $P = 0.01$ W, $N_0 = 10^{-7}$ W/Hz, $\alpha = 1$, $\beta = 2\pi 10^4$ rad/s.

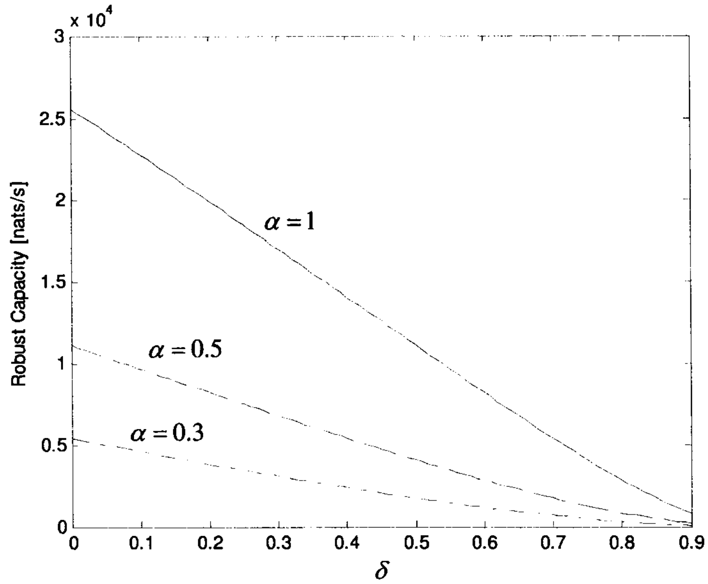

Figure 3.2.3: Robust capacity versus the size of channel uncertainty set δ

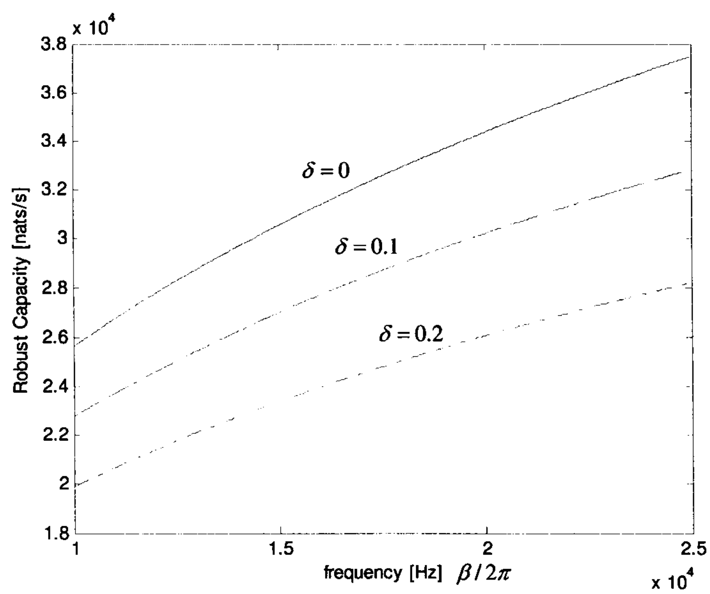

Figure 3.2.4: Robust capacity versus channel bandwidth

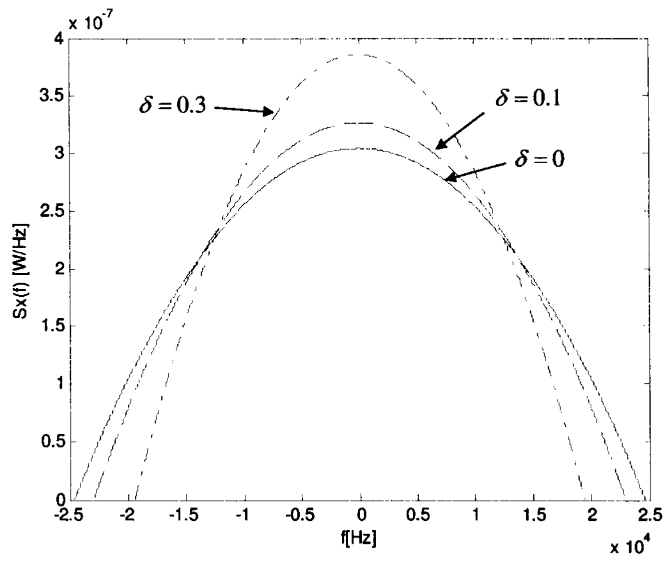

Figure 3.2.5: Optimal transmitter's PSD parameterized by the size of uncertainty set δ

B. Uncertain Channel and Colored Gaussian Noise

To illustrate the effect of the channel uncertainty on the capacity in the presence of a colored noise, we consider the following example. The channel is modeled by a second order transfer function $\tilde{H}(s) = \frac{\omega_n^2}{s^2+2\xi\omega_n s+\omega_n^2}$, $s = j\omega = j2\pi f$. It is assumed that the damping ratio ξ, while unknown, belongs to a certain interval, $0 < \xi_{low} \leq \xi \leq \xi_{up} < 1$. This set is approximated by using the following procedure. We choose the natural frequency to be $\omega_n = 2\ \pi\ 10^4$ rad/s, nominal damping ratio $\xi_{nom} = 0.3$, and $0.2 \leq \xi \leq 0.5$ (see Fig. 3.2.6). Further, the size of the uncertainty set is defined by $|W_1| = |H_{nom}| - |H_{low}|$, where $H_{low}(s) = \frac{\omega_n^2}{s^2+2\xi_{up}\omega_n s+\omega_n^2}$, $H_{nom}(s) = \frac{\omega_n^2}{s^2+2\xi_{nom}\omega_n s+\omega_n^2}$. The values of $\xi_{low} = 0.2$ and $\xi_{up} = 0.5$ are deliberately chosen such that $|H_{ub}| = |H_{nom}| + |W_1|$ is a good approximation of $H_{up}(s) = \frac{\omega_n^2}{s^2+2\xi_{low}\omega_n s+\omega_n^2}$. Thus, the frequency response uncertainty set is roughly described by $|H_{nom}| \pm |W_1|$. However, $|W_1| = |H_{nom}| - |H_{low}|$ implies $|H_{low}| = |H_{nom}| - |W_1|$. This means that the robust capacity is determined by the transfer function H_{low}. The uncertainty set is identified by the range of the damping ratio, $\Delta\xi = \xi_{up} - \xi_{low}$, $\Delta\xi = 0.30$ for this particular case. The value $\Delta\xi = 0$ corresponds to the nominal channel model. The PSD of the noise is given by the first order transfer function $S_n(f)|W_{nom}(f)|^2 = |\frac{\alpha}{s+\beta}|^2$, $s = j\omega = j2\pi f$, where $\alpha = 1$, $\beta = 5\ 10^4$ rad/s. The power of the transmitted signal is limited to P =0.01 W. Fig. 3.2.7 depicts the robust capacity for different sizes of the channel frequency response uncertainty sets. It can be seen that the robust capacity decreases as the channel uncertainty increases. Fig. 3.2.8 shows the effect of the channel uncertainty on the optimal PSD of the transmitter. The change of the optimal bandwidth is negligible, which is expected because the uncertainty in the damping ratio does not affect the channel bandwidth. Thus, the channel uncertainty may affect the robust capacity, but this does not contribute to a significant change in the optimal PSD of the transmitted signal.

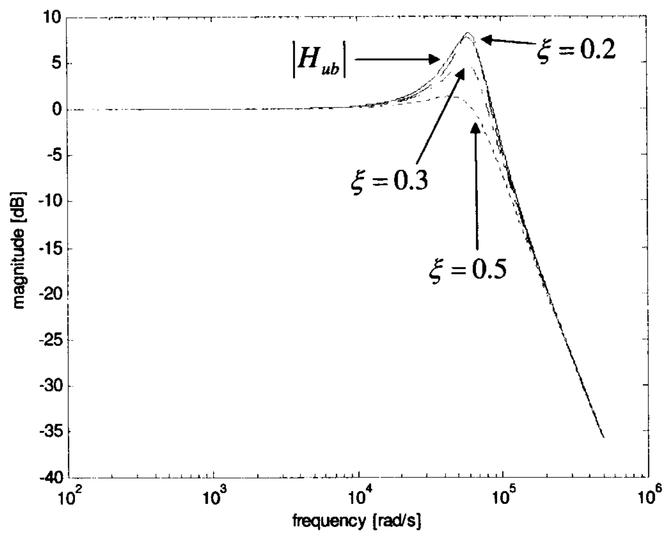

Figure 3.2.6: Approximation of channel uncertainty set

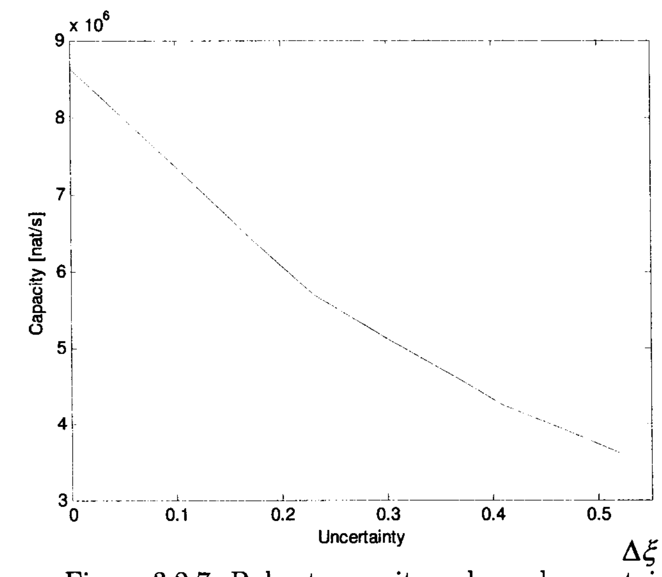

Figure 3.2.7: Robust capacity - channel uncertainty

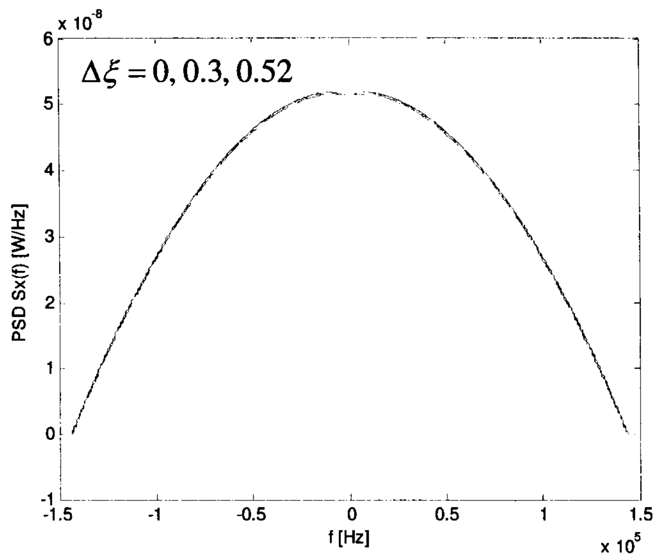

Figure 3.2.8: PSD - channel uncertainty

C. Channel Known, Noise Unknown

It is assumed that the channel frequency response is completely known, $|W_1| = 0$, and $S_{\mathbf{n}} = 1$ W/Hz. The noise uncertainty set is defined by $\tilde{W}(f) = \frac{\xi_p(f)}{j2\pi f/\beta+1}$, where $\xi_p(f) = \xi + \Delta_2(f)\delta\xi$, $\xi = \alpha/\beta$, $0 \leq \delta < 1$, and $W_{nom}(f) = \frac{\xi}{j2\pi f/\beta+1}$. Thus, $|\frac{\tilde{W}(f)}{W_{nom}(f)} - 1| = |\Delta_2(f)\delta| \leq |\delta| = |W_2(f)|$, and the uncertainty set is described by a ball in the frequency domain, centered at $|W_{nom}(f)|$ with radius $W_2(f)$. The radius, i.e., the size of uncertainty set is determined by the parameter δ. The channel is modeled by a second order transfer function $\tilde{H}(s) = \frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$, $s = j\omega = j2\pi f$. The parameters are chosen as follows: $\alpha = 1$, $\beta = 50000$ rad/s, $\zeta = 0.3$, and $P = 0.01W$. Fig. 3.2.9 shows that the robust capacity decreases as the size of the uncertainty set determined by δ increases, while the slope is larger for small uncertainty. Fig. 3.2.10 shows the optimal PSD of the transmitted signal $S_{\mathbf{x}}^o(f)$ for different values of δ and constant P. It appears that the transmitter tries to combat the uncertainty by reducing its bandwidth, and by regrouping the power towards lower frequencies.

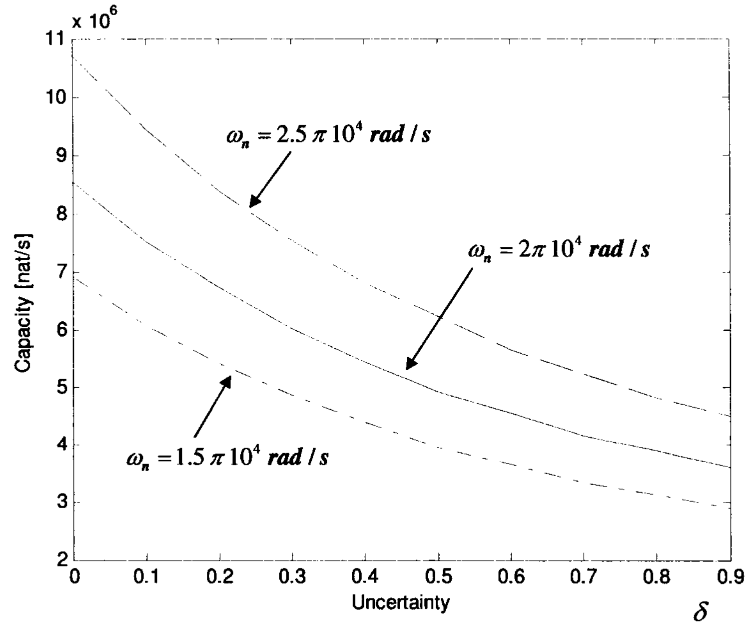

Figure 3.2.9: Robust capacity - noise uncertainty

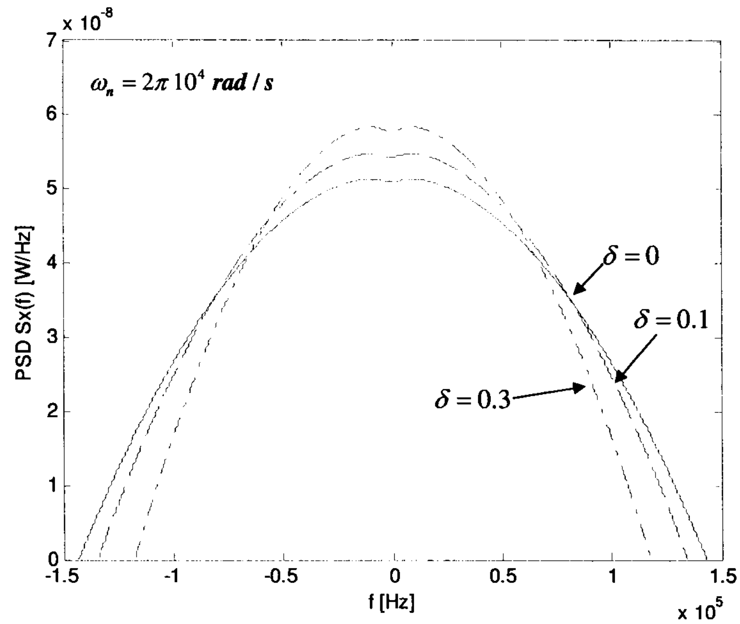

Figure 3.2.10: Optimal PSD - noise uncertainty

D. Channel Unknown, Noise Unknown

To illustrate the effect of the channel and the noise uncertainty on the capacity, we consider the following example. The channel is modeled by a second order transfer function $\tilde{H}(s) = \frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$, $s = j\omega = j2\pi f$, where it is assumed that the damping ratio ζ although unknown, takes a value from the certain interval, $0 < \zeta_{low} \leq \zeta \leq \zeta_{up} < 1$. This set is approximated by using the procedure described in Example B of this section. We choose the natural frequency to be $\omega_n = 2\ \pi\ 10^4$ rad/s and a nominal damping ratio $\zeta_{nom} = 0.3$. The uncertainty sets are identified by the range of the damping ratios $\Delta\zeta = \zeta_{up} - \zeta_{low}$, $\Delta\zeta = 0$, $\Delta\zeta = 0.15$, and $\Delta\zeta = 0.30$, while $\Delta\zeta = 0$ corresponds to the nominal channel model. The noise uncertainty set is the one considered in Example C. The power is constrained to P =0.01 W. Fig. 3.2.11 depicts the effect of the noise uncertainty for different sizes of the channel frequency response uncertainty sets. Similarly to the previous example, the channel uncertainty tends to affect the capacity more for lower values of uncertainty. For instance, the distance between the curves $\Delta\zeta = 0$ and $\Delta\zeta = 0.15$ is larger than the distance between the curves $\Delta\zeta = 0.15$ and $\Delta\zeta = 0.30$.

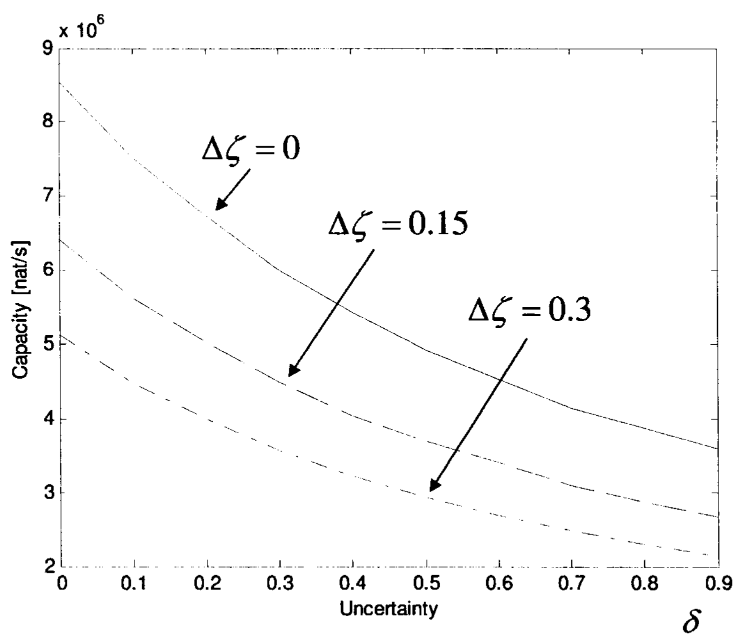

Figure 3.2.11: Robust capacity - channel-noise uncertainty

3.3 The Capacity of Uncertain Wireless Fading Channels

In this section, the results from the previous sections are applied to compute the capacity of uncertain wireless channels.

Consider a received signal, consisting of a two-ray model given by

$$y(t) = z_1(t)x(t-\tau_1) + z_2(t)x(t-\tau_2) + n(t). \tag{3.3.33}$$

Here, $\mathbf{z_1} \triangleq \{z_1(t) : -\infty < t < +\infty\}$ and $\mathbf{z_2} \triangleq \{z_2(t) : -\infty < t < +\infty\}$ are the complex amplitudes of the two random scattered components, which are realizations of independent zero-mean complex, stationary Gaussian processes having independent real and imaginary parts. The constants τ_1 and τ_2 are the nominal delays which are assumed to be known. The signal $\mathbf{n} \triangleq \{n(t) : -\infty < t < +\infty\}$ is a complex white Gaussian noise with two-sided PSD, $S_{\mathbf{n}}(f) = 2N_0$. It is assumed that the channel is band-limited to bandwidth W, and that the transmitter has finite power, i.e., $S_{\mathbf{x}}(f) \in B_1 \triangleq \{S_{\mathbf{x}}(f) : \int_{-W/2}^{+W/2} S_{\mathbf{x}}(f)df \leq 2P_{\mathbf{x}}\}$. Without loss of generality, time division multiple access (TDMA) scheme is assumed, where the message is conveyed in K time slots. Each time slot has length T, and each is transmitted in a different frame of length T_F. Further, it is assumed that $2WT >> 1$ and $T_c >> T$, where T_c is a coherence time [60]. It follows that the change of the scattered components $\mathbf{z_1}$ and $\mathbf{z_2}$ during one time slot of duration T is minor. Thus, within one time slot, $\mathbf{z_1}$ and $\mathbf{z_2}$ are slowly varying as functions of time, and, hence, they are represented by random variables. On the other hand, it is assumed that $\mathbf{z_1}$ and $\mathbf{z_2}$ may change considerably over the whole frame T_F.

Under the above assumptions, the mutual information rate conditioned on the specific realization of the channel, $\mathbf{z_1}$, $\mathbf{z_2}$, τ_1, and τ_2, is a random variable, given by [57]

$$I(\mathbf{x};\mathbf{y}|\mathbf{z_1} = z_1, \mathbf{z_2} = z_2, \tau_1, \tau_2) = \frac{1}{2}\int_{-W/2}^{W/2} \log\Big(1 + \frac{S_{\mathbf{x}}(f)|H(f)|^2}{2N_0}\Big)df, \tag{3.3.34}$$

where

$$H(f) = z_1 e^{j2\pi f\tau_1} + z_2 e^{j2\pi f\tau_2}. \tag{3.3.35}$$

Here, the sample paths of the random scattered components $\mathbf{z_1}$ and $\mathbf{z_2}$ are approximated by complex valued random variables z_1 and z_2, respectively, as explained in [57]. $H(f)$ is the channel frequency response in one particular time slot out of K, with the spatial frequency parameter f. In [57], it was shown that $\beta = |H(f)|^2$ is exponentially distributed random variable, independent on spatial frequency f.

Since (3.3.34) is of the form considered in the previous section, we can further proceed by identifying the nominal channel and the perturbed using an H^∞ uncertainty model. The uncertainty regarding the channel parameters z_1, z_2, τ_1, and τ_2, is described through the multiplicative uncertainty description, $B_2 \triangleq \{\tilde{H}(f) \in H^\infty : \tilde{H}(f) = H_{nom}(f)(1 + \Delta(f)W_1(f)), H_{nom} \in H^\infty, W_1 \in H^\infty, \Delta \in H^\infty, \|\Delta\|_\infty \leq 1\}$, where $H_{nom}(f) = H(f)$. The multiplicative model is chosen because it is convenient to model the uncertainty in time delays. Suppose that a more realistic model is given by

$$\begin{aligned} \tilde{H}(f) &= (1+\Delta\alpha)e^{-j2\pi f(1+\Delta\tau)}(z_1 e^{j2\pi f\tau_1} + z_2 e^{j2\pi f\tau_2}) && (3.3.36)\\ &= (1+\Delta\alpha)e^{-j2\pi f(1+\Delta\tau)}H_{nom}(f), && (3.3.37) \end{aligned}$$

where the uncertainties in time delays and attenuations are described through the parameters $\Delta\alpha$, $\Delta\alpha_{low} \leq \Delta\alpha \leq \Delta\alpha_{up}$, and $\Delta\tau$, $\Delta\tau_{low} \leq \Delta\tau \leq \Delta\tau_{up}$. Then, the multiplicative description of the channel frequency response is described in terms of $W_1(f)$, from the following condition

$$\Big|\frac{\tilde{H}(f)}{H_{nom}(f)} - 1\Big| = |(1+\Delta\alpha)e^{-j2\pi f(1+\Delta\tau)} - 1| \leq |W_1(f)|, \ \forall\ \Delta\tau,\ \Delta\alpha,\ f. \qquad (3.3.38)$$

This condition is implied by the definition of the multiplicative uncertainty representation. Next, the following two cases are investigated: 1) When the CSI is present at the receiver, 2) When the CSI is present at both, the transmitter and receiver.

3.3.1 Capacity of Wireless Channels with Imperfect Knowledge of CSI at the Receiver

The channel capacity of the wireless channels, with no restriction on the decoding delay ($K \to \infty$), is known as ergodic capacity. Following [57], the ergodic capacity for the wireless channel described by the model (3.3.33), with the CSI at the receiver side, is given by

$$C_E = \sup_{S_{\mathbf{x}} \in B_1} \inf_{\tilde{H} \in B_2} E\, \frac{1}{2}\int_{-W/2}^{W/2} \log\Big(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{2N_0}\Big) df, \qquad (3.3.39)$$

where the expectation is taken over the random variable $\beta = |H_{nom}(f)|^2$. Since the CSI is not available to the transmitter, it follows that the optimal PSD of the transmitter $S^o_{\mathbf{x}}(f)$ should not depend on the nominal transfer function $H_{nom}(f)$. The following theorem gives the value of the channel capacity for low and high SNR.

Theorem 3.3.1 *The ergodic capacity of the wireless communication channel (3.3.33), subject to the channel frequency response uncertainty description defined by B_2, when the nominal CSI is available at the receiver, but not at the transmitter, is given as follows*

1. *Low SNR*

$$C_E = \frac{1}{2}\, k \int_{-W/2}^{W/2} \left[\frac{(1-|W_1(f)|)^2}{2N_0}\right]^2 df, \tag{3.3.40}$$

where

$$k = \frac{2P_{\mathbf{x}}}{\int_{-W/2}^{W/2} \frac{(1-|W_1(f)|)^2}{2N_0} df}. \tag{3.3.41}$$

The optimal PSD of the transmitter that maximizes the mutual information, is given by

$$S_{\mathbf{x}}^o(f) = k\, \frac{(1-|W_1(f)|)^2}{2N_0}. \tag{3.3.42}$$

2. *High SNR*

$$C_E = \frac{1}{2} \int_{\mathcal{S}} \left[\log\left(\frac{\nu^0(1-|W_1(f)|)^2}{2N_0}\right) - \mathcal{C}\right] df, \tag{3.3.43}$$

where

$$\int_{\mathcal{S}} S_{\mathbf{x}}^o(f) df = \int_{\mathcal{S}} \left(\nu^o - \frac{2N_0}{(1-|W_1(f)|)^2}\right) df = 2P_{\mathbf{x}}, \tag{3.3.44}$$

and $\mathcal{S} \triangleq \{f : \nu^o - \frac{2N_0}{(1-|W_1(f)|)^2} > 0\}$, and $\mathcal{C} \approx 0.577...$ is the Eüler constant.

Proof. The proof is given in Appendix B.

Remark 3.3.2 *$S_{\mathbf{x}}^o(f)$ given by (3.3.42), suggests that the optimal PSD, for low SNR, behaves like a matched filter which is matched to the "uncertain part" of the SNR, $\frac{(1-|W_1(f)|)^2}{2N_0}$. For high SNR, $S_{\mathbf{x}}^o(f)$ given by (3.3.44), water-fills the "uncertain part" of the SNR, $\frac{(1-|W_1(f)|)^2}{2N_0}$.*

Although in the previous derivations, it is assumed that the nominal frequency response $H_{nom}(f)$ is not available to the transmitter, it is assumed that the transmitter has the information regarding the size of the uncertainty set in the form of $|W_1(f)|$.

3.3.2 Capacity of Wireless Channels with Imperfect Knowledge of CSI at the Receiver and Transmitter

1. Ergodic capacity. When the CSI is present at both sides of the communication link, the ergodic capacity is given by [18]

$$C_E = E \sup_{S_{\mathbf{x}} \in B_1} \inf_{\tilde{H} \in B_2} \frac{1}{2} \int_{-W/2}^{W/2} \log\Big(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{2N_0}\Big) df. \tag{3.3.45}$$

In this case, the optimal PSD of the transmitter $S_{\mathbf{x}}^o(f)$ is the function of the realization of the nominal channel frequency response $\beta = |H_{nom}(f)|^2$.

Theorem 3.3.3 *The ergodic capacity of the wireless communication channel (3.3.33), subject to the channel frequency response uncertainty description defined by B_2, when the nominal CSI is available to the transmitter and receiver, is given by*

$$C_E = E_\beta \frac{1}{2} \int_S \log\left(\frac{\nu^0 \beta (1 - |W_1(f)|)^2}{2N_0}\right) df, \tag{3.3.46}$$

where

$$\int_S S_{\mathbf{x}}^o(f) df = \int_S \Big(\nu^o - \frac{2N_0}{\beta(1-|W_1(f)|)^2}\Big) df = 2P_{\mathbf{x}} \quad P. \ \ a.s., \tag{3.3.47}$$

where $\mathcal{S} \triangleq \{f : \nu^o - \frac{2N_0}{\beta(1-|W_1(f)|)^2}, \ P. \ a.s. > 0\}$.

Proof. The proof is given in Appendix B.

2. Capacity vs. outage. When there is a restriction on the decoding delay (i.e., K is finite), the Shannon capacity does not exists in the strict sense [57]. This means that there is a non-zero probability, which is independent of the code length, such that the instantaneous channel parameters could take values such that the coding rate, no matter how small, cannot be supported over the channel with an error probability which exponentially decreases with the codeword length. In this kind of applications, the outage probability is considered, which is defined by

$$\Pr\big(\sup_{S_{\mathbf{x}} \in B_1} \inf_{\tilde{H} \in B_2} I(\mathbf{x}; \mathbf{y} | \mathbf{z_1} = z_1, \mathbf{z_2} = z_2, \tau_1, \tau_2) \leq \epsilon\big). \tag{3.3.48}$$

This is the probability that the random variable $\sup_{S_{\mathbf{x}} \in B_1} \inf_{\tilde{H} \in B_2} I(\mathbf{x}; \mathbf{y} | \mathbf{z_1} = z_1, \mathbf{z_2} = z_2, \tau_1, \tau_2)$, which represents the instantaneous robust capacity, is smaller than ϵ. The expression for $\sup_{S_{\mathbf{x}} \in B_1} \inf_{\tilde{H} \in B_2} I(\mathbf{x}; \mathbf{y} | \mathbf{z_1} = z_1, \mathbf{z_2} = z_2, \tau_1, \tau_2)$ is given in Appendix B. It can be seen that the outage probability depends on the size of the uncertainty set $|W_1(f)|$. For the case of a completely known channel, $|W_1(f)| = 0$, and the above probability is evaluated by using the Lobachevsky's functions [57].

The main benefit of the approach, shown in this and the previous section, is that it enables the estimation of the cost of robustness with respect to uncertainties of each parameter separately (for instance, with respect to attenuations only) or with respect to combined uncertainties (of both, delays and attenuations). The cost of robustness, which is the loss of the capacity due to uncertainties, comes naturally from the capacity formulas since they depend directly on the size of the uncertainty sets. It should be noted that the important part of the capacity solution is the construction of the transfer function $W_1(f)$, which is not a trivial problem [30].

Ergodic capacity numerical example. This example illustrates the application of a multiplicative uncertainty description for the computation of the ergodic capacity subject to the uncertainty in the fading wireless channel parameters. It is assumed that the transmission bandwidth is $W = 30$ kHz, $P_x = 5$ mW, $N_0 = 5\ 10^{-8}$ W/Hz. Further, it is assumed that the uncertainty due to the attenuation estimation error is given by $0 \leq \Delta\alpha \leq 0.9$, and an uncertainty in the delay is given by $0 \leq \Delta\tau \leq 0.1$. Fig. 3.3.12 shows the decrease of the ergodic capacity versus the attenuation uncertainty, $\Delta\alpha$.

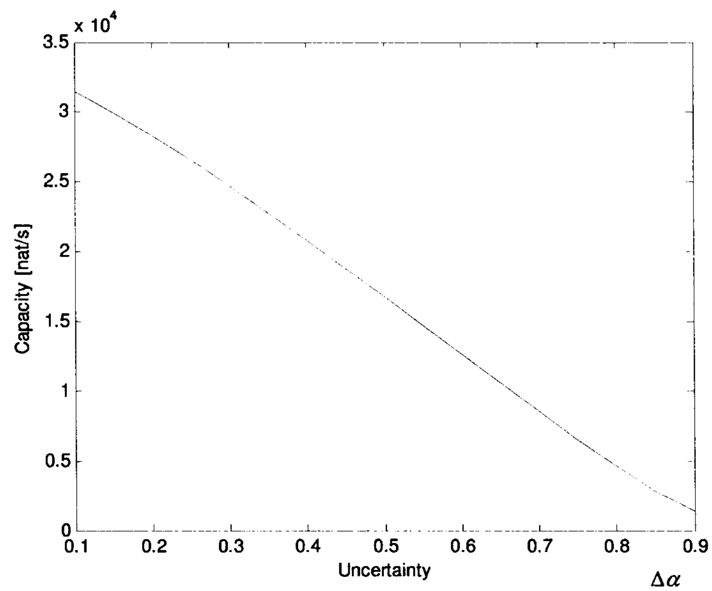

Figure 3.3.12: Ergodic capacity versus uncertainty in attenuations

3.4 Robust transmission in the Presence of Jamming

In this section, the capacity of continuous-time Gaussian channels is computed, when the noise uncertainty is described by a subset of L_1 function space. Specifically, it is assumed that the noise uncertainty is described by the set, $A_4 \triangleq \{S_{\mathbf{n}} : (f) \int_{-\infty}^{+\infty} S_{\mathbf{n}}(f)df \leq P_{\mathbf{n}}\}$, and, hence, the noise power is finite. The transmitter has some information about the jammer's bandwidth or the channel (such as the range of frequencies, which is modeled through the filter $\tilde{W}(f)$), but has no information about the PSD of the noise $S_{\mathbf{n}}(f)$, aside from knowing that it is finite in power as depicted by A_4. The size of the admissible set A_4 is determined by the constant $P_{\mathbf{n}}$. In addition, it is assumed that the channel is uncertain, and the channel uncertainty is described by the set A_2. The advantage of the approach taken here is that the optimal transmitter's and jammer's strategies are given in terms of optimal PSD's as opposed to [6], where the strategies are given in terms of the operator's eigenvalues, which could be more desirable from the engineering point of view.

Subject to the above uncertainty description, the channel capacity is defined by

$$C = \sup_{S_{\mathbf{x}} \in A_1} \inf_{S_{\mathbf{n}} \in A_4} \inf_{\tilde{H} \in A_2} \int_{-\infty}^{\infty} \log(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2})df. \tag{3.4.49}$$

If $W_1(f) = 0$, the channel capacity in the presence of solely noise uncertainty is obtained.

The solution of (3.4.49) is obtained by first applying the result of Corollary 3.2.3 to resolve the minimum with respect to the channel uncertainty. Then, the maximin optimization problem remains to be solved, where the minimum is with respect to the noise uncertainty, and the maximum is with respect to the transmitted signal PSD. In the sequel, without loss of generality, it is assumed that $W_1(f) = 0$ and $\tilde{W}(f) = 1$.

The channel capacity is formulated as a two player game with the mutual information rate as a pay-off function, where the transmitter, i.e., communicator chooses its strategy from the set A_1 to maximize the pay-off, and the jammer chooses its strategy from the set A_4 to minimize the pay-off. Thus, the strategies of the two players are the PSD's which belong to admissible sets A_1 and A_4.

However, before we proceed, it is important to distinguish, depending on the application, which player chooses its strategy first from the corresponding set. In order to that, we define the maximin and minimax problems, also known as the lower and upper values, respectively, as follows

$$\text{Lower Value: } C^- = \sup_{S_{\mathbf{x}} \in A_1} \inf_{S_{\mathbf{n}} \in A_4} J(S_{\mathbf{x}}, S_{\mathbf{n}}), \tag{3.4.50}$$

$$\text{Upper Value: } C^{+} = \inf_{S_{\mathbf{n}} \in A_4} \sup_{S_{\mathbf{x}} \in A_1} J(S_{\mathbf{x}}, S_{\mathbf{n}}). \tag{3.4.51}$$

The Lower Value corresponds to the case when the communicator announces his strategy $S_{\mathbf{x}}$ in advance, while the jammer announces his strategy $S_{\mathbf{n}} = S_{\mathbf{n}}(S_{\mathbf{x}})$ after knowing his opponent's choice. The Upper Value is defined in an analogous way. In general, $C^{-} \leq C^{+}$ (also known as weak max-min inequality [15]). If the inequality holds with equality then $C = C^{-} = C^{+}$, and C is called the value of the zero-sum game, provided there is a saddle point. If a saddle point exists then the supremum and infimum can be interchanged implying independence of communicator's and jammer's strategies. The saddle point property is defined as follows. Assume that $S_{\mathbf{x}}^{o}$ and $S_{\mathbf{n}}^{o}$ are the optimal strategies of the communicator and jammer, respectively. Then the saddle point property is described as

$$J(S_{\mathbf{x}}, S_{\mathbf{n}}^{o}) \leq J(S_{\mathbf{x}}^{o}, S_{\mathbf{n}}^{o}) \leq J(S_{\mathbf{x}}^{o}, S_{\mathbf{n}}), \tag{3.4.52}$$

where $S_{\mathbf{x}} \in A_1$ and $S_{\mathbf{n}} \in A_4$. Thus, a saddle value $J(S_{\mathbf{x}}^{o}, S_{\mathbf{n}}^{o})$ represents the optimal value of the cost function for both players.

From the definition of the problem, it is not clear whether a saddle point exists [56] (because this optimization problem is over infinite dimensional function spaces, and the sets A_1 and A_4 may not be compact). Hence, the Lower and the Upper values C^{-}, C^{+} will be explicitly computed by using convex optimization techniques and the Lagrange duality principle. It turns out that the optimal strategies of the communicator and the jammer are the same for maximin and minimax problems, implying the existence of the saddle point.

The Maximin Capacity. Here, we provide the solution of (3.4.50) (i.e., Lower value), C^{-}.

Theorem 3.4.1 *Consider the additive uncertainty description of $\tilde{H}(f)$, A_1, and noise description given by A_4. Suppose that $\frac{S_{\mathbf{x}}(f)(|H_{nom}(f)|+|W_1(f)|)^2}{S_{\mathbf{n}}(f)}$ is bounded, integrable. Then the following holds*

$$C^{-} = \frac{1}{2} \int_{-\infty}^{+\infty} \log\Big(1 + \frac{\lambda_1^o}{\lambda_2^o}(|H_{nom}(f)| - |W_1(f)|)^2\Big) df, \tag{3.4.53}$$

where

$$S_{\mathbf{n}}^{o}(f) = \frac{(|H_{nom}(f)| - |W_1(f)|)^2}{2(\lambda_1^o(|H_{nom}(f)| - |W_1(f)|)^2 + \lambda_2^o)}, \tag{3.4.54}$$

$$\begin{aligned} S_{\mathbf{x}}^{o}(f) &= \frac{\lambda_1^o(|H_{nom}(f)| - |W_1(f)|)^2}{2\lambda_2^o(\lambda_1^o(|H_{nom}(f)| - |W_1(f)|)^2 + \lambda_2^o)} \\ &= \frac{\lambda_1^o}{\lambda_2^o} S_{\mathbf{n}}^{o}(f), \end{aligned} \tag{3.4.55}$$

$$\int_{-\infty}^{+\infty} S_{\mathbf{x}}^{o}(f)df = P_{\mathbf{x}}, \quad \int_{-\infty}^{+\infty} S_{\mathbf{n}}^{o}(f)df = P_{\mathbf{n}}, \tag{3.4.56}$$

in which λ_1^o and λ_2^o are Lagrange multipliers of the two constraint sets A_4, A_1, respectively.

Proof. The proof is given in Appendix C.

Remark 3.4.2 *The statement of Theorem 3.4.1 can be easily extended to the case when there is a fixed white noise term $\mathbf{s} \triangleq \{s(t) : -\infty < t < +\infty\}$, in addition to $\mathbf{n}$. The PSD of overall noise $z(t) = n(t) + s(t)$ is given by $S_{\mathbf{z}}(f) = S_{\mathbf{n}}(f) + N_{\mathbf{s}}$, where $N_{\mathbf{s}}$ is a constant over all frequencies, and $S_{\mathbf{n}}(f) \in A_4$. Then, the optimal jammer's PSD is $S_{\mathbf{n}}^o(f) = S_{\mathbf{z}}(f) - N_{\mathbf{s}}$, where $S_{\mathbf{z}}(f)$ is given by (3.4.54).*

From (3.4.55) is deduced that the optimal PSD's of the transmitter and jammer are proportional. This is a conclusion that cannot be seen in the finite dimensional cases (for instance, see [41]) and infinite dimensional case [6]. It appears that both signals try to mimic each other and that the best possibility for both is to create a white noise like situation. This is concluded from the formula for the channel capacity (3.4.53), which is illustrated in Fig 3.4.13, $|R(f)| = |H_{nom}(f)| - |W_1(f)|$. Also, by manipulating (3.4.54) and (3.4.55), it is shown that the optimal PSD's satisfy so-called water-filling equation

$$S_{\mathbf{x}}^o(f) + \frac{S_{\mathbf{n}}^o(f)}{(|H_{nom}(f)| - |W_1(f)|)^2} = \frac{1}{2\lambda_2^o}, \tag{3.4.57}$$

where $S_{\mathbf{n}}^o(f)$ is given by (3.4.54). This means that both players try to put most of their power at frequencies where the opponent's power is small. This is consistent with the finite dimensional solution of vector channels, provided each frequency is viewed as a separate channel [41]. Finally, it is important to note that it is possible to distinguish and measure the impact of two types of uncertainties on the capacity, namely, the channel uncertainty and the noise uncertainty, and the impact of the channel uncertainty on the optimal strategies.

The Minimax Capacity. The upper value C^+ is computed by first performing the maximization of $J(S_{\mathbf{x}}, S_{\mathbf{n}})$ over $S_{\mathbf{x}}(f) \in A_1$, for a given $S_{\mathbf{n}}(f) \in A_4$. It is well-known that the maximization over A_1 gives the classical water-filling formulae $S_{\mathbf{x}}^o(S_{\mathbf{n}}) + \frac{S_{\mathbf{n}}(f)}{(|H_{nom}(f)| - |W_1(f)|)^2} = \frac{1}{2\lambda_2^o}$, where λ_2^o is a Lagrange multiplier of the set A_1. Next, $\inf_{S_{\mathbf{n}} \in A_4} J(S_{\mathbf{x}}^o(S_{\mathbf{n}}), S_{\mathbf{n}})$ is found by using Lagrange duality technique as before. The optimal PSD's turn out to be equal to the optimal PSD's in the case of the maximin problem, demonstrating the existence of the saddle point (see Appendix C). Thus, C^+ is equal to C^-, and they correspond to the value of the zero sum game C.

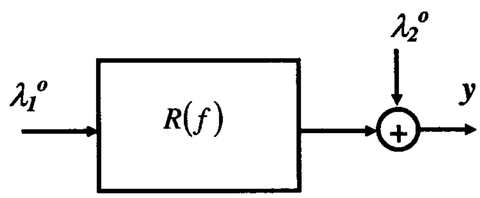

Figure 3.4.13: Explanation of jamming capacity formula

3.4.1 Examples for the Capacity Subject to Jamming

A. Jamming with Known Channel

This example is based on the results of Theorem 3.4.1 which gives the maximin channel capacity C^-, when $\tilde{H}(s) = \omega_n^2/(s^2 + 2\xi\omega_n s + \omega_n^2)$ is completely known ($W_1(s) = 0$), $\xi = 0.2$. The SNR is defined as $SNR = 10\log(P_{\mathbf{x}}/P_{\mathbf{n}})$, $P_{\mathbf{x}} = 0.1$ W, which implies that the SNR will depend on the noise uncertainty. As the uncertainty increases (i.e. $P_{\mathbf{n}}$ increases), the SNR decreases, and hence the capacity decreases. Fig. 3.4.14 shows the channel capacity as a function of the SNR for different values of the natural frequency ω_n (in rad/s) (which affects the bandwidth of the communication channel). As the bandwidth increases, the capacity also increases, indicating how the channel capacity is affected by the channel frequency responses $\tilde{H}(s)$. From Fig. 3.4.15, it can be seen that $|\tilde{H}(s)|^{-2}S_{\mathbf{n}}^o(s)$ water-fills $S_{\mathbf{x}}^o(s)$. The parameters are chosen as follows: $\xi = 0.2$, $P_{\mathbf{n}} = 0.055$ W.

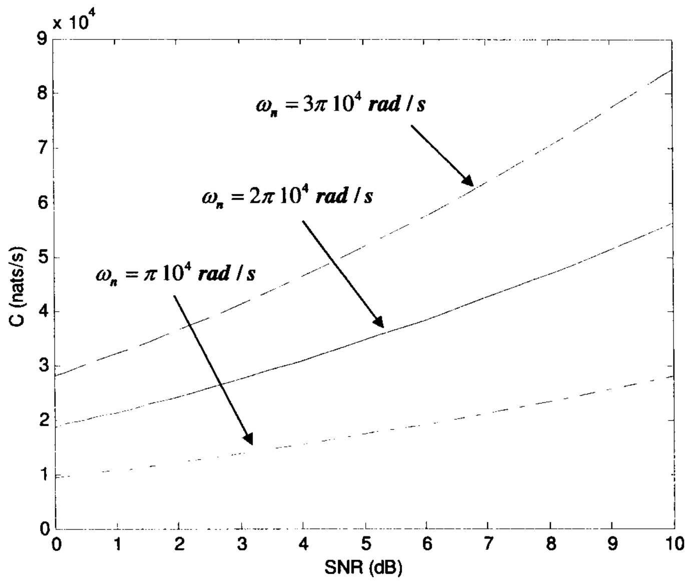

Figure 3.4.14: Channel capacity vs. SNR, jamming with known channel

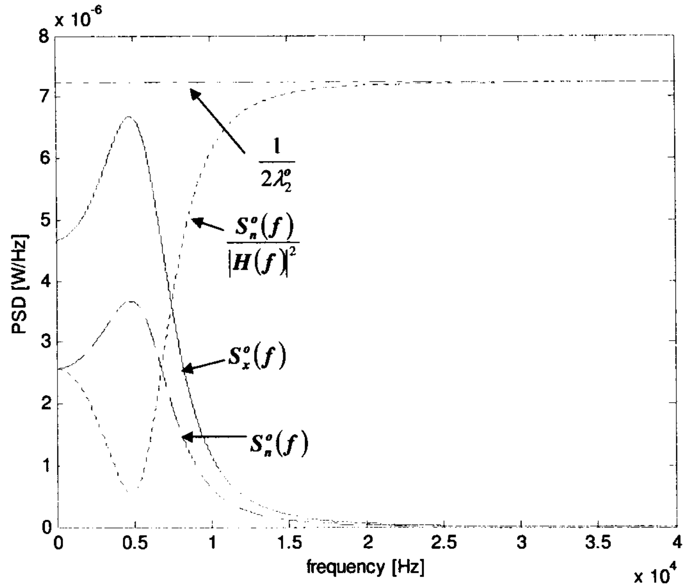

Figure 3.4.15: Water-filling effect

B. Partially Known Channel with Jamming

This example is an application of Theorem 3.4.1, in the presence of the channel uncertainty and jamming. The channel is represented by the second order frequency response $H(s) = \frac{\omega_n^2}{s^2+2\xi\omega_n s+\omega_n^2}$, where a damping ration ξ is uncertain, and $\omega_n = 2\pi 10^4$ rad/s. It is assumed that ξ belongs to the interval $\Delta\xi = \xi_{low} - \xi_{up}$. The nominal channel frequency response $H_{nom}(f)$ is determined by the nominal damping ration $\xi_{nom} = 0.3$ (see Example B in Section 3.2.3). The transmitted power is limited by $P_{\mathbf{x}} = 0.1$W. Fig. 3.4.16 shows the change in the capacity vs. the SNR ration, defined as $SNR = 10 \log P_{\mathbf{x}}/P_{\mathbf{n}}$, for different values of $\Delta\xi$. For instance, the channel capacity decreases by 3 knat/s due to the noise uncertainty, when the SNR decreases from 2dB to 1dB for $\Delta\xi = 0$, and decreases for additional 3 knat/s, due to channel uncertainty $\Delta\xi = 0.15$. Fig. 3.4.17 shows that the optimal transmitter's strategy to combat the noise uncertainty is to shrink the bandwidth and to regroup the power towards the lower frequencies.

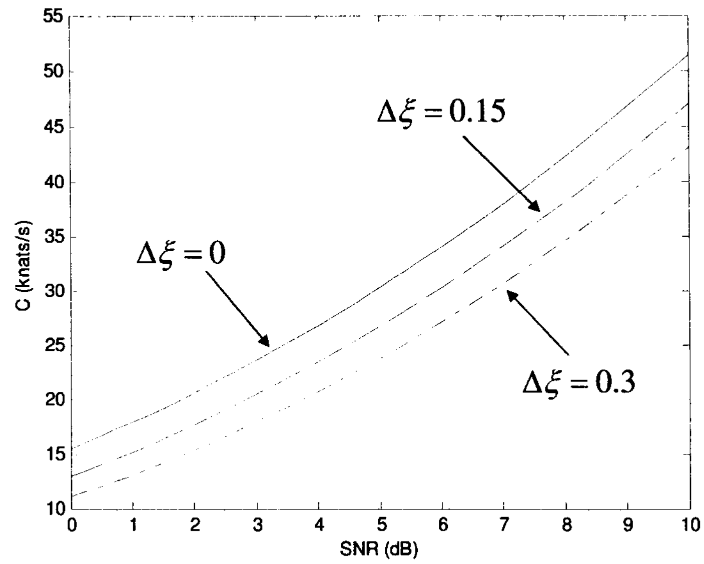

Figure 3.4.16: Capacity vs. uncertainty, jamming and channel uncertain

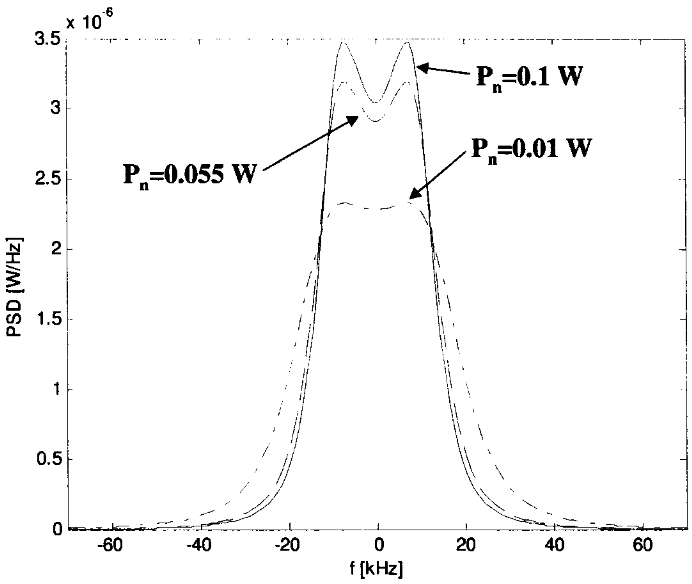

Figure 3.4.17: Optimal PSD of the communicator, jamming and channel uncertain

3.5 Channel Coding and Converse to Channel Coding Theorem

In this section, it is shown that all the capacities derived in the previous sections are indeed the operational capacities of the corresponding compound channels. That is, the channel coding theorem and its converse are derived. The channel coding theorem states that there exist encoders and decoders such that the decoding error probabilities tend to zero exponentially, uniformly over the uncertainty sets, as the codeword lengths tend to infinity provided the code rates are smaller than the derived capacities. The converse of the channel coding theorem states that if the code rates are larger than the corresponding capacities, the decoding error probabilities cannot be made arbitrary small as the codeword lengths tends to infinity. The channel coding theorem and its converse are derived by extending the approach of Root and Varaiya [63], who derived the channel coding theorem and its converse for the class of white Gaussian noise channels, to colored Gaussian noise.

Next, the definitions for a channel code, an attainable rate, and operational channel capacity are given.

Definition 3.5.1 *A channel code (M, ϵ, T) for a set of communication channels $\mathcal{B}$ is defined as the set consisting of M distinct time-functions $\{x_1(t), ..., x_M(t)\}$, in the interval $-T/2 \leq t \leq T/2$ and the set of M disjoint sets $\{D_1, ..., D_M\}$ from the space of a received signal $\mathbf{y}$, such that*

$$\frac{1}{T}\int_{-T/2}^{T/2} x_k^2(t)dt \leq P, \tag{3.5.58}$$

for each k, and such that the error probability for each codeword satisfies

$$\Pr\left(\mathbf{y} \in D_k^c | x_k(t) \ sent\right) \leq \epsilon, \tag{3.5.59}$$

$k = 1, ..., M$, uniformly over the set $\mathcal{B}$.

Definition 3.5.2 *A positive constant R is called an attainable coding rate if there exists a sequence of codes $\{(M, \epsilon_n, T_n)\}$, $M = \exp[T_n R]$, such that when $n \to +\infty$, then $T_n \to +\infty$ and $\epsilon_n \to 0$ uniformly for all communication channels in $\mathcal{B}$. Here ϵ_n is the codeword probability of error as previously defined in (3.5.59), and T_n is a codeword time duration.*

Definition 3.5.3 *The operational capacity C represents the supremum of all attainable rates R.*

Next, we introduce the additional conditions on the sets A_2, A_3, and A_4, which are sufficient for proving the channel coding theorem and its converse.

Define the frequency response of an equivalent communication channel by

$$K(f) = \left(\frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}\right)^{1/2}, \tag{3.5.60}$$

and denote its inverse Fourier transform by $k(t)$. Further define sets $\mathcal{A}_i$ and $\mathcal{B}_i$, $i = 1, 2$, as follows

$$\mathcal{A}_1 \triangleq \{K(f) : S_{\mathbf{x}}(f) \in A_1, \tilde{H}(f) \in A_2, \tilde{W}(f) \in A_3\}, \tag{3.5.61}$$

$$\mathcal{A}_2 \triangleq \{K(f) : S_{\mathbf{x}}(f) \in A_1, \tilde{H}(f) \in A_2, S_{\mathbf{n}}(f) \in A_4\}, \tag{3.5.62}$$

$$\mathcal{B}_i \triangleq \{k(t) : K(f) \in \mathcal{A}_i, k(t) \text{ satisfies } 1.,2.,3.\},\ i = 1, 2, \tag{3.5.63}$$

where $\mathcal{A}_i$ corresponds to a particular problem, and

1) $k(t)$ has finite duration δ,

2) $k(t)$ is square integrable ($k \in L_2$),

3) $\int_{-\infty}^{-\alpha} |K(f)|^2 df + \int_{\alpha}^{+\infty} |K(f)|^2 df \to 0$ when $\alpha \to +\infty$.

Subject to these conditions the sets $\mathcal{B}_i$, $i = 1, 2$, are conditionally compact subsets of L_2 (see [63]). Moreover, the compactness is sufficient for the proof of coding theorem and its converse. Note that the condition 1) can be relaxed (see Lemma 4 [33]). Finally, denote the operational capacities of the codes associated with the sets of channels $\mathcal{B}_i$ by C_i, $i = 1, 2$, respectively. Then the following theorem holds.

Theorem 3.5.4 *The operational capacities C_i for the sets of communication channels $\mathcal{B}_i$, $i = 1, 2$, are equal to corresponding robust capacities, and they are given by (3.2.7) and (3.4.53), respectively.*

Proof. The proof is given in Appendix D.

3.6 Coding Theorem Using Worst Case Channel Notion

The proof of the coding theorem can be simplified if the following fact is observed (communicated to the author by Prof. Kschischang, external examiner).

From Theorem 3.2.1, it is clear that the robust channel capacities are equal to the channel capacities of the worst case channels and the worst case noises from the uncertainty sets. Based on the definitions of the uncertainty sets, the worst case channel corresponds to the channel with the minimum gain and the maximal noise for each frequency, out of all possible channel frequency responses and noise PSD's.

This suggests that a universal code (the one which works simultaneously for all channels belonging to the uncertainty set) is obtained by constructing a code for the worst case channel. If this code performs well on the worst case channel, it will perform even better for "better" channels.

Here, it should be noted that the channel capacity depends on the magnitude of the channel but not on its phase. This can be explained in the following way. A continuous time channel, considered in this chapter, may be understood as the set of infinite number of narrow band channels, each centered at one frequency. For each of these channels, the worst case channel is the one that has the smallest magnitude and the largest noise. The phase does not matter because it can be estimated with arbitrary accuracy, which does not affect the capacity.

3.7 Summary

In this chapter, the information theoretic limits of compound SISO continuous-time Gaussian channels are considered. Specifically, the uncertainty in the channel frequency response is described by a subset of H^∞ space, while the uncertainty in PSD of the noise is described either by using H^∞ space or L_1 space. First, the problems when the channel frequency response and the noise uncertainties are defined in H^∞ space, are considered. It is shown that the robust capacity formulas are explicitly given in terms of the size of uncertainties sets. New water-filling formulas are derived which show how the optimal PSD's of the transmitted signals are affected by the size of uncertainty sets. The above results are applied to wireless fading channels. The second class of problems considered, is based on the assumption that the PSD of the noise is described by a subset of L_1 space. The channel capacity is formulated by using a game theoretical framework, where one player, the communicator (transmitter), wishes to maximize the mutual information rate over the set of finite power PSD's, while the other player, the jammer (noise), wishes to minimize it over its constraint. Although the problems are defined over infinite dimensional spaces, and, hence, the application of the von Neumann minimax theorem is not possible, the existence of a saddle point is shown. Moreover, the solution states that the optimal PSD of the transmitter is proportional to the optimal PSD of the noise, and that the water-filling equation holds for the optimal transmitter's and jammer's strategies. Finally, it is shown that all computed capacities are equal to the corresponding operational capacities.

Chapter 4

Robust Capacity of MIMO Gaussian Channels

4.1 Introduction

This chapter is concerned with the information theoretic bounds and optimal transmission strategies for compound MIMO Gaussian channels, when the channel and noise uncertainties are modeled in the frequency domain. It generalizes the results for SISO compound channels found in Chapter 3. One of the reasons for uncertainty modeling in the frequency domain is motivated by the mutual information rate formula, which is given in the frequency domain (see (4.2.7) below).

The organization of the chapter is as follows. In Section 4.2, definitions of information MIMO channel capacities subject to uncertainties are given. Again, as in Chapter 3, the expression "information capacity" is used to denote the maximin of the mutual information rate, which is different from the operational capacity, i.e., the channel capacity. In Section 4.3, the solution of the information channel capacity problem is obtained, when the channel uncertainty is described by a subset of H^{∞} space. In Section 4.4, the solution of the information channel capacity problem is provided, when the noise uncertainty is described by a subset of L_1 space. In Section 4.5, two examples of information channel capacity computations are presented. In Section 4.6, the operational meaning for the computed information capacities is discussed.

4.2 Problem Definition - MIMO Case

Consider a discrete-time MIMO channel defined by

$$y(t) = \sum_{j=-\infty}^{+\infty} h(t-j)x(j) + n(t), \tag{4.2.1}$$

where $\mathbf{x} \triangleq \{x(t) : t \in Z\}$ is a m-component complex stationary stochastic process representing the transmitted signal, $\mathbf{n} \triangleq \{n(t) : t \in Z\}$ is a p-component complex Gaussian stochastic process representing the noise, $\mathbf{y} \triangleq \{y(t) : t \in Z\}$ is a p-component complex stationary stochastic process representing the received signal, and $\mathbf{h} \triangleq \{h(t) : t \in Z\}$ is a sequence of complex $p \times m$ matrices representing the impulse response of the MIMO communication channel. It is assumed that $\mathbf{x}$ generates a Hilbert space [17]. Here, $H(e^{j\theta})$ represents the channel frequency response matrix, which is the discrete Fourier transform of $\mathbf{h}$ given by

$$H(e^{j\theta}) = \sum_{t=-\infty}^{+\infty} h(t)e^{-j\theta t}, \tag{4.2.2}$$

where θ is the normalized frequency (as used, for instance, in [16] or [60]), and it is assumed that $\sum_{t=-\infty}^{+\infty} h(t)e^{-j\theta t}$ converges in $L_2(F_{\mathbf{x}})$, to a limit in $L_2(F_{\mathbf{x}})$, denoted as $H(e^{j\theta})$. $L_2(F_{\mathbf{x}})$ is a Hilbert space of complex-valued Lebesgue-Stieltjes measurable functions $H(e^{j\theta})$ of a finite norm

$$\|H(e^{j\theta})\|_{L_2(F_{\mathbf{x}})} \triangleq \text{Trace} \int_0^{2\pi} H(e^{j\theta})dF_{\mathbf{x}}(\theta)H^*(e^{j\theta}). \tag{4.2.3}$$

$F_{\mathbf{x}}$ denotes the matrix spectral distribution of $\mathbf{x}$ [17], which is assumed to be absolutely continuous with respect to the Lebesgue measure on $[0, 2\pi]$. Hence, $dF_{\mathbf{x}}(\theta) = W_{\mathbf{x}}(\theta)d\theta$, where $W_{\mathbf{x}}(\theta)$ represents the PSD matrix of $\mathbf{x}$. $W_{\mathbf{n}}(\theta)$ represents the PSD matrix of $\mathbf{n}$.

This model can be used to represent a multi-pair telephone cable that includes the effect of far-end crass-talk [16].

Next, two information capacity problems are defined; first, when the channel frequency response matrix $H(e^{j\theta})$ is uncertain, and second, when the PSD matrix of the noise $W_{\mathbf{n}}(\theta)$ is uncertain.

First problem: Channel unknown, noise known. As in the SISO case, the uncertainty of the channel frequency response matrix can be described by using two low-structured uncertainty representations, additive $H(e^{j\theta}) = H_{nom}(e^{j\theta}) + W_1(e^{j\theta})\Delta(e^{j\theta})W_2(e^{j\theta})$ or multiplicative $H(e^{j\theta}) = (I + W_1(e^{j\theta})\Delta(e^{j\theta})W_2(e^{j\theta}))H_{nom}(e^{j\theta})$ [80]. Similarly to SISO case, $H_{nom}(e^{j\theta})$ is the nominal frequency response, and $W_1(e^{j\theta})\Delta(e^{j\theta})W_2(e^{j\theta})$ is the perturbation. $W_1(e^{j\theta})$ and $W_2(e^{j\theta})$ are stable transfer matrices that characterize the frequency structure of the uncertainty. In this section, we use the additive model in the derivation. Further, introduce two sets

$$A_1 \triangleq \{W_{\mathbf{x}}(\theta) : \int_0^{2\pi} \text{Trace}(W_{\mathbf{x}}(\theta))d\theta \leq P_{\mathbf{x}}\}, \tag{4.2.4}$$

$$A_2 \triangleq \{H \in H^\infty : H = H_{nom} + W_1 \Delta W_2, H, H_{nom}, W_1, W_2, \Delta \in H^\infty, \|\Delta\|_\infty \leq 1\}. \tag{4.2.5}$$

More specifically, let $W_1(e^{j\theta}) = I_p$, and $W_2(e^{j\theta}) = w(e^{j\theta})I_m$, where $w(e^{j\theta})$ is a scalar fixed function, and I_k stands for the identity matrix of dimension k. Thus, $H(e^{j\theta})$ describes the disk centered at $H_{nom}(e^{j\theta})$ with radius determined by $w(e^{j\theta})$. This specific type of uncertainty is chosen because is more tractable in computations. The information channel capacity subject to uncertainty in the frequency response matrix $H(e^{j\theta})$ is defined by

$$C \triangleq \sup_{W_{\mathbf{x}} \in A_1} \inf_{H \in A_2} J(W_{\mathbf{x}}, W_{\mathbf{n}}), \tag{4.2.6}$$

where

$$J(W_{\mathbf{x}}, W_{\mathbf{n}}) = \frac{1}{4\pi} \int_0^{2\pi} \log\det(I_p + H(e^{j\theta})W_{\mathbf{x}}(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^{-1}(\theta))d\theta, \tag{4.2.7}$$

is the mutual information rate between processes $\mathbf{x}$ and $\mathbf{y}$ [17]. The notation $(.)^*$ means complex conjugate transpose, and I_p stands for identity matrix of dimension p. The conditions for the existence of the mutual information rate formula are the following [17]: 1) $\mathbf{x}$ and $\mathbf{y}$ are wide-sense stationary Gaussian processes, 2) $\mathbf{x}$ and $\mathbf{n}$ are mutually independent, 3) $\mathbf{x}$ and $\mathbf{y}$ are full-ranked processes (for the definition of full ranked processes see [17], page 26, Definition 3.1), which is needed to apply Szegö formula ([17], page 36, Theorem 3.4) in the derivation of the mutual information ([17], page 146), 4) Joint PSD matrix of $\mathbf{x}$ and $\mathbf{y}$

$$\begin{bmatrix} W_{\mathbf{x}}(\theta) & W_{\mathbf{xy}}(\theta) \\ W_{\mathbf{xy}}^*(\theta) & W_{\mathbf{y}}(\theta) \end{bmatrix} \tag{4.2.8}$$

is nonsingular, where $W_{\mathbf{y}}(\theta)$ is the PSD matrix of $\mathbf{y}$, and $W_{\mathbf{xy}}(\theta)$ is the cross-PSD matrix of $\mathbf{x}$ and $\mathbf{y}$. Formula (4.2.7) has been already used in [54]. Here, it is assumed that the noise is white, $W_{\mathbf{n}}(\theta)) = I$. The effect of the colored noise can be taken into account by factorizing $W_{\mathbf{n}}(\theta)$ and including it in the channel frequency response matrix $H(e^{j\theta})$.

Second problem: Channel known, noise unknown. The noise uncertainty is defined through the uncertainty of the PSD matrix $W_{\mathbf{n}}(\theta)$. It is assumed that although unknown, $W_{\mathbf{n}}(\theta)$ belongs to the set $A_3 \triangleq \{W_{\mathbf{n}}(\theta) : \int_0^{2\pi} \text{Trace}(W_{\mathbf{n}}(\theta))d\theta \leq P_{\mathbf{n}}\}$. The same constraint is introduced for the transmitter such that $A_1 \triangleq \{W_{\mathbf{x}}(\theta) : \int_0^{2\pi} \text{Trace}(W_{\mathbf{x}}(\theta))d\theta \leq P_{\mathbf{x}}\}$. The information capacity of an uncertain MIMO channel is defined similarly to its SISO counterpart by

$$C \triangleq \sup_{W_{\mathbf{x}} \in A_1} \inf_{W_{\mathbf{n}} \in A_3} J(W_{\mathbf{x}}, W_{\mathbf{n}}). \tag{4.2.9}$$

In the next two sections, the solutions of (4.2.6) and (4.2.9) are presented.

4.3 First MIMO Problem : Channel Unknown, Noise Known

Here, it is assumed that the transmitter and the receiver know the nominal channel frequency response matrix $H_{nom}(e^{j\theta})$, as well as the size of the uncertainty set $w(e^{j\theta})$. This implies that both, the transmitter and the receiver, are aware of uncertainty set A_2. The solution of the MIMO capacity problem, in the case of the channel frequency response uncertainty, uses the singular value decomposition of the nominal frequency response matrix $H_{nom}(e^{j\theta}) = U(e^{j\theta})\Sigma(\theta)V^*(e^{j\theta})$, where $U(e^{j\theta})$ and $V(e^{j\theta})$ are $p \times p$ and $m \times m$ unitary matrices, respectively, and

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}_{p \times m}, \tag{4.3.10}$$

where $\Sigma_1(\theta) = \text{diag}(\sigma_1(\theta), ..., \sigma_n(\theta))$ is a diagonal matrix whose elements are called singular values of $H_{nom}(e^{j\theta})$, where $n \leq \min(p, m)$. Also, the derivation assumes a specific structure of the matrix $\Delta(e^{j\theta})$,

$$\Delta(e^{j\theta}) = U(e^{j\theta}) \begin{bmatrix} \Delta_1(e^{j\theta}) & 0 \\ 0 & 0 \end{bmatrix} V^*(e^{j\theta}). \tag{4.3.11}$$

$\Delta_1(e^{j\theta})$ is an $n \times n$ matrix, where n is the rank of $H_{nom}(e^{j\theta})$, $n \leq \min(p, m)$. The reason for introducing this assumption is that it enables the application of Hadamard's inequality for maximization of the mutual information with respect to the PSD matrix $W_{\mathbf{x}}(\theta)$. This means that the most general problem defined by set A_2 in (4.2.5) still remains unsolved and it is the part of the future work. In Section 4.6, it will be shown that the information capacity computed in Theorem 4.3.1 represents an upper bound to the achievable rate in the most general case, for arbitrary $\Delta(e^{j\theta})$.

The information channel capacity is given by the following theorem.

Theorem 4.3.1 *The information capacity of a discrete-time compound MIMO Gaussian channel modeled by (4.2.1), subject to a channel frequency response uncertainty described by H^∞ norm constraint through the set A_2 and subject to (4.3.11), is given by*

$$C = \frac{1}{4\pi}\sum_{i=1}^{n}\int_{\mathcal{S}_i}\log[\mu(\sigma_i(\theta) - |w(e^{j\theta})|)^2]d\theta, \tag{4.3.12}$$

$$\sum_{i=1}^{n}\int_{\mathcal{S}_i}(\mu - (\sigma_i(\theta) - |w(e^{j\theta})|)^{-2})d\theta = P_{\mathbf{x}}, \tag{4.3.13}$$

where

$$\mathcal{S}_i \triangleq \{\theta : \mu - (\sigma_i(\theta) - |w(e^{j\theta})|)^{-2} > 0\},\ i = 1, ..., n,\ \mu > 0, \tag{4.3.14}$$

and μ is related to the Lagrange multiplier of the power constraint associated with the set A_1.

A matrix $\Delta_1^o(e^{j\theta})$, which minimizes the mutual information rate, is given by

$$\Delta_1^0(e^{j\theta}) = \begin{pmatrix} e^{-jarg(w)+j\pi} & 0 & \cdots & 0 \\ 0 & e^{-jarg(w)+j\pi} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-jarg(w)+j\pi} \end{pmatrix}. \tag{4.3.15}$$

Notation $arg(w)$ stands for the phase of the transfer function $w(e^{j\theta})$.

Proof. The proof is given in Appendix E.

Corollary 4.3.2 *If $|w(e^{j\theta})|$ is equal to zero, the formula for the channel capacity when there is no uncertainty is obtained [16].*

Remark 4.3.3 *The information capacity formula implies that the capacity will be different from zero only if there exists an interval of frequencies such that $\sigma_i(\theta) - |w(e^{j\theta})| > 0$, i=1,...,n.*

From the derived information capacity, it can be seen that if the uncertainty $|w(e^{j\theta})|$ grows over the whole frequency range, at one point it will reach the lowest singular value $\sigma_n(\theta)$ of the nominal frequency response matrix $H_{nom}(e^{j\theta})$. Then, the optimal way of transmission is to concentrate the power on the rest of the channel modes, and not to allocate the power to the mode that vanished because of uncertainty. By the same logic, if the uncertainty keeps growing, the modes will vanish one by one, and at the end only the strongest mode will remain.

Also, it should be noted that the information capacity for memoryless channels is obtained from the above capacity, by formally removing the integration, in the problem statement and final answer. Then, $W_{\mathbf{x}}$ represents the covariance matrix of the transmitted signal, $W_{\mathbf{n}}$ represents the covariance matrix of the noise, and H is the channel matrix in time domain.

4.4 Second MIMO Problem: Channel Known, Noise Unknown

In this section, the problem of computing the information channel capacity of a Gaussian MIMO channel when the PSD matrix of the noise $W_{\mathbf{n}}(\theta)$ is unknown, is considered. Again, as in the SISO case, the problem may be treated as a game between the transmitter $\mathbf{x}$ and the noise/jammer $\mathbf{n}$. Using the same reasoning as in Chapter 3, the standard result of the game theory cannot be applied to claim a saddle point existence [56]. Therefore, the problem is solved by directly solving the maximin problem. The finite-dimensional version of the problem has been solved in [71] by using the saddle point existence and a linear matrix inequality technique. Thus, the explicit dependence between the optimal transmitter's and jammer's strategies cannot be seen. The obtained results for MIMO capacity resemble the results derived for the SISO case. The information channel capacity of the MIMO Gaussian channel with the noise uncertainty is given by the following theorem.

Theorem 4.4.1 *The information capacity of a discrete-time MIMO Gaussian channel modeled by (4.2.1), subject to a noise uncertainty described by the L_1 norm constraint posed on the PSD matrix $W_{\mathbf{n}}(\theta)$, A_3, is given by*

$$C = \frac{1}{4\pi}\int_0^{2\pi} \log\det(I_p + H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^{o-1}(\theta))d\theta, \qquad (4.4.16)$$

where the optimal PSD matrices satisfy

$$W_{\mathbf{n}}^{o2}(\theta) + \frac{1}{2}W_{\mathbf{n}}^o(\theta)H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta}) + \frac{1}{2}H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^o(\theta)$$
$$-\frac{1}{4\pi\lambda_1^o}H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta}) = 0, \qquad (4.4.17)$$
$$H^*(e^{j\theta})W_{\mathbf{n}}^{o-1}(\theta)(I_p + H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^{o-1}(\theta))^{-1}H(e^{j\theta}) = 4\pi\lambda_2^o I_m, \qquad (4.4.18)$$

and $\lambda_1^o > 0$ and $\lambda_2^o > 0$ are the Lagrange multipliers associated with the constraint sets A_1 and A_2, respectively, which are computed from

$$\int_0^{2\pi} Trace(W_{\mathbf{x}}^o(\theta))d\theta = P_{\mathbf{x}}, \quad \int_0^{2\pi} Trace(W_{\mathbf{n}}^o(\theta))d\theta = P_{\mathbf{n}}. \tag{4.4.19}$$

Proof. The proof is given in Appendix F.

Remark 4.4.2 *From the solution given in Appendix F, it can be seen that the optimal solutions $W_{\mathbf{x}}^o(\theta)$ and $W_{\mathbf{n}}^o(\theta)$ constitute a saddle point of the optimization problem given by (4.2.9), meaning*

$$J(W_{\mathbf{x}}, W_{\mathbf{n}}^o) \leq J(W_{\mathbf{x}}^o, W_{\mathbf{n}}^o) \leq J(W_{\mathbf{x}}^o, W_{\mathbf{n}}), \tag{4.4.20}$$

for any $W_{\mathbf{x}}(\theta) \neq W_{\mathbf{x}}^o(\theta)$ and any $W_{\mathbf{n}}(\theta) \neq W_{\mathbf{n}}^o(\theta)$. In addition, it is true that

$$\sup_{W_{\mathbf{x}} \in A_1} \inf_{W_{\mathbf{n}} \in A_3} J(W_{\mathbf{x}}, W_{\mathbf{n}}) = \inf_{W_{\mathbf{n}} \in A_3} \sup_{W_{\mathbf{x}} \in A_1} J(W_{\mathbf{x}}, W_{\mathbf{n}}). \tag{4.4.21}$$

Further, note that (4.4.17) and (4.4.18) correspond to SISO equations (C.0.3) and (3.4.57). When $H(e^{j\theta})$ is square and invertible, after some manipulation of (4.4.17) and (4.4.18), it is shown that

$$W_{\mathbf{x}}^o(\theta) = \frac{\lambda_1^o}{\lambda_2^o} H^*(e^{j\theta}) W_{\mathbf{n}}^o(\theta) H(e^{j\theta}) (H^*(e^{j\theta}) H(e^{j\theta}))^{-1}, \tag{4.4.22}$$

which is MIMO equivalent of SISO (3.4.55). Thus, when the transmitter does not know the PSD of the noise, it will use the PSD matrix $W_{\mathbf{x}}^o(\theta)$, which guarantees the transmission rate of at least $J(W_{\mathbf{x}}^o, W_{\mathbf{n}}^o)$ according to (4.4.20). Similarly, "nature" tends to use the PSD matrix $W_{\mathbf{n}}^o(\theta)$, which is a scaled version of the optimal PSD matrix $W_{\mathbf{x}}^o(\theta)$. Hence, (4.4.22) shows the advantage of a direct solution of optimization problem (4.2.9), since it directly relates the optimal transmitter's and jammer's strategies, which cannot be seen if the problem is solved through numerical techniques.

In addition, the derived formula can be used to model the communication channel subject to jamming in the following scenario. Assume that the jammer knows the number of antennas that is used by a wireless system. One such example is a public cellular system subject to the jammer's attack. If the channel matrix between the jammer and the base station of the cellular system is denoted by $G(e^{j\theta})$ (of dimension $p \times p$), then the mutual information subject to jamming is given by

$$\begin{aligned} J(W_{\mathbf{x}}, W_{\mathbf{n}}) &= \frac{1}{4\pi}\int_0^{2\pi} \log\det(I_p + H(e^{j\theta})W_{\mathbf{x}}(\theta)H^*(e^{j\theta})(G(e^{j\theta})W_{\mathbf{n}}(\theta)G^*(e^{j\theta}))^{-1})d\theta \\ &= \frac{1}{4\pi}\int_0^{2\pi} \log\det(I_p + H_1(e^{j\theta})W_{\mathbf{x}}(\theta)H_1^*(e^{j\theta})W_{\mathbf{n}}^{-1}(\theta))d\theta, \end{aligned} \tag{4.4.23}$$

where $H_1(e^{j\theta}) \triangleq G^{-1}(e^{j\theta})H(e^{j\theta})$. Here, it is assumed that $G(e^{j\theta})$ is an invertible matrix.

4.5 Examples

4.5.1 Partially Known Channel

The following example illustrates the computation of the channel capacity, when the uncertainty with respect to the channel frequency response matrix is described by the subset of H^∞ space.

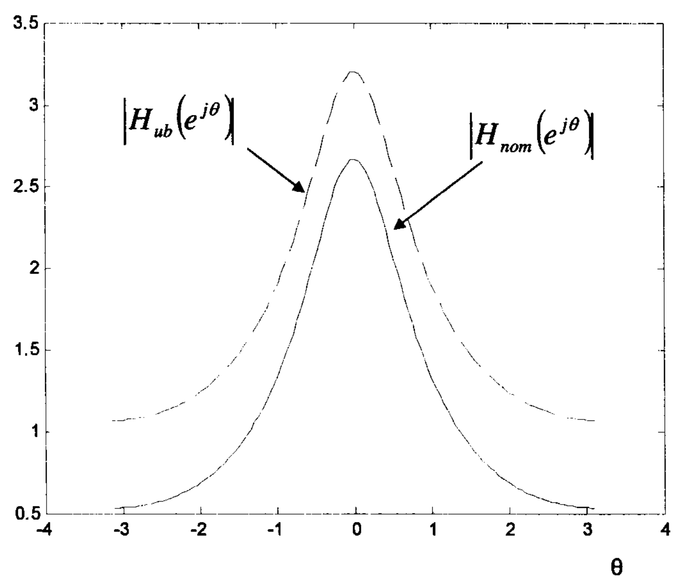

Figure 4.5.1: Magnitudes of $[H_{nom}(e^{j\theta})]_{11}$ and $[H_{ub}(e^{j\theta})]_{11}$

The uncertainty is represented by the additive description $H(e^{j\theta}) = H_{nom}(e^{j\theta}) + \Delta(e^{j\theta})w(e^{j\theta})$. The nominal channel frequency response matrix is given by

$$H_{nom}(e^{j\theta}) = \begin{bmatrix} \frac{1}{1-\frac{1}{2}e^{-j\theta}}\frac{1}{1-\frac{1}{4}e^{-j\theta}} & \frac{1}{1-\frac{1}{4}e^{-j\theta}} \\ \frac{1}{1-\frac{1}{4}e^{-j\theta}} & \frac{1}{1-\frac{1}{2}e^{-j\theta}}\frac{1}{1-\frac{1}{4}e^{-j\theta}} \end{bmatrix}, \tag{4.5.24}$$

while the upper bound on the uncertainty set is given by

$$H_{ub}(e^{j\theta}) = \begin{bmatrix} \frac{\delta}{1-\frac{1}{2}e^{-j\theta}} & \frac{1.4}{1-\frac{1}{4}e^{-j\theta}} \\ \frac{1.4}{1-\frac{1}{4}e^{-j\theta}} & \frac{\delta}{1-\frac{1}{2}e^{-j\theta}} \end{bmatrix}, \tag{4.5.25}$$

where δ is a known constant. The magnitudes of $[H_{nom}(e^{j\theta})]_{11}$ and $[H_{ub}(e^{j\theta})]_{11}$ are shown in Fig. 4.5.1. The size of the uncertainty set is determined by observing that $\bar{\sigma}(H_{ub}(e^{j\theta}) - H_{nom}(e^{j\theta})) = \bar{\sigma}(\Delta(e^{j\theta})w(e^{j\theta})I_m) \leq |w(e^{j\theta})|$, where $\bar{\sigma}(A)$ is the maximal singular value of a matrix A. It is chosen that $|w(e^{j\theta})| = \bar{\sigma}(H_{ub}(e^{j\theta}) - H_{nom}(e^{j\theta}))$ for each frequency $\theta \in [-\pi, \pi]$.

In Fig. 4.5.2, 4.5.3, the size of the uncertainty set $|w(e^{j\theta})|$ is compared to the singular values $\sigma_1(\theta)$ and $\sigma_2(\theta)$ of the nominal frequency response matrix $H_{nom}(e^{j\theta})$, respectively, for $\delta = 1.6$. Fig. 4.5.3 suggests that the transmission over the second mode is optimal just over the part of the frequency bandwidth, where $\sigma_2(\theta) > |w(e^{j\theta})|$. This implies that for a larger size of the uncertainty set $|w(e^{j\theta})|$, the weaker mode will not be used in the transmission.

The capacity is computed by employing (4.3.12) and (4.3.13). It is interesting to observe (4.3.13). It requires the positivity of $\mu - (\sigma_i(\theta) - |w(e^{j\theta})|)^{-2}$. Fig. 4.5.4 shows $(\sigma_1(\theta) - |w(e^{j\theta})|)^{-2}$, while Fig. 4.5.6 and 4.5.5 show $(\sigma_2(\theta) - |w(e^{j\theta})|)^{-2}$. In order to have a transmission over the i^{th} mode, the constant μ, which is chosen in accordance with the power constraint (4.3.13), must be larger than $(\sigma_i(\theta) - |w(e^{j\theta})|)^{-2}$ over at least one frequency $\theta \in [-\pi, \pi]$. Because $\sigma_i(\theta) - |w(e^{j\theta})|$ is smaller for smaller singular values of a nominal channel frequency response matrix, $(\sigma_i(\theta) - |w(e^{j\theta})|)^{-2}$ is larger for smaller singular values. This can be verified from Fig. 4.5.4 and 4.5.6. Thus, for small values of transmitted power $P_{\mathbf{x}}$, the level of power poured into $(\sigma_i(\theta) - |w(e^{j\theta})|)^{-2}$ may be small. Hence, for small transmitted power $P_{\mathbf{x}}$, it could happen that the weaker modes could be left without power. In fact, the larger uncertainty makes the weaker modes even weaker, which may contribute to the transmission only over the strongest mode for small transmitted power.

Fig. 4.5.7 shows the channel capacity versus the parameter δ, which determines the size of the uncertainty set, for a fixed value of P/N_0. As expected, the channel capacity decreases, as the size of uncertainty set increases.

4.5.2 Partially Known Noise

Next, it will be shown how the channel capacity of a MIMO communication channel can be computed when the PSD matrix of the noise, $W_{\mathbf{n}}(\theta)$, belongs to the set A_3, which is a subset

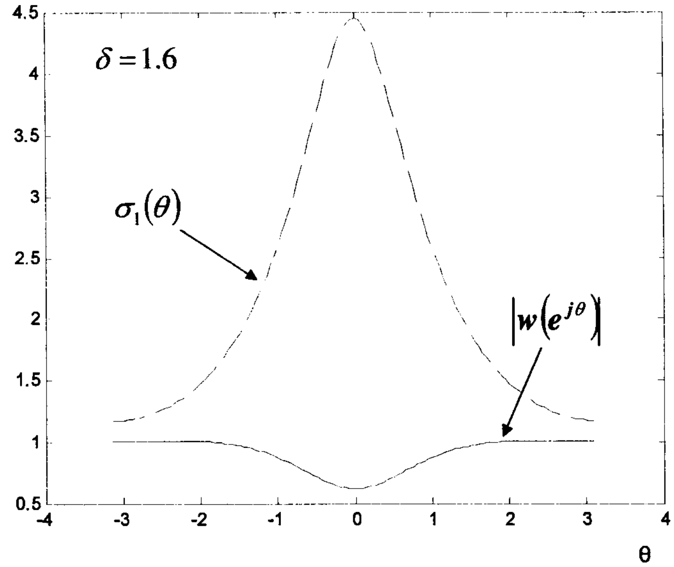

Figure 4.5.2: $\sigma_1(\theta)$ and $|w(e^{j\theta})|$

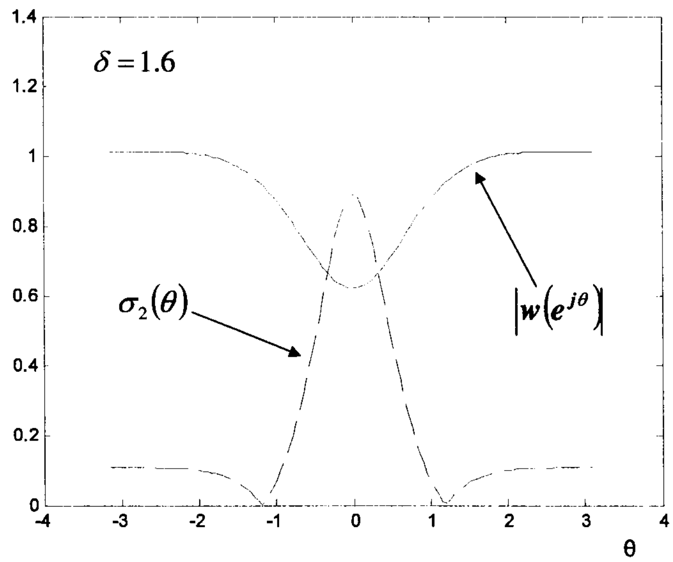

Figure 4.5.3: $\sigma_2(\theta)$ and $|w(e^{j\theta})|$

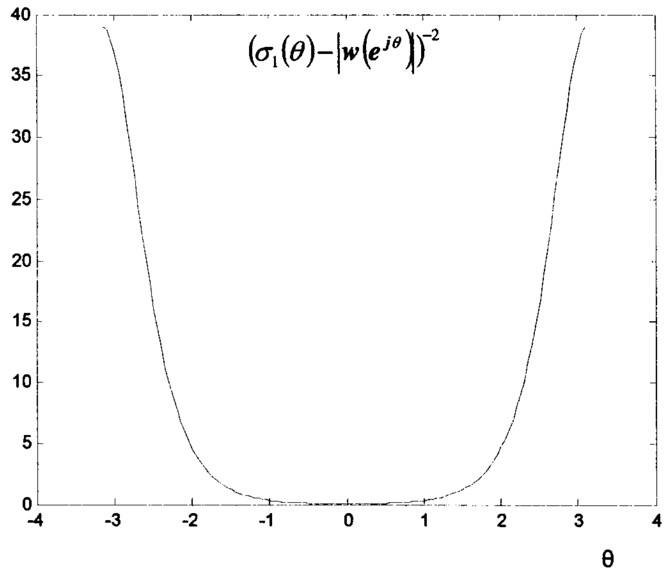

Figure 4.5.4: $(\sigma_1(\theta) - |w(e^{j\theta})|)^{-2}$

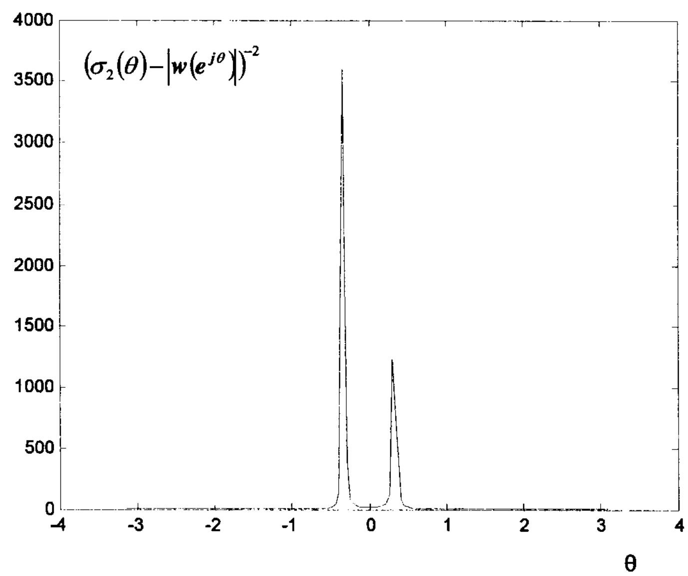

Figure 4.5.5: $(\sigma_2(\theta) - |w(e^{j\theta})|)^{-2}$

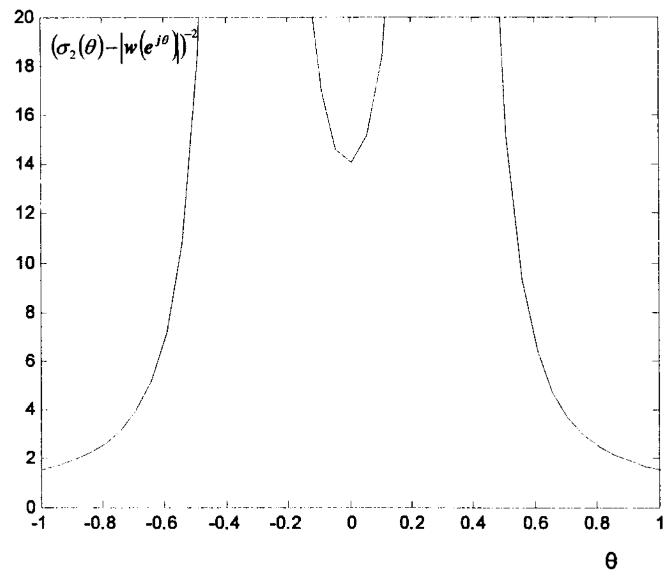

Figure 4.5.6: $(\sigma_2(\theta) - |w(e^{j\theta})|)^{-2}$

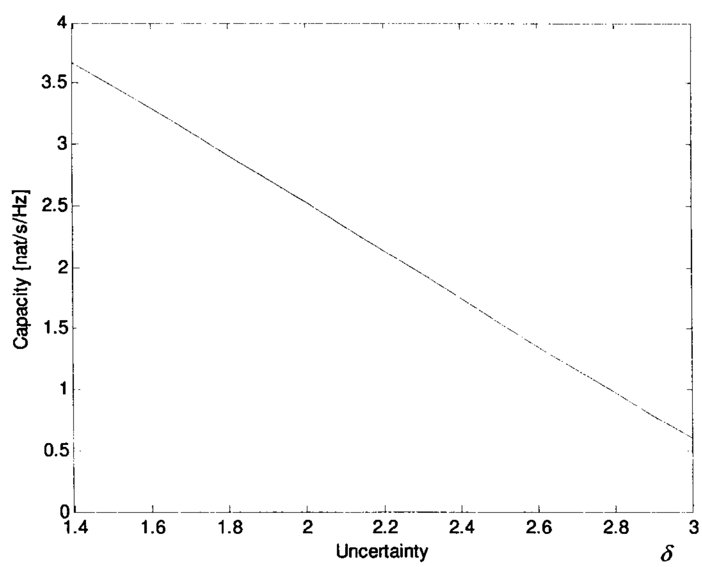

Figure 4.5.7: Channel capacity C vs. δ

of L_1. The channel frequency response is completely known, and it is given by

$$H(e^{j\theta}) = \begin{bmatrix} \frac{1}{1-\frac{1}{2}e^{-j\theta}} & \frac{1}{1-\frac{1}{4}e^{-j\theta}} \\ \frac{1}{1-\frac{1}{4}e^{-j\theta}} & \frac{1}{1-\frac{1}{2}e^{-j\theta}} \end{bmatrix}. \tag{4.5.26}$$

In this case, the channel frequency response matrix $H(e^{j\theta})$ is square and invertible. Then, the optimal PSD matrices of the communicator and the jammer are given by

$$W_{\mathbf{x}}^o(\theta) = \frac{\lambda_1^o}{\lambda_2^o} H^*(e^{j\theta})(4\pi\lambda_2^o I_p + 4\pi\lambda_1^o H(e^{j\theta})H^*(e^{j\theta}))^{-1}H(e^{j\theta}), \tag{4.5.27}$$

$$W_{\mathbf{n}}^o(\theta) = (4\pi\lambda_2^o I_p + 4\pi\lambda_1^o H(e^{j\theta})H^*(e^{j\theta}))^{-1}H(e^{j\theta})H^*(e^{j\theta}). \tag{4.5.28}$$

The formula for a channel capacity becomes

$$C = \frac{1}{4\pi}\int_0^{2\pi} \log \det(I_p + \frac{\lambda_1^o}{\lambda_2^o} H(e^{j\theta})H^*(e^{j\theta}))d\theta. \tag{4.5.29}$$

To find all required quantities, the Lagrangian multipliers λ_1^o and λ_2^o are numerically computed. Fig. 4.5.8 shows the channel capacity vs. $SNR = 10\log P_{\mathbf{x}}/P_{\mathbf{n}}$. The transmitted power is limited by $P_{\mathbf{x}} = 0.1$W.

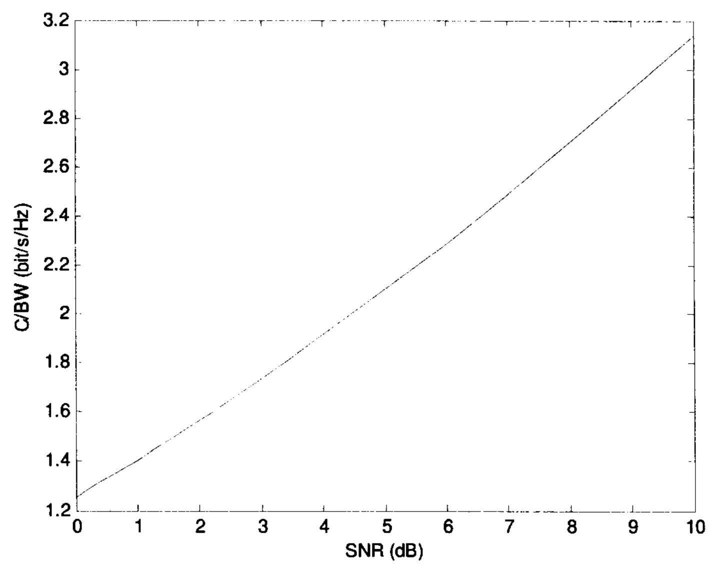

Figure 4.5.8: MIMO channel capacity subject to uncertain noise

It is interesting to notice that the Lagrange multipliers λ_1^o and λ_2^o represent the derivatives of the capacity C with respect to $P_{\mathbf{n}}$ and $P_{\mathbf{x}}$, respectively. The multipliers are shown in Fig.

4.5.9 and 4.5.10. It can be seen that the capacity is more sensitive to variations in $P_{\mathbf{n}}$ and $P_{\mathbf{x}}$ for high SNR. But, a better picture regarding the sensitivity of the capacity with respect to $P_{\mathbf{n}}$ and $P_{\mathbf{x}}$ is obtained by comparing the relative perturbation of the capacity, $\frac{\Delta C}{C}$, to the relative perturbations of $P_{\mathbf{n}}$, $\frac{\Delta P_{\mathbf{n}}}{P_{\mathbf{n}}}$, and $P_{\mathbf{x}}$, $\frac{\Delta P_{\mathbf{x}}}{P_{\mathbf{x}}}$. Viewing C as a function of $P_{\mathbf{n}}$ and $P_{\mathbf{x}}$, and taking the limit with respect to $P_{\mathbf{n}}$ and $P_{\mathbf{x}}$, we get

$$S_{\mathbf{n}} = \lim_{P_{\mathbf{n}} \to 0} \frac{\Delta C/C}{\Delta P_{\mathbf{n}}/P_{\mathbf{n}}} = \frac{P_{\mathbf{n}}}{C}\frac{dC}{dP_{\mathbf{n}}}, \tag{4.5.30}$$

$$S_{\mathbf{x}} = \lim_{P_{\mathbf{x}} \to 0} \frac{\Delta C/C}{\Delta P_{\mathbf{x}}/P_{\mathbf{x}}} = \frac{P_{\mathbf{x}}}{C}\frac{dC}{dP_{\mathbf{x}}}. \tag{4.5.31}$$

The ratios of relative perturbations are shown in Fig. 4.5.11 and Fig. 4.5.12. For this particular example, the relative perturbation of the capacity is between 1/3 and 1/2 of the relative perturbations of $P_{\mathbf{n}}$ and $P_{\mathbf{x}}$. Thus, the sensitivity of the capacity with respect to $P_{\mathbf{n}}$ and $P_{\mathbf{x}}$ is not significant.

Other important conclusions come from the optimal PSD matrices $W_{\mathbf{x}}^o(\theta)$ and $W_{\mathbf{n}}^o(\theta)$ depicted in Fig. 4.5.13, 4.5.14. Two transmitted signal PSD functions, $[W_{\mathbf{x}}^o(\theta)]_{11} = [W_{\mathbf{x}}^o(\theta)]_{22}$ and $[W_{\mathbf{x}}^o(\theta)]_{12} = [W_{\mathbf{x}}^o(\theta)]_{21}$, and two noise PSD functions $[W_{\mathbf{n}}^o(\theta)]_{11} = [W_{\mathbf{n}}^o(\theta)]_{22}$ and $[W_{\mathbf{n}}^o(\theta)]_{12} = [W_{\mathbf{n}}^o(\theta)]_{21}$ are presented. The first conclusion is that the off diagonal elements are different from zero, which implies that the optimal transmission and the optimal jamming require cooperations between different transmitters and different jammers in MIMO communication jamming scheme. Another interesting observation is that diagonal and off-diagonal PSD functions are, in a sense, complemental. For instance, the minimum of $[W_{\mathbf{x}}^o(\theta)]_{12}$ corresponds to the maximum of $[W_{\mathbf{x}}^o(\theta)]_{11}$, and vice versa. This might be the case for this particular example, but it remains to be explored if this is the rule and under which conditions, or merely it is a coincidence.

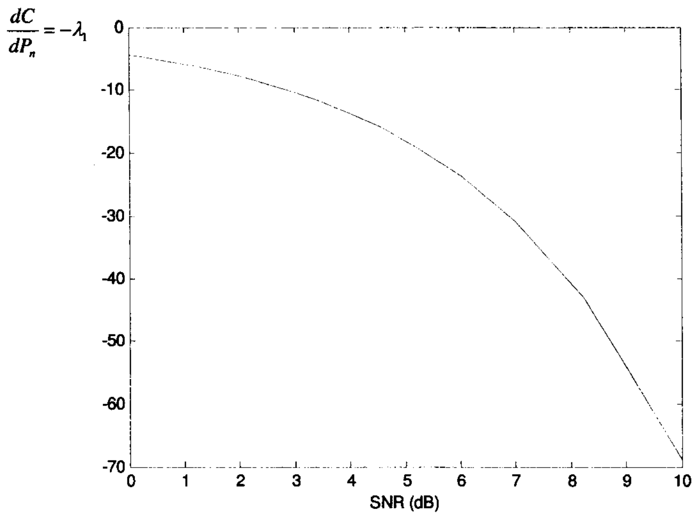

Figure 4.5.9: Derivative of capacity w.r.t. $P_{\mathbf{n}}$

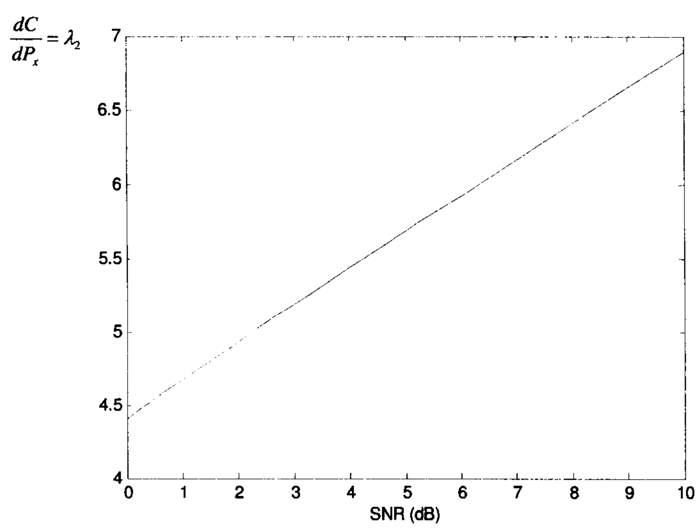

Figure 4.5.10: Derivative of capacity w.r.t. $P_{\mathbf{x}}$

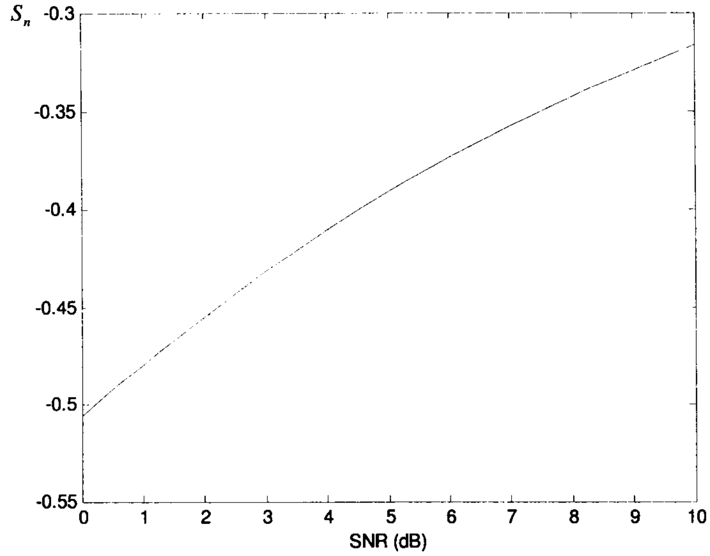

Figure 4.5.11: Sensitivity w.r.t. $P_{\mathbf{n}}$

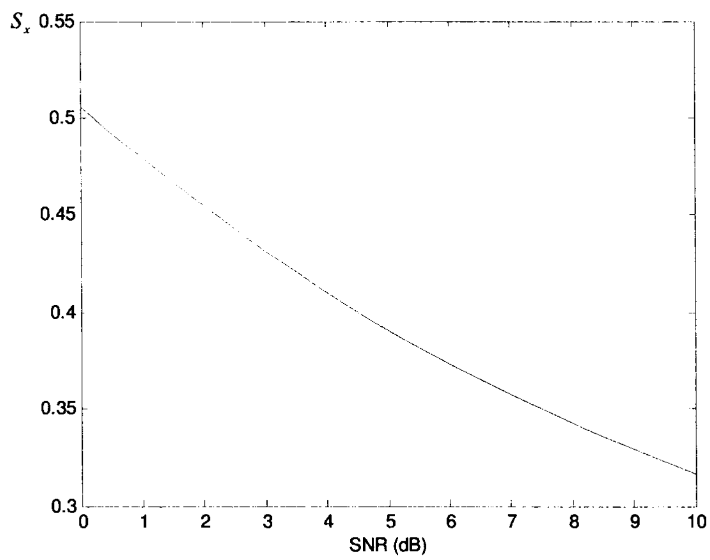

Figure 4.5.12: Sensitivity w.r.t. $P_{\mathbf{x}}$

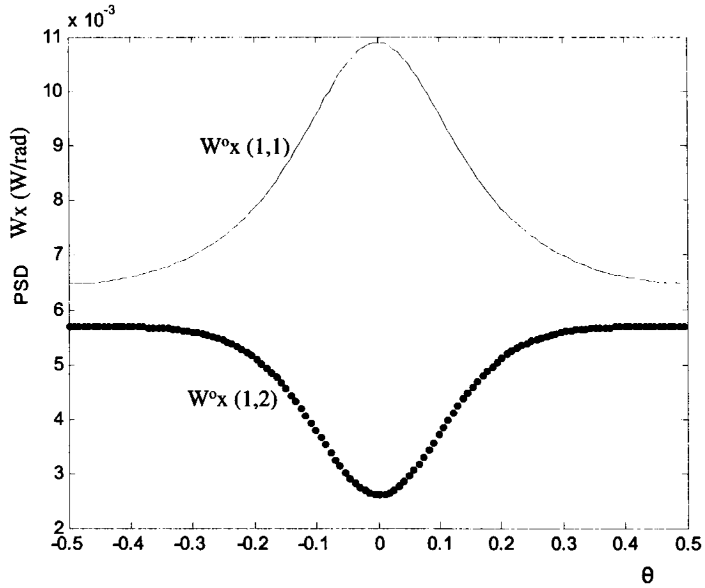

Figure 4.5.13: Optimal communicator's PSD matrix

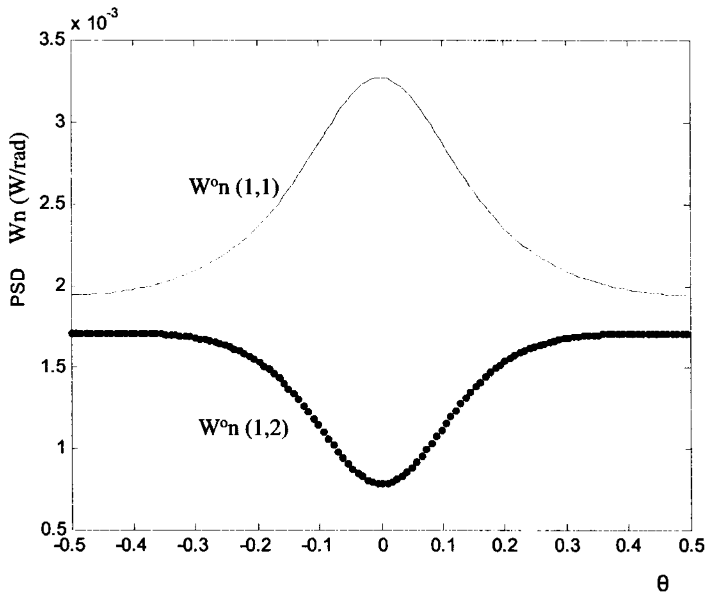

Figure 4.5.14: Optimal jammer's PSD matrix

4.6 Achievable Rates for MIMO Channels

In Chapter 3, it is proved that the operational capacities of corresponding SISO compound channels are equal to the worst case channel capacities. In this section, it is shown that in the case of MIMO compound channels, certain additional conditions have to be imposed in order to prove the achievability of the worst case information capacities computed in Theorem 4.3.1 and Theorem 4.4.1.

4.6.1 Achievable Rates for Uncertain Channel

When the channel uncertainty is described by the set A_2, it is shown that the supremum and the infimum of the problem defined by (4.2.6) (see Appendix E), can be interchanged. Moreover, the form of the information capacity formula given by (4.3.12) and (4.3.13), and subject to (4.3.11), suggests that we are dealing with the information capacity of the worst case channel, as in the SISO case. Namely, from (4.3.12) and (4.3.13), the worst case channel is determined by the singular values of the nominal channel matrix, $H_{nom}(e^{j\theta})$, that are reduced by the size of the uncertainty set $|w(e^{j\theta})|$. Similarly to the SISO solution, where the capacity is the function of the magnitude but not the phase, the MIMO capacity depends only on the singular values of the channel frequency response and not on the unitary matrices $U(e^{j\theta})$ and $V(e^{j\theta})$ [16]. Therefore, all channel matrices, which have the same singular values, constitute an equivalent class, where all members of the class have the same channel capacity.

Next, it is demonstrated that that the worst case channel capacity, given by (4.3.12) and (4.3.13), can be achieved, subject to (4.3.11)

$$\Delta(e^{j\theta}) = U(e^{j\theta})\begin{bmatrix} \Delta_1(e^{j\theta}) & 0 \\ 0 & 0 \end{bmatrix} V^*(e^{j\theta}) = U(e^{j\theta})\Delta_s(e^{j\theta})V^*(e^{j\theta}), \tag{4.6.32}$$

where $U(e^{j\theta})$ and $V(e^{j\theta})$ are the unitary matrices that correspond to the singular value decomposition of the the nominal channel frequency response matrix

$$H_{nom}(e^{j\theta}) = U(e^{j\theta})\Sigma(\theta)V^*(e^{j\theta}). \tag{4.6.33}$$

Note that, even when

$$\Delta_s(e^{j\theta}) \triangleq \begin{bmatrix} \Delta_1(e^{j\theta}) & 0 \\ 0 & 0 \end{bmatrix} \tag{4.6.34}$$

is an unknown diagonal matrix such that $\|\Delta_s(e^{j\theta})\|_\infty \leq 1$, in general, $\Delta(e^{j\theta})$ is not diagonal. Thus, the uncertainty is placed on each element of the matrix $H_{nom}(e^{j\theta})$. Any channel frequency response matrix from A_2 subject to (4.6.32) is given by

$$H(e^{j\theta}) = U(e^{j\theta})(\Sigma(\theta) + \Delta_s(e^{j\theta})w(e^{j\theta}))V^*(e^{j\theta}), \tag{4.6.35}$$

where $\Sigma(\theta) + \Delta_s(e^{j\theta})w(e^{j\theta})$ is diagonal. In order to achieve the worst case channel capacity, it is enough to diagonalize the channel matrix $H(e^{j\theta})$ by precoding the transmitted signal by $V(e^{j\theta})$ and by shaping the received signal by $U^*(e^{j\theta})$ as shown in Fig. 4.6.15. Consequently, n parallel channels are obtained, which enables the use of SISO codes. Here, each of n codes is tuned on the worst case channel $\sigma_i(\theta) - |w(e^{j\theta})|$, $i = 1, ..., n$, for each of n parallel channels.

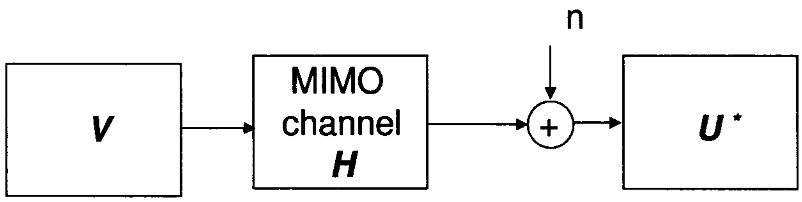

Figure 4.6.15: Channel capacity C vs. δ

In the general case, $\Delta(e^{j\theta})$ does not have a form (4.6.32), where $\Delta_s(e^{j\theta})$ is diagonal. Next, look at the following form of the additive uncertainty description given by

$$\begin{aligned} H(e^{j\theta}) &= H_{nom}(e^{j\theta}) + \Delta(e^{j\theta})w(e^{j\theta}) && (4.6.36)\\ &= U(e^{j\theta})\Sigma(\theta)V^*(e^{j\theta}) + \Delta(e^{j\theta})w(e^{j\theta}) && (4.6.37)\\ &= U(e^{j\theta})(\Sigma(\theta) + U^*(e^{j\theta})\Delta(e^{j\theta})V(e^{j\theta})w(e^{j\theta}))V^*(e^{j\theta}). && (4.6.38) \end{aligned}$$

Thus, the second term within parentheses corresponds to (4.6.32), i.e.,

$$\Delta_s(e^{j\theta}) \triangleq U^*(e^{j\theta})\Delta(e^{j\theta})V(e^{j\theta}). \tag{4.6.39}$$

It follows that the most general case (4.6.38) is equivalent to the case when $\Delta_s(e^{j\theta})$ is not diagonal. Then, $\Delta_s(e^{j\theta})$ can be written as the sum of diagonal and off-diagonal part, namely, $\Delta_s(e^{j\theta}) = \Delta_{s,diag}(e^{j\theta}) + \Delta_{s,off-diag}(e^{j\theta})$. The effect of off-diagonal elements of $\Delta_s(e^{j\theta})$ may be represented by an additive noise, which will be explained next. The equivalent channel frequency response, obtained by precoding the transmitted signal by $V(e^{j\theta})$ and by shaping the received signal by $U^*(e^{j\theta})$ (see Fig. 4.6.15), is given by

$$\begin{aligned} &H_{eq}(e^{j\theta}) = U^*(e^{j\theta})H(e^{j\theta})V(e^{j\theta}) && (4.6.40)\\ &= \Sigma(\theta) + \Delta_{s,diag}(e^{j\theta})w(e^{j\theta}) + \Delta_{s,off-diag}(e^{j\theta})w(e^{j\theta}), && (4.6.41) \end{aligned}$$

with the output of the i^{th} channel, represented in the frequency domain by

$$y_i(e^{j\theta}) = (\sigma_i(\theta) + \delta_{s,ii}(e^{j\theta})w(e^{j\theta}))x_i(e^{j\theta}) + \sum_{k\neq i} \delta_{s,ik}(e^{j\theta})w(e^{j\theta})x_k(e^{j\theta}) + n_i(e^{j\theta}). \quad (4.6.42)$$

Here, $\delta_{s,ij}(e^{j\theta})$, $1 \leq i \leq p$, $1 \leq j \leq m$, are the entries of $\Delta_s(e^{j\theta})$. The off-diagonal elements $\delta_{s,ij}(e^{j\theta})$, $i \neq j$, contribute to the equivalent noise

$$n_{eq,i}(e^{j\theta}) = n_i(e^{j\theta}) + \sum_{k\neq i} \delta_{s,ik}(e^{j\theta})w(e^{j\theta})x_k(e^{j\theta}). \quad (4.6.43)$$

Thus, one can use n codes, each tuned to a corresponding $\sigma_i(\theta) - |w(e^{j\theta})|$ channel, and has to include the additional noise term created by off-diagonal elements of $\Delta_s(e^{j\theta})$. This means that the operational capacity in the most general case will be less then the value found in Theorem 4.3.1.

4.6.2 Achievable Rates for Uncertain Noise

When the noise uncertainty is described by the set A_3, the existence of the saddle point is proved in Appendix F. Thus, as in the SISO case, it suggests that the notion of the worst case noise can be employed.

To prove the achievability of an information capacity computed in Theorem 4.4.1, in addition, it is assumed that the receiver has the knowledge of the noise PSD matrix $W_{\mathbf{n}}(\theta)$, while the transmitter does not have to know it. If the transmitter uses a "Gaussian codebooks" (the definition of "Gaussian codebooks" is given further in the text) and the PSD $W_{\mathbf{x}}^o(\theta)$, obtained in Theorem 4.4.1, the achievable code rate is $J(W_{\mathbf{x}}^o(\theta), W_{\mathbf{n}}(\theta))$, given by (4.2.7), for any choice of the PSD noise matrix $W_{\mathbf{n}}(\theta)$ (guaranteed by [16]). From the saddle point property, which is proved in Appendix F, $J(W_{\mathbf{x}}^o(\theta), W_{\mathbf{n}}^o(\theta)) \leq J(W_{\mathbf{x}}^o(\theta), W_{\mathbf{n}}(\theta))$, implying that the achievable transmission rate is lower bounded by $J(W_{\mathbf{x}}^o(\theta), W_{\mathbf{n}}^o(\theta))$. This transmission rate is achievable when the worst case noise affects the transmission (having PSD $W_{\mathbf{n}}^o(\theta)$). Hence, the compound capacity is given by $J(W_{\mathbf{x}}^o(\theta), W_{\mathbf{n}}^o(\theta))$ subject to the assumption that the receiver knows $W_{\mathbf{n}}(\theta)$, but the transmitter does not have to know it. The transmitter has to have knowledge of $J(W_{\mathbf{x}}^o(\theta), W_{\mathbf{n}}^o(\theta))$ in order to choose a transmission rate. Therefore, "Gaussian codebooks" provide the robustness when the transmitter does not know the noise. In the rest of the section, the computation of the probability of the decoding error is given when the maximum likelihood decoding is used, to illustrate the point made. Here, the intention is not to give a rigorous derivation of the probability of the decoding error, but to show that an upper bound on the probability of the decoding error decreases as the function of the mutual information rate. This will imply that the worst case channel has the largest probability of error, meaning that for all other noises from the uncertainty set, the performance of "Gaussian codebooks" has to be better. The upper bound is obtained by using a similar procedure given by Gallager ([35], pages 379-381) and Brandenburg and Wyner [16].

Assume that a "Gaussian codebook" is used. Thus, the codewords $\{\mathbf{x}_i\}_{i=1}^{M}$ of length N, where $\mathbf{x}_i = [x_i(0), ..., x_i(N-1)]$, $x_i(k) \in \mathbf{R}^m$, $k = 0, ..., (N-1)$, $i = 1, ..., M$, are the sample paths of M independent discrete-time Gaussian processes, having the same PSD matrices $W_{\mathbf{x}}^o(\theta)$, the one that is optimal with respect to the worst case noise PSD, $W_{\mathbf{n}}^o(\theta)$ [47]. It is assumed that stochastic processes, which generate the codewords, are zero-mean and ergodic.

Next, the equivalent representation of the channel, given by (4.2.1), is introduced. Based on the Gallager's work [35], the equivalent representation of (4.2.1) is given by a convolution type operator T_N, having the frequency response $H_1(e^{j\theta}) = W_{\mathbf{n}}^{-1/2}(\theta)H(e^{j\theta})$, and additive white noise having PSD which is an identity matrix (see also [16]). Here, $W_{\mathbf{n}}^{-1/2}(\theta)$ is obtained by factorizing $W_{\mathbf{n}}^{-1}(\theta)$.

Further, we compute the probability of decoding error for $M = 2$ codewords, assuming that the maximum likelihood decoding is used. Denote by $\mathbf{u}_1 = [u_1(0), ..., u_1(N-1)]$ and $\mathbf{u}_2 = [u_2(0), ..., u_2(N-1)]$, $u_i(k) \in \mathbf{R}^p$, $i = 1, 2$, $k = 0, ..., N-1$, the outputs of the filter T_N for two input codewords, $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Thus, $\mathbf{u}_i = T_N\mathbf{x}_i$, $i = 1, 2$. By adding the equivalent Gaussian noise $\mathbf{z} = [z(0), ..., z(N-1)]$, $z(k) \in \mathbf{R}^p$, $k = 0, ..., N-1$, whose PSD matrix is identity matrix, the equivalent representation of the channel is obtained

$$\mathbf{y}_i = T_N\mathbf{x}_i + \mathbf{z}, \; i = 1, 2. \tag{4.6.44}$$

Also, we define a norm for a codeword or the channel output as follows

$$|||\mathbf{x}||| \triangleq \sqrt{\sum_{k=0}^{N-1} \|x(k)\|^2}, \tag{4.6.45}$$

where $\|\cdot\|$ is Euclidean norm. Then, the probability of the decoding error when a maximum likelihood decoding is applied, subject to the assumption that $\mathbf{x}_1$ is sent, is given by

$$P_{e,1} \triangleq \Pr\{|||\mathbf{y}_1 - T_N\mathbf{x}_2||| < |||\mathbf{y}_1 - T_N\mathbf{x}_1|||\}. \tag{4.6.46}$$

After some manipulations of (4.6.46), we get [16]

$$P_{e,1} = \Phi\left(\frac{1}{2}|||T_N(\mathbf{x}_2 - \mathbf{x}_1)|||\right) = \Phi\left(\frac{1}{2}|||\mathbf{u}_2 - \mathbf{u}_1|||\right), \tag{4.6.47}$$

where $\Phi(u) = \int_u^{+\infty} \exp[-\frac{x^2}{2}]dx$. Next, we are going to use the fact that the codewords $\mathbf{x_1}$ and $\mathbf{x_2}$ are random and independent. First, observe that the argument in (4.6.47) has three terms, $\sum_{k=0}^{N-1} \|u_1(k)\|^2$, $\sum_{k=0}^{N-1} \|u_2(k)\|^2$ and $2\sum_{k=0}^{N-1} u_1^*(k)u_2(k)$. We multiply and divide the argument of Φ by N. For N large enough, due to the ergodic assumption and because $\mathbf{u}_1$ and $\mathbf{u}_2$ are independent, term

$$\lim_{N\to+\infty} \frac{2}{N} \sum_{k=0}^{N-1} u_1^*(k)u_2(k) = \frac{1}{2\pi}\text{Trace} \int_0^{2\pi} H_1(e^{j\theta})W_{\mathbf{x}_1\mathbf{x}_2}(\theta)H_1^*(e^{j\theta})d\theta \tag{4.6.48}$$

vanishes. Here, $W_{\mathbf{x}_1\mathbf{x}_2}(\theta)$ represents the cross-PSD of $\mathbf{x}_1$ and $\mathbf{x}_2$, which is equal to zero, because $\mathbf{x}_1$ and $\mathbf{x}_2$ are obtained from independent and zero-mean Gaussian processes. Two other terms $\frac{1}{N}\sum_{k=0}^{N-1} \|u_1(k)\|^2$ and $\frac{1}{N}\sum_{k=0}^{N-1} \|u_2(k)\|^2$ are asymptotically equal, because two independent processes, whose sample paths represent two codewords $\mathbf{x}_1$ and $\mathbf{x}_2$, have the same PSD matrix $W_{\mathbf{x}}^o(\theta)$. In other words

$$\begin{aligned}
\lim_{N\to+\infty} \frac{1}{N} \sum_{k=0}^{N-1} \|u_1(k)\|^2 &= \lim_{N\to+\infty} \frac{1}{N} \sum_{k=0}^{N-1} \|u_2(k)\|^2 &\quad (4.6.49)\\
&= \lim_{N\to+\infty} \frac{1}{N} \sum_{k=0}^{N-1} \|u(k)\|^2 &\quad (4.6.50)\\
&= \frac{1}{2\pi}\text{Trace} \int_0^{2\pi} H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta})d\theta. &\quad (4.6.51)
\end{aligned}$$

Therefore, for very large N, the probability of the decoding error is transformed into

$$\begin{aligned}
P_{e,1} &= \Phi\left(\frac{1}{2}\sqrt{N\frac{1}{N}\sum_{k=0}^{N-1} \|u_1(k) - u_2(k)\|^2}\right) &\quad (4.6.52)\\
&= \Phi\left(\sqrt{\frac{N}{4\pi}\text{Trace} \int_0^{2\pi} H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta})d\theta}\right), &\quad (4.6.53)
\end{aligned}$$

or

$$P_{e,1} = \Phi\left(\sqrt{\frac{N}{2}SNR}\right), \tag{4.6.54}$$

where

$$SNR \triangleq \frac{1}{2\pi}\text{Trace} \int_0^{2\pi} H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta})d\theta. \tag{4.6.55}$$

Here, the goal is to obtain an upper bound on the probability of the decoding error as a function of the mutual information. Observe that $H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta})$ is Hermitian and therefore diagonalizable. Thus, $H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta}) = V(e^{j\theta})\Lambda(\theta)V^*(e^{j\theta})$, where $\Lambda(\theta)$ is a diagonal matrix, containing the eigenvalues of $H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta})$, $\lambda_i(\theta)$, i=1,...,n. It follows that

$$\begin{aligned} SNR &= \frac{1}{2\pi}\text{Trace}\int_0^{2\pi} H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta})d\theta && (4.6.56)\\ &= \frac{1}{2\pi}\int_0^{2\pi}\sum_{i=1}^{n}\lambda_i(\theta)d\theta && (4.6.57)\\ &\geq \frac{1}{2\pi}\int_0^{2\pi}\sum_{i=1}^{n}\log(1+\lambda_i(\theta))d\theta && (4.6.58)\\ &= \frac{1}{2\pi}\int_0^{2\pi}\log\det(I_p + H_1(e^{j\theta})W_{\mathbf{x}}^o(\theta)H_1^*(e^{j\theta}))d\theta. && (4.6.59)\end{aligned}$$

The inequality follows from $\log(1+u) \leq u$ for base e. Consequently, having in mind that $\Phi(u)$ is a decreasing function of the argument, the upper bound on the probability of error is given by

$$P_{e,1} = \Phi\left(\sqrt{\frac{N}{2}SNR}\right) \leq \Phi\left(\sqrt{NJ(W_{\mathbf{x}}^o, W_{\mathbf{n}})}\right). \tag{4.6.60}$$

If the Chernoff bound is applied, for $M \geq 2$, the upper bound on the probability of the decoding error is given by

$$P_{e,1} \leq (M-1)e^{-\frac{N}{2}J(W_{\mathbf{x}}^o, W_{\mathbf{n}})}. \tag{4.6.61}$$

Thus, from (4.6.60), it is clear that the upper bound on the probability of decoding error, for a specific channel and noise, is obtained by substituting, the specific channel and noise, and the optimal transmitter's PSD matrix. Because $\Phi(u)$ is a decreasing function, this implies that the worst case noise, characterized by $W_{\mathbf{n}}^o$ derived in Theorem 4.4.1, achieves the largest upper bound on the decoding error, out of all noises from the uncertainty set. The same conclusion is obtained by Diggavi and Cover [28], where the authors considered a similar problem (in their work, they considered covariance matrices of the transmitter and noise) with the same assumptions on the receiver and transmitter knowledge, but where Mahalanobis distance decoding is used.

As for the future work, it remains to be seen if the achievability of the information capacity derived in Theorem 4.4.1 can be proved for the case when both, the transmitter and the receiver, do not know the noise PSD. One approach would be to generalize the derivation of Root and Varaiya [63] to MIMO case. To see this, introduce the following matrix spaces: 1) the linear space of $p \times m$ matrices, $l_{p,m}^1 \triangleq \{h_n : \mathbf{Z} \to \mathbf{R}^{p\times m}, \sum_{n=-\infty}^{+\infty} |[h_n]_{ij}| < +\infty, 1 \leq$

$i \leq p, 1 \leq j \leq m\}$, where $[h_n]_{ij}$ is an entry of the matrix h_n, 2) the linear space of square summable sequences of $p \times m$ matrices, $l^2_{p,m} \triangleq \{h_n : \mathbf{Z} \to \mathbf{R}^{p\times m}, \sum_{n=-\infty}^{+\infty} \|h_n\|^2 < +\infty\}$, where $\|h_n\|^2 \triangleq Trace(h_n h_n^T)$, and T denotes a matrix transpose. Then, the coding theorem might be obtained by paralleling the approach given in Section 3.5. Namely, by using these spaces, it is possible to redefine the conditions for conditional compactness of the set A_3, found in Section 3.5. It should be observed that condition 3) is not needed, because we deal with a discrete-time case where the integration in the frequency domain is between 0 and 2π.

4.7 Summary

This chapter deals with the information bounds of compound MIMO Gaussian channels with memory for two different types of uncertainty. The uncertainty of the channel matrix frequency response is described through a subset of H^∞ space, while the uncertainty of the noise PSD is represented through a subset of L_1 function space. For both problems, the explicit information channel capacity formulas are derived, as well as the optimal PSD matrices of the transmitted signals. For the case when the noise is uncertain, the optimal PSD of the noise/jammer is derived, and it is shown that it is proportional to the optimal PSD of the transmitted signal. For the case when the channel frequency response matrix is uncertain, the channel capacity formula and the optimal PSD of the transmitted signal depend on the size of the uncertainty set. From these two formulas, it can be concluded that the transmission over the strongest mode of the nominal channel frequency response matrix is the optimal strategy when the size of the uncertainty set increases. At one point, the size of the uncertainty set will dominate all weaker modes of the channel matrix, and, hence, the transmitter will send information only over the strongest mode.

Chapter 5

Conclusion

5.1 Synopsis

This thesis illustrates the effect of the channel and noise uncertainties on the performance of communication systems from an information theoretic point of view, when uncertainties are modeled by subsets of H^{∞} and L_1 normed spaces, in frequency domain. SISO and MIMO compound Gaussian channels are considered.

H^{∞} uncertainty descriptions are used to represent the uncertainties in the channel frequency response and the PSD of the noise, when there exists some a priori knowledge regarding the channel and the noise. L_1 uncertainty models are employed to represent the uncertainty of the PSD of the noise, when only the knowledge regarding the average power constraint on the noise is available.

In the case of SISO communication channels, when the channel and the noise uncertainties are represented by subsets of H^{∞} space, it is shown that the channel capacities and the optimal PSD's of the transmitted signals depend on the nominal channel and noise models, and on the sizes of uncertainty sets. The optimal PSD's of the transmitted signals are obtained in the form of the water-filling equations. The H^{∞} modeling technique is also applied to derive the capacities of ergodic wireless fading channels. When the uncertainty in the PSD of the noise is represented by a subset of L_1 space, the channel capacity problem is formulated by using a game theoretic framework. For this infinite dimensional optimization problem, it is shown that a saddle point exists. It is proved that the capacity achieving PSD of the transmitted signal is proportional to the optimal PSD of the noise (jammer).

In the case of MIMO communication channels, two problems are solved. When the channel uncertainty is represented by a subset of H^{∞} space, it is shown that the channel capacity and the optimal PSD matrix of the transmitted signal depend on the size of the uncertainty set. The channel capacity formula suggests that the one-mode transmission is optimal for a large uncertainty set. When the noise uncertainty is represented by a subset of L_1 space, the optimal PSD matrix of the noise (jammer) is proportional to the optimal PSD matrix of the transmitted signal. In general, the two optimal PSD matrices are not diagonal, which implies that there should exist a cooperation between the transmitters of multiple-input systems. Moreover, as in the case of SISO communication systems, it is shown that a saddle point exists for L_1 noise uncertainty model.

5.2 Directions for Future Research

The new approach for modeling, analysis, and design of uncertain communication systems, which is introduced in this thesis, raises a few issues related to

1. Encoding and decoding for uncertain channels;
2. Generalization of proposed models;
3. Relations between frequency and time domain models;
4. Applications of the models in network information theory.

This thesis introduces the description of a compound channel using a nominal and perturb part. From the point of view of encoding and decoding, it would be interesting to design such an encoder and decoder that would operate on the compound channel, defined by the nominal channel and the size of the uncertainty set. Further, this concept could be extended to a corresponding adaptive scheme, where the nominal model and the size of the uncertainty set could be updated as more and more data are collected by the receiver. This would be an extension of universal encoders and decoders. It is predicted that this scheme would require less information rate for a feedback channel than existing adaptive schemes, that convey the CSI to the transmitter continuously in time. A new robust adaptive scheme would feed back the CSI consisting of the nominal model and the size of the uncertainty set, only when it is found that the current description is too conservative or when the size of the uncertainty set grows.

Another direction for further research is a generalization of uncertainty models proposed in the thesis. It was shown that H^{∞} models can be used for uncertainty modeling of ergodic and non-ergodic wireless fading channel. Here, it is assumed that the nominal frequency response is a random variable exponentially distributed. The ergodic capacity is found by taking the expectation over the exponentially distributed nominal part. The H^∞ description is used to model a short term uncertainty within the coherence time T_c. The long term uncertainty can be modeled by introducing uncertainty in the distribution of the nominal frequency response. This uncertainty might be introduced as a ball of distributions. The center of the ball can be defined by the nominal distribution of the nominal frequency response, and the radius would be determined by the relative entropy constraint. For this kind of a model the channel capacity can be computed, and corresponding universal encoders and decoders can be built.

In this thesis, the emphasis is on the frequency domain uncertainty models. It would be interesting if the connection between the frequency domain models and the state space models could be established. [73].

A natural extension of the proposed models is to consider their application in network information theory.

Appendix A

Robust Capacity - Additive Channel Uncertainty

Proof of Theorem 3.2.1. The condition for boundedness and integrability of

$$\frac{(|H_{nom}(f)| + |W_1(f)|)^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)| - |W_2(f)|)^2} \tag{A.0.1}$$

comes from Lemma 8.5.7, [35] (page 423). Namely, a sufficient condition for the existence of the mutual information in the frequency domain is the integrability of $\frac{|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}$ for all $\tilde{H}(f) \in A_2$ and all $\tilde{W}(f) \in A_3$. This is satisfied when $\frac{(|H_{nom}(f)|+|W_1(f)|)^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)|-|W_2(f)|)^2}$ is integrable. Thus,

$$\int_{-\infty}^{+\infty} \frac{|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2} df \leq \int_{-\infty}^{+\infty} \frac{(|H_{nom}(f)| + |W_1(f)|)^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)| - |W_2(f)|)^2} df. \tag{A.0.2}$$

Next,

$$\inf_{\tilde{H} \in A_2} \int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)|H_{nom}(f) + \Delta_1(f)W_1(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}\right) df \tag{A.0.3}$$

is computed. From the fact that $|\Delta_1(f)| \leq 1$, and because log is a monotonically increasing function, it is clear that (A.0.3) is equivalent to

$$\int_{-\infty}^{+\infty} \log\left(1 + \frac{S_{\mathbf{x}}(f)(|H_{nom}(f)| - |W_1(f)|)^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}\right) df, \tag{A.0.4}$$

when $\Delta_1(f) = \Delta_1^o(f) = \exp[-j\arg(W_1(f)) + j\arg(H_{nom}(f)) + j\pi]$. The infimum over the set A_3 can be found in a similar way.

The problem of finding supremum is exactly the same as in the classical case [35]. This leads to the modified water-filling equation

$$S_{\mathbf{x}}^o(f) + \frac{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}{(|H_{nom}(f)| - |W_1(f)|)^2} = \nu^o, \tag{A.0.5}$$

where ν^o is a positive constant equal to $-1/2\lambda$, such that

$$\int_{\mathcal{S}} S^o_{\mathbf{x}}(f) = P, \tag{A.0.6}$$

where $\mathcal{S} \triangleq \{f : \nu^o - \frac{S_{\mathbf{n}}(f)(|W_{nom}(f)|+|W_2(f)|)^2}{(|H_{nom}(f)|-|W_1(f)|)^2} > 0\}$.

Next, it will be shown that the capacity formula for the corresponding discrete-time channel is just the special case of the previously derived formula for the channels with memory. A channel with a memory can be seen as a group of infinite number of channels, each assigned to one frequency. Thus, we have to find the channel capacity that corresponds to just one frequency. First, assume that the signal is multiplied by a window in the frequency domain $G(f) = 1$, $f \in [-W, W]$, and $G(f) = 0$ otherwise. Then the channel capacity per Hz is given by

$$C = \frac{1}{2 \times 2W} \int_{-W}^{+W} \log\Big(\frac{\nu^o(|H_{nom}(f)| - |W_1(f)|)^2}{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}\Big) df \tag{A.0.7}$$

$$\frac{1}{2W} \int_{-W}^{+W} \Big(\nu^o - \frac{S_{\mathbf{n}}(f)(|W_{nom}(f)| + |W_2(f)|)^2}{(|H_{nom}(f)| - |W_1(f)|)^2}\Big) df = P. \tag{A.0.8}$$

When W tends to zero

$$C = \frac{1}{2 \times 2W} 2W \log\Big(\frac{\nu^o(|H_{nom}(0)| - |W_1(0)|)^2}{S_{\mathbf{n}}(0)(|W_{nom}(0)| + |W_2(0)|)^2}\Big) \tag{A.0.9}$$

$$S^o_{\mathbf{x}}(0) = \frac{1}{2W} 2W\Big(\nu^o - \frac{S_{\mathbf{n}}(0)(|W_{nom}(0)| + |W_2(0)|)^2}{(|H_{nom}(0)| - |W_1(0)|)^2}\Big), \tag{A.0.10}$$

giving

$$C = \frac{1}{2}\log\Big(1 + \frac{S^o_{\mathbf{x}}(0)(|H_{nom}(0)| - |W_1(0)|)^2}{S_{\mathbf{n}}(0)(|W_{nom}(0)| + |W_2(0)|)^2}\Big). \tag{A.0.11}$$

Further, denote the autocorrelation function of $\mathbf{x}$ by $R_{\mathbf{x}}(\tau)$. For $\tau = 0$, $R_{\mathbf{x}}(0) = E[x^2(t)]$ is the variance of a random variable $x(t)$. Also, for band limited case

$$R_{\mathbf{x}}(0) = \int_{-W}^{+W} S^o_{\mathbf{x}}(f) df. \tag{A.0.12}$$

Then, when W tends to zero, the variance per Hz is given by

$$\frac{1}{2W} R_{\mathbf{x}}(0) = S^o_{\mathbf{x}}(0) = \frac{1}{2W} E[x^2(t)]. \tag{A.0.13}$$

By the same line of reasoning

$$\frac{1}{2W} R_{\mathbf{n}}(0) = S^o_{\mathbf{n}}(0) = \frac{1}{2W} E[n^2(t)]. \tag{A.0.14}$$

Appendix B

Capacity of Wireless Fading Channels

Proof of Theorem 3.3.1. The capacity is given by

$$C_E = \sup_{S_{\mathbf{x}} \in B_1} \inf_{\|\Delta\|_\infty \leq 1} E \, \frac{1}{2} \int_{-W/2}^{W/2} \log\left(1 + \frac{S_{\mathbf{x}}(f)|H_{nom}(f)(1+\Delta(f)W_1(f))|^2}{2N_0}\right) df. \quad \text{(B.0.1)}$$

The infimum can be switched with the expectation, because it is taken over $\Delta(f)$, which is not random. The infimum with respect to the multiplicative uncertainty is resolved similarly to the infimum found in Appendix A which gives

$$\begin{aligned} &\inf_{\tilde{H} \in B_2} \int_{-W/2}^{W/2} \log\left(1 + \frac{S_{\mathbf{x}}|H_{nom}(f)(1+\Delta(f)W_1(f))|^2}{2N_0}\right) df \\ &= \int_{-W/2}^{W/2} \log\left(1 + \frac{S_{\mathbf{x}}(f)[|H_{nom}(f)|(1-|W_1(f)|)]^2}{2N_0}\right) df, \end{aligned} \quad \text{(B.0.2)}$$

for $\Delta^o(f) = \exp[-j \arg(W(f)) + j\pi]$.

Here, $\beta = |H_{nom}(f)|^2$ is the exponentially distributed random variable, with density $p(\beta) = \exp(-\beta)$, $\beta \geq 0$. By Fubini's theorem, we are able to exchange the expected value and integration with respect to frequency in (B.0.1). Then, (B.0.1) transforms into

$$\begin{aligned} C_E &= \sup_{S_{\mathbf{x}} \in B_1} \frac{1}{2} \int_{-W/2}^{W/2} E_\beta \, \log(1 + \beta\rho(f)) df && \text{(B.0.3)} \\ &= \sup_{S_{\mathbf{x}} \in B_1} \frac{1}{2} \int_{-W/2}^{W/2} -\exp(\rho^{-1}(f)) E_i(-\rho^{-1}(f)) df, && \text{(B.0.4)} \end{aligned}$$

where $\rho(f) = \frac{S_{\mathbf{x}}(f)(1-|W_1(f)|)^2}{2N_0}$, and $E_i(x) = \int_{-\infty}^{x} \frac{e^u}{u} du$ is the exponential integral function. Because, it is not possible to find the closed form solution in general, the expressions for low and high SNR will be given. By using the expressions found in [57], $-\exp(\rho^{-1}(f))E_i(-\rho^{-1}(f))$ can be approximated by $\rho(f)$, for $\rho(f) << 1$, and by $\log(1+\rho(f)) - \mathcal{C}$, for $\rho(f) >> 1$, where $\mathcal{C}$ is the Eüler constant.

1. (Low SNR, $\rho(f) << 1$). The ergodic capacity is given by

$$C_E = \sup_{S_{\mathbf{x}} \in B_1} \frac{1}{2} \int_{-W/2}^{W/2} \rho(f) df \quad \text{(B.0.5)}$$

$$= \sup_{S_{\mathbf{x}} \in B_1} \frac{1}{2} \int_{-W/2}^{W/2} \frac{S_{\mathbf{x}}(f)(1 - |W_1(f)|)^2}{2N_0} df. \quad \text{(B.0.6)}$$

The upper bound on the integral can be found by using the Cauchy-Schwartz inequality as follows

$$\left(\int_{-W/2}^{W/2} \frac{S_{\mathbf{x}}(f)(1 - |W_1(f)|)^2}{2N_0} df \right)^2 \leq \int_{-W/2}^{W/2} S_{\mathbf{x}}^2(f) df \int_{-W/2}^{W/2} \left[\frac{(1 - |W_1(f)|)^2}{2N_0} \right]^2 df, \quad \text{(B.0.7)}$$

where the equality over $S_{\mathbf{x}}(f) \in B_1$ is achieved for

$$S_{\mathbf{x}}^o(f) = k \, \frac{(1 - |W_1(f)|)^2}{2N_0}. \quad \text{(B.0.8)}$$

Here, k is a positive constant. The optimal PSD $S_{\mathbf{x}}^o(f)$ is computed by substituting (B.0.8) into the power constraint $\int_{-W/2}^{W/2} S_{\mathbf{x}}^o(f) df = 2P_{\mathbf{x}}$ and solving for k.

2. (High SNR, $\rho(f) >> 1$). The ergodic capacity for this case is given by the formula

$$C_E = \sup_{S_{\mathbf{x}} \in B_1} \frac{1}{2} \int_{-W/2}^{W/2} \left[\log \left(1 + \frac{S_{\mathbf{x}}(f)(1 - |W_1(f)|)^2}{2N_0} \right) - C \right] df. \quad \text{(B.0.9)}$$

Thus, this is a standard formula for the channel capacity except that the frequency response of the communication channel is replaced by the part $(1 - |W_1(f)|)^2$ that describes the uncertainty. The optimal PSD is in the form of water-filling formula such that $S_{\mathbf{x}}^o(f)$ water-fills $\frac{2N_0}{(1-|W_1(f)|)^2}$.

Proof of Theorem 3.3.3. The capacity is given by formula

$$E \sup_{S_{\mathbf{x}} \in B_1} \inf_{\tilde{H} \in B_2} \frac{1}{2} \int_{-W/2}^{W/2} \log \left(1 + \frac{S_{\mathbf{x}}(f) |\tilde{H}(f)|^2}{2N_0} \right) df. \quad \text{(B.0.10)}$$

The infimum is given by (B.0.2). Further, the optimal transmitted PSD is given by a standard water-filling formula

$$S_{\mathbf{x}}^o(f) = \nu^o - \frac{2N_0}{[|H_{nom}(f)|(1 - |W_1(f)|)]^2} \quad \text{P. a.s.,} \quad \text{(B.0.11)}$$

where $\nu^o = \frac{1}{2\lambda^o} > 0$. The constant ν^o is computed from the constraint equation

$$\int_{\mathcal{S}} S^o_{\mathbf{x}}(f)df = 2P_{\mathbf{x}} \text{ P. a.s.}, \tag{B.0.12}$$

where $\mathcal{S} = \{f : \nu^o - \frac{2N_0}{[|H_{nom}(f)|(1-|W_1(f)|)]^2} > 0 \text{ P. a.s.}\}$. The capacity formula is obtained by setting $|W_1(f)| = |(1+\Delta\alpha)e^{-j2\pi f(1+\Delta\tau)} - 1|$.

Appendix C

Capacity of Gaussian Channels Subject to Jamming

Proof of Theorem 3.4.1. Without loss of generality, the proof of the theorem will be given for $W_1(f) = 0$, $W_{nom}(f) = 1$. The conditions of integrability and boundedness of $\frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)}$ are sufficient conditions for the pay-off functional

$$J(S_{\mathbf{x}}, S_{\mathbf{n}}) = \int_{-\infty}^{+\infty} \log\Big(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)}\Big) df \tag{C.0.1}$$

to exist [35]. To find the infimum, the following Lagrangian is used

$$J_1(S_{\mathbf{x}}, S_{\mathbf{n}}, \lambda_1) = \frac{1}{2}\int_{-\infty}^{+\infty} \log\Big(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)}\Big) df + \lambda_1\Big(\int_{-\infty}^{+\infty} S_{\mathbf{n}}(f) df - P_{\mathbf{n}}\Big). \tag{C.0.2}$$

For a given $S_{\mathbf{x}}(f) \in A_1$, $\inf_{S_{\mathbf{n}} \in A_4} J(S_{\mathbf{x}}, S_{\mathbf{n}})$ is found by $\sup_{\lambda_1 \geq 0} \inf_{S_{\mathbf{n}} \in A_4} J_1(S_{\mathbf{x}}, S_{\mathbf{n}}, \lambda_1)$. The variation with respect to $S_{\mathbf{n}}(f)$ gives the quadratic equation

$$S_{\mathbf{n}}^{o2}(S_{\mathbf{x}}(f)) + S_{\mathbf{n}}^{o}(S_{\mathbf{x}}(f)) S_{\mathbf{x}}(f)|\tilde{H}(f)|^2 - \frac{1}{2\lambda_1} S_{\mathbf{x}}(f)|\tilde{H}(f)|^2 = 0. \tag{C.0.3}$$

The positive solution of the quadratic equation (C.0.3)

$$S_{\mathbf{n}}^{o}(S_{\mathbf{x}}(f)) = \frac{1}{2}\Big(- S_{\mathbf{x}}(f)|\tilde{H}(f)|^2 \pm \sqrt{S_{\mathbf{x}}^2(f)|\tilde{H}(f)|^4 + 4\frac{1}{2\lambda_1} S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}\Big) \tag{C.0.4}$$

is replaced in the Lagrangian $J_1(S_{\mathbf{x}}, S_{\mathbf{n}}, \lambda_1)$. Thus

$$\sup_{S_{\mathbf{x}} \in A_1} \inf_{S_{\mathbf{n}} \in A_4} J(S_{\mathbf{x}}, S_{\mathbf{n}}) = \sup_{S_{\mathbf{x}} \in A_1} \sup_{\lambda_1 \geq 0} J_1(S_{\mathbf{x}}, S_{\mathbf{n}}^o, \lambda_1). \tag{C.0.5}$$

Further, the supremum with respect to $S_{\mathbf{x}}$ is resolved by defining the Lagrangian

$$\begin{aligned} J_2(S_{\mathbf{x}}, S_{\mathbf{n}}^o, \lambda_1, \lambda_2) &= \frac{1}{2}\int_{-\infty}^{+\infty} \log\Big(1 + \frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}^o(S_{\mathbf{x}}(f))}\Big) df \\ +\lambda_1\Big(\int_{-\infty}^{+\infty} S_{\mathbf{n}}^o(S_{\mathbf{x}}(f)) df - P_{\mathbf{n}}\Big) &- \lambda_2\Big(\int_{-\infty}^{+\infty} S_{\mathbf{x}}(f) df - P_{\mathbf{x}}\Big). \end{aligned} \tag{C.0.6}$$

Thus,

$$\sup_{S_{\mathbf{x}}\in A_1}\inf_{S_{\mathbf{n}}\in A_4} J(S_{\mathbf{x}},S_{\mathbf{n}}) = \sup_{S_{\mathbf{x}}\in A_1}\sup_{\lambda_1\geq 0} J_1(S_{\mathbf{x}},S^o_{\mathbf{n}},\lambda_1) = \inf_{\lambda_2\geq 0}\sup_{\lambda_1\geq 0}\sup_{S_{\mathbf{x}}\in A_1} J_2(S_{\mathbf{x}},S^o_{\mathbf{n}},\lambda_1,\lambda_2). \quad \text{(C.0.7)}$$

The optimal transmitter PSD $S^o_{\mathbf{x}}(f)$ and the constraint equations are found by application of Kuhn-Tucker conditions. By varying $J_2(S_{\mathbf{x}},S^o_{\mathbf{n}},\lambda_1,\lambda_2)$ with respect to $S^o_{\mathbf{x}}(f)$, we get

$$S^o_{\mathbf{x}}(f) = \frac{\lambda^o_1(|H_{nom}(f)|-|W(f)|)^2}{2\lambda^o_2(\lambda^o_1(|H_{nom}(f)|-|W(f)|)^2+\lambda^o_2)}. \quad \text{(C.0.8)}$$

Then, $S^o_{\mathbf{x}}(f)$ is substituted in (C.0.3) to find the optimal PSD of the noise $S^o_{\mathbf{n}}(f)$.

Minimax capacity. In order to show the existence of a saddle point, we need to prove that

$$\sup_{S_{\mathbf{x}}\in A_1}\inf_{S_{\mathbf{n}}\in A_4} J(S_{\mathbf{x}},S_{\mathbf{n}}) = \inf_{S_{\mathbf{n}}\in A_4}\sup_{S_{\mathbf{x}}\in A_1} J(S_{\mathbf{x}},S_{\mathbf{n}}). \quad \text{(C.0.9)}$$

The easiest way to prove this is by assuming that a saddle point exists and by showing that the optimal strategies in this case are equal to the optimal strategies derived in the case of maximin capacity. Next, define a Lagrangian

$$J_1(S_{\mathbf{x}},S_{\mathbf{n}},\lambda_1,\lambda_2) = \frac{1}{2}\int_{-\infty}^{+\infty}\log\Big(1+\frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)}\Big)df + \lambda_1\Big(\int_{-\infty}^{+\infty}S_{\mathbf{n}}(f)df - P_{\mathbf{n}}\Big)$$
$$-\lambda_2\Big(\int_{-\infty}^{+\infty}S_{\mathbf{x}}(f)df - P_{\mathbf{x}}\Big), \quad \text{(C.0.10)}$$

where $\lambda_1,\ \lambda_2 > 0$. The Lagrangian is varied with respect to both $S_{\mathbf{x}}(f)$ and $S_{\mathbf{n}}(f)$, assuming the independence of two strategies. By varying $J_1(S_{\mathbf{x}},S_{\mathbf{n}},\lambda_1,\lambda_2)$ with respect to $S_{\mathbf{x}}(f)$, the water-filling formula

$$S^o_{\mathbf{x}}(f) + \frac{S_{\mathbf{n}}(f)}{(|H_{nom}(f)|-|W_1(f)|)^2} = \frac{1}{2\lambda^o_2}, \quad \text{(C.0.11)}$$

is obtained. By varying $J_1(S_{\mathbf{x}},S_{\mathbf{n}},\lambda_1,\lambda_2)$ with respect to $S_{\mathbf{n}}(f)$ gives (C.0.3). By combining (C.0.3) and (C.0.11), exactly the same strategies are obtained as ones found in Theorem 3.4.1.

Appendix D

Colored Noise Coding Theorem

In Section 3.5, it is claimed that the channel coding theorem and its converse, which were proved by Root and Varaiya [63] for a white Gaussian noise, may be adapted for the case of a colored noise.

The problem of computing the channel capacity and proving the channel coding theorem in the presence of a colored Gaussian noise for a known channel frequency response of a continuous time Gaussian channel and known PSD of the noise was first defined and considered by Gallager in [35]. At the same time, Root and Varaiya [63] proved the coding theorem for the class of continuous-time Gaussian channels but in the presence of a white noise. The idea is to combine these two approaches to prove the coding theorem for the class of Gaussian channels but in the presence of a colored Gaussian noise, or to be more precise, it is shown that Root's and Varaiya's coding theorem still applies when the two key lemmas (Lemma 8.5.1 and Lemma 8.5.6), given by Gallager [35], are rederived for the class of Gaussian channels $\mathcal{B}_i$, $i = 1, 2$, as defined in Section 3.5. Why these two lemmas are important for the proof will become clear from the discussion below. Further, $\mathcal{B}$ is used to represent any uncertainty set to ease the notation.

The Gallager's idea is sometimes called the whitening of the colored noise. It consists of dividing the frequency response of the channel $\tilde{H}(f)$ by PSD of the noise $S_{\mathbf{n}}(f)\tilde{W}^2(f)$, as it was done in (3.5.60). Then instead of working with a channel frequency response $\tilde{H}(f)$, one deals with an equivalent frequency response $K(f) = \tilde{H}(f)/\sqrt{S_{\mathbf{n}}(f)}\tilde{W}(f)$ and equivalent white noise with PSD equals to 1 W/Hz. But, Gallager assumed that $\tilde{H}(f)$ and $S_{\mathbf{n}}(f)$ are completely known. Thus, the whitening approach will be used to prove that the result from [63] can still be applied when some of the components of $K(f)$, $\tilde{H}(f)$ or $S_{\mathbf{n}}(f)\tilde{W}^2(f)$, or both are unknown, i.e., belong to an uncertainty set.

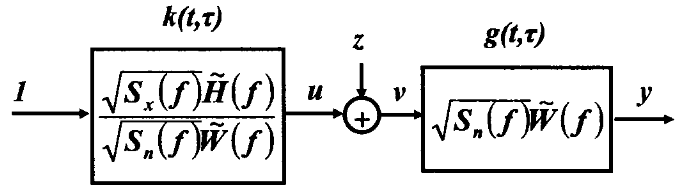

Figure D.0.1: Equivalent channel model

Consider a model of a communication channel shown in Fig. 3.1.1. By applying the transformation from [35], this model can be transformed into equivalent one, shown in Fig. D.0.1, in the sense that two models for the same input have the same output. There are some differences between two equivalent models. First, the equivalent noise $\mathbf{z} = \{z(t); -T/2 \leq t \leq T/2\}$ in Fig. D.0.1 is a white Gaussian noise with the PSD equal to 1; second, there exists additional filter with the impulse response $g(t)$ with the Fourier transform equal to the square root of the PSD of a colored Gaussian noise $G(f) = \sqrt{S_{\mathbf{n}}(f)\tilde{W}(f)}$; third, the impulse response of the equivalent channel $k(t)$ has a Fourier transform $K(f) = \sqrt{S_{\mathbf{x}}(f)}\tilde{H}(f)/\sqrt{S_{\mathbf{n}}(f)\tilde{W}(f)}$; fourth, the PSD of the transmitted signal is introduced in the channel frequency response, but this does not change anything with respect to mutual information (recall the formula for the mutual information in the frequency domain). If one could, for a moment, neglect the additional filter $g(t)$, and with the assumption that the impulse response of the equivalent channel $k(t)$ satisfies compactness conditions (1),2), and 3)) given in Section 3.5 (which implies that the channel capacity would depend on the equivalent frequency response $K(f)$ with the equivalent additive white noise $\mathbf{z}$), the coding theorem derived in [63] would be applicable for a colored noise also. In other words, the communication problem with a colored Gaussian noise is reduced to a problem with a white Gaussian noise. Then the operational capacity of the class of Gaussian channels is given by

$$C = \sup_{S_{\mathbf{x}}(f)\in A_1} \inf_{K(f)\in\mathcal{B}} \frac{1}{2}\int \log\left(1 + \frac{S_{\mathbf{x}}(f)|K(f)|^2}{N}\right)df, \tag{D.0.1}$$

where $K(f)$ is the frequency response of an equivalent channel with impulse response $k(t)$ (follows from Theorem 6, 7, and 8 [63]). N is a constant representing the PSD of a white Gaussian noise, which in our case is equal to 1. Thus, to prove the applicability of approach found in [63], it is enough to show that the filter $g(t)$ does not destroy any information about $\mathbf{x}$ in $\mathbf{v}$ if the equivalent frequency response belongs to an uncertainty set $\mathcal{B}$.

Before presenting the main result, the Theorem 8.4.1 from [35] (page 392) will be cited because it is extensively used in the derivation. The theorem is a consequence of some functional analysis results on linear operators with symmetric kernels (see for instance [46]).

Theorem D.0.1 *(Theorem 8.4.1) Let $h(t,\tau)$ be nonzero, and let $\int\int h^2(t,\tau)dtd\tau < +\infty$ be satisfied. Then there exists a sequence (infinite or finite) of decreasing positive numbers, called eigenvalues, $\lambda_1 \geq ... \geq \lambda_i \geq ... > 0$, and in one to one correspondence with these numbers, there exist two sets of orthonormal functions $\phi_i(\tau)$ and $\theta_i(t)$, called input eigenfunctions and output eigenfunctions respectively; these numbers and functions have the following properties*

1.

$$\int R(\tau_1,\tau_2)\phi_i(\tau_2)d\tau_2 = \lambda_i\phi_i(\tau_1) \tag{D.0.2}$$

$$R(\tau_1,\tau_2) = \int h(t,\tau_1)h(t,\tau_2)dt \tag{D.0.3}$$

2.

$$\sqrt{\lambda_i}\theta_i(t) = \int h(t,\tau)\phi_i(\tau)d\tau \tag{D.0.4}$$

$$\sqrt{\lambda_i}\phi_i(t) = \int h(t,\tau)\theta_i(\tau)dt \tag{D.0.5}$$

3.

$$\int R_0(t_1,t_2)\theta_i(t_2)dt_2 = \lambda_i\theta_i(t_1) \tag{D.0.6}$$

$$R_0(t_1,t_2) = \int h(t_1,\tau)h(t_2,\tau)d\tau \tag{D.0.7}$$

4. Let $x(\tau)$ be an arbitrary L_2 function and let $(x,\phi_i) = \int x(\tau)\phi_i(\tau)d\tau$. Then the following three statements are equivalent
i) $(x,\phi_i) = 0$
ii) $\int R(\tau_1,\tau_2)x(\tau_2)d\tau_2 = 0$
iii) $\int h(t,\tau)x(\tau)d\tau = 0$
Moreover if $x(\tau)$ is expanded as

$$x(\tau) = \sum_i x_i\phi_i(\tau) + x_r(\tau); x_i = (x,\phi_i) \tag{D.0.8}$$

Then

$$\int h(t,\tau)x(\tau)d\tau = \sum_i x_i\sqrt{\lambda_i}\theta_i(t) \tag{D.0.9}$$

$$\int R(\tau_1,\tau_2)x(\tau_2)d\tau_2 = \sum_i x_i\lambda_i\phi_i(t) \tag{D.0.10}$$

5. *Let $u(t)$ be an arbitrary L_2 function and let $(u,\theta_i) = \int u(t)\theta_i(t)dt$. Then the following three statements are equivalent*

   i) $(u,\theta_i) = 0$

   ii) $\int R_0(t_1,t_2)u(t_2)dt_2 = 0$

   iii) $\int h(t,\tau)u(t)dt = 0$

   Moreover if $u(t)$ is expanded as

$$u(t) = \sum_i u_i\theta_i(t) + u_r(t); u_i = (u,\theta_i) \tag{D.0.11}$$

Then

$$\int h(t,\tau)u(t)dt = \sum_i u_i\sqrt{\lambda_i}\phi_i(\tau) \tag{D.0.12}$$

$$\int R_0(t_1,t_2)u(t_2)dt_2 = \sum_i u_i\lambda_i\theta_i(t) \tag{D.0.13}$$

6.

$$h(t,\tau) = \sum_i \sqrt{\lambda_i}\phi_i(\tau)\theta_i(t) \tag{D.0.14}$$

$$\int\int h^2(t,\tau)dtd\tau = \sum_i \lambda_i \tag{D.0.15}$$

$$R(\tau_1,\tau_2) = \sum_i \lambda_i\phi_i(\tau_1)\phi_i(\tau_2) \tag{D.0.16}$$

$$R_0(t_1,t_2) = \sum_i \lambda_i\theta_i(t_1)\theta_i(t_2) \tag{D.0.17}$$

7. *λ_i and $\phi_i(\tau)$ are the solutions of the following maximization problems*

$$\lambda_i = \max \|\int h(t,\tau)x(\tau)d\tau\|_2^2 \tag{D.0.18}$$

$$\lambda_i = \max \|R(\tau_1,\tau_2)x(\tau_2)d\tau_2\|_2 \tag{D.0.19}$$

where the maximization is over all $x(\tau)$ satisfying $\|x\|_2 = 1$, and $(x,\phi_j) = 0$ for $1 \le j < i$. $\phi_i(\tau)$ corresponds to $x(\tau)$ that maximizes the above quantities.

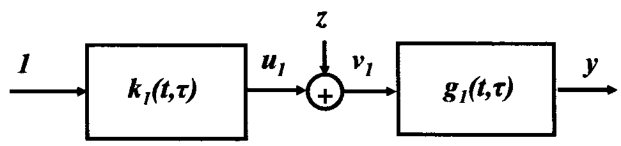

Figure D.0.2: Equivalent channel model 1

This theorem represents a useful tool in analyzing continuous time channels. It transforms a continuous time channel into the countable number of parallel random variable type channels. To see this, Theorem D.0.1 is applied on the system shown in Fig. D.0.1. In the next derivation, the approach given by Gallager [35] is closely followed. At the beginning, it is assumed that the transmitted signal has finite duration T. To avoid any ambiguity, which might emerge about the limits of integration in the convolution integral, the time-invariant filters $k(t)$ and $g(t)$ are replaced with time-varying filters (see Fig. D.0.2)

$$k_1(t,\tau) = \begin{cases} k(t-\tau), & |\tau| \le T/2 \\ 0, & \text{otherwise} \end{cases} \tag{D.0.20}$$

$$g_1(t,\tau) = \begin{cases} g(t-\tau), & |t| \le T/2 \\ 0, & \text{otherwise} \end{cases} \tag{D.0.21}$$

Let $\{\mu_i(T)\}$, $\{\xi_i(t)\}$, and $\{\eta_i(t)\}$ represent the sets of eigenvalues, input eigenfunctions, and output eigenfunctions of the filter $k_1(t,\tau)$, respectively. Similarly, $\{\lambda_{i,g}\}$, $\{\phi_{i,g}(t)\}$, and $\{\theta_{i,g}(t)\}$ represent the eigenvalues, input eigenfunctions, and output eigenfunctions of the filter $g_1(t,\tau)$, respectively. Then according to the part 4 and 5 of the Theorem D.0.1

$$v_1(t) = \sum_i (x_i\sqrt{\mu_i(T)} + z_i)\eta_i(t) + z_r(t), \tag{D.0.22}$$

where $\{x_i\}$ and $\{z_i\}$ are the coefficients of the expansions of $x(\tau)$ and $z(t)$ with respect to the input eigenfunctions $\{\xi_i(t)\}$ and output eigenfunctions $\{\eta_i(t)\}$, respectively. The term $z_r(t)$ is the part of the white noise $z(t)$ orthogonal to the output eigenfunctions $\{\eta_i(t)\}$. At this point, it is interesting to note that white Gaussian noise does not have representation with respect to time variable, i.e., it cannot be represented as a collection of random variables such that each corresponds to a particular time instant. On the other hand $z_i = (z, \theta_i) = \int z(t)\eta_i(t)dt$ represents Gaussian random variable, and there is no problem in dealing with it. Further, introduce the following notation $v_{1,i} = x_i\sqrt{\mu_i} + z_i$.

As noted in [35], if $\{v_{1,i}\}$ could be computed by the receiver from the observed signal $\mathbf{y}$, the mutual information between the transmitted signal $\mathbf{x}$ and received signal $\mathbf{y}$ would

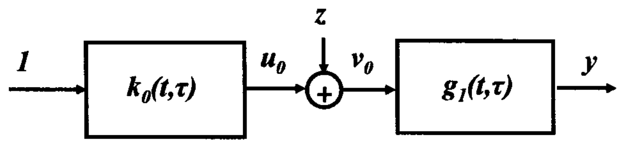

Figure D.0.3: Equivalent channel model 2

correspond to the mutual information between the random variables $\{x_i\}$ and $\{v_{1,i}\}$. But there exists a problem because the output of the filter $g_1(t,\tau)$ does not uniquely determine the input $v_1(t)$. Actually, there could be found different signals $v_1(t)$ that would give zero at the output of the filter $g_1(t,\tau)$ [35]. That is why the channel filter $k_1(t,\tau)$ has to be modified such that the component of the signal $u(t)$ at the output of the channel filter does not contain the components that are orthogonal to the set of input eigenfunctions $\{\phi_{i,g}(t)\}$ of the filter $g_1(t,\tau)$ (see [35]). A modified filter is given by

$$k_0(t,\tau) = \sum_i \phi_{i,g}(t) \int k_1(t_1,\tau)\phi_{i,g}(t_1)dt_1. \tag{D.0.23}$$

From equation (D.0.23), the impulse response $k_0(t,\tau)$ is the expansion of the impulse response $k_1(t,\tau)$ on the set of input eigenfunctions $\{\phi_{i,g}(t)\}$. At this point, it is important to emphasize that the received signal $\mathbf{y}$ remains the same. This can be seen from part 4 of the Theorem D.0.1, and it is the consequence of the fact that the output of the filter does not contain the part of the input signal that is orthogonal to the set of input eigenfunctions. Define the eigenvalues, input eigenfunctions, and output eigenfunctions of a modified filter $k_0(t,\tau)$ as $\lambda_i(T)$, $\phi_i(\tau)$, and $\theta_i(t)$, respectively. The modified equation (D.0.22) is given as

$$v_0(t) = \sum_i (x_i\sqrt{\lambda_i(T)} + z_i)\theta_i(t) + z_r(t) \tag{D.0.24}$$

where the explanation for $\{x_i\}$, $\{z_i\}$, and $z_r(t)$ is similar to the previous one, except that the expansions of the signals are with respect to the output eigenfunctions $\{\theta_i(t)\}$ of a modified filter $k_0(t,\tau)$. Also, as it is done in the previous consideration, introduce the following notation $v_{0,i} = x_i\sqrt{\lambda_i} + z_i$. Fig. D.0.3 shows the model of transformed communication system.

Thus, by this two steps, the original continuous-time channel is reduced to the set of countable number of parallel random variable channels. By using Theorem 7.5.1 (see [35]), which gives the capacity of parallel Gaussian random variable channels, the capacity of continuous-time additive Gaussian channel is given by

$$\lim_{T\to\infty}\frac{1}{2T}\sum_i \log\Big(1+\lambda_i(T)\Big). \tag{D.0.25}$$

Gallager proved that this limit can be represented in terms of the frequency response of the channel $\tilde{H}$ and PSD's $S_{\mathbf{x}}(f)$ and $S_{\mathbf{n}}(f)$ (Lemma 8.5.3, Lemma 8.5.7) as

$$\int_{-\infty}^{+\infty}\log\Big(1+\frac{S_{\mathbf{x}}(f)|\tilde{H}(f)|^2}{S_{\mathbf{n}}(f)|\tilde{W}(f)|^2}\Big)df, \tag{D.0.26}$$

by relating the sum of eigenvalues $\{\mu_i(T)\}$ of the original model and the sum of eigenvalues of transformed model $\{\lambda_i(T)\}$ given in Fig. D.0.2 and Fig. D.0.3, respectively, in Lemma 8.5.6. Also, he proved that for the model given in Fig. D.0.3, it is possible to reconstruct the sequence $\{v_{0,i}\}$ from the received signal $\mathbf{y}$ in Lemma 8.5.1. But, his proves are valid when the equivalent channel frequency response $K(f)$ is completely known. In order to be able to use Root's and Varaiya's result [63], one needs two prove that these two lemmas are true for all equivalent impulse responses $k(t)$ from the uncertainty set $\mathcal{B}$.

Now, we are ready to discuss the main part of the proof. First, we discuss Lemma 8.5.6 from [35]. It relates the sums of eigenvalues of the filters $k_0(t,\tau)$ and $k_1(t,\tau)$, as T tends to infinity. We want to show that Lemma 8.5.6 holds for all $k(t)\in\mathcal{B}$ uniformly over $\mathcal{B}$, where $\mathcal{B}$ is the set of all impulse responses $k(t)$ that satisfy the conditions given in Section 3.5. To prove that, Lemma D.0.2 is used. It relies on a well-known Arzela-Ascoli theorem [46]. To apply that theorem, the following notation is introduced. Here, (D.0.23) can be seen as a mapping from the set $\mathcal{B}$ to $\mathbf{R}$ for fixed parameters t and τ. Denote this mapping as $q_{T,\tau}(k(t)) = k_0(t,\tau)$. By Lemma 8.5.5 [35], it is known that $\lim_{T\to\infty} q_{T,0}(k(t)) = \lim_{T\to\infty} k_0(t,0) = k(t)$. So, the limit of $q_{T,0}(k(t))$ is known. Actually, one has to show that the limit is uniform over $\mathcal{B}$.

Lemma D.0.2 *Assume that $k(t)\in\mathcal{B}$. Then*

$$\lim_{T\to\infty} q_{T,0}(k(t)) = \lim_{T\to\infty} k_0(t,0) = k(t) \tag{D.0.27}$$

uniformly over $\mathcal{B}$.

Proof. It should be noted that the set $\mathcal{B}$ is a conditionally compact subset of L_2 that follows from its definition (see [31]). Thus, to prove a uniform convergence in (D.0.27), by

Arzela-Ascoli theorem it is enough to show the equicontinuity of the family of mappings $\{q_{T,0}(k(t))\}, 0 < T < \infty$. Assume that $k_a(t), k_b(t) \in \mathcal{B}$. It follows from the definition of $\mathcal{B}$ that their truncated versions $k_{1,a}(t,\tau), k_{1,b}(t,\tau)$ belong to L_2 with respect to variable t. Then

$$\begin{aligned} |q_{T,\tau}(k_a(t)) - q_{T,\tau}(k_b(t))| &= |\sum_i \phi_{i,g}(t) \int (k_{1,a}(t_1,\tau) - k_{1,b}(t_1,\tau))\phi_{i,g}(t_1)dt_1| && \text{(D.0.28)} \\ &\leq \sum_i |\phi_{i,g}(t)| \|k_{1,a}(t_1,\tau) - k_{1,b}(t_1,\tau))\|_2 \|\phi_{i,g}(t_1)\|_2 && \text{(D.0.29)} \\ &= \sum_i |\phi_{i,g}(t)| \|k_{1,a}(t_1,\tau) - k_{1,b}(t_1,\tau))\|_2 && \text{(D.0.30)} \end{aligned}$$

which holds for any τ, and therefore for $\tau = 0$ as well. The inequality follows from the Schwartz inequality, while the last equality comes from the assumption that the input eigenfunctions $\{\phi_{i,g}(t)\}$ are orthonormal. Thus, when $\|k_{1,a}(t_1,\tau) - k_{1,b}(t_1,\tau))\|_2 \to 0$, $|q_{T,0}(k_a(t)) - q_{T,0}(k_b(t))|$ tends to zero which proves the equicontinuity of the family of mappings $\{q_{T,0}(k(t))\}$, and that in turn proves the assertion of the lemma.
In the next step Lemma 8.5.6 is rederived for the set of channels $k(t) \in \mathcal{B}$. The derivation exactly follows the steps of the original lemma up to the point when Lemma D.0.2 is applied.

Lemma D.0.3 *(Rederived Lemma 8.5.6) Suppose that the impulse responses $k_1(t,\tau)$ and $k_0(t,\tau)$, which correspond to the impulse response $k(t) \in \mathcal{B}$, have the following sets of eigenvalues $\mu_i(T)$, $\lambda_i(T)$, respectively. Then*

$$\lim_{T\to\infty} \frac{1}{T} \sum_i \mu_i(T) = \lim_{T\to\infty} \frac{1}{T} \sum_i \lambda_i(T) \tag{D.0.31}$$

uniformly over the set B.

Proof. Having in mind that $k(t)$ is a time invariant impulse response, from (D.0.15) follows that

$$\frac{1}{T} \sum_i \mu_i(T) = \frac{1}{T} \int \int k_1^2(t,\tau) dt d\tau = \int k^2(t) dt \tag{D.0.32}$$

Further, observe that $\lambda_i(T) \leq \mu_i(T)$ (see Lemma 8.5.4 [35]). So, to prove the lemma, it is enough to show

$$\lim_{T\to\infty} \frac{1}{T} \int \int k_0^2(t,\tau) dt d\tau \geq \int k^2(t) dt - \epsilon \tag{D.0.33}$$

for arbitrary ϵ. By Lemma D.0.2, for any ϵ, one can choose T_ϵ such that

$$\int k_0^2(t,0) dt \geq \int k^2(t) dt - \epsilon, T \geq T_\epsilon \tag{D.0.34}$$

uniformly over the set $\mathcal{B}$. Further, (D.0.23) can be rewritten as

$$k_0(t+\tau,\tau) = \sum_i \phi_{i,T}(t+\tau) \int k_1(u)\phi_{i,T}(u+\tau)du \tag{D.0.35}$$

From (D.0.35), $k_0(t+\tau,\tau)$ represents an expansion of $k(t)$ by means of the orthonormal set $\{\phi_{i,T}(t+\tau)\}$. As explained in [35], by Lemma 8.5.5, the reminder of the expansion, i.e., the part of $k(t)$ orthogonal to all $\{\phi_{i,T}(t+\tau)\}$ is also orthogonal to all $\{\phi_{i,T_\epsilon}(t)\}$ for $T \geq T_\epsilon + 2|\tau|$. It means that the reminder of the expansion of $k(t)$ with respect to $\{\phi_{i,T}(t+\tau)\}$ can be upper bounded by the reminder of the expansion of $k(t)$ with respect to $\{\phi_{i,T_\epsilon}(t)\}$, and the letter can be upper bounded by ϵ giving

$$\int k_0^2(t+\tau,\tau)dt \geq \int k^2(t)dt - \epsilon, T \geq T_\epsilon + 2|\tau|. \tag{D.0.36}$$

This follows from (D.0.34) and the fact that $k_0(t,0)$ is nondecreasing while converging to $k(t)$ when T increases. In other words, the reminder is nonincreasing as T tends to infinity. The previous statement is the result of Lemma 8.5.5. Now, due to

$$\frac{1}{T}\int_{\tau=-T/2}^{T/2}\int_{t=-\infty}^{+\infty} k_0^2(t,\tau)dtd\tau \geq \frac{1}{T}\int_{|\tau|\leq\frac{T-T_\epsilon}{2}}\int k_0^2(t,\tau)dtd\tau \tag{D.0.37}$$

and by noticing that (D.0.36) is true for $T \geq T_\epsilon + 2|\tau|$, equation (D.0.37) becomes

$$\frac{1}{T}\int\int k_0^2(t,\tau)dtd\tau \geq \frac{T-T_\epsilon}{T}\left[\int k^2(t)dt - \epsilon\right] \tag{D.0.38}$$

Letting $T \to \infty$, the required result is obtained. Thus, (D.0.33) holds uniformly for all $k(t) \in \mathcal{B}$, proving that the summation of the eigenvalues of two impulse responses are the same in the limit.

The next step in relating (D.0.25) and (D.0.26) for the set of impulse responses $\mathcal{B}$, is to rederive Lemma 8.5.1 [35] such that it holds uniformly over $\mathcal{B}$. To rederive Lemma 8.5.1, we need a sequence of lemmas whose proofs are given below. The lemmas are concerned with the properties of the impulse responses $k_a(t), k_b(t) \in \mathcal{B}$ when $\|k_a(t) - k_b(t)\|_2 \to 0$. Just have in mind that by the definition of $\mathcal{B}$, $k_a(t), k_b(t) \in L_2$.

Lemma D.0.4 *Assume that $k_a(t)$, $k_b(t) \in \mathcal{B}$, and that $k_{0,a}(t,\tau)$, $k_{0,b}(t,\tau)$ are the corresponding modified filters. Whenever $\|k_a(t) - k_b(t)\|_2 \to 0$ then $|k_{0,a}(t,\tau) - k_{0,b}(t,\tau)| \to 0$ for any τ.*

Proof. Using the definition (D.0.23)

$$|k_{0,a}(t,\tau) - k_{0,b}(t,\tau)| = |\sum_i \phi_{i,g}(t) \int (k_a(t_1-\tau) - k_b(t_1-\tau))\phi_{i,g}(t_1)dt_1| \quad \text{(D.0.39)}$$

$$\leq \sum_i |\phi_{i,g}(t)| \|k_a(t) - k_b(t)\|_2 \quad \text{(D.0.40)}$$

The inequality follows from the Schwartz inequality and orthonormal property of the eigenfunctions. So, from (D.0.40), for any value of τ, $|\tau| \leq T/2$, whenever $\|k_a(t) - k_b(t)\|_2 \to 0$ then $|k_{0,a}(t,\tau) - k_{0,b}(t,\tau)| \to 0$.

Corollary D.0.5 *If $\|k_a(t) - k_b(t)\|_2 \to 0$ then $|u_{0,a}(t) - u_{0,b}(t)| \to 0$ for any value of τ.*

Proof. This follows from $u_0(t) = \int k_0(t,\tau)x(\tau)d\tau$, Lemma D.0.4, and the fact that $x(\tau) \in L_2$. Apply the Schwartz inequality, and the result is immediate.

Next, we use some of properties of Theorem D.0.1 to show what occurs with the eigenvalues $\{\lambda_{a,i}(T)\}$ and $\{\lambda_{b,i}(T)\}$, output eigenfunctions $\{\theta_{a,i}(t)\}$ and $\{\theta_{b,i}(t)\}$, and functions $R_{0,a}(t_1,t_2)$ and $R_{0,b}(t_1,t_2)$ of the corresponding filters $k_{0,a}(t)$, $k_{0,b}(t)$ when $\|k_a(t) - k_b(t)\|_2 \to 0$.

Lemma D.0.6 *Assume that $k_a(t)$, $k_b(t) \in \mathcal{B}$, and $k_{0,a}(t,\tau)$, $k_{0,b}(t,\tau)$ are the corresponding modified filters. If $\|k_a(t) - k_b(t)\|_2 \to 0$ then*

1. $|\lambda_{a,i}(T) - \lambda_{b,i}(T)| \to 0$ *for each* i
2. $|R_{0,a}(t_1,t_2) - R_{0,b}(t_1,t_2)| \to 0$ *for each* t_1, t_2
3. $|\theta_{a,i}(t) - \theta_{b,i}(t)| \to 0$ *for each* i.

Proof.

1. From equation (D.0.18) of Theorem D.0.1, an eigenvalue λ_i can be understood as the norm of the linear operator Q_i on the space of L_2 functions defined as $Q_i\mathbf{x} = \int k_0(t,\tau)x(\tau)d\tau$, where $k_0(t,\tau)$ is called the kernel of the linear operator Q_i. Thus

$$|\lambda_{a,i}(T) - \lambda_{b,i}(T)| = |\|Q_{a,i}\| - \|Q_{b,i}\|| \quad \text{(D.0.41)}$$

$$\leq \|Q_{a,i} - Q_{b,i}\| \quad \text{(D.0.42)}$$

$$= \sup_{\|\mathbf{x}\|\leq 1} \| \int (k_{0,a}(t,\tau) - k_{0,b}(t,\tau))x(\tau)d\tau\|^2 \quad \text{(D.0.43)}$$

$$\leq \sup_{\|\mathbf{x}\|\leq 1} \|\|k_{0,a}(t,\tau) - k_{0,b}(t,\tau)\|_2 \|x(t)\|_2\|^2 \quad \text{(D.0.44)}$$

where the last inequality follows from the Schwartz inequality. Now, the proof follows from Lemma D.0.4.

2. Start from the formula (D.0.7) that relates $R_0(t_1,t_2)$ and $k_0(t,\tau)$. Then

$$\begin{aligned} |R_{0,a}(t_1,t_2) - R_{0,b}(t_1,t_2)| &= \left|\int (k_{0,a}(t_1,\tau)k_{0,a}(t_2,\tau) - k_{0,b}(t_1,\tau)k_{0,b}(t_2,\tau))d\tau\right| \\ &\leq \int |k_{0,a}(t_2,\tau)||(k_{0,a}(t_1,\tau) - k_{0,b}(t_1,\tau))|d\tau \\ &+ \int |k_{0,b}(t_1,\tau)||(k_{0,a}(t_2,\tau) - k_{0,b}(t_2,\tau))|d\tau \end{aligned} \tag{D.0.45}$$

When $\|k_a(t) - k_b(t)\|_2 \to 0$, then by Lemma D.0.4 both integrals on the right-hand side of (D.0.45) tends to zero that establishes the result.

3. This part is implied by *1* and *2*. Gronwall's lemma is the key ingredient of the proof. By using (D.0.6)

$$|\theta_{a,i}(t_1) - \theta_{b,i}(t_1)| = \left|\int \Big(\frac{R_{0,a}(t_1,t_2)}{\lambda_{a,i}(T)}\theta_{a,i}(t_2) - \frac{R_{0,b}(t_1,t_2)}{\lambda_{b,i}(T)}\theta_{b,i}(t_2)\Big)dt_2\right| \tag{D.0.46}$$

$$\begin{aligned} &\leq \int \left|\frac{R_{0,a}(t_1,t_2)}{\lambda_{a,i}(T)}(\theta_{a,i}(t_2) - \theta_{b,i}(t_2))\right|dt_2 \\ &+ \int \left|\theta_{b,i}(t_2)\Big(\frac{R_{0,a}(t_1,t_2)}{\lambda_{a,i}(T)} - \frac{R_{0,b}(t_1,t_2)}{\lambda_{b,i}(T)}\Big)\right|dt_2 \end{aligned} \tag{D.0.47}$$

Note that any eigenvalue $\lambda_i(T)$ is strictly greater than zero by Theorem D.0.1. Further, introduce the following notation

$$f_1(t) = |\theta_{a,i}(t_1) - \theta_{b,i}(t_1)| \tag{D.0.48}$$

$$f_2(t) = \left|\frac{R_{0,a}(t_1,t_2)}{\lambda_{a,i}(T)}\right| \tag{D.0.49}$$

$$K = \int \left|\theta_{b,i}(t_2)\Big(\frac{R_{0,a}(t_1,t_2)}{\lambda_{a,i}(T)} - \frac{R_{0,b}(t_1,t_2)}{\lambda_{b,i}(T)}\Big)\right|dt_2 \tag{D.0.50}$$

Gronwall's lemma states that if $f_1(t),\ f_2(t) \geq 0$ are continuous functions on the interval of integration, and constant $K > 0$ then it is true that $f_1(t) \leq K\exp\Big[\int f_2(s)ds\Big]$ giving

$$|\theta_{a,i}(t_1) - \theta_{b,i}(t_1)| \leq K\exp\Big[\int \left|\frac{R_{0,a}(t_1,t_2)}{\lambda_{a,i}(T)}\right|dt_2\Big] \tag{D.0.51}$$

When $\|k_a(t) - k_b(t)\|_2 \to 0$ constant K tends to zero by *1* and *2* of this Lemma so the result follows.

The preceding two lemmas are the useful results that will help in the rederivation of Lemma 8.5.1 [35].

Lemma D.0.7 *(Rederived Lemma 8.5.1) Assume that $k(t) \in \mathcal{B}$. Then the set of random variables $\{v_{0,i}\}$ found in (D.0.24) can be uniquely determined from received signal* $\mathbf{y}$ *uniformly for all $k(t) \in \mathcal{B}$.*

Proof. As for Lemma D.0.3, the main tool is Arzela-Ascoli theorem. Original Lemma 8.5.1 states that the error due to the omission of the filter $g_1(t,\tau)$ tends to zero if enough number of terms of the series expansion of $\mathbf{y}$ with respect to the output eigenfunctions $\{\theta_{i,g}(t)\}$ of the filter $g_1(t,\tau)$ are taken. By using Arzela-Ascoli theorem, it will be shown that this error tends to zero uniformly over the set $\mathcal{B}$. Again, the derivation closely follows the steps of the original Lemmma 8.5.1.

Because the output of the filter $k_0(t,\tau)$ suppresses the components of the input that are orthogonal to the set of input eigenfunctions $\{\phi_{i,g}(t)\}$ of the filter $g_1(t,\tau)$ it is possible to expand any output eigenfunction $\theta_i(t)$ of the filter $k_0(t,\tau)$ without a reminder

$$\theta_i(t) = \sum_j \beta_{i,j}\phi_{j,g}(t), \text{ where } \beta_{i,j} = \int \theta_i(t)\phi_{j,g}(t)dt \tag{D.0.52}$$

The consequences of original lemma are also

$$v_{0,i} = \sum_{j=1}^{m} \beta_{i,j} \int v_0(t)\phi_{j,g}(t)dt + \int v_0(t) \sum_{j=m+1}^{\infty} \beta_{i,j}\phi_{j,g}(t)dt \tag{D.0.53}$$

$$\int y(t)\theta_{j,g}(t)dt = \sqrt{\lambda_{j,g}} \int v_0(\tau)\phi_{j,g}(\tau)d\tau \tag{D.0.54}$$

Replace now (D.0.54) into (D.0.53)

$$v_{0,i} = \sum_{j=1}^{m} \frac{\beta_{i,j}}{\sqrt{\lambda_{j,g}}} \int y(t)\theta_{j,g}(t)dt + \int v_0(t) \sum_{j=m+1}^{\infty} \beta_{i,j}\phi_{j,g}(t)dt \tag{D.0.55}$$

that shows how the random variable $v_{0,i}$ can be computed if $\mathbf{y}$ is observed. Here, the main problem is to show that the second term in (D.0.55) tends to zero as $m \to \infty$ uniformly over $\mathcal{B}$. The second term can be rewritten as

$$\int v_0(t) \sum_{j=m+1}^{\infty} \beta_{i,j}\phi_{j,g}(t)dt = \int u_0(t) \sum_{j=m+1}^{\infty} \beta_{i,j}\phi_{j,g}(t)dt + \int z(t) \sum_{j=m+1}^{\infty} \beta_{i,j}\phi_{j,g}(t)dt \tag{D.0.56}$$

It was proven by original lemma that both terms tends to zero when $m \to \infty$. First consider the first term in (D.0.56). To prove the lemma we need the following result

$$\begin{aligned} |u_{0,a}(t)\beta_{i,j,a} - u_{0,b}(t)\beta_{i,j,b}| &\leq |u_{0,a}(t)(\beta_{i,j,a} - \beta_{i,j,b})| + |\beta_{i,j,b}(u_{0,a}(t) - u_{0,b}(t))| \\ &\leq |u_{0,a}(t)(\beta_{i,j,a} - \beta_{i,j,b})| + \epsilon_1, \ \epsilon_1 > 0 \end{aligned} \tag{D.0.57}$$

where the second inequality holds when $\|k_a(t) - k_b(t)\|_2 \to 0$, because of Corollary D.0.5. The signals $u_{0,a}(t)$ and $u_{0,b}(t)$ are the output signals from the filters $k_{0,a}(t,\tau)$ and $k_{0,b}(t,\tau)$, respectively, when $k_a(t),\ k_b(t) \in \mathcal{B}$. Here, the same approach applies as in Lemma D.0.2. Denote the first term of (D.0.56) as $q_m(k(t))$, the functional from $\mathcal{B}$ to $\mathbf{R}$. The original lemma shows that $\lim_{m\to\infty} q_m(k(t)) = 0$. Further, the equicontinuity is proved for the family of mappings $\{q_m(k(t))\}$ as follows

$$
\begin{aligned}
|q_m(k_a(t)) - q_m(k_a(t))| &= |\int \sum_{j=m+1}^{\infty} (\beta_{i,j,a}u_{0,a}(t) - \beta_{i,j,b}u_{0,b}(t))\phi_{j,g}(t)dt| && \text{(D.0.58)}\\
&\leq \int \sum_{j=m+1}^{\infty} |u_{0,a}(t)||\beta_{i,j,a} - \beta_{i,j,b}||\phi_{j,g}(t)|dt && \text{(D.0.59)}\\
&\leq \left\{\int |u_{0,a}(t)|^2 dt \int (\sum_{j=m+1}^{\infty} |\beta_{i,j,a} - \beta_{i,j,b}||\phi_{j,g}(t)|)^2 dt\right\}^{1/2} && \text{(D.0.60)}
\end{aligned}
$$

where the first inequality comes from (D.0.57), and the second comes from the Schwartz inequality. To prove that $|q_m(k_a(t)) - q_m(k_a(t))| \to 0$ whenever $\|k_a(t) - k_b(t)\|_2 \to 0$, we have to show that $|\beta_{i,j,a} - \beta_{i,j,b}| \to 0$ if $\|k_a(t) - k_b(t)\|_2 \to 0$. Thus

$$
\begin{aligned}
|\beta_{i,j,a} - \beta_{i,j,b}| &= |\int (\theta_{a,i}(t) - \theta_{b,i}(t))\phi_{j,g}(t)dt| && \text{(D.0.61)}\\
&\leq \|\theta_{a,i}(t) - \theta_{b,i}(t)\|_2 && \text{(D.0.62)}
\end{aligned}
$$

The inequality is the consequence of Schwartz inequality and orthonormality of eigenfunctions $\{\phi_{j,g}(t)\}$. The equicontinuity follows from (D.0.62) and Lemma D.0.6, part *3*. So, by Arzela-Ascoli theorem the limit $\lim_{m\to\infty} q_m(k(t)) = 0$ is uniform for all $k(t) \in \mathcal{B}$. The proof that the second term in (D.0.56) tends to zero uniformly over $\mathcal{B}$ is similar to one give for the first term.

Thus, it is proved that Lemma 8.5.1 and Lemma 8.5.6 (Lemmas D.0.3 and D.0.7) given in [35] are true for all equivalent impulse responses $k(t) \in \mathcal{B}$, uniformly over $\mathcal{B}$. This means that (D.0.25) tends to (D.0.26) uniformly over $\mathcal{B}$, which is the condition required by Theorems 7 and 8 in [63] to prove a channel coding theorem and converse to channel coding theorem for class of channels $\mathcal{B}$.

Appendix E

MIMO Channels with Channel Uncertainty

Proof of Theorem 4.3.1. Theorem 4.3.1 provides the solution of the problem (4.2.6) for the specific structure of the matrix $\Delta(e^{j\theta})$ given by (4.3.11). Otherwise, the maximization step of the mutual information with respect to the PSD matrix $W_{\mathbf{x}}(\theta)$ would be much harder to solve because we would not be able to use Hadamard's inequality as in [69].

The idea behind the proof is the following. First, the necessary conditions in terms of $W_{\mathbf{x}}(\theta)$ and $W_{\mathbf{n}}(\theta)$ are found for the problem defined by (4.2.6) (supinf optimization problem), for a general form of $\Delta(e^{j\theta})$. Thereafter, the necessary conditions for the inverse problem are found (infsup optimization problem), subject to a general form of $\Delta(e^{j\theta})$. It turns out that the necessary conditions are equivalent, implying that

$$\sup_{W_{\mathbf{x}}\in A_1} \inf_{H\in A_2} J(W_{\mathbf{x}}(\theta), W_{\mathbf{n}}(\theta)) = \inf_{H\in A_2} \sup_{W_{\mathbf{x}}\in A_1} J(W_{\mathbf{x}}(\theta), W_{\mathbf{n}}(\theta)). \tag{E.0.1}$$

This means that sup and inf problem may be interchanged. Second, by considering infsup problem, we show the reason for taking the specific structure of $\Delta(e^{j\theta})$. In the third step, supinf problem is solved for that specific structure of $\Delta(e^{j\theta})$.

To show that supinf and infsup of $J(W_{\mathbf{x}}(\theta), W_{\mathbf{n}}(\theta))$ are equal, first, the supinf problem is solved explicitly. Notice that the constraint $H(e^{j\theta}) \in A_2$ is the same as $\|\Delta(e^{j\theta})\|_\infty \leq 1$. Instead of working with this constraint, we will work with the constraint $\Delta^*(e^{j\theta})\Delta(e^{j\theta}) - I_m \leq 0$, for $\theta \in [0, 2\pi]$. This notation means that $\Delta^*\Delta - I_m$ is non-positive definite. The latter constraint implies the former. This modification does not change the problem, because it turns out that the optimal $\Delta^o(e^{j\theta})$ lies on the boundary, $\|\Delta^o(e^{j\theta})\|_\infty = 1$.

Next, define the Lagrangian

$$J_1(W_{\mathbf{x}}, \Delta, K) = \frac{1}{4\pi}\int_0^{2\pi} \log \det(I_p + (H_{nom} + w\Delta)W_{\mathbf{x}}(H_{nom} + w\Delta)^*)d\theta$$

$$+\int_0^{2\pi} \text{Trace}[K(\Delta^*\Delta - I_m)]d\theta, \tag{E.0.2}$$

where K is non-negative definite matrix, which is a Kuhn-Tucker condition [15]. The variation of J_1 with respect to $\Delta(e^{j\theta})$ gives

$$\frac{1}{4\pi} wW_{\mathbf{x}}(H_{nom} + w\Delta^o)^*(I_p + (H_{nom} + w\Delta^o)W_{\mathbf{x}}(H_{nom} + w\Delta^o)^*)^{-1} + K\Delta^{o,*} = 0, \tag{E.0.3}$$

which cannot be solved explicitly. Therefore, it has to be accounted as the equality constraint in order to resolve the supremum part. The original problem $\sup_{W_{\mathbf{x}} \in A_1} \inf_{\|\Delta\|_\infty \leq 1} J(W_{\mathbf{x}}, W_{\mathbf{n}})$ is equivalent to $\sup_{W_{\mathbf{x}} \in A_1} \sup_{K \geq 0} J_1(W_{\mathbf{x}}, \Delta^o, K)$ as its dual [15].

Further, introduce the Lagrangian

$$\begin{aligned}
J_2(W_{\mathbf{x}}, \Delta^o, K, \lambda_1) &= \frac{1}{4\pi} \int_0^{2\pi} \log \det(I_p + (H_{nom} + w\Delta^o)W_{\mathbf{x}}(H_{nom} + w\Delta^o)^*)d\theta \\
&+ \int_0^{2\pi} \text{Trace}[K(\Delta^*\Delta - I_m)]d\theta \\
&- \int_0^{2\pi} \text{Tr.}[S(\frac{1}{4\pi} wW_{\mathbf{x}}(H_{nom} + w\Delta^o)^* + K\Delta^{o*}(I_p + (H_{nom} + w\Delta^o)W_{\mathbf{x}}(H_{nom} + w\Delta^o)^*))]d\theta \\
&- \lambda_1\Big(\int_0^{2\pi} \text{Trace}(W_{\mathbf{x}})d\theta - P_{\mathbf{x}}\Big),
\end{aligned} \tag{E.0.4}$$

where λ_1 is a positive constant, which is a Kuhn-Tucker condition [15]. Then

$$\sup_{W_{\mathbf{x}} \in A_1} \sup_{K \geq 0} J_1(W_{\mathbf{x}}, \Delta^o, K) \tag{E.0.5}$$

is equivalent to

$$\inf_{\lambda_1 \geq 0} \sup_{K \geq 0} \sup_{W_{\mathbf{x}} \in A_1} J_2(W_{\mathbf{x}}, \Delta^o, K, \lambda_1), \tag{E.0.6}$$

because (E.0.6) is a dual problem of (E.0.5) [15]. Since $W_{\mathbf{x}}(\theta)$ and $\Delta^o(e^{j\theta})$ are related through the equality constraint (E.0.3), the Lagrangian J_2 has to be varied with respect to both, $W_{\mathbf{x}}(\theta)$ and $\Delta^o(e^{j\theta})$. By varying J_2 with respect to $\Delta^o(e^{j\theta})$, the following equation is obtained

$$\begin{aligned}
&\frac{1}{4\pi} wW_{\mathbf{x}}(H_{nom} + w\Delta^o)^*(I_p + (H_{nom} + w\Delta^o)W_{\mathbf{x}}(H_{nom} + w\Delta^o)^*)^{-1} + K\Delta^{o,*} = \\
&W_{\mathbf{x}}H_{nom}^* SK\Delta^{o,*}w + W_{\mathbf{x}}\Delta^{o,*}SK\Delta^{o,*}|w|^2,
\end{aligned} \tag{E.0.7}$$

where the term on the left hand side is equal to zero (see (E.0.3)). This observation implies that

$$wW_{\mathbf{x}}(\theta)(H_{nom}^*(e^{j\theta}) + w^*(e^{j\theta})\Delta^{o,*}(e^{j\theta}))SK\Delta^{o,*}(e^{j\theta}) = 0. \tag{E.0.8}$$

From (E.0.8) follows that either $H_{nom}(e^{j\theta})+w(e^{j\theta})\Delta^o(e^{j\theta})=0$, or $S=0$, or $K=0$, or some combination of the previous conditions is true. If $H_{nom}(e^{j\theta})+w(e^{j\theta})\Delta^o(e^{j\theta})=0$, then from (E.0.3) follows that $K=0$, meaning that the constraint imposed on $\Delta^o(e^{j\theta})$ vanishes, and the channel capacity C is equal to zero, which is a trivial solution. Thus, the possibility that remains is $S=0$. This and Lagrangian J_2 indicate that sup and inf problems may be solved independently. To verify this claim, observe that when $S=0$, the Lagrangian J_2, (E.0.4), corresponds to the Lagrangian when a saddle point exists. When a saddle point exists, the Lagrangian consists of the pay-off function, the term that describes the constraint on $\Delta(e^{j\theta})$, and the term that describes the constraint on $W_{\mathbf{x}}(\theta)$, while the Lagrangian is varied with respect to $\Delta(e^{j\theta})$ and $W_{\mathbf{x}}(\theta)$ as they were independent. This is exactly the case of (E.0.4) for $S=0$. Hence, the conclusion is

$$\sup_{W_{\mathbf{x}}\in A_1}\inf_{H\in A_2} J(W_{\mathbf{x}}(\theta),W_{\mathbf{n}}(\theta))=\inf_{H\in A_2}\sup_{W_{\mathbf{x}}\in A_1} J(W_{\mathbf{x}}(\theta),W_{\mathbf{n}}(\theta)). \tag{E.0.9}$$

Thus, for $S=0$, we vary the Lagrangian J_2 with respect to $W_{\mathbf{x}}(\theta)$ to get

$$\begin{aligned}(H_{nom}+w\Delta^o)^*(I+(H_{nom}+w\Delta^o)W_{\mathbf{x}}(H_{nom}+w\Delta^o)^*)^{-1}(H_{nom}+w\Delta^o)&\\ =4\pi\lambda_1 I_m,&\end{aligned} \tag{E.0.10}$$

which represents the equation that is satisfied by the optimal $W^o_{\mathbf{x}}(\theta)$. From Kuhn-Tucker conditions [15]

$$\lambda_1\Big[\int_0^{2\pi}\mathrm{Trace}(W^o_{\mathbf{x}}(\theta))d\theta-P_{\mathbf{x}}\Big]=0, \tag{E.0.11}$$

$$\int_0^{2\pi}\mathrm{Trace}[K(\Delta^*\Delta-I_m)]d\theta=0, \tag{E.0.12}$$

and observing that $\lambda_1\neq 0$ and $K\neq 0$ (because the opposite conditions imply trivial solution $C=0$), it is obtained that

$$\int_0^{2\pi}\mathrm{Trace}(W^o_{\mathbf{x}}(\theta))d\theta=P_{\mathbf{x}}, \tag{E.0.13}$$

$$\|\Delta^o(e^{j\theta})\|_\infty=1. \tag{E.0.14}$$

The previous equations do not provide us with the complete solution for $\Delta^o(e^{j\theta})$. The reason for this is that $\|\cdot\|_\infty$ norm puts the constraint on $\Delta^*(e^{j\theta})\Delta(e^{j\theta})$, but not on $\Delta(e^{j\theta})$ itself. So, we have to make an additional step to find $\Delta^o(e^{j\theta})$.

Next, it is shown why we deal with a particular case of $\Delta(e^{j\theta})$. We find what conditions should be satisfied such that the integrand

$$\det(I_p+(H_{nom}(e^{j\theta})+w(e^{j\theta})\Delta(e^{j\theta}))W_{\mathbf{x}}(\theta)(H_{nom}(e^{j\theta})+w(e^{j\theta})\Delta(e^{j\theta}))^*) \tag{E.0.15}$$

can be maximized in $W_{\mathbf{x}}(\theta)$ by using Hadamard's inequality. Having in mind that $H_{nom}(e^{j\theta}) = U(e^{j\theta})\Sigma(e^{j\theta})V^*(e^{j\theta})$, and by using the fact that $\det(I_n + AB) = \det(I_m + BA)$ (where A is $n \times m$ matrix, and B is $m \times n$ matrix), the integrand can be expressed as

$$\det(I_p + V^* W_{\mathbf{x}} V(\Sigma + wU^*\Delta V)^*(\Sigma + wU^*\Delta V)) \tag{E.0.16}$$

$$= \det(I_p + \tilde{W}_{\mathbf{x}} R^* R), \tag{E.0.17}$$

where $R \triangleq \Sigma + wU^*\Delta V$, and $\tilde{W}_{\mathbf{x}} \triangleq V^* W_{\mathbf{x}} V$. Further, by using Hadamard's inequality (as in [69]), the determinant can be upper bounded by the product of the diagonal elements,

$$\det(I_p + \tilde{W}_{\mathbf{x}} R^* R) \leq \Pi_{i=1}^{n}(1 + [Q]_{ii}), \tag{E.0.18}$$

where

$$Q = \tilde{W}_{\mathbf{x}} R^* R. \tag{E.0.19}$$

The equality in (E.0.18) is achieved when Q is diagonal. If R^*R were diagonal, Hadamard's inequality could be used to maximize the mutual information rate in $\tilde{W}_{\mathbf{x}}$. Then maximizing $\tilde{W}_{\mathbf{x}}$ would be diagonal. But, if $\Delta(e^{j\theta})$ is chosen as in (4.3.11), R^*R is diagonal. Further, one could solve for $\tilde{W}_{\mathbf{x}}(\theta)$ by employing the Lagrange multiplier method. But, this approach would lead to some ambiguities because it can be shown that such determined Lagrange multiplier increases with uncertainty, and the next step of minimization with respect to $\Delta(e^{j\theta})$ will not be clear. Therefore, we will get back to the original supinf problem and use this information concerning the diagonal property of $\tilde{W}_{\mathbf{x}}$ and R^*R, to determine $\Delta^o(e^{j\theta})$, which minimizes $J(W_{\mathbf{x}}(\theta), W_{\mathbf{n}}(\theta))$. Thus, assume $\tilde{W}_{\mathbf{x}}$ is diagonal and choose the matrix $\Delta(e^{j\theta})$ as follows

$$\Delta^o(e^{j\theta}) = U(e^{j\theta})\begin{bmatrix} \Delta_1(e^{j\theta}) & 0 \\ 0 & 0 \end{bmatrix} V^*(e^{j\theta}), \tag{E.0.20}$$

and $\Delta_1(e^{j\theta})$ is $n \times n$ matrix, $\Delta_1(e^{j\theta}) = \mathrm{diag}(\delta_1, ..., \delta_n)$, where n is the rank of $H_{nom}(e^{j\theta})$, $n \leq \min(p, m)$. This ensures that R^*R is diagonal. Then

$$\det(I_p + \tilde{W}_{\mathbf{x}} R^* R) = \Pi_{i=1}^{n}(1 + [\tilde{W}_{\mathbf{x}}]_{ii}|\sigma_i + \delta_i w|^2), \tag{E.0.21}$$

where $\{\sigma_i\}_{i=1}^{n}$ are singular values of $H_{nom}(e^{j\theta})$. Note that $\Delta_1^*\Delta_1$ is identity matrix, which comes from $\Delta^*\Delta = I_m$ (see (E.0.12)). It follows that $|\delta_i| = 1$, $i = 1, ..., n$. The determinant in (E.0.21) is minimized when $\xi_i = |\sigma_i + \delta_i w|^2$ is minimized. ξ_i is lower bounded as follows

$$\begin{aligned} \xi_i &= \sigma_i^2 + 2\sigma_i|\delta_i||w|\cos(\arg(\delta_i) + \arg(w)) + |\delta_i|^2|w|^2 && \text{(E.0.22)} \\ &\geq (\sigma_i - |w|)^2, && \text{(E.0.23)} \end{aligned}$$

where $\delta_i = |\delta_i|e^{\mathrm{j}\arg(\delta_i)} = e^{\mathrm{j}\arg(\delta_i)}$, and $w = |w|e^{\mathrm{j}\arg(w)}$. The matrix $\Delta_1(e^{j\theta})$ that achieves the lower bound is given by

$$\Delta_1(e^{j\theta}) = \begin{pmatrix} e^{-\mathrm{j}\arg(w)+\mathrm{j}\pi} & 0 & \cdots & 0 \\ 0 & e^{-\mathrm{j}\arg(w)+\mathrm{j}\pi} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-\mathrm{j}\arg(w)+\mathrm{j}\pi} \end{pmatrix}. \qquad \text{(E.0.24)}$$

The optimal PSD matrix $W_{\mathbf{x}}(\theta)$ is found by substituting (E.0.20) into (E.0.10).

Appendix F

Capacity of MIMO Channels with Noise Uncertainty

Proof of Theorem 4.4.1. The solution of the MIMO channel capacity in the presence of the noise uncertainty requires the solution of the following optimization problem

$$\sup_{W_{\mathbf{x}} \in A_1} \inf_{W_{\mathbf{n}} \in A_2} J(W_{\mathbf{x}}(\theta), W_{\mathbf{n}}(\theta)), \tag{F.0.1}$$

where $J(W_{\mathbf{x}}(\theta), W_{\mathbf{n}}(\theta)) = \frac{1}{4\pi}\int_0^{2\pi} \log\, \det(I_p + H(e^{j\theta})W_{\mathbf{x}}(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^{-1}(\theta))d\theta$. As in the SISO case, we formulate the Lagrangian

$$\begin{aligned} J_1(W_{\mathbf{x}}, W_{\mathbf{n}}, \lambda_1) &= \frac{1}{4\pi}\int_0^{2\pi} \log\, \det(I + HW_{\mathbf{x}}H^*W_{\mathbf{n}}^{-1})d\theta \\ &+\lambda_1\Big(\int_0^{2\pi} \mathrm{Trace}(W_{\mathbf{n}})d\theta - P_{\mathbf{n}}\Big), \end{aligned} \tag{F.0.2}$$

where $\lambda_1 \geq 0$ is a Kuhn-Tucker condition [15]. The variation of J_1 with respect to $W_{\mathbf{n}}$ gives the quadratic, Riccati type equation

$$\begin{aligned} &W_{\mathbf{n}}^{o2}(\theta) + \frac{1}{2}W_{\mathbf{n}}^o(\theta)H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta}) + \frac{1}{2}H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^o(\theta) \\ &-\frac{1}{4\pi\lambda_1^o}H(e^{j\theta})W_{\mathbf{x}}^o(\theta)H^*(e^{j\theta}) = 0, \end{aligned} \tag{F.0.3}$$

which cannot be solved explicitly, in contrast to the SISO case. Therefore, it has to be accounted as the equality constraint in order to resolve the supremum part. The original problem $\sup_{W_{\mathbf{x}} \in A_1} \inf_{W_{\mathbf{n}} \in A_2} J(W_{\mathbf{x}}, W_{\mathbf{n}})$ is equivalent to $\sup_{W_{\mathbf{x}} \in A_1} \sup_{\lambda_1 \geq 0} J_1(W_{\mathbf{x}}, W_{\mathbf{n}}^o, \lambda_1)$ (dual problem [15]). Further, introduce the Lagrangian

$$\begin{aligned} J_2(W_{\mathbf{x}}, W_{\mathbf{n}}^o, \lambda_1, \lambda_2) &= \frac{1}{4\pi}\int_0^{2\pi} \log\, \det(I_p + HW_{\mathbf{x}}H^*W_{\mathbf{n}}^{o-1})d\theta \\ &+\lambda_1\Big(\int_0^{2\pi} \mathrm{Trace}(W_{\mathbf{n}}^o)d\theta - P_{\mathbf{n}}\Big) \end{aligned}$$

$$-\lambda_2\Big(\int_0^{2\pi}\text{Trace}(W_{\mathbf{x}})d\theta - P_{\mathbf{x}}\Big)$$
$$-\int_0^{2\pi}\text{Trace}[K(W_{\mathbf{n}}^{o2}+\frac{1}{2}HW_{\mathbf{x}}H^*W_{\mathbf{n}}^o+\frac{1}{2}W_{\mathbf{n}}^oHW_{\mathbf{x}}H^*-\frac{1}{4\pi\lambda_1^o}HW_{\mathbf{x}}H^*)]d\theta, \quad (\text{F.0.4})$$

where λ_2 is a positive constant, and K is a positive semi definite matrix, which are Kuhn-Tucker conditions [15]. Then,

$$\sup_{W_{\mathbf{x}}\in A_1}\sup_{\lambda_1\geq 0} J_1(W_{\mathbf{x}},W_{\mathbf{n}}^o,\lambda_1) \quad (\text{F.0.5})$$

is equivalent to

$$\inf_{\lambda_2\geq 0}\sup_{\lambda_1\geq 0}\sup_{W_{\mathbf{x}}\in A_1} J_2(W_{\mathbf{x}},W_{\mathbf{n}}^o,\lambda_1,\lambda_2), \quad (\text{F.0.6})$$

(dual problem [15]). Since $W_{\mathbf{x}}$ and $W_{\mathbf{n}}^o$ are related through equality constraint (F.0.3), the Lagrangian J_2 is varied with respect to both, $W_{\mathbf{x}}$ and $W_{\mathbf{n}}^o$. By varying J_2 with respect to $W_{\mathbf{x}}$, the following equation is obtained

$$H^*W_{\mathbf{n}}^{o-1}(I+HW_{\mathbf{x}}H^*W_{\mathbf{n}}^{o-1})^{-1}H-4\pi H^*W_{\mathbf{n}}KH+\frac{1}{\lambda_1}H^*KH=4\pi\lambda_2 I_m. \quad (\text{F.0.7})$$

By varying J_2 with respect to $W_{\mathbf{n}}$, we have

$$-W_{\mathbf{n}}^{o-1}(I_p+HW_{\mathbf{x}}H^*W_{\mathbf{n}}^{o-1})^{-1}HW_{\mathbf{x}}H^*W_{\mathbf{n}}^{o-1}+4\pi\lambda_1 I_p=4\pi K(2W_{\mathbf{n}}^o+HW_{\mathbf{x}}H^*). \quad (\text{F.0.8})$$

(F.0.8) can be further massaged to give

$$\begin{aligned}&W_{\mathbf{n}}^{o2}+HW_{\mathbf{x}}H^*W_{\mathbf{n}}^o-\frac{1}{4\pi\lambda_1^o}HW_{\mathbf{x}}H^*\\&=\frac{1}{\lambda_1^o}(I+HW_{\mathbf{x}}H^*W_{\mathbf{n}}^{o-1})W_{\mathbf{n}}^oK(2W_{\mathbf{n}}^o+HW_{\mathbf{x}}H^*)W_{\mathbf{n}}^o\\&=0,\end{aligned} \quad (\text{F.0.9})$$

which follows from (F.0.3). Hence, one or more terms on the right hand side must be equal to zero. From the setting of the problem, the only possibility that remains is $K=0$. This implies that the two constraints imposed on $W_{\mathbf{x}}$ and $W_{\mathbf{n}}$ can be decoupled, and that the saddle point exists. Thus,

$$\sup_{W_{\mathbf{x}}\in A_1}\inf_{W_{\mathbf{n}}\in A_3} J(W_{\mathbf{x}},W_{\mathbf{n}})=\inf_{W_{\mathbf{n}}\in A_3}\sup_{W_{\mathbf{x}}\in A_1} J(W_{\mathbf{x}},W_{\mathbf{n}}), \quad (\text{F.0.10})$$

as well as

$$J(W_{\mathbf{x}},W_{\mathbf{n}}^o)\leq J(W_{\mathbf{x}}^o,W_{\mathbf{n}}^o)\leq J(W_{\mathbf{x}}^o,W_{\mathbf{n}}). \quad (\text{F.0.11})$$

In addition, (F.0.7) and (F.0.8) transform into

$$H^*(e^{j\theta})W_{\mathbf{n}}^{o-1}(\theta)(I + He^{j\theta}W_{\mathbf{x}}(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^{o-1}(\theta))^{-1}H(e^{j\theta}) = 4\pi\lambda_2 I \tag{F.0.12}$$

$$W_{\mathbf{n}}^{o-1}(\theta)(I + H(e^{j\theta})W_{\mathbf{x}}(\theta)H^*e^{j\theta}W_{\mathbf{n}}^{o-1}(\theta))^{-1}H(e^{j\theta})W_{\mathbf{x}}(\theta)H^*(e^{j\theta})W_{\mathbf{n}}^{o-1}(\theta) = 4\pi\lambda_1 I \tag{F.0.13}$$

These two equations represent the necessary conditions (4.4.18) and (4.4.17), as given in the theorem.

Bibliography

[1] R. Ahlswede, "The capacity of a channel with arbitrary varying Gaussian channel probability functions," *Trans. 6th Prague Conf. Information Theory, Statistical Decision Functions and Random Processes*, Sept. 1971, pp. 13-21.

[2] C. R. Baker, "Capacity of Gaussian channel without feedback," *Information and control*, vol. 37, pp. 70-89, 1978.

[3] C. R. Baker, "Capacity of mismatched Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 33, pp. 802-812, no. 6, Nov. 1987.

[4] C. R. Baker, and S. Ihara, "Information capacity of the stationary Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1314-1326, no. 5, Sep. 1991.

[5] C. R. Baker and I.-F. Chao, "Information capacity of channels with partially unknown noise. I. Finite dimensional channels," *SIAM J. Appl. Math.*, vol. 56, no. 3, pp. 946-963, Jun. 1996.

[6] C. R. Baker and I.-F. Chao, "Information capacity of channels with partially unknown noise. II. Infinite dimensional channels," *SIAM J. Control and Optimization*, vol. 34, no. 4, pp. 1461-1472, Jul. 1996.

[7] T. Basar, "A complete characterization of minimax, and maximin encoder-decoder policies for communication channels with incomplete statistical description," *IEEE Trans. Inform. Theory*, vol. 31, pp. 482-489, Jan. 1985.

[8] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels - Information-Theoretic and Communications Aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619-2692, Oct. 1998.

[9] N. M. Blachman, "Communication as a game," IRE Wescon 1957 Conference Record, vol. 2, pp. 61-66, 1957.

[10] N. M. Blachman, "The effect of statistically dependent interference upon channel capacity," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 553557, Sep. 1962.

[11] N. M. Blachman, "On the capacity of a band-limited channel perturbed by statistically dependent interference," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 4855, Jan. 1962.

[12] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, pp. 1229-1241, Dec. 1959.

[13] H. Boche and E. A. Jorsweick, "Multiuser MIMO systems, worst case noise, and transmitter cooperation," *Proc. of the 3rd IEEE Int. Symp. Signal Processing and Information Technology, ISSPIT 2003*, 2003.

[14] H. Bolcskei, D. Gesbert, and A. Paulraj, "On the capacity of OFDM- based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225-234, Feb. 2002.

[15] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2003.

[16] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: the multivariate case," *Bell Syst. Tech. J.*, vol. 53, no. 5, 1974.

[17] P. E. Caines, *Linear stochastic systems.* New York: John Wiley & Sons, 1988.

[18] G. Caire and S. Shamai, "On the Capacity of Some Channels with Channel State Information," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2007-2019, Sep. 1999.

[19] C. D. Charalambous, S. Z. Denic and S. M. Djouadi, "Robust Capacity of White Gaussian Noise Channels with Uncertainty," *Proc. 43th IEEE Conf. Decision and Control*, Bahamas, 2004.

[20] C. D. Charalambous, S. Denic and S.M. Djouadi, "Robust Capacity of Uncertain Gaussian Noise Channel with Channel and Noise Uncertainty," presented at ACC 2005.

[21] C. D. Charalambous, S. Denic and S.M. Djouadi, "Robust Capacity for Additive Colored Gaussian Uncertain Channels," presented at IFAC Congress, Praha, 2005.

[22] T. Cover, J. Thomas, *Elements of Information Theory.* John Wiley & Sons, 1991.

[23] I. Csiszár, J. Körner, *Information theory: Coding theorems for discrete memoryless systems*, New York: Academic Press, 1981.

[24] I. Csiszár, P. Narayan, "Capacity of the Gaussian arbitrary varying channels," *IEEE Trans. Inform. Theory*, vol. 37, no. 1, pp. 18-26, Jan., 1991.

[25] S. Z. Denic, C. D. Charalambous and S.M. Djouadi, "Capacity of Gaussian channels with noise uncertainty", *Proc. IEEE CCECE 2004*, Canada, May 2004.

[26] S. Z. Denic, C. D. Charalambous and S.M. Djouadi, "Robust capacity for additive colored Gaussian uncertain channels," preprint.

[27] S. Z. Denic, C. D. Charalambous,S. M. Djouadi, and O. Milenkovic, "Capacity of MIMO Gaussian channels with H normed channel uncertainties," accepted for presentation at 9th Canadian Workshop on Information Theory, 2005.

[28] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint", *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 3072-3081, Nov. 2001.

[29] L. Dobrushin, "Optimal information transmission through a channel with unknown parameters", *Radiotekhnika i Electronika*, vol. 4, pp. 1951-1956, 1959.

[30] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory.* New York: McMillan Publishing Company, 1992.

[31] N. Dunford and J. T. Schwartz, *Linear Operators, General Theory* Wiley-Interscience; New Ed edition, 1988

[32] M. Feder and A. Lapidoth, "Universal Decoding for Channels with Memory", *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1726-1745, Sep. 1998.

[33] L. J. Forys, P. P. Varaiya, "The ϵ-capacity of classes of unknown channels", *Inform. Contr.*, vol. 44, pp. 376-406, 1969.

[34] G. J Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas," Bell Labs Tech. J., pp. 4159, 1996.

[35] R. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[36] I. M. Gel'fand, and A. M. Yaglom, "Calculation of amount of information about a random function contained in another such function," *Amer. Math. Soc. Transl.*, vol. 12, no. 2, pp. 199-246, 1959.

[37] A. Goldsmith and P. Varaiya, "Capacity of Fading Channels with Channel Side Information", *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1986-1992, Nov. 1997.

[38] A.J. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE J. Select. Areas Commun.*, vol. 21, No. 5, pp. 684-702, Jun. 2003.

[39] D. Hosly and A. Lapidoth, "The capacity of MIMO Ricean channels is monotonic in the singular values of the mean," 5^{th} *International ITG Conference on Source and Channel Coding*, Jan. 2004.

[40] B. Hughes and P. Narayan, "Gaussian arbitrary varying channels," *IEEE Trans. Inform. Theory*, vol. 33, no. 2, pp. 267-284, Mar. 1987.

[41] B. Hughes and P. Narayan, "The capacity of vector Gaussian arbitrary varying channel", *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 995-1003, Sep. 1988.

[42] S. Ihara, "On the capacity of Channels with additive non-Gaussian noise," *Inform. Contr.*, vol. 37, pp. 34-39, 1978.

[43] S. A. Jafar, S. Vishwanatah and A. J. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. Int. Conf. Commun.*, vol. 7, pp. 2266-2270, 2001.

[44] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," *IEEE Trans. Wireless Commun.*, Jul. 2004, vol. 3, no. 4, pp. 1165-1175.

[45] A. N. Kolmogorov, "On the Shannon theory of information in the case of continuous signals," *IEEE Trans. Inform. Theory*, vol. 2, pp. 102-108, 1956.

[46] E. Kreyszig, *Introductory functional analysis with applications*, John Wiley, 1991.

[47] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels", *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1520-1529, Sep. 1996.

[48] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty", *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148-2177, Oct. 1998.

[49] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple antenna systems on flat fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426-2467, Oct. 2003.

[50] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Oct. 1999.

[51] R. J. McElice, "Communications in the presence of jamming An information theoretic approach", in *Secure Digital Commun.*, G. Longo, ed., Springer-Verlang, New York, 1983, pp. 127-166.

[52] M. Médard, "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel", *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933-946, May 2000.

[53] M. Medard, "Channel uncertainty in communications", *IEEE Information Theory Society Newsletters*, vol. 53, no. 2, p. 1, pp. 10-12, Jun. 2003.

[54] A. F. Molisch, M. Steinbauer, M. Toeltsch, E. Bonek, and R. S. Thomä, "Capacity of MIMO systems based on measured wireless channels," *IEEE J. Select. Areas Commun.*, vol. 20, no. 3, pp. 561-569, Apr. 2002.

[55] A. Narula, M. Trott, and G. Wornel, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Transactions on Information Theory*, vol. 45, pp. 2418-2433, Nov. 1999.

[56] M. J. Osborne, A. Rubinstein, *A course in game theory*, MIT Press, 1994.

[57] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information Theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359-378, May 1994.

[58] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uinform power allocation in MIMO channels: a game theoretic approach," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1707-1727, Jul. 2003.

[59] M. S. Pinsker, *Information and information stability of random variables and processes,* Holden-Day, San Francisco, 1964.

[60] J. Proakis, *Digital communications*, McGraw-Hill, 1995.

[61] T. Rapaport, *Wireless communications: principle and practice*, Prentice Hall, 2001.

[62] G. G. Raleigh and J. M. Cioffi, "Spatio temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, Mar. 1998.

[63] W. L. Root, P. P. Varaiya, "Capacity of Classes of Gaussian Channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350-1393, Nov. 1968.

[64] C. E. Shannon, "Mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, pp. 623-656, 1948.

[65] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289-293, Jul. 1958.

[66] C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, no. 1, pp. 10-21, Jan. 2003.

[67] S. Simon and A. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 406-417, Apr. 2003.

[68] S. Simon and A. Moustakas, "Optimality of beamforming in multiple transmitter multiple receiver communication systems with partial channel knowledge," *Proc. DIMACS Workshop Signal Processing Wireless Communications*, DIMACS center, Rutgers Univ., Oct. 7-9, 2002.

[69] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585596, Nov. 1999.

[70] E. Telatar and D. Tse, "Capacity and Mutual Information of Wideband Multipath Fading Channels", *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1384-1400 Jul. 2000.

[71] S. Vishwanath, S. Boyd, and A. Goldsmith, "Worst-case capacity of Gaussian vector channels," *Proc. 2003 Canadian Workshop on Information Theory*, 2003.

[72] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632-2639, Sep. 2001.

[73] A. A. Stoorvogel and J. H. van Schuppen, "System identification with information theoretic criteria," *Proc. NATO advanced study institute 'From identification to learning'*, Como, Italy, 1994.

[74] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. Select. Areas Commun.*, vol. 5, pp. 871878, Jun. 1987.

[75] J. Wolfowitz, *Coding Theorems of Information Theory*, Springer Verlang, Belin Heildelberg, 1978.

[76] T. Yoo, E. Yoon, and A. J. Goldsmith, "MIMO Capacity with Channel Uncertainty: Does Feedback Help?", *IEEE GlobeCom 2004*, Dallas, Texas, Dec. 2004.

[77] T. Yoo and A. Goldsmith, "Capacity and optimal power allocation for fading MIMO channels with channel estimation error," *http://wsl.stanford.edu/Publications.html#Taesang%20Yoo*

[78] G. Zames, "Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Trans. Automat. Control*, vol. 26, pp. 301-320, 1981.

[79] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359-383, Feb. 2002.

[80] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control.* New Jersey: Prentice hall, 1996.